بسم الله الرحمن الرحیم

قابل استفاده دانشجویان کارشناسی و داوطلبین آزمون کارشناسی ارشد
رشته‌های اقتصاد، مدیریت، حسابداری، برنامه ریزی شهری و منطقه‌ای،
برنامه ریزی محیط زیست، سنجش از دور(GIS)

آمار و احتمال

شامل:

- شرح کامل مباحث درسی همراه با مثال‌های تألیفی
- آزمون‌های کارشناسی ارشد همراه با حل تشریحی
- خودآزمایی‌های طبقه‌بندی شده موضوعی

مؤلف: محسن طورانی

سرشناسه	:	طورانی، محسن، ۱۳۵۱ -
عنوان و نام پدیدآور	:	آمار و احتمال قابل استفاده دانشجویان کارشناسی و داوطلبین آزمون کارشناسی‌ارشد/ مولف محسن طورانی.؛ [برای موسسه آموزش عالی آزاد پارسه].
مشخصات نشر	:	تهران: حرکت نو، ۱۳۹۰.
مشخصات ظاهری	:	۸۰۰ ص.: مصور (رنگی)، جدول (رنگی)، نمودار (رنگی).؛ ۲۲×۲۹ س‌م.
شابک	:	۲۱۰۰۰۰ ریال 978-600-6347-01-1
موضوع	:	دانشگاه‌ها و مدارس عالی -- ایران -- آزمون‌ها
موضوع	:	آمار ریاضی -- آزمون‌ها و تمرین‌ها (عالی)
موضوع	:	آمار ریاضی -- راهنمای آموزشی (عالی)
موضوع	:	احتمالات -- راهنمای آموزشی (عالی)
موضوع	:	احتمالات -- آزمون‌ها و تمرین‌ها (عالی)
موضوع	:	آزمون دوره‌های تحصیلات تکمیلی -- ایران
شناسه افزوده	:	مؤسسه آموزش عالی آزاد پارسه
رده‌بندی کنگره	:	۱۳۹۰ ۸۳آ۷۸۵ط/ LB۲۳۵۳
رده‌بندی دیویی	:	۳۷۸/۱۶۶۴
شماره کتابشناسی ملی	:	۲۴۳۱۹۱۴

عنوان کتاب	:	آمار و احتمال
مؤلف	:	محسن طورانی
ناشر	:	حرکت نو
سال ویرایش	:	۱۳۹۰
نوبت چاپ	:	چهارم ـ ۱۳۹۰
چاپ اول	:	۱۳۸۶
شمارگان	:	۸۰۰۰
چاپ	:	ناجی نشر
طرح جلد	:	موسسه آموزش عالی آزاد پارسه
قیمت	:	۲۱۰ ۰۰۰ ریال
شابک	:	978-600-6347-01-1

تلفن: ۸۴۳۸۸

تقدیم به

دانشجویان عزیز

یادداشت ناشر

خـرد رهنمـای و خـرد دلگشـای
خرد دســت گیرد به هـر دو سرای
خرد چشم جان اســت چون بنگری
تو بـی‌چشـم شـادان جهان نسپری
بــه گفتـار داننــدگــان راه جــوی
بــه گیتی بپوی و بــه هر کس بگوی
ز هـــر دانـــش چون سخن بشنوی
از آمـوختـن یـک زمـان نغنـوی
فردوسی

در دنیایی زندگی می‌کنیم که تکنولوژی به پیش‌پاافتاده‌ترین جنبه‌های زندگیمان سـرک می‌کشـد و همــه مرزهایــش را درنوردیده، در این میانه دسترســی به متون و منابع نگارشــی با یک کلیک میســر می‌شــود و هر فرد می‌تواند صاحب کتابخانه‌ای بزرگ باشــد، اما با تمام این تفاسیر، آنچه شــگفت‌انگیز است میل انسانها به داشــتن و خواندن کتاب‌های واقعی است، کتاب‌هایی از جنس کاغذ و جوهر که سالهاســت از نســلی به نسل دیگر در ســفرند و چرخ‌های سترگ تکنولوژی را به حرکت درمی‌آورند. شــاید این خاصیت کاغذ و جوهر اســت که گویی با علم و دانش پیمانی ناگسستنی بسته است، اتاق اساتید بزرگ دانشگاه را که نگاه می کنی نخستین و برجسته‌ترین نکته‌ای که نظرت را جلب می کند کتابخانه‌های غنی و کتب منبع آنهاست که همگی شاید یادگار روزگار دانشجویی و علم‌آموزیشان بوده است.

مؤسسه آموزش عالی آزاد پارسه نیز آرزو دارد کتابهایش همواره همراه شما باشند از این روست که حداکثر تلاش خود را کرده تا این کتابها علاوه بر دربرگیری همه نکاتی که برای یک دانشجوی کنکوری مورد نیاز است با بیانی روان و قابل فهم کلیه مطالب و مباحث آکادمیک را شرح و بسط دهد تا خواننده را از رجوع به کتابهای مشــابه بی‌نیاز کند و ســالها در کتابخانه‌های شــما به عنوان مرجعی ماندگار و جامع حضور داشته باشد.

پارســه دغدغه ای دیگر نیز دارد، امروز که اکثریت نســل جوان جامعه ما دارای تحصیلات عالی هســتند، جای تفکر و تامل بسیار اســت که ســرانه کتاب و کتابخوانی هر ایرانی از جایگاه نازلی در میان دیگر کشــورهای دنیا برخوردار باشد.این حقیقت تلخ تر می نماید وقتی نگاهی به پیشینه فرهنگ و هنر در این مرز و بوم بیاندازیم که هیچ کس نمی تواند منکر برجســتگی و شــکوهش باشــد. شاید وظیفه هر انسان دلسوز و نهاد آموزشــی و فرهنگی این است که در راه اعتلای مجدد این فرهنگ گام بردارد چرا که این قدم ها هر چند کوچک هم که باشد می تواند مقدمه‌ای گردد برای خیزی بلند و سرافرازانه به جایگاه مرتفع اسلاف بزرگ و ماندگارمان. پارسه نیز بر خود بایسته می داند با رویکردهای فرهنگی، مانند اهدا کتاب و تجهیز کتابخانه ها به سهم خود در پیشبرد این جریان همت گمارد و امیدوار است در آینده نزدیک شاهد شکوفایی دوباره این فرهنگ فاخر در کشورمان باشد.

مؤسسه آموزش عالی آزاد پارسه از شما تقاضا دارد تا نظرات و پیشنهادات خود را درمورد کتابهای این مؤسسه به پست الکترونیکی به آدرس CRM@parseh.ac.ir ارسال کنید تا هر روز و با یکدیگر به سمت بهترشدن پیش رویم.

مقدمه

جهان هر کس به اندازه وسعت فکر اوست.
ادیسون

هدف از انتشار این کتاب، ارائه مجموعه‌ای کامل از درس و تست است که با بهره‌گیری از معتبرترین منابع درسی و بر اساس سرفصل‌های مصوب شورای عالی برنامه‌ریزی و عناوین درسی، برای رشته‌های اقتصاد، مدیریت، حسابداری، برنامه‌ریزی شهری و منطقه‌ای، برنامه‌ریزی محیط زیست و سنجش از دور (GIS) تدوین شده است.

این کتاب در برگیرنده ۶ فصل در قالب مفاهیم کامل و نکات مهم درسی است؛ نحوه بیان مطالب هر فصل مبتنی بر تجربه مؤلف از دوره‌های متعدد تدریس آمار در کلاس‌های دانشگاهی و آمادگی آزمون کارشناسی ارشد است، به‌گونه‌ای که سعی بر آن بوده تا در عین سادگی از سطح علمی مطلوبی نیز برخوردار باشد.

در پایان هر فصل سؤالات آزمون‌های کارشناسی ارشد سال‌های ۸۶ تا ۸۸ دانشگاه سراسری با تقسیم‌بندی موضوعی و سؤالات سال ۸۹ در انتهای کتاب، همراه با حل تشریحی آورده شده است تا داوطلب بتواند پس از مطالعه هر فصل، آن‌ها را تحلیل و بررسی کند؛ توصیه می‌شود که داوطلبان هر رشته، سؤالات مربوط به تمام رشته‌ها را بررسی کنند.

به‌علاوه، خودآزمایی‌های طبقه‌بندی‌شده بر اساس موضوع، از مجموعه سؤالات کتب مرجع به‌گونه‌ای انتخاب شده است که شامل تمام مطالب فصل باشد.

بودجه‌بندی سؤالات آزمون‌های سراسری سال‌های گذشته، با توجه به رشته و بر اساس فصل‌های این کتاب، در جدول زیر آمده است:

رشته / فصل	اقتصاد و محیط زیست	مدیریت و حسابداری	برنامه‌ریزی شهری و منطقه‌ای و سنجش از دور (GIS)
اول	۴ تا ۵ سؤال	۵ تا ۷ سؤال	۵ تا ۸ سؤال
دوم	۳ تا ۴ سؤال	۲ تا ۴ سؤال	۱ تا ۲ سؤال
سوم	۴ تا ۵ سؤال	۲ تا ۳ سؤال	۲ تا ۳ سؤال
چهارم	۳ تا ۴ سؤال	۳ تا ۴ سؤال	۳ تا ۴ سؤال
پنجم	۷ تا ۸ سؤال	۰ تا ۱ سؤال	۱ تا ۲ سؤال
ششم	۵ تا ۶ سؤال	۰ تا ۱ سؤال	۰ تا ۲ سؤال

این بودجه‌بندی در کنکور دانشگاه آزاد اسلامی وجود ندارد و دانشجویان تمام رشته‌ها باید تمام فصل‌های کتاب را مطالعه کنند.

تهیه و تدوین این مجموعه مدیون پیگیری‌ها و زحمات ریاست محترم مؤسسه پارسه، جناب آقای مهندس کاوه عابدینی‌زاده است که در تمام مراحل اینجانب را همراهی کرده‌اند.

بی‌تردید هیچ نوشته ای خالی از اشکال نیست؛ دانشجویان گرامی می‌توانند نظرات و پیشنهادات خود را به آدرس M.Tourani@Parseh.ac.ir ارسال کنند.

محسن طورانی
تابستان ۱۳۹۰

فهرست

فصل ۱ آمار توصیفی

علم آمار

واژه آمار دو مفهوم کاملاً متفاوت دارد؛ در اصطلاح عام، آمار عبارت است از اعداد و ارقام معناداری که بیان‌کننده اطلاعاتی خاص مانند جمعیت یک کشور، تعداد خانوارهای یک شهر یا درآمد سرانه هستند. اما مفهوم تخصصی آن یعنی علم آمار، مجموعه‌ای از روش‌هاست که برای جمع‌آوری، طبقه‌بندی، تلخیص، تجزیه و تحلیل و تفسیر اطلاعات مورد استفاده قرار می‌گیرد.

جامعه و نمونه

اطلاعات آماری، از یک جامعه آماری حاصل می‌شود. **جامعه** (Population) مجموعه‌ای از عناصر مورد نظر است که حداقل یک صفت مشترک (صفت مشخصه) داشته باشند. برای مثال دانشجویانی که در درس آمار نمره بالاتر از 17 گرفته‌اند، جامعه‌ای را تشکیل می‌دهند که نمره بالاتر از 17، صفت مشترک آن‌هاست.

جامعه آماری، محدود یا نامحدود است. **جامعه آماری محدود** جامعه‌ای است که عناصر آن، مجموعه‌ای محدود و پایان‌پذیر باشد؛ مانند تعداد کارمندان یک شرکت. **جامعه آماری نامحدود** جامعه‌ای است که عناصر آن، مجموعه‌ای نامحدود یا بی‌پایان باشد؛ مانند تعداد ماهی‌های تمام اقیانوس‌ها.

سرشماری یا نمونه‌گیری؟

جمع‌آوری اطلاعات از جامعه (آمارگیری) به محدود یا نامحدود بودن جامعه بستگی دارد؛ اگر جامعه محدود باشد، اغلب می‌توان از سرشماری (جمع‌آوری اطلاعات مورد نظر از تمام عناصر جامعه) استفاده کرد؛ اما درصورتی‌که جامعه نامحدود باشد، اغلب سرشماری انجام نمی‌شود زیرا یا دسترسی به تمام عناصر جامعه ممکن نیست و یا سرشماری بسیار پرهزینه و وقت‌گیر است؛ بنابراین **پارامتر** (**شاخص مورد نظر**) همواره در جامعه **مجهول** است، هرچند در هر لحظه **مقدار ثابتی** دارد. در چنین شرایطی، جمع‌آوری اطلاعات با نمونه‌گیری انجام می‌شود. **نمونه** (Sample) تعداد محدودی از عناصر جامعه است که تمام خصوصیات (ویژگی‌های) جامعه را دارد.

✓ **دقت کنید!**

اندازه جامعه را با N و اندازه نمونه را با n نشان می‌دهند.

اگر شاخصی از جامعه با آمارگیری از تمام عناصر آن (سرشماری) به دست آید، به آن **پارامتر** (Parameter) گویند. برای مثال اگر میانگین قد دانش‌آموزان یک کشور با استفاده از اندازه‌گیری قد تمام دانش‌آموزان آن کشور حاصل شود، پارامتر جامعه به دست آمده است.

اگر شاخصی از جامعه با آمارگیری از بخشی از عناصر آن (**نمونه‌گیری**) به دست آید، به آن **آماره** (**برآوردکننده، تخمین‌زننده**) (Statistic) گویند. برای مثال اگر میانگین قد دانش‌آموزان یک کشور با اندازه‌گیری قد بعضی از دانش‌آموزان آن کشور حاصل شود، آماره‌ای به دست می‌آید که برآوردکننده‌ای برای پارامتر جامعه است.

طبیعی است مقدار آماره همواره **معلوم** است، اما ازآنجاکه نمونه به صورت تصادفی از نقاط مختلف جامعه اختیار می‌شود، مقدار آن (آماره) از یک نمونه به نمونه دیگر تغییر می‌کند و بنابراین یک **متغیر تصادفی** است. درنتیجه بدیهی است اگر بخواهیم شاخص‌های به‌دست‌آمده از نمونه‌گیری (آماره‌ها) را به پارامتر جامعه تعمیم دهیم، باید آن‌ها را **آزمون** کنیم تا اولاً بتوانیم میزان خطا در هر یک از آماره‌ها را محاسبه کنیم و ثانیاً بتوانیم از بین آن‌ها بهترین شاخصی را پیدا کنیم که بتواند تقریب نزدیکی از پارامتر جامعه باشد.

نتیجه:

گروه	آمارگیری	شاخص	مشخصات شاخص	آزمون کردن
جامعه	سرشماری	پارامتر (θ)	ثابت (عدد) ـ مجهول	نیاز ندارد.
نمونه	نمونه‌گیری	آماره (برآوردکننده) $(\hat{\theta})$	متغیر ـ معلوم	نیاز دارد $(n \geq 1)$

مثال پارامترها به کمک کدام‌یک از موارد زیر قابل محاسبه هستند؟

۱) جامعه
۲) جامعه و نمونه
۳) نمونه
۴) به شرایط تحقیق بستگی دارد.

حل: گزینه ۱ درست است.

مسیر توسعه علم آمار

از نظر موضوعی مسیر توسعه علم آمار به سه مرحله تقسیم می‌شود:

۱ـ آمار توصیفی (Descriptive Statistics)

آمار توصیفی عبارت است از روش‌هایی که در جمع‌آوری، طبقه‌بندی، تلخیص و نمایش داده‌های آماری مورد استفاده قرار می‌گیرد؛ به طوری که برای استفاده‌کنندگان به سادگی قابل فهم باشد. در این بخش از علم آمار، با محاسبه **شاخص‌های جامعه** (**پارامترها**) از طریق **سرشماری** فقط به **توصیف جامعه** پرداخته می‌شود بدون آنکه استنباطی درباره اطلاعات صورت گیرد.

بنابراین در **آمار توصیفی**:

موضوع: توصیف جامعه	**هدف**: محاسبه پارامترهای جامعه	**نوع آمارگیری**: سرشماری

مثال برای بررسی فرضیه‌ای از تمام مشاهدات جامعه استفاده شده است، آن را چه می‌نامند؟

۱) آمار توصیفی
۲) آمار استنباطی
۳) آمار پارامتریک
۴) همبستگی

حل: گزینه ۱ درست است.

۲ـ آمار استنباطی (Inferential Statistics)

در آمار استنباطی ابتدا شاخص‌های نمونه (آماره‌ها) از طریق نمونه‌گیری از جامعه محاسبه شده، سپس به کمک برآورد (تخمین) و آزمون، نتایج حاصل از نمونه (آماره‌ها) به پارامترهای جامعه تعمیم داده می‌شوند. به طور کلی در مباحث آماری، هر جا سخن از استنباط و استنتاج باشد، آن را آمار استنباطی می‌خوانند.

بنابراین در **آمار استنباطی**:

موضوع: استنباط پارامتر جامعه از آماره نمونه	**هدف:** محاسبه آماره‌ها و تعمیم آن‌ها به پارامترها	**نوع آمارگیری:** نمونه‌گیری

تفاوت در نوع آمارگیری (سرشماری و نمونه‌گیری) و نتایج حاصل از آن (پارامتر و آماره) باعث تقسیم‌بندی علم آمار به «آمار توصیفی» و «آمار استنباطی» شده است.

۳ـ آمار ناپارامتریک (Non-Parametric Statistics)

این نوع آمار در مقابل آمار پارامتریک بیان می‌شود. در **آمار پارامتریک** متغیرها دارای مقیاس **کمّی (پیوسته)** هستند و مشاهدات از **منحنی (توزیع) نرمال** تبعیت می‌کنند (منحنی نرمال نمایشی از قرار گرفتن عناصر جامعه است که در آن تمام عناصر، حول مرکز (میانگین جامعه) متقارن هستند)، اما در **آمار ناپارامتریک** بیشتر متغیرها دارای **مقیاس کیفی** هستند و چون دقیقاً قابل اندازه‌گیری نیستند، از **هیچ توزیع آماری پیروی نمی‌کنند**. این نوع آمار، **آزاد از توزیع** خوانده می‌شود.

بنابراین در **آمار ناپارامتریک**:

موضوع: متغیرهای کیفی

مشخصات: آزاد از توزیع (به توزیع نرمال و هیچ توزیع مشخصی وابسته نیستند.)

و در **آمار پارامتریک**:

موضوع: متغیرهای کمّی

مشخصات: از توزیع نرمال تبعیت می‌کند.

مثال چه نوع آماری آزاد از توزیع است؟

۱) توصیفی ۲) پارامتریک ۳) ناپارامتریک ۴) استنتاجی

حل: گزینه ۳ درست است.

آمار توصیفی در این فصل و آمار استنباطی در فصل‌های ۵ و ۶ بررسی می‌شود. آمار ناپارامتریک خارج از حوزه بحث این کتاب است.

صفت (Attribute)

تعریف: کمیت یا کیفیتی که متعلق به عناصر جامعه آماری است، صفت نامیده شده و به دو بخش تقسیم می‌شود: صفت ثابت (مشخصه، مشترک) و صفت متغیر.

۱ـ صفت ثابت (مشخصه، مشترک)

صفت یا موضوعی است که میان افراد جامعه مشترک است و افراد جامعه را از جوامع دیگر متمایز می‌کند، مانند صفت دانش‌آموز بودن برای جامعه دانش‌آموزان یک کشور.

مثال کدام تعریف برای صفت مشخصه درست است؟

۱) صفتی است که اندازه‌گیری آن از فردی به فرد دیگر تغییر می‌کند.

۲) صفتی است که قابل اندازه‌گیری است.

۳) صفتی مشترک برای افراد جامعه است.

۴) صفتی که قابل شمارش باشد.

حل: گزینه ۳ درست است.

۲ـ صفت متغیر

صفت یا موضوعی است که اندازه‌گیری آن در یک جامعه آماری، اعضای جامعه را از هم متمایز می‌کند، مانند قد دانش‌آموزان، حقوق کارمندان، تعداد غایبین کلاس و گروه خونی افراد شهر. در مباحث آماری، این صفت، **متغیر** (Variable) نامیده می‌شود.

انواع متغیرها

متغیرها، بر این اساس که **قابل اندازه‌گیری** باشند یا نباشند، به صورت زیر دسته‌بندی می شوند:

۱ـ متغیر کمّی (قابل اندازه‌گیری یا شمارش)

۲ـ متغیر کیفی (غیر قابل اندازه‌گیری)

۱ـ متغیرهای کمّی

تعریف: متغیرهایی که قابل اندازه‌گیری یا شمارش هستند و همواره می‌توان عددی (اعشاری یا صحیح) را به آن‌ها نسبت داد، متغیرهای کمّی نامیده می‌شوند.

متغیرهای کمّی **قابل اندازه‌گیری** مانند **قد** افراد یک شهر، **وزن** افراد یک کلاس، **درجه** حرارت شهرهای یک منطقه، **سن** افراد بازنشسته، **متراژ** خانه‌های شهر، **شدت** زلزله و **میزان** آلودگی هوا.

متغیرهای کمّی **قابل شمارش** مانند **تعداد** افراد خانواده، **تعداد** زنبورهای یک کندو، **تعداد** تخلفات رانندگی و **تعداد** غایبین یک کلاس.

در این متغیرها عدد حاصل از اندازه‌گیری یا شمارش حتماً دامنه معینی دارد؛ برای مثال، بعد از اندازه‌گیری، قد افراد یک شهر می‌تواند بین 150 تا 200 سانتی‌متر باشد یا بعد از شمارش، تعداد غایبین یک کلاس می‌تواند بین 10 تا 20 نفر باشد.

انواع متغیرهای کمّی

متغیرهای کمّی، بر این اساس که بتوانند هر مقداری را در یک فاصله محدود (بین دو عدد دلخواه) اختیار کنند یا خیر، به دو دسته تقسیم می‌شوند: متغیرهای کمّی پیوسته و متغیرهای کمّی گسسته.

الف) متغیرهای کمّی پیوسته

یک متغیر کمّی را پیوسته گویند، درصورتی‌که هرگاه دو مقدار a و b را اختیار کرد، بتواند هر مقدار بین آن دو را نیز اختیار کند؛ به عبارت دیگر، عدد اعشاری برای آن مفهوم داشته باشد. برای مثال اگر وزن افراد یک کلاس بین 50 تا 60 کیلوگرم باشد، آن‌گاه وزن یک فرد می‌تواند هر عددی بین 50 تا 60 کیلوگرم (مانند 52.5 یا 53.75 یا 53) باشد؛ بنابراین وزن، یک متغیر کمّی پیوسته است.

نتیجه: تمام **متغیرهای کمّی قابل اندازه‌گیری** مانند قد، وزن، درجه حرارت، سن و متراژ که می‌توانند تمام مقادیر یک بازه $[a,b]$ را اختیار کنند، متغیرهای **کمّی پیوسته** هستند.

ب) متغیرهای کمّی گسسته

یک متغیر کمّی را گسسته گویند، درصورتی‌که هرگاه دو مقدار a و b را اختیار کرد، نتواند هر مقدار بین آن دو را اختیار کند؛ به عبارت دیگر، عدد اعشاری برای آن مفهوم نداشته باشد. آن دسته از متغیرهای کمّی که پیوسته نباشند، گسسته هستند. برای مثال، اگر تعداد غایبین یک کلاس در ابتدای هفته 3 نفر و در پایان هفته 7 نفر باشد، ممکن است در یکی از روزهای هفته کلاس 4 نفر غایب داشته باشد، اما ممکن نیست تعداد غایبین 4.5 نفر باشد؛ بنابراین تعداد غایبین، یک متغیر کمّی گسسته است.

نتیجه: تمام **متغیرهای کمّی قابل شمارش** مانند **تعداد** افراد خانواده، **تعداد** تخلفات رانندگی و **تعداد** غایبین کلاس که نمی‌توانند تمام مقادیر بازه $[a,b]$ را اختیار کنند، متغیر **کمّی گسسته** هستند.

۲ـ متغیرهای کیفی

تعریف: متغیرهایی که قابل اندازه‌گیری یا شمارش نیستند و فقط ممکن است به نوع یا ترتیب خاصی تعلق داشته باشند، متغیرهای کیفی نامیده می‌شوند.

مانند گروه خونی (O, AB, B, A)، نوع کشت (دیم یا آبی)، کیفیت کالا (سالم یا معیوب)، شغل (آزاد یا دولتی)، مراحل زندگی (کودکی، نوجوانی، جوانی و ...)، مراحل تحصیل (دبستان، راهنمایی، دبیرستان، ...)، سطح درآمد (پردرآمد، متوسط‌درآمد، کم درآمد).

انواع متغیرهای کیفی

متغیرهای کیفی بر این اساس که یک ترتیب طبیعی بین آن‌ها وجود داشته باشد یا خیر، به دو دسته تقسیم می‌شوند: متغیرهای کیفی ترتیبی و متغیرهای کیفی اسمی.

الف) متغیرهای کیفی ترتیبی

در این متغیرها یک نوع ترتیب طبیعی وجود دارد؛ مانند «بالا، وسط، پایین»، «خوب، متوسط، ضعیف»، «پردرآمد، متوسط‌درآمد، کم درآمد»، «کودکی، نوجوانی، ... »، «دبستان، راهنمایی، ...».

ب) متغیرهای کیفی اسمی

در این نوع متغیرها هیچ نوع ترتیبی وجود ندارد و هیچ مقایسه‌ای بین متغیرها امکان‌پذیر نیست؛ مانند گروه خونی (O, AB, B, A)، شغل (آزاد یا دولتی).

انواع متغیرها در یک نگاه

متغیرها
- ۱ـ کمّی
  - الف) پیوسته
  - ب) گسسته
- ۲ـ کیفی
  - الف) ترتیبی
  - ب) اسمی

مقایسه متغیرهای کمّی و کیفی

متغیرهای کمّی به دلیل قابلیت اندازه‌گیری، دقت بیشتری دارند و نتایج حاصل از آن‌ها قابل تعمیم به جامعه است.

مثال نوع هر یک از متغیرهای زیر را مشخص کنید.

حل:

الف) رنگ اتومبیل‌های موجود در یک نمایشگاه اتومبیل ⟵ کیفی اسمی

ب) مقاومت یک ترانزیستور ⟵ کمّی پیوسته

ج) وضع سواد (باسواد، بی‌سواد) ⟵ کیفی ترتیبی

د) تعداد شکایات رسیده به یک پاسگاه پلیس ⟵ کمّی گسسته

ه) درآمد دانشجویان شاغل ⟵ کمّی پیوسته

و) وضعیت تأهل کارمندان یک شرکت ⟵ کیفی اسمی

ز) سن دانشجویان شرکت‌کننده در یک دوره هنری ⟵ کمّی پیوسته

ح) تعداد بیماران مراجعه‌کننده به یک پزشک در طول روز ⟵ کمّی گسسته

ط) میزان تحصیلات افراد یک شهر (دیپلم، کارشناسی، کارشناسی ارشد) ⟵ کیفی ترتیبی

ی) گنجایش آب یک تانکر ⟵ کمّی پیوسته

مقیاس (Scale)

انواع متغیرها (کمّی و کیفی) با توجه به مقیاس‌های مختلف، قابل اندازه‌گیری هستند؛ این مقیاس‌ها عبارت‌اند از:

۱ـ مقیاس اسمی (طبقه‌ای)

۲ـ مقیاس ترتیبی (رتبه‌ای)

۳ـ مقیاس فاصله‌ای

۴ـ مقیاس نسبتی (نسبی ـ کسری)

۱ـ مقیاس اسمی (طبقه‌ای) (Nominal Scale)

ساده‌ترین روش در اندازه‌گیری که از علایم یا اعداد برای گروه‌بندی (طبقه‌بندی) افراد یا اشیا استفاده می‌کند، مقیاس اسمی نامیده می‌شود. در این مقیاس، تنها عناصر جامعه از یکدیگر جدا می‌شوند و به هیچ وجه نمی‌توان فاصله میان آن‌ها را مشخص کرد.

مانند گروه خونی افراد (O, AB, B, A) یا شماره لباس بازیکنان بسکتبال

خصوصیات:

✓ ترتیب ندارد.

✓ فاصله را مشخص نمی‌کند.

✓ مبدأ صفر (مطلق یا قراردادی) ندارد.

کاربرد: برای اندازه‌گیری متغیرهای کیفی به کار می‌رود.

۲ـ مقیاس ترتیبی (رتبه‌ای) (Rank Scale)

درصورتی‌که بین داده‌ها با مقیاس اسمی یک نوع ترتیب طبیعی وجود داشته باشد، آن‌گاه یک مقیاس ترتیبی خواهیم داشت.

مانند سطوح تحصیلی«دبستان، راهنمایی، ...»، سطح درآمد «کم درآمد، متوسط‌درآمد، پردرآمد»، طبقات یک ساختمان «بالا، وسط، پایین».

خصوصیات:

- ✓ ترتیب دارد.
- ✓ فاصله را مشخص نمی‌کند.
- ✓ مبدأ صفر (مطلق یا قراردادی) ندارد.
- ✓ نسبت به مقیاس اسمی، مقیاس قوی‌تری است.

کاربرد: هرچند این مقیاس ترتیب بین داده‌ها را مشخص می‌کند، اما چون نمی‌تواند فاصله بین داده‌ها را مشخص کند، برای اندازه‌گیری متغیرهای کیفی به کار می‌رود.

۳ـ مقیاس فاصله‌ای (Interval Scale)

اگر یک مقیاس، تمام مشخصات مقیاس ترتیبی را داشته باشد و علاوه بر آن بتواند فاصله بین هر دو مقدار دلخواه را با معیار مشخصی به دست آورد، مقیاسی قوی‌تر از مقیاس ترتیبی خواهیم داشت که مقیاس فاصله‌ای نامیده می‌شود.

خصوصیات:

- ✓ ترتیب دارد.
- ✓ فاصله را مشخص می‌کند.
- ✓ صفر قراردادی دارد (برای مثال، در دو مقیاس سانتی‌گراد و فارنهایت که برای اندازه‌گیری درجه حرارت استفاده می‌شود، نقطه صفر یکسان نیست و در هر یک به طور اختیاری و قراردادی انتخاب شده است.)
- ✓ صفر مطلق ندارد.
- ✓ نسبت هر دو عدد، مستقل از واحد و مستقل از نقطه صفر است. برای مثال:

$$\frac{10^\circ \text{ سانتی‌گراد}}{5^\circ \text{ سانتی‌گراد}} = 2 \quad \text{و} \quad \frac{10^\circ \text{ فارنهایت}}{5^\circ \text{ فارنهایت}} = 2$$

کاربرد: با توجه به قابلیت اندازه‌گیری فاصله بین داده‌ها، برای اندازه‌گیری متغیرهای کمّی به کار می‌رود.

۴ـ مقیاس نسبتی (نسبی ـ کسری) (Ratio Scale)

در شرایطی که یک مقیاس تمام مشخصات مقیاس فاصله‌ای را داشته باشد و علاوه بر آن دارای نقطه صفر مطلق (واقعی) باشد، به آن مقیاس نسبتی گفته می‌شود.

واحدهای پوند، گرم، کیلوگرم، متر، سانتی‌متر و ... دارای مقیاس نسبتی بوده که همگی دارای صفر مطلق هستند.

خصوصیات:

- ✓ ترتیب دارد.
- ✓ فاصله را مشخص می‌کند.
- ✓ صفر قراردادی و مطلق دارد.
- ✓ نسبت هر دو عدد، مستقل از واحد و مستقل از نقطه صفر است. برای مثال:

$$\frac{10\text{cm}}{5\text{cm}} = 2 \ , \ \frac{10\text{km}}{5\text{km}} = 2$$

کاربرد: به عنوان قوی‌ترین مقیاس برای اندازه‌گیری متغیرهای کمّی به‌کار می‌رود.

رابطه بین مقیاس‌ها

متغیرهای کمّی — متغیرهای کیفی

مقیاس نسبتی >> مقیاس فاصله‌ای >> مقیاس ترتیبی >> مقیاس اسمی

↓ قوی‌ترین مقیاس — ↓ ضعیف‌ترین مقیاس

مراتب / مقیاس	ترتیب	فواصل	صفر قراردادی	صفر مطلق (واقعی)
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه‌ای	دارد	ندارد	ندارد	ندارد
فاصله‌ای	دارد	دارد	دارد	ندارد
نسبی	دارد	دارد	دارد	دارد

مثال ۱ کدام‌یک از مقیاس‌های زیر صفر مطلق دارد؟

۱) نسبی ۲) اسمی ۳) فاصله‌ای ۴) رتبه‌ای

حل: گزینه ۱ درست است.

مثال ۲ کدام مقیاس از ویژگی‌های بهتری برای اندازه‌گیری برخوردار است؟ (حسابداری ـ ۸۳ ، ۸۴)

۱) نسبی ۲) فاصله‌ای ۳) اسمی ۴) رتبه‌ای

حل: گزینه ۱ درست است.

مثال ۳ وزن محصولات یک کارخانه دارای چه نوع مقیاسی است؟

۱) رتبه‌ای ۲) نسبی ۳) اسمی ۴) فاصله‌ای

حل: گزینه ۲ درست است.

مراحل تحقیقات آماری

مجموعه مراحل اساسی هر پژوهش آماری یا تحقیق علمی را می‌توان به ترتیب زیر بررسی کرد:

مرحله اول: هدف‌گذاری

مرحله دوم : جمع‌آوری داده‌ها

مرحله سوم: تجزیه و تحلیل داده‌ها

مرحله چهارم: بیان یافته‌ها

مثال اولین مرحله در هر تحقیق علمی یا پژوهش آماری کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) جمع‌آوری داده‌ها ۲) فرضیه‌سازی

۳) هدف‌گذاری ۴) تحلیل یافته‌ها

حل: گزینه ۳ درست است.

عناصر اصلی در پژوهش‌های مدیریتی

دو عنصر اصلی در پژوهش‌های مدیریتی وجود دارد:

۱ـ فرضیه‌های تحقیق: این فرضیه‌ها از پژوهش‌ها و نظریات گذشته استنتاج می‌شود و راهنمای مطالعات جدید است.

۲ـ متغیرهای لازم برای آزمایش فرضیه‌ها: این متغیرها امکان اندازه‌گیری و مشاهده فرضیه‌های تحقیق را برای پژوهشگران فراهم می‌کنند.

متغیرهای لازم برای پژوهش

الف) متغیر خصیصه: متغیری که مقدار آن ممکن است از یک عضو به عضو دیگر تغییر کند، متغیر خصیصه نامیده می‌شود.

ب) متغیر وابسته: متغیری که معلول احتمالی یا فرضی است، متغیر وابسته، پاسخ یا برونداد نامیده می‌شود.

ج) متغیر مستقل: متغیری که علت احتمالی یا فرضی متغیر وابسته است، متغیر مستقل، محرک یا درونداد نامیده می‌شود.

در حالت کلی متغیر مستقل را به عنوان متغیری تعریف می‌کنند که توسط محقق انتخاب می‌شود، تحت کنترل درمی‌آید و مورد سنجش قرار می‌گیرد.

د) متغیر تعدیل‌کننده: متغیری که بخواهیم تأثیر آن را در متغیر مستقل و وابسته ببینیم، متغیر تعدیل‌کننده نامیده می‌شود.

متغیر تعدیل‌کننده در تحقیق اعمال می‌شود و ارتباط بین متغیر مستقل و متغیر وابسته را تعدیل می‌کند.

ه) متغیر کنترل: متغیری که تأثیرش در پژوهش‌ها خنثی شده و از بین می‌رود، متغیر کنترل نامیده می‌شود.

تفاوت متغیرهای تعدیل‌کننده و کنترل

در متغیرهای تعدیل‌کننده تأثیر متغیر همواره مورد مطالعه و بررسی قرار می‌گیرد، اما در متغیرهای کنترل لازم است تأثیر متغیرها در مطالعه خنثی شود و از بین برود.

مثال کدام‌یک از متغیرهای زیر از این لحاظ که تأثیرات آن‌ها از بین می‌رود با متغیرهای تعدیل‌کننده تفاوت دارد؟

۱) مستقل ۲) وابسته ۳) پاسخ ۴) کنترل

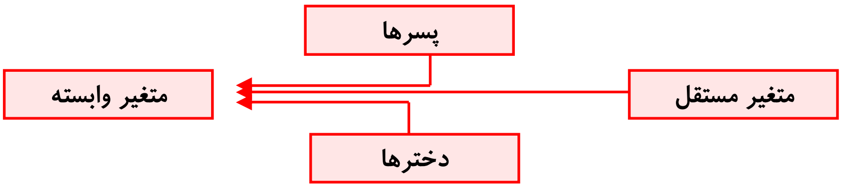
حل: گزینه ۴ درست است.

مثال مفهومی: برای درک بیشتر متغیرها بهتر است به مثال زیر توجه کنید:

فرض کنید می‌خواهیم تأثیر محل سکونت را بر معدل دانشجویان بررسی کنیم. برای این‌کار معدل دانشجویان روزانه رشته مدیریت را می‌یابیم. می‌دانیم که عده‌ای از دانشجویان در خوابگاه و عده‌ای خارج از خوابگاه زندگی می‌کنند و همچنین عده‌ای از دانشجویان دختر و عده‌ای پسر هستند. در این تحقیق متغیر **محل سکونت** دانشجویان (خوابگاهی، غیرخوابگاهی)، **متغیر مستقل** است و متغیری که مستقیماً تحت تأثیر متغیر مستقل است یعنی «**معدل دانشجویان**»، **متغیر وابسته** است.

متغیر مستقل **(محل سکونت)** **[علت]**		**متغیر وابسته** **(معدل)** **[معلول]**

متغیری که بنا بر فرض، اثر متغیر مستقل بر متغیر وابسته را تغییر می‌دهد، یعنی **جنسیت** (دختر و پسر) متغیر **تعدیل‌کننده** است؛ مثلاً ممکن است محل سکونت در معدل دختران تأثیر بیشتری داشته باشد تا در معدل پسران.



اما متغیری که در طول تحقیق ثابت نگه داشته می‌شود و تأثیرش در طول تحقیق خنثی است، یعنی **دانشجوی روزانه رشته مدیریت** بودن، متغیر **کنترل** است.

توجه کنید که در این تحقیق، **معدل** دانشجویان متغیر **خصیصه** نیز است.

تفاوت و شباهت متغیر تعدیل‌کننده با متغیر مستقل

متغیر تعدیل‌کننده بر دستاورد تحقیق تأثیر می‌گذارد، درست همان‌طورکه متغیر مستقل چنین عمل می‌کند. از این نظر، متغیر تعدیل‌کننده شبیه متغیر مستقل است، اما متغیر تعدیل‌کننده کنترل ندارد؛ یعنی محقق می‌تواند متغیر مستقل را تحت کنترل درآورد، اما متغیر تعدیل‌کننده را نمی‌تواند. برای مثال، محقق نمی‌تواند جنسیت دانشجویان را تحت کنترل درآورد؛ بنابراین متغیر تعدیل‌کننده شبیه متغیر مستقل است؛ زیرا بر نتیجه تحقیق تأثیر می‌گذارد و با متغیر مستقل تفاوت دارد زیرا نمی‌تواند توسط محقق کنترل شود.

گروه گواه (شاهد) و گروه آزمایش

بهتر است با طرح یک مثال این دو گروه را معرفی کنیم. فرض کنید که می‌خواهیم تأثیر کلاس تقویتی ریاضی را در دانش‌آموزان یک کلاس بررسی کنیم. برای این کار دانش‌آموزان کلاس را به دو گروه مساوی تقسیم می‌کنیم و سپس گروهی را به کلاس تقویتی می‌فرستیم و گروهی دیگر را ثابت نگه داشته و به کلاس تقویتی نمی‌فرستیم. گروهی که به کلاس تقویتی رفته‌اند درواقع مورد آزمایش واقع شده‌اند و گروهی که ثابت نگه داشته شده و به کلاس تقویتی نرفته‌اند، گروه گواه و یا شاهد هستند.

ما از این طریق می‌توانیم تغییرات گروه آزمایش‌شده را با گروه گواه که از ابتدا ثابت مانده و تغییری روی آن اعمال نشده، مشاهده کنیم.

توجه کنید که دو گروه آزمایش و گواه کاملاً **مستقل** از یکدیگرند.

مثال در بررسی اثربخشی یک دوره آموزش مدیریت، از یک گروه گواه و یک گروه آزمایش استفاده شده است. گروه فرضیه‌های این نوع تحقیق چگونه‌اند؟ (حسابداری و مدیریت ـ ۸۵)

۱) همبسته ۲) مستقل ۳) جورشده ۴) توصیفی

حل: گزینه ۲ درست است.

داده‌های آماری (Statistics Data)

اندازه‌های صفت متغیر عناصر جامعه یا نمونه آماری که با استفاده از اندازه‌گیری، آزمایش، مشاهده و ... به دست می‌آیند، داده‌های آماری نامیده می‌شوند.

داده‌های آماری در دو گروه طبقه‌بندی شده و طبقه‌بندی نشده قرار می‌گیرند.

۱ـ داده‌های طبقه‌بندی نشده

تعریف: داده‌های خام (اولیه) $x_1, x_2, \ldots, x_N$ که از طریق اندازه‌گیری صفت متغیر عناصر جامعه به دست می‌آید و گروه‌بندی نشده‌اند، داده‌های طبقه‌بندی نشده نامیده می‌شوند.

برای مثال، سن 7 نفر از اهالی یک محله 10 , 20 , 15,10 , 10 , 30 , 10 است؛ این داده‌ها طبقه‌بندی نشده‌اند.

۲ـ داده‌های طبقه‌بندی شده

در طبقه‌بندی مشاهدات، عوامل مختلفی از جمله تعداد مشاهدات و تجربیات آمارگر تأثیر دارند. یک روش تقریبی برای طبقه‌بندی مشاهدات به شرح زیر است:

۱ـ دامنه تغییرات مشاهدات (R) به روش زیر محاسبه می‌شود:

$$R = \max(x_i) - \min(x_i) = \text{بزرگ‌ترین مشاهده} - \text{کوچک‌ترین مشاهده}$$

۲ـ تعداد طبقات (k) با یکی از دو روش تجربی زیر محاسبه می‌شود: (N = تعداد مشاهدات)

$$1)\ k = 1 + 3.32 \log N$$

$$2)\ k = \sqrt{N}$$

دقت کنید که هر دو رابطه تقریبی هستند و بر یکدیگر برتری ندارند.

۳ـ فاصله طبقات (I) به روش زیر محاسبه می‌شود:

$$I = \frac{R}{k} = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}}$$

اگر حاصل تقسیم بالا اعشاری شود، برای راحتی محاسبات، آن را به نزدیک‌ترین عدد صحیح گرد می‌کنیم. (بالاتر از 0.5 به عدد صحیح بعدی و پایین‌تر از 0.5 به عدد صحیح قبلی گرد می‌شود.)

مثال اگر حداقل و حداکثر مشاهدات به ترتیب 200 و 400 و فاصله طبقات 25 باشد، تعداد طبقات جدول طبقه‌بندی داده‌ها کدام است؟

۱) 16　　۲) 8　　۳) 5　　۴) 25

حل: گزینه ۲ درست است.

$$\begin{cases} (\text{فاصله طبقات})\ I = \dfrac{(\text{دامنه تغییرات})\ R}{(\text{تعداد طبقات})\ k} \quad \rightarrow \quad k = \dfrac{R}{I} = \dfrac{200}{25} = 8 \\ R = \max(x_i) - \min(x_i) = 400 - 200 = 200 \end{cases}$$

حدود طبقات

چنانچه مشاهدات جامعه به k طبقه با فاصله I تقسیم شده باشند، آن‌گاه می‌توان طبقات را به صورت زیر نشان داد:

$C-L$ (حدود طبقات)	L_1-U_1	$\dots$	L_i-U_i	$\dots$	L_k-U_k

در هر طبقه مانند طبقه iام حدود طبقه شامل L_i (حد پایین طبقه) و U_i (حد بالای طبقه) است و به صورت $\left[L_i-U_i\right)$ در نظر گرفته می‌شود.

همان‌طور که دیده می‌شود حد پایین طبقه، بسته و حد بالای طبقه، باز است. درنتیجه اطلاعات مربوط به هر طبقه شامل حد پایین طبقه $\left(L_i\right)$ تا قبل از حد بالای طبقه $\left(U_i\right)$ است.

برای مثال، سن افراد یک شرکت به صورت زیر طبقه‌بندی شده است:

$C-L$ (حدودطبقات)	$20-30$	$30-40$	$40-50$
تعداد	5 نفر	6 نفر	9 نفر

در این جدول $30>$ 5 نفر ≤ 20، $40>$ 6 نفر ≤ 30 و $50>$ 9 نفر ≤ 40 قرار دارند.

طبقات گسسته و پیوسته

اگر در طبقات متوالی یک جدول، حد بالای هر طبقه همان حد پایین طبقه بعد باشد **طبقات**، **پیوسته** و در غیر این صورت، **طبقات**، **گسسته** هستند.

برای مثال:

طبقات پیوسته	طبقات گسسته
$1-3\ ,\ 3-5\ ,\ 5-7$	$1-3\ ,\ 4-6\ ,\ 7-9$

پیوسته کردن طبقات

برای پیوسته کردن طبقات گسسته کافی است فاصله بین دو طبقه متوالی را بر 2 تقسیم کرده و به ترتیب آن را از حد پایین هر طبقه کم و به حد بالای طبقه اضافه کنیم.

برای مثال:

$$1-3\ ,\ 4-6\ ,\ 7-9 \xrightarrow{\frac{4-3}{2}=\frac{7-6}{2}=0.5} 0.5-3.5\ ,\ 3.5-6.5\ ,\ 6.5-9.5$$

✓ **دقت کنید!**

در محاسبه شاخص‌های آماری ابتدا باید طبقات پیوسته شوند، سپس شاخص مورد نظر بررسی شود.

فاصله طبقات، طول طبقات، عرض طبقات

$$\text{فاصله طبقات}: I=\frac{R\ \text{(دامنه تغییرات)}}{k\ \text{(تعداد طبقات)}}$$

طول طبقات (دسته) = تفاوت دو حد پایین یا دو حد بالا در دو طبقه متوالی

عرض طبقات = تفاوت حد بالای طبقه از حد پایین طبقه

برای مثال:

$$\text{طبقات گسسته}: 8-10\ ,\ 11-13 \longrightarrow \begin{cases} \text{عرض طبقات} = 10-8 = 13-11 = 2 \\ \text{طول طبقات} = 11-8 = 13-10 = 3 \\ \text{فاصله طبقات} = \dfrac{13-8}{2} = \dfrac{5}{2} = 2.5 \simeq 3 \end{cases}$$

$$\text{طبقات پیوسته}: 4-7\ ,\ 7-10 \longrightarrow \begin{cases} \text{عرض طبقات} = 7-4 = 10-7 = 3 \\ \text{طول طبقات} = 7-4 = 10-7 = 3 \\ \text{فاصله طبقات} = \dfrac{10-4}{2} = 3 \end{cases}$$

نتیجه:

۱ـ همواره «فاصله طبقات» = «طول طبقات» است.

۲ـ اگر طبقات پیوسته باشند، «فاصله طبقات» = «طول طبقات» = «عرض طبقات» است.

مثال اگر $90-99$ و $80-89$ دو طبقه متوالی از یک جدول طبقه‌بندی شده باشند، فاصله طبقات کدام است؟ (حسابداری ـ ۸۳)

۱) 9 ۲) 9.5 ۳) 10 ۴) هیچ‌کدام

حل: گزینه ۳ درست است.

$$\text{فاصله طبقات} = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}} = \frac{99-80}{2} = 9.5 \simeq 10$$

$$\text{طول طبقات} = \begin{cases} \text{تفاضل حد پایین دو طبقه متوالی} = 90-80 = 10 \\ \text{یا} \\ \text{تفاضل حد بالای دو طبقه متوالی} = 99-89 = 10 \end{cases}$$

مرکز طبقات (متوسط طبقات) (Middle of Classes)

در طبقه‌بندی مشاهدات، سعی بر آن است اعضایی که در یک طبقه قرار می‌گیرند از نظر اندازه صفت متغیر مورد نظر تفاوت چندانی با هم نداشته باشند؛ به عبارت دیگر با کمی اغماض می‌توان اعضای یک طبقه را دارای یک اندازه مشترک دانست. حال اگر بخواهیم این اندازه مشترک را به دست آوریم بهترین اندازه‌ای که می‌توان به هر یک از اعضای طبقه نسبت داد، وسط طبقه است که **مرکز طبقه** نامیده می‌شود.

اگر یک جدول طبقه‌بندی شده با طبقاتی به صورت $L_i - U_i$ (حد بالای طبقه i − حد پایین طبقه i) مفروض باشد، آن‌گاه برای محاسبه مرکز هر یک از طبقات به صورت زیر عمل می‌کنیم:

$$\textbf{مرکز طبقه } i\textbf{ام}: x_i = \frac{(\text{حد بالای طبقه } i) + (\text{حد پایین طبقه } i)}{2} = \frac{L_i + U_i}{2}$$

به عبارت دیگر:

(مرکز طبقه) x_i	$\frac{L_1+U_1}{2}$	$\frac{L_2+U_2}{2}$	$\dots$	$\frac{L_k+U_k}{2}$
(حدود طبقات) $C-L$	L_1-U_1	L_2-U_2	$\dots$	L_k-U_k
(فراوانی مطلق) F_i	F_1	F_2	$\dots$	F_k

برای مثال:

(مرکز طبقه) x_i	$\frac{1+4}{2}=2.5$	$\frac{4+7}{2}=5.5$	$\frac{7+10}{2}=8.5$	$\frac{10+13}{2}=11.5$
$C-L$	$1-4$	$4-7$	$7-10$	$10-13$

(مرکز طبقه) x_i	$\frac{1+3}{2}=2$	$\frac{4+6}{2}=5$	$\frac{7+9}{2}=8$
$C-L$	$1-3$	$4-6$	$7-9$

طبقات با فواصل نامساوی

گاهی ممکن است مجموعه اطلاعات آماری طوری باشد که نتوان مسئله مساوی بودن فاصله‌های طبقات را در آن رعایت کرد. برای مثال، فرض کنید بخواهیم حقوق پرداختی به کارمندان شرکتی را در یک جدول توزیع فراوانی خلاصه کنیم و برای این جدول فاصله طبقه را 200 هزار تومان در نظر بگیریم. ممکن است طبقات را به صورت زیر تشکیل دهیم:

100 هزار تومان تا کمتر از 300 هزار تومان

300 هزار تومان تا کمتر از 500 هزار تومان

500 هزار تومان تا کمتر از 700 هزار تومان

حدود طبقات	$100-300$	$300-500$	$500-700$
فراوانی	21	62	17

ولی ازآنجاکه بخش وسیعی از کارمندان شرکت در طبقه 300 هزار تومان تا کمتر از 500 هزار تومان قرار می‌گیرند، این طبقه‌بندی، توزیع حقوق پرداختی را به درستی نشان نمی‌دهد. برای بهتر آشکار شدن توزیع حقوق پرداختی شاید ناچار باشیم فاصله طبقات را نامساوی در نظر بگیریم، مانند:

100 هزار تومان تا کمتر از 300 هزار تومان

300 هزار تومان تا کمتر از 350 هزار تومان

350 هزار تومان تا کمتر از 400 هزار تومان

400 هزار تومان تا کمتر از 450 هزار تومان

450 هزار تومان تا کمتر از 500 هزار تومان

500 هزار تومان تا کمتر از 700 هزار تومان

حدود طبقات	$100-300$	$300-350$	$350-400$	$400-450$	$450-500$	$500-700$
فراوانی	23	11	19	18	14	17

در مواردی نیز ممکن است یک یا چند مشاهده از توده مشاهدات آماری آن‌قدر فاصله بگیرند که در تنظیم جدول توزیع فراوانی مجبور شویم حد پایین طبقه اول یا حد بالای طبقه آخر را باز بگذاریم. برای مثال، فرض کنید بخواهیم درآمد کارکنان شرکتی را با فاصله طبقات 50 هزار تومان طبقه‌بندی کنیم. در این شرکت ممکن است مدیرانی وجود داشته باشند که حقوق آنان حدود 200 هزار تومان بیشتر از سایر کارکنان آن شرکت باشد. به کار بردن فواصل طبقاتی مساوی در ساختن جدول توزیع فراوانی حقوق پرداختی این شرکت باعث می‌شود که فراوانی 4 طبقه جدول برابر صفر شود:

حدود طبقات	$250-300$	$300-350$	$350-400$	$400-450$	$450-500$	$500-550$	$550-600$	$600-650$	$650-700$
فراوانی	25	16	18	0	0	0	0	4	3

برای اجتناب از ایجاد چنین حالتی، می‌توان حد بالای طبقه آخر را باز گذاشت و آن را به صورت زیر نوشت:

حدود طبقات	$250-300$	$300-350$	$350-400$	≥ 400
فراوانی	25	16	18	7

همان‌طور که دیده می‌شود، طبقات بیشتر یا مساوی 400 تعداد کمی از جامعه را دربرمی‌گیرند.

در جداول با حدود باز، ازآنجاکه حد پایین در طبقه اول، حد بالا در طبقه آخر وجود ندارد و یا هر دو وجود ندارند، نمی‌توان مرکز طبقه را برای طبقات ابتدا، انتها یا هر دو محاسبه کرد؛ درنتیجه، همان‌طور که در پایان فصل خواهیم دید، امکان محاسبه بعضی از شاخص‌ها برای این جداول وجود ندارد.

فراوانی (Frequency)

با توجه به آنکه جدول طبقه‌بندی بر اساس «تعداد مشاهدات هر طبقه (فراوانی مطلق)» یا «نسبت مشاهدات هر طبقه به کل مشاهدات (فراوانی نسبی)» یا ... تهیه شده باشد، انواع فراوانی را می‌توان به شرح زیر بررسی کرد:

فراوانی مطلق

تعریف: تعداد مشاهدات (دفعات تکرار) طبقه یا دسته iام از کل مشاهدات جامعه را فراوانی مطلق می‌نامند و با F_i نشان می‌دهند.

نکته:

(مجموع فراوانی‌های داده‌ها برابر با حجم کل جامعه است؛ k، تعداد طبقات یا دسته‌ها در جدول است.) $1)\ \sum_{i=1}^{k} F_i = N$

(فراوانی مطلق همواره عددی صحیح و بزرگ‌تر یا مساوی صفر است.) $2)\ F_i \ge 0$

فراوانی نسبی

تعریف: نسبت فراوانی مطلق هر طبقه یا دسته (F_i) به کل مشاهدات جامعه (N) را فراوانی نسبی می‌نامند و آن را با f_i نشان می‌دهند. $f_i = \frac{F_i}{N}$

نکته:

(مجموع فراوانی‌های نسبی داده‌ها یا طبقات برابر یک است.) $1)\ \sum_{i=1}^{k} f_i = f_1 + \dots + f_k = 1$

(فراوانی نسبی همواره مقداری بین صفر و یک است.) $2)\ 0 \le f_i \le 1$

درصد فراوانی نسبی

تعریف: حاصل‌ضرب فراوانی نسبی هر طبقه یا دسته در 100، درصد فراوانی نسبی نامیده می‌شود.

درصد فراوانی نسبی داده یا طبقه iام $= f_i \times 100$

فراوانی تجمعی

تعریف: برای هر دسته یا طبقه، مجموع فراوانی مطلق طبقه و طبقات قبلی را فراوانی تجمعی می‌نامند و آن را با F_{c_i} نشان می‌دهند.

$$F_{c_i} = F_1 + F_2 + \dots + F_i$$

نکته: فراوانی تجمعی طبقه آخر، برابر با حجم کل جامعه است (k، تعداد طبقات یا دسته‌ها در جدول است).

$$F_{c_k} = \sum_{i=1}^{k} F_i = N$$

فراوانی نسبی تجمعی

تعریف: برای هر دسته یا طبقه، نسبت تجمع مشاهدات طبقه و طبقات قبلی به کل مشاهدات را فراوانی نسبی تجمعی می‌نامند و آن را با f_{c_i} نشان می‌دهند. f_{c_i} از حاصل تقسیم فراوانی تجمعی طبقه به کل مشاهدات به دست می‌آید:

$$f_{c_i} = \frac{F_{c_i}}{N} \quad \rightarrow \quad f_{c_i} = \frac{F_1 + \ldots + F_i}{N} = \sum f_i$$

نکته: فراوانی نسبی تجمعی آخرین طبقه، یک است (k، تعداد طبقات یا دسته‌ها در جدول است).

$$f_{c_k} = \sum_{i=1}^{k} f_i = 1$$

برای مثال، به جدول زیر توجه کنید:

$C-L$	$0-5$	$5-10$	$10-15$	$15-20$	
F_i	10	20	30	40	$\rightarrow \sum_{i=1}^{4} F_i = N = 100$
F_{c_i}	10	30	60	$100 = N$	$\rightarrow$ حجم کل جامعه یا فراوانی تجمعی آخرین طبقه
f_i	$\frac{10}{100}$	$\frac{20}{100}$	$\frac{30}{100}$	$\frac{40}{100}$	$\rightarrow \sum_{i=1}^{k} f_i = \frac{10+20+30+40}{100} = 1$
f_{c_i}	$\frac{10}{100}$	$\frac{30}{100}$	$\frac{60}{100}$	$\frac{100}{100} = 1$	$\rightarrow$ فراوانی نسبی تجمعی آخرین طبقه

✓ **دقت کنید!**

در بعضی از منابع، فراوانی مطلق را با f_i، فراوانی تجمعی را با F_i و فراوانی نسبی را با rf نشان می‌دهند. تشخیص اینکه داده‌های جدول، چه نوعی از فراوانی را نشان می‌دهد، ساده است زیرا فراوانی نسبی عددی بین صفر و یک $\left(0 \le f_i \le 1\right)$ و فراوانی مطلق عدد صحیحی بزرگ‌تر یا مساوی صفر است.

مثال ۱ در جدول فراوانی تجمعی داده‌های دسته‌بندی شده زیر، اگر درصد فراوانی نسبی دسته وسط 24 باشد، فراوانی مطلق دسته چهارم کدام است؟ (برنامه ریزی شهری ـ ۸۶)

مرکز دسته	13	15	17	19	21
فراوانی تجمعی	5	14	a	41	50

۱) 14 ۲) 15 ۳) 16 ۴) 17

حل: گزینه ۲ درست است.

با توجه به نکته گفته‌شده، فراوانی تجمعی دسته آخر برابر حجم داده‌ها (N) است، بنابراین در این سؤال $N = 50$ است.

$$\begin{cases} f_i = \dfrac{F_i}{N} & \text{(فراوانی نسبی)} \\ f_i \times 100 = \dfrac{F_i}{N} \times 100 & \text{(درصد فراوانی نسبی)} \end{cases}$$

(فراوانی مطلق دسته سوم) $F_3 = 12 \quad \leftarrow \quad \frac{F_3}{50} \times 100 = 24 \quad \leftarrow \quad$ درصد فراوانی نسبی دسته وسط $= f_3 \times 100 = 24$

حال با داشتن F_3 می‌توان مقدار a (فراوانی تجمعی دسته سوم = F_{c_3}) را به دست آورد.

با توجه به اینکه $F_{c_i} = F_i + F_{c_{i-1}}$، فراوانی تجمعی هر دسته برابر است با فراوانی مطلق آن دسته به‌علاوه فراوانی تجمعی دسته قبل، داریم:

$$a = F_{c_3} = F_3 + F_{c_2} = 12 + 14 = 26$$

حال با کامل شدن ردیف فراوانی تجمعی جدول به راحتی ردیف فراوانی مطلق به دست می‌آید، زیرا:

$$F_i = F_{c_i} - F_{c_{i-1}}$$

$$F_4 = F_{c_4} - F_{c_3} = 41 - 26 = 15$$

دسته	اول	دوم	سوم	چهارم	پنجم
F_{c_i}	5	14	$\boxed{a = 26}$	41	50 = N
F_i	5	9	12	$\boxed{15}$	9
$f_i \times 100$			24		

مثال ۲ اگر 50 داده آماری در یک جدول توزیع فراوانی دسته‌بندی شده باشند و 20% داده‌ها در طبقه آخر قرار داشته باشند، فراوانی تجمعی دسته ماقبل آخر کدام است؟

۱) 10 ۲) 40 ۳) 45 ۴) 30

حل: گزینه ۲ درست است.

نکته:
۱) فراوانی تجمعی دسته آخر با « N : تعداد داده‌ها» برابر است. ($F_{c\,\text{آخر}} = N$)
۲) فراوانی تجمعی نسبی دسته آخر برابر 1 است. ($f_{c\,\text{آخر}} = 1$)
۳) $F_i = F_{c_i} - F_{c_{i-1}}$
۴) $f_i = f_{c_i} - f_{c_{i-1}}$

	آخر	ماقبل آخر	...	
$\sum F = N = 50$	$\boxed{10}$		...	F_i
	50	x	...	F_{c_i}
	$\boxed{0.2}$		...	f_i

راه حل اول:

$$\begin{cases} f_{\text{آخر}} = f_{c\,\text{آخر}} - f_{c\,\text{ماقبل آخر}} \rightarrow 0.2 = 1 - f_{c\,\text{ماقبل آخر}} \rightarrow f_{c\,\text{ماقبل آخر}} = 0.8 \\ f_c = \frac{F_c}{N} \rightarrow F_{c\,\text{ماقبل آخر}} = N \times f_{\text{ماقبل آخر}} = 50 \times 0.8 = 40 \\ F_{c\,\text{آخر}} = N = 50 \end{cases}$$

راه حل دوم:

$$\begin{cases} f_c = \dfrac{F_c}{N} \;\rightarrow\; f_{\text{آخر}} = 0.2 \;\rightarrow\; F_{\text{آخر}} = N \times 0.2 = 50 \times 0.2 = 10 \\ F_{\text{آخر}} = F_{c\,\text{آخر}} - F_{c\,\text{ماقبل آخر}} \;\rightarrow\; 10 = 50 - F_{c\,\text{ماقبل آخر}} \;\rightarrow\; F_{c\,\text{ماقبل آخر}} = 40 \\ N = F_{\text{آخر}} = 50 \end{cases}$$

مثال۳ 20 داده آماری که کمترین آن‌ها 12 و بیشترین آن‌ها 27 است، در 3 طبقه دسته‌بندی شده‌اند. اگر 15% داده‌ها از 22 بیشتر و 25% داده‌ها از 17 کوچک‌تر باشند، فراوانی مطلق طبقه وسط کدام است؟

۱) 17 ۲) 20 ۳) 15 ۴) 12

حل: گزینه ۴ درست است.

ابتدا باید جدول توزیع فراوانی را با توجه به اطلاعات مسئله رسم کنیم سپس ببینیم طبقه وسط در کجای جدول قرار دارد.

$$\begin{cases} \text{فاصله طبقات} = \dfrac{R}{k} = \dfrac{27-12}{3} = \dfrac{15}{3} = 5 \\ R : \text{دامنه تغییرات} = \text{بیشترین داده} - \text{کمترین داده} \\ k : \text{تعداد طبقات} \end{cases}$$

	12 − 17	17 − 22 (طبقه وسط)	22 − 27	
f_i	0.25	$\boxed{0.6}$	0.15	$\sum f_i = 1$

ازآنجاکه مجموع فراوانی نسبی داده‌ها برابر یک است $(\Sigma f_i = 1)$، داریم:

$$\begin{cases} f_i = \dfrac{F_i}{N} \;\rightarrow\; F_{\text{وسط}} = f_{\text{وسط}} \times N = 0.6 \times 20 = 12 \\ \sum f_i = 1 \;\rightarrow\; 0.25 + f_{\text{وسط}} + 0.15 = 1 \;\rightarrow\; f_{\text{وسط}} = 0.6 \end{cases}$$

مشخص‌کننده‌های عددی

مشخص‌کننده‌های عددی، معیارهایی هستند که برای مقایسه چند جامعه به کار می‌روند و به سه دسته تقسیم می‌شوند:

۱ـ معیارهای مرکزی

۲ـ معیارهای پراکندگی

۳ـ معیارهای پراکندگی نسبی

معیارهای مرکزی (Measures of Central)

یکی از مهم‌ترین مسایل در مطالعه و بررسی هر جامعه آماری، تعیین مقادیری است که نماینده (مرکز) جامعه بوده و داده‌ها حول آن توزیع شده‌اند.

تعریف: هر شاخص عددی را که نشانگر مرکز داده‌های جامعه است و مشاهدات جامعه اطراف آن قرار دارند، معیار مرکزی می‌نامند.

انواع معیارهای مرکزی

برای محاسبه مرکز جامعه‌های آماری شاخص‌های متعددی وجود دارد که بر اساس آن شاخص‌ها، به انواع مختلف معیارهای مرکزی می‌رسیم که عبارت‌اند از:

۱ـ مد (Mode)

۲ـ میانه (Median)

۳ـ چندک (Quantile)

۴ـ میانگین (Mean)

مد (نما) (Mode)

تعریف: در میان مشاهدات هر جامعه، داده‌ای (x_i) که بیشترین فراوانی (تکرار) را داشته باشد، مد (نما) می‌نامند و آن را با Mo نشان می‌دهند.

✓ **دقت کنید!**

داده‌ای که بیشترین تکرار را داشته باشد، مطمئناً بیشترین فراوانی مطلق (F_i) و درنتیجه بیشترین فراوانی نسبی (f_i) را خواهد داشت.

مثال ۱ مد (نمای) مربوط به مشاهدات $7,4,13,5,7,8,7,8$ کدام است؟

۱) 7 ۲) 8 ۳) $8,7$ ۴) وجود ندارد.

حل: گزینه ۱ درست است.

در این مشاهدات، داده $x_i = 7$ بیشترین تکرار را دارد (3 بار) بنابراین $Mo = 7$.

مثال ۲ مد (نما) مربوط به مشاهدات جدول زیر کدام است؟

x_i	-1	0	1	2	
F_i	30	40	20	10	$N = \sum F_i = 100$
f_i	0.3	0.4	0.2	0.1	$\sum f_i = 1$

۱) -1 ۲) 0

۳) 1 ۴) 2

حل: گزینه ۲ درست است.

در این جدول داده $x_i = 0$ بیشترین فراوانی مطلق و درنتیجه بیشترین فراوانی نسبی را دارد $Mo = 0$.

مشخصات مد (نما)

۱ــ شاخص مرکزی مد (نما) لزوماً منحصربه‌فرد نیست، اگر چند داده بیشترین تکرار را داشته باشند، همه آن داده‌ها به عنوان مد در نظر گرفته می‌شوند.

۲ــ در شرایطی که تمام داده‌ها (x_i ها) به یک اندازه تکرار شده باشند (فراوانی‌های مطلق یا نسبی یکسان داشته باشند)، آن‌گاه جامعه فاقد شاخص مرکزی مد (Mo) خواهد بود.

۳ــ هرگاه تمام داده‌ها را در یک عدد ثابت ضرب یا بر یک عدد ثابت تقسیم کنیم و همچنین با یک عدد ثابت جمع و یا از یک عدد ثابت کم کنیم، مد داده‌ها نیز به همان نسبت تغییر می‌کند.

$$x_1,\dots,x_n \to ax_i \pm b \to Mo(ax_i \pm b) = aMo_x \pm b$$

خاصیت مهم مد (نما)

ازآنجاکه مد مشاهده‌ای است با بیشترین فراوانی، هنگامی که به عنوان نماینده و تخمین کمیت تک‌تک مشاهدات آماری مورد استفاده قرار گیرد، تعداد خطاها در تخمین، حداقل است؛ به عبارت دیگر، اگر N_e تعداد تخمین‌های اشتباه باشد، N_e مینیمم است.

برای مثال، داده‌های $2,2,9,9,9,9,5$ را در نظر بگیرید:

اگر به‌جای تمام داده‌ها مقدار 2 را قرار دهیم، در تخمین 5 داده اشتباه کرده‌ایم.

اگر به‌جای تمام داده‌ها مقدار 9 را قرار دهیم، در تخمین 3 داده اشتباه کرده‌ایم.

اگر به‌جای تمام داده‌ها مقدار 5 را قرار دهیم، در تخمین 6 داده اشتباه کرده‌ایم.

مشاهده می‌شود که با قرار دادن 9 (مد) به جای سایر داده‌ها، تعداد تخمین‌های اشتباه، حداقل است.

مثال۳ مد (نما) مربوط به مشاهدات $7\,,2\,,9\,,2\,,7\,,2\,,4\,,3\,,7\,,3$ کدام است؟

۱) 2 ۲) 7 ۳) 7,2 ۴) وجود ندارد.

حل: گزینه ۳ درست است.

در این مشاهدات، داده‌های $x_i = 7,2$ بیشترین تکرار را نسبت به بقیه داده‌ها دارند (3 بار)، بنابراین $Mo = 2,\ 7$.

مثال۴ مد (نما) مربوط به مشاهدات $4,7,4,2,7,4,2,2,7$ کدام است؟

۱) 7 ۲) 3 ۳) 4 ۴) وجود ندارد.

حل: گزینه ۴ درست است.

داده‌های $7,4,2$ هر یک 3 بار تکرار شده‌اند و فراوانی یکسان دارند، بنابراین مد وجود ندارد؛ یعنی $Mo = \varnothing$.

مثال۵ برای داده‌های $X_1,X_2,\dots,X_n$ مقدار مد برابر 10 است. اگر هر یک از داده‌ها را نصف کرده و از هر کدام 1 واحد کم کنیم، مقدار مد چقدر می‌شود؟

۱) 5 ۲) 10 ۳) 9 ۴) 4

حل: گزینه ۴ درست است.

$$Mo\left(\frac{1}{2}X_i - 1\right) = \frac{1}{2}Mo(X_i) - 1 = \frac{1}{2}(10) - 1 = 4$$

کاربرد مد (نما)

هرگاه در یک جامعه بیشترین فراوانی (تکرار)، معیار سنجش و انتخاب باشد (نظرسنجی، رأی‌گیری برای انتخابات و ...)، آن‌گاه از شاخص مرکزی مد (نما) استفاده می‌شود.

مثال ۶ نظر گروهی از سوادآموزان راجع به زمان پخش برنامه نهضت سوادآموزی از سیمای جمهوری اسلامی جمع‌آوری شده است. کدام شاخص مرکزی برای آن داده‌ها مناسب‌تر است؟ (اقتصاد ـ ۷۳)

۱) میانگین ۲) میانه ۳) نما ۴) چارک اول

حل: گزینه ۳ درست است.

با توجه به آنکه معیار انتخاب، زمان پخش برنامه بر اساس بیشترین درخواست سوادآموزان است، مناسب‌ترین معیار مرکزی برای درخواست‌ها، شاخص مد (نما) است.

محاسبه مد در داده‌های طبقه‌بندی شده

در داده‌های طبقه‌بندی شده با داشتن حدود واقعی طبقات، مد به صورت زیر محاسبه می‌شود.

الف) انتخاب طبقه‌ای که بیشترین فراوانی مطلق $\left(F_i\right)$ یا فراوانی نسبی $\left(f_i\right)$ را دارد (طبقه مددار).

ب) محاسبه مقدار مد از رابطه زیر:

$$\boxed{Mo = \text{حد پایین طبقه مددار} + \frac{d_1}{d_1+d_2} \times \text{طول طبقه}}$$

$$\begin{cases} d_1 = F_i - F_{i-1} \ \text{یا}\ f_i - f_{i-1} = \text{فراوانی مطلق یا نسبی طبقه مد دار} - \text{فراوانی مطلق یا نسبی طبقه قبل} \\ d_2 = F_i - F_{i+1} \ \text{یا}\ f_i - f_{i+1} = \text{فراوانی مطلق یا نسبی طبقه مد دار} - \text{فراوانی مطلق یا نسبی طبقه بعد} \end{cases}$$

مثال ۱ مد در داده‌های جدول زیر کدام است؟

حدود طبقات	$2-4$	$4-6$	$6-8$	$8-10$
فراوانی	10	14	8	3

۱) 5.5 ۲) 4.8 ۳) 4 ۴) 5.2

حل: گزینه ۲ درست است.

الف) طبقه $(4-6)$ بیشترین فراوانی را دارد و طبقه مددار است، بنابراین مد عددی بین 4 تا 6 است.

ب) مقدار مد از رابطه زیر محاسبه می‌شود:

$$Mo = 4 + \frac{(14-10)}{(14-10)+(14-8)} \times 2 = 4.8$$

مثال ۲ مد در داده‌های جدول زیر کدام است؟

حدود طبقات	$2-4$	$5-7$	$8-10$	$11-13$
فراوانی	10	14	8	3

۱) 5.2 ۲) 5.7 ۳) 4.5 ۴) 6.5

حل: گزینه ۲ درست است.

الف) طبقه $(5-7)$ بیشترین فراوانی را دارد و طبقه مددار است، اما حدود واقعی طبقه به علت گسسته بودن طبقات باید محاسبه شود:

$$\text{طبقه } (5-7) \xrightarrow[\frac{1}{2} \text{ واحد از حد پایین کم کرده و به حد بالا اضافه می‌شود.}]{\frac{5-4}{2} = \frac{8-7}{2} = \frac{11-10}{2} = \frac{1}{2}} (4.5-7.5)$$

با توجه به حدود واقعی طبقه، طول طبقه برابر 3 است.

ب) مقدار مد به از رابطه محاسبه می‌شود:

$$Mo = 4.5 + \frac{(14-10)}{(14-10)+(14-8)} \times 3 = 5.7$$

میانه (Median)

تعریف: اگر مشاهدات (x_i) در یک جامعه به صورت صعودی مرتب شده باشند، آن‌گاه داده‌ای را که در وسط داده‌ها قرار دارد (یعنی 50 درصد مشاهدات قبل از آن و 50 درصد مشاهدات بعد از آن قرار دارند)، میانه می‌نامند و آن را با Md یا Me نشان می‌دهند.

نکته: بنا بر تعریف، شاخص میانه (Md) جزیی از 50 درصد مشاهدات قبل از خود است.

به عبارت دیگر:

$$\underbrace{x_{min} \xleftrightarrow{\quad x_i \le Md \quad}}_{\text{50 درصد مشاهدات کمتر یا مساوی میانه (Md) هستند.}} Md \underbrace{\xleftrightarrow{\quad x_i > Md \quad} x_{max}}_{\text{50 درصد مشاهدات بیشتر از میانه (Md) هستند.}}$$

مثال اگر میانه نمره دانش‌آموزان یک کلاس 12 باشد، آن‌گاه نمرات 50 درصد دانش‌آموزان .... است.

۱) کمتر یا مساوی 12 ۲) مساوی 12 ۳) بیشتر از 12 ۴) گزینه‌های ۱ و ۳

حل: گزینه ۴ درست است.

اگر میانه نمرات دانش‌آموزان یک کلاس 12 باشد $(Md = 12)$، به آن مفهوم است که اگر نمرات دانش‌آموزان کلاس را به صورت صعودی مرتب کنیم، داریم:

$$x_{min} \xleftrightarrow[\le Md=12]{\frac{1}{2}=\%50} Md = 12 \xleftrightarrow[>Md=12]{\frac{1}{2}=\%50} x_{max}$$

50 درصد نمرات کلاس کمتر یا مساوی (حداکثر) 12 و 50 درصد بیشتر از 12 هستند.

✓ دقت کنید!

در بیان میانه فقط می‌توان از عبارت‌های زیر استفاده کرد:

۱ـ 50 درصد مشاهدات کمتر یا مساوی میانه هستند؛ به مفهوم حداکثر $(\le)$ است.

۲ـ 50 درصد مشاهدات بیشتر از میانه هستند $(>)$.

در بیان میانه به هیچ وجه نمی‌توان از عبارت‌های زیر استفاده کرد:

۱ـ 50 درصد مشاهدات بیشتر یا مساوی میانه هستند؛ بیشتر یا مساوی به مفهوم حداقل $(\ge)$ است.

۲ـ 50 درصد مشاهدات کمتر از میانه هستند $(<)$.

محاسبه میانه در داده‌های طبقه‌بندی نشده

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل میانه را از رابطه زیر به دست می‌آوریم (N = تعداد مشاهدات):

$$\text{محل میانه} = \frac{N}{2} + \frac{1}{2}$$

ج) داده متناظر با محل میانه را از بین داده‌های صعودی مرتب‌شده به دست می‌آوریم.

مثال میانه داده‌های زیر را محاسبه کنید.

الف) $6 , 7 , 9 , 0 , -1$ ب) $9 , 7 , 5 , 0 , 4 , -1$

حل:

الف)

مرتب‌سازی : $-1 , 0 , \overset{\text{سومی}}{\overbrace{6}} , 7 , 9$

محل میانه $= \frac{N}{2} + \frac{1}{2} = \frac{5}{2} + \frac{1}{2} = 3 \rightarrow$ مقدار میانه $= 6 \rightarrow Md = 6$

ب)

مرتب‌سازی : $-1 , 0 , 4 , 5 , 7 , 9$

محل میانه $= \frac{N}{2}+\frac{1}{2}=\frac{6}{2}+\frac{1}{2}=\frac{7}{2}=3.5$

مقدار میانه $= \underset{\text{سومی}}{4}+0.5 \quad (\underset{\text{چهارمی}}{5} - \underset{\text{سومی}}{4}) = 4.5 \rightarrow Md = 4.5$

نکته: اگر تعداد داده‌ها فرد باشد، میانه بعد از مرتب‌سازی، داده وسط است؛ اما اگر تعداد داده‌ها زوج باشد، میانه بعد از مرتب‌سازی، میانگین دو عدد وسط خواهد بود.

مثال قبل را مجدداً مرور می‌کنیم:

$-1 , 0 , 6 , 7 , 9$; تعداد داده‌ها فرد است. $\rightarrow$ میانه $= 6$

$-1 , 0 , 4 , 5 , 7 , 9$; تعداد داده‌ها زوج است. $\rightarrow$ میانه $= \frac{4 + 5}{2} = 4.5$

محاسبه میانه در داده‌ها با مرکز دسته (طبقه)

اگر مشاهدات با جدول فراوانی که حدود دسته‌ها در آن با مرکز دسته‌ها (x_i) مشخص شده است، نمایش داده شوند، برای محاسبه میانه به صورت زیر عمل می‌کنیم:

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل میانه $\frac{N}{2}+\frac{1}{2}$ است.

ج) برای پیدا کردن محل میانه از چپ به راست اولین مرکز دسته‌ای را انتخاب می‌کنیم که در آن $F_{c_i} \geq \frac{N}{2}+\frac{1}{2}$ باشد.

مثال ۱ میانه در جدول زیر کدام است؟

x_i	4	1	0	7
F_i	3	8	5	9

۱) 7.5 ۲) 6.5

۳) 4 ۴) 1

حل: گزینه ۴ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل میانه را تعیین می‌کنیم:

$N = 25 \rightarrow \frac{N}{2}+\frac{1}{2}=13$

	(اول تا 5ام)		(6ام تا 13ام)		
		↓		↓	
x_i	0	1	4	7	
F_i	5	8	3	9	$\sum F_i = N = 25$
F_{c_i}	5	13	16	25	

ج) اولین مرکز دسته‌ای که در آن $\frac{N}{2}+\frac{1}{2}=13$ $F_{c_i} \geq$ باشد، دسته $x=1$ است؛ بنابراین:

$$\text{محل میانه} = \frac{N}{2}+\frac{1}{2}=13 \xrightarrow[\text{عدد 1 است.}]{\text{داده 6ام تا 13ام}} Md=1$$

مثال ۲ میانه در جدول زیر کدام است؟

x_i	4	1	0	7
F_i	8	4	5	9

۱) 6.5 ۲) 4

۳) 0.5 ۴) 1

حل: گزینه ۲ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل میانه را تعیین می‌کنیم:

$$N=26 \rightarrow \frac{N}{2}+\frac{1}{2}=13.5$$

	(اول تا 5ام)	(6ام تا 9ام)	(10ام تا 17ام)		
	↓	↓	↓		
x_i	0	1	4	7	
F_i	5	4	8	9	$\sum F_i = N = 26$
F_{c_i}	5	9	17	26	

ج) اولین مرکز دسته‌ای که در آن $\frac{N}{2}+\frac{1}{2}=13.5$ $F_{c_i} \geq$ باشد، دسته $x=4$ است؛ بنابراین:

$$\text{محل میانه} = \frac{N}{2}+\frac{1}{2}=13.5 \xrightarrow[\text{عدد 4 است.}]{\text{داده 10ام تا 17ام}} Md=4$$

مثال ۳ میانه در جدول زیر کدام است؟

x_i	4	1	0	7
F_i	3	8	5	10

۱) 4 ۲) 0.5

۳) 2.5 ۴) 6

حل: گزینه ۳ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل میانه را تعیین می‌کنیم:

$$N=26 \rightarrow \frac{N}{2}+\frac{1}{2}=13.5$$

(14ام تا 16ام) (6ام تا 13ام) (اول تا 5ام)

x_i	0	1	4	7	
F_i	5	8	3	10	$\sum F_i = N = 26$
F_{c_i}	5	13	16	26	

ج) اولین مرکز دسته‌ای که در آن $F_{c_i} \geq \frac{N}{2}+\frac{1}{2}=13.5$ باشد، دسته $x=4$ است، بنابراین:

$$\text{محل میانه} = \frac{N}{2}+\frac{1}{2}=13.5 \xrightarrow[\text{داده 14ام عدد 4است}]{\text{داده 13ام عدد 1است}} Md = \underset{13\text{ام}}{1} + 0.5(\underset{14\text{ام}}{4} - \underset{13\text{ام}}{1}) = 2.5$$

تبصره: درصورتی‌که در جدول داده‌های دسته‌بندی شده با مرکز دسته، فراوانی نسبی مشاهدات داده شود، برای محاسبه میانه به صورت زیر عمل می‌کنیم:

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) فراوانی تجمعی نسبی دسته‌ها را به دست می‌آوریم.

ج) اولین مرکز دسته‌ای که فراوانی تجمعی نسبی آن بیشتر از $\frac{1}{2}=0.5$ باشد، $\left(f_{c_i} \geq 0.5\right)$ میانه داده‌ها است.

مثال ۴ میانه در جدول داده‌های زیر کدام است؟

x_i	4	1	0	7
f_i	0.3	0.1	0.4	0.2

۱) 4 ۲) 2.5

۳) 1 ۴) 0

حل: گزینه ۳ درست است.

الف) مرکز دسته‌ها را به ترتیب صعودی مرتب می‌کنیم.

ب) فراوانی تجمعی نسبی دسته‌ها را به دست می‌آوریم:

x_i	0	1	4	7
f_i	0.4	0.1	0.3	0.2
f_{c_i}	0.4	0.5	0.8	1

ج) اولین مرکز دسته‌ای که در آن $f_{c_i} \geq 0.5$ باشد، دسته $x=1$ است؛ بنابراین میانه، عدد یک $(Md=1)$ خواهد بود.

محاسبه میانه در داده‌های طبقه‌بندی شده

الف) فراوانی تجمعی $\left(F_{c_i}\right)$ جدول را محاسبه می‌کنیم.

ب) اولین طبقه‌ای (از چپ به راست) که فراوانی تجمعی‌اش بیشتر یا مساوی $\frac{N}{2}$ باشد $\left(F_{c_i} \geq \frac{N}{2}\right)$، طبقه میانه‌دار است $\left(L_i - U_i\right)$.

ج) مقدار میانه برابر است با:

$$\boxed{Md = L_i + \frac{\frac{N}{2} - F_{c_{i-1}}}{F_i} \times I}$$

L_i : حد پایین طبقه میانه‌دار

$F_{c_{i-1}}$: فراوانی تجمعی طبقه ماقبل میانه‌دار

F_i : فراوانی مطلق طبقه میانه‌دار

I : طول طبقات (فاصله طبقات)

توجه: اگر طبقات گسسته باشند، باید بعد از اجرای مرحله (ب)، طبقه میانه‌دار را پیوسته کرده، سپس مرحله (ج) را اجرا کنیم.

مثال ۱ میانه داده‌های جدول زیر کدام است؟ (مدیریت ـ ۷۳)

(فاصله طبقات) $C-L$	$20-29$	$30-39$	$40-49$
(فراوانی) F_i	3	6	7

۱) 34.6 ۲) 34.5
۳) 37.8 ۴) 37.3

حل: گزینه ۳ درست است.

الف) فراوانی تجمعی $\left(F_{c_i}\right)$ جدول را محاسبه می‌کنیم:

$C-L$	$20-29$	$30-39$	$40-49$	
F_i	3	6	7	$N=\sum F_i=16$
F_{c_i}	3	9	16	

ب) میانه در اولین طبقه‌ای است که $F_{c_i} \geq \frac{N}{2}=\frac{16}{2}=8$ باشد؛ بنابراین طبقه $(30-39)$ طبقه میانه‌دار است.

با توجه به گسسته بودن طبقات، ابتدا دسته دوم را پیوسته می‌کنیم $\leftarrow$ $(29.5-39.5)$

ج) مقدار میانه برابر است با:

$$Md=L_i+\frac{\frac{N}{2}-F_{c_{i-1}}}{F_i}\times I=29.5+\frac{8-3}{6}\times 10=37.8$$

دقت کنید که فاصله طبقات، وقتی که حدود طبقات پیوسته باشد، مورد نظر است.

مثال ۲ با توجه به جدول طبقه‌بندی شده زیر، مقدار میانه کدام است؟

$C-L$	$10-20$	$20-30$	$30-40$	$40-50$
f_i	0.1	0.4	0.3	0.2

۱) 25 ۲) 30
۳) 20 ۴) 40

حل: گزینه ۲ درست است.

الف) فراوانی تجمعی نسبی $\left(f_{c_i}\right)$ جدول را محاسبه می‌کنیم:

$C-L$	$10-20$	$20-30$	$30-40$	$40-50$	
f_i	0.1	0.4	0.3	0.2	$\sum f_i=N=1$
f_{c_i}	0.1	0.5	0.8	1	

ب) میانه در اولین طبقه‌ای است که $f_{c_i} \geq \frac{N}{2}=\frac{1}{2}$ باشد؛ بنابراین طبقه $(20-30)$ طبقه میانه‌دار است.

ج) مقدار میانه برابر است با:

$$Md=L_i+\frac{\frac{N}{2}-F_{c_{i-1}}}{F_i}\times I=20+\frac{\frac{1}{2}-0.1}{0.4}\times 10=30$$

نتیجه:

در شرایطی که مشاهدات جدول با فراوانی نسبی $\left(f_i\right)$ داده شوند، برای محاسبه میانه به همان روش اصلی عمل کرده و فقط به‌جای $\frac{N}{2}$ از $\frac{1}{2}$ استفاده می‌کنیم.

مثال۳ میانه توزیع آماری 40 مشاهده 32.5 است. اگر $I=5$ (فاصله طبقات) و فراوانی طبقه میانه‌دار 10 و مجموع فراوانی‌های ماقبل طبقه میانه‌دار 14 باشد، حدود کرانه طبقه میانه‌دار کدام است؟

۱) $29.5-34.5$ ۲) $29-39$ ۳) $30-40$ ۴) $35-39$

حل: گزینه ۱ درست است.

$$Md=L_i+\frac{\frac{N}{2}-F_{c_{i-1}}}{F_i}\times I \rightarrow 32.5=L_i+\frac{\frac{40}{2}-14}{10}\times 5 \rightarrow L_i=29.5$$

حد پایین طبقه : $L_i=29.5 \xrightarrow{I=5}$ حد بالای طبقه : $U_i=29.5+5=34.5$

مشخصات میانه

۱ـ هر جامعه آماری فقط یک میانه دارد (میانه برخلاف مد مانند میانگین منحصربه‌فرد است).

۲ـ اگر به بزرگ‌ترین عدد یک سری داده، عدد ثابتی را اضافه کنیم یا از کوچک‌ترین عدد یک سری داده، عدد ثابتی را کم کنیم، ازآنجاکه ترتیب داده‌ها تغییر نمی‌کند، این تغییرات هیچ تأثیری روی میانه ندارند و درنتیجه مقدار آن تغییری نمی‌کند.

$$\begin{cases}\boxed{5},8,12,17,\boxed{34}\\ Md=12\end{cases}\xrightarrow[x_{max}=34\xrightarrow{+7}41]{x_{min}=5\xrightarrow{-3}2}\begin{cases}\boxed{2},8,12,17,\boxed{41}\\ Md=12\end{cases}$$

همان‌طور که دیده می‌شود کاهش مقدار 3 از کوچک‌ترین مشاهده $(x_{min}=5)$ و اضافه کردن مقدار 7 به بزرگ‌ترین مشاهده $(x_{max}=34)$، تغییری در ترتیب داده‌ها ایجاد نکرده و مقدار میانه $(Md=12)$ نیز تغییر نمی‌کند.

مثال درصورتی‌که به بزرگ‌ترین عدد یک سری داده، عدد ثابتی را اضافه یا از کوچک‌ترین عدد یک سری داده، عدد ثابتی را کم کنیم، این افزایش یا کاهش بر کدام معیار تأثیر نمی‌گذارد؟ (اقتصاد ـ ۸۰)

۱) ضریب پراکندگی ۲) میانه ۳) میانگین ۴) واریانس

حل: گزینه ۲ درست است.

۳ـ طبق خاصیت ۲، مشاهده می‌شود که برخلاف مد (نما) که تابع فراوانی (بیشترین فراوانی) بود، میانه تابع ترتیب داده‌هاست؛ به عبارت دیگر، تا زمانی که تغییر در داده‌ها ترتیب داده‌ها را تغییر ندهد، مقدار میانه تغییر نمی‌کند.

۴ـ هرگاه کل داده‌ها را در عدد ثابتی ضرب یا بر عدد ثابتی تقسیم کنیم و یا با عدد ثابتی جمع یا از عدد ثابتی کم کنیم، میانه داده‌ها نیز به همان نسبت تغییر می‌کند.

$$x_1,\dots,x_n \rightarrow ax_i\pm b \rightarrow Md(ax_i\pm b)=aMd_x\pm b$$

مثال اگر میانه مشاهدات $x_6,\dots,x_2,x_1$ برابر 4 باشد، میانه مشاهدات زیر برابر است با:

$$-2x_6+3,\dots,-2x_2+3,-2x_1+3$$

۱) -8 ۲) 4 ۳) -5 ۴) 7

حل: گزینه ۳ درست است.

$$Md(-2x+3)=-2Md_x+3=-2\times 4+3=-5$$

خاصیت مهم میانه

مهم‌ترین خاصیت میانه می‌تواند به یکی از حالت‌های زیر بیان شود:

۱ـ مجموع **قدرمطلق** انحرافات (تفاضلات) از میانه همیشه **حداقل** است.

$$\sum |x_i - Md| = \min$$

۲ـ مجموع **قدرمطلق** انحرافات از **میانه** از هر مجموع **قدرمطلق** انحرافات نسبت به هر نقطه دلخواه $(a \neq Md)$ **کوچک‌تر** است.

$$\sum |x_i - Md| < \sum |x_i - a|$$

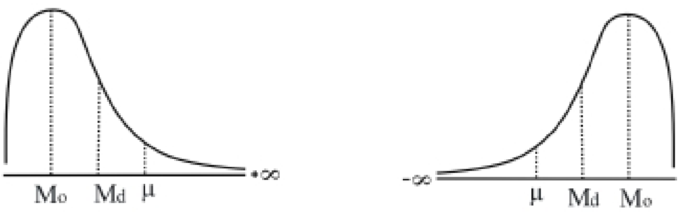
۳ـ هرگاه مجموع **قدرمطلق** انحرافات از نقطه دلخواهی (مانند c) حداقل باشد، آن‌گاه آن نقطه میانه $(c = Md)$ است و برعکس.

$$\sum |x_i - c| = \min \rightleftarrows c = Md$$

کاربرد میانه

در شرایط زیر استفاده از میانه (Md) به عنوان شاخص مرکزی مناسب‌تر به نظر می‌رسد:

۱ـ در توزیع‌های نامتقارن، از میانه برای نمایش اندازه تمایل توزیع‌ها به مرکز توزیع استفاده می‌شود؛ زیرا میانه نسبت به میانگین و مد (نما) کمتر تحت تأثیر مقادیر انتهای توزیع قرار می‌گیرد و همیشه در وسط قرار دارد.



۲ـ در توزیع‌هایی که تعداد اندکی مشاهده در ابتدا یا انتهای آن‌ها وجود دارد (اندازه داده‌هایی که در ابتدا و انتها توزیع واقع شده‌اند، با سایر مشاهدات به طور قابل ملاحظه‌ای تفاوت داشته باشند).

برای مثال، $N = 18$ مشاهده زیر را در نظر بگیرید:

$\boxed{-200}\ ,\ 2,2,2,3,3,4,4,4,5,5,5,6,6,6,6,\ \boxed{100}\ ,\ \boxed{400}$

همان‌طورکه دیده می‌شود داده‌های -200 (در ابتدا) و 100 و 400 (در انتها) فاصله زیادی با سایر مشاهدات دارند و تعدادشان نیز اندک است (3 مشاهده). در این شرایط **پارامتر مرکزی میانه نسبت به میانگین کمتر تحت تأثیر مقادیر ابتدایی و انتهایی** $(-200\ ,\ 100\ ,\ 400)$ **قرار می‌گیرد**. برای همین استفاده از میانه مناسب‌تر به نظر می‌رسد.

۳ـ در جدول داده‌های طبقه‌بندی شده زمانی که حدود ابتدا یا انتها باز (نامشخص) باشد.

برای مثال:

$C-L$	<2	$2-5$	$5-8$	≥ 8	
فراوانی	1	8	7	2	$\sum F_i = N = 18$

با توجه به داده‌های مثال قبل دیده می‌شود که تنها 1 مشاهده قبل از $x = 2$ وجود دارد (-200)، از این رو حدود ابتدا باز است، همچنین تنها 2 مشاهده بعد از $x = 8$ وجود دارد $(400{,}100)$، بنابراین حدود انتها نیز باز است.

✓ دقت کنید!

در جدول، امکان محاسبه مرکز طبقه (x_i) در طبقه اول و آخر وجود ندارد و نمی‌توان شاخص مرکزی میانگین $\left(\mu = \frac{\sum x_i}{N}\right)$ را محاسبه کرد؛ درنتیجه تنها شاخص مرکزی مناسب، میانه است.

چندک (Quantile)

چندک‌ها مقادیری از مشاهدات هستند که جامعه آماری را به فواصل مساوی با نسبت‌های $\frac{1}{4}$ (چارک)، $\frac{1}{10}$ (دهک) و $\frac{1}{100}$ (صدک) تقسیم می‌کنند به طوری که فراوانی در هر یک از این فواصل، درصد معینی از فراوانی کل را تشکیل می‌دهد.

چارک

اگر جامعه آماری را به 4 قسمت مساوی تقسیم کنیم، چارک‌های اول (Q_1)، دوم (Q_2) و سوم (Q_3) به وجود می‌آیند. بعد از مرتب کردن داده‌ها به صورت صعودی، داریم:

$$x_{min} \xleftrightarrow{\frac{1}{4}=0.25} Q_1 \xleftrightarrow{\frac{1}{4}=0.25} Q_2 \xleftrightarrow{\frac{1}{4}=0.25} Q_3 \xleftrightarrow{\frac{1}{4}=0.25} x_{max}$$

$Q_1 = \frac{1}{4}$ (چارک اول)
- مقداری که $\frac{1}{4} = \%25$ مشاهدات، کوچک‌تر یا مساوی $(\le)$ آن هستند.
- یا
- مقداری که $\frac{3}{4} = \%75$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

$Q_2 = \frac{2}{4}$ (چارک دوم = میانه)
- مقداری که $\frac{2}{4} = \%50$ مشاهدات، کوچک‌تر یا مساوی $(\le)$ آن هستند.
- یا
- مقداری که $\frac{2}{4} = \%50$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

$Q_3 = \frac{3}{4}$ (چارک سوم)
- مقداری که $\frac{3}{4} = \%75$ مشاهدات، کوچک‌تر یا مساوی $(\le)$ آن هستند.
- یا
- مقداری که $\frac{1}{4} = \%25$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

دهک

اگر جامعه آماری را به 10 قسمت مساوی تقسیم کنیم، دهک‌های اول (D_1) تا نهم (D_9) به وجود می‌آیند. بعد از مرتب کردن داده‌ها به صورت صعودی داریم:

$$x_{min} \xleftrightarrow{\frac{1}{10}=0.10} D_1 \xleftrightarrow{\frac{1}{10}=0.10} D_2 \;\ldots\; D_5 \;\ldots \xleftrightarrow{\frac{1}{10}=0.10} D_9 \xleftrightarrow{\frac{1}{10}=0.10} x_{max}$$

$D_1 = \frac{1}{10}$ (دهک اول)
- مقداری که $\frac{1}{10} = \%10$ مشاهدات، کوچک‌تر یا مساوی $(\le)$ آن هستند.
- یا
- مقداری که $\frac{9}{10} = \%90$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

$D_5=\frac{5}{10}$ (دهک پنجم = میانه)
- مقداری که $\frac{5}{10}=\%50$ مشاهدات، کوچکتر یا مساوی $(\leq)$ آن هستند.
- یا
- مقداری که $\frac{5}{10}=\%50$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

$D_9=\frac{9}{10}$ (دهک نهم)
- مقداری که $\frac{9}{10}=\%90$ مشاهدات، کوچک‌تر یا مساوی $(\leq)$ آن هستند.
- یا
- مقداری که $\frac{1}{10}=\%10$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

صدک

اگر جامعه آماری را به 100 قسمت مساوی تقسیم کنیم، صدک‌های اول (P_1) تا نود و نهم (P_{99}) به وجود می‌آیند. بعد از مرتب کردن داده‌ها به صورت صعودی داریم:

$$x_{min} \xleftrightarrow{\frac{1}{100}=\%1} P_1 \xleftrightarrow{\frac{1}{100}=\%1} P_2 \ \ldots\ P_{50}\ \ldots \xleftrightarrow{\frac{1}{100}=\%1} P_{99} \xleftrightarrow{\frac{1}{100}=\%1} x_{max}$$

$P_1=\frac{1}{100}$ (صدک اول)
- مقداری که $\frac{1}{100}=\%1$ مشاهدات، کوچک‌تر یا مساوی $(\leq)$آن هستند.
- یا
- مقداری که $\frac{99}{100}=\%99$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

$P_{50}=\frac{50}{100}$ (صدک پنجاهم = میانه)
- مقداری که $\frac{50}{100}=\%50$ مشاهدات، کوچکتر یا مساوی $(\leq)$ آن هستند.
- یا
- مقداری که $\frac{50}{100}=\%50$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

$P_{99}=\frac{99}{100}$ (صدک نود و نهم)
- مقداری که $\frac{99}{100}=\%99$ مشاهدات، کوچک‌تر یا مساوی $(\leq)$ آن هستند.
- یا
- مقداری که $\frac{1}{100}=\%1$ مشاهدات، بزرگ‌تر $(>)$ از آن هستند.

تطبیق چندک‌ها

صدک (P_a)	دهک (D_a)	چارک (Q_a)	صدک (P_a)	دهک (D_a)	چارک (Q_a)
$P_1 = \frac{1}{100}$	–	–	$P_{50} = \frac{50}{100}$	$D_5 = \frac{5}{10}$	$Q_2 = Md = \frac{50}{100}$
⋮	–	–	⋮	–	–
$P_{10} = \frac{10}{100}$	$D_1 = \frac{1}{10}$	–	$P_{60} = \frac{60}{100}$	$D_6 = \frac{6}{10}$	–
⋮	–	–	⋮	–	–
$P_{20} = \frac{20}{100}$	$D_2 = \frac{2}{10}$	–	$P_{70} = \frac{70}{100}$	$D_7 = \frac{7}{10}$	–
⋮	–	–	⋮	–	–
$P_{25} = \frac{25}{100}$	–	$Q_1 = \frac{25}{100}$	$P_{75} = \frac{75}{100}$	–	$Q_3 = \frac{75}{100}$
⋮	–	–	⋮	–	–
$P_{30} = \frac{30}{100}$	$D_3 = \frac{3}{10}$	–	$P_{80} = \frac{80}{100}$	$D_8 = \frac{8}{10}$	–
⋮	–	–	⋮	–	–
$P_{40} = \frac{40}{100}$	$D_4 = \frac{4}{10}$	–	$P_{90} = \frac{90}{100}$	$D_9 = \frac{9}{10}$	–
⋮	–	–	⋮	–	–
			$P_{99} = \frac{99}{100}$	–	–

مشاهده می شود که:

ـ بیشترین تطبیق درصد بین چندک‌ها، در میانه رخ می‌دهد $(Md = Q_2 = D_5 = P_{50})$.

ـ صدک شامل تمام چندک‌هاست.

مثال ۱ دهک چهارم حقوق کارمندان 125 هزار تومان است. این به آن مفهوم است که .... 125 هزار تومان حقوق می‌گیرند.

۱) 40 درصد کارمندان کمتر یا مساوی
۲) 60 درصد کارمندان بیشتر از
۳) 40 درصد کارمندان
۴) 60 درصد کارمندان حداقل

حل: گزینه‌های ۱ و ۲ درست هستند.

$$x_{min} \xleftarrow[\le 125000]{\frac{4}{10}=\%40} D_4 = 125000 \xleftarrow[>125000]{\frac{60}{100}=\%60} x_{min}$$

کمتر یا مساوی (حداکثر) بیشتر

مثال ۲ چارک اول و سوم قد دانش‌آموزان به ترتیب 1.20 و 1.50 متر است. این به آن مفهوم است که ....

۱) 25% دانش‌آموزان قدشان کمتر و یا مساوی 1.20 متر است.
۲) 50% دانش‌آموزان قدشان بین 1.20 و 1.50 متر است.
۳) 25% دانش‌آموزان قدشان بیشتر از 1.50 متر است.
۴) هر سه گزینه درست است.

حل: گزینه ۴ درست است.

$$x_{min} \xleftrightarrow{\frac{1}{4}} Q_1 = 1.20 \xleftrightarrow{\frac{1}{4}} Q_2 \xleftrightarrow{\frac{1}{4}} Q_3 = 1.50 \xleftrightarrow{\frac{1}{4}} x_{max}$$

جملات دیگری که می‌توانند درست باشند عبارت‌اند از:

75% دانش‌آموزان قدشان بیشتر از 1.20 متر است.

75% دانش‌آموزان قدشان کمتر یا مساوی 1.50 متر است.

محاسبه چندک‌ها در داده‌های طبقه‌بندی نشده

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل چندک مورد نظر را از روابط زیر به دست می‌آوریم (N : تعداد مشاهدات):

$$\text{محل چارک } a\text{ام} = \frac{aN}{4} + \frac{1}{2} \qquad a = 1,2,3$$

$$\text{محل دهک } a\text{ام} = \frac{aN}{10} + \frac{1}{2} \qquad a = 1,2,\ldots,9$$

$$\text{محل صدک } a\text{ام} = \frac{aN}{100} + \frac{1}{2} \qquad a = 1,2,\ldots,99$$

ج) داده متناظر با محل چندک را از بین داده‌های صعودی مرتب‌شده به دست می‌آوریم.

مثال چارک سوم، دهک ششم و صدک هفتاد و هشتم را در مشاهدات زیر محاسبه کنید.

73 , 10 , 15 , 30 , 84 , 45 , 73 , 90 , 100 , 20

حل:

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم:

تعداد مشاهدات : $N = 10$

$$10, 15, 20, 30, \underset{\text{پنجم}}{\underset{\downarrow}{45}}, \underset{\text{ششم}}{\underset{\downarrow}{73}}, \underset{\text{هفتم}}{\underset{\downarrow}{73}}, \underset{\text{هشتم}}{\underset{\downarrow}{84}}, \underset{\text{نهم}}{\underset{\downarrow}{90}}, 100$$

$$\text{محل چارک سوم} = \frac{3N}{4} + \frac{1}{2} = 8 \longrightarrow Q_3 = 84$$

$$\text{محل دهک ششم} = \frac{6N}{10} + \frac{1}{2} = 6.5 \longrightarrow D_6 = \underset{\text{ششم}}{\underset{\downarrow}{73}} + 0.5\ (\underset{\text{هفتم}}{\underset{\downarrow}{73}} - \underset{\text{ششم}}{\underset{\downarrow}{73}}) = 73$$

$$\text{محل صدک هفتاد و هشتم} = \frac{78N}{100} + \frac{1}{2} = 8.3 \longrightarrow P_{87} = \underset{\text{هشتم}}{\underset{\downarrow}{84}} + 0.3\ (\underset{\text{نهم}}{\underset{\downarrow}{90}} - \underset{\text{هشتم}}{\underset{\downarrow}{84}}) = 85.8$$

نکته: در بعضی منابع، محل چندک a ام از روابط زیر به دست می‌آید:

$$\text{محل چارک } a\text{ام} = \frac{aN}{4} + \frac{1}{4} \qquad a = 1,2,3$$

$$\text{محل دهک } a\text{ام} = \frac{aN}{10} + \frac{1}{10} \qquad a = 1,2,\ldots,9$$

$$\text{محل صدک } a\text{ام} = \frac{aN}{100} + \frac{1}{100} \qquad a = 1,2,\ldots,99$$

ممکن است محل چندک به‌دست‌آمده از این روابط با محل چندک به‌دست‌آمده از روابطی که در ابتدا مطرح شد، متفاوت باشد؛ در این صورت با توجه به گزینه‌ها باید رابطه مناسب را به کار برد.

محاسبه چندک‌ها در داده‌ها با مرکز دسته (طبقه)

اگر مشاهدات با جدول فراوانی نمایش داده شوند که حدود دسته‌ها در آن با مرکز دسته‌ها (x_i) مشخص شده است، برای محاسبه چندک‌ها به صورت زیر عمل می‌کنیم:

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل چارک، $\frac{1}{2}+\frac{aN}{4}$، محل دهک، $\frac{1}{2}+\frac{aN}{10}$ و محل صدک $\frac{1}{2}+\frac{aN}{100}$ است.

ج) برای تعیین کردن محل چارک، دهک، صدک از چپ به راست اولین مرکز دسته‌ای را انتخاب می‌کنیم که به ترتیب برای چارک، $F_{c_i} \geq \frac{aN}{4}+\frac{1}{2}$، برای دهک، $F_{c_i} \geq \frac{aN}{10}+\frac{1}{2}$ و برای صدک، $F_{c_i} \geq \frac{aN}{100}+\frac{1}{2}$ باشد.

مثال ۱ چارک اول در جدول داده‌های زیر کدام است؟

x_i	4	1	0	7
F_i	9	3	4	10

۱) 1
۲) 4
۳) 0
۴) 0.25

حل: گزینه ۱ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل چارک اول را تعیین می‌کنیم:

$$N=26 \quad \rightarrow \quad \frac{aN}{4}+\frac{1}{2}=\frac{1\times 26}{4}+\frac{1}{2}=7$$

	(داده اول تا 4ام)	(داده 5ام تا 7ام)			
x_i	0	1	4	7	
F_i	4	3	9	10	$\sum F_i = N = 26$
F_{c_i}	4	7	16	26	

ج) اولین مرکز دسته‌ای که در آن $F_{c_i} \geq \frac{aN}{4}+\frac{1}{2}=7$ باشد، دسته $x=1$ است؛ بنابراین:

$$\text{محل چارک اول} = \frac{aN}{4}+\frac{1}{2}=7 \xrightarrow[\text{عدد 1 است.}]{\text{داده 5ام تا 7ام}} Q_1=1$$

مثال ۲ دهک ششم در جدول داده‌های زیر کدام است؟

x_i	4	1	0	7
F_i	7	5	6	8

۱) 0.6
۲) 4
۳) 2.4
۴) 0

حل: گزینه ۲ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل دهک ششم را تعیین می‌کنیم:

$$N=26 \quad \to \quad \frac{aN}{10}+\frac{1}{2}=\frac{6\times 26}{10}+\frac{1}{2}=16.1$$

	(داده اول تا 6 ام)	(داده 7 ام تا 11 ام)	(داده 12ام تا 18ام)		
x_i	0	1	4	7	
F_i	6	5	7	8	$\sum F_i = N = 26$
F_{c_i}	6	11	18	26	

ج) اولین مرکز دسته‌ای که در آن $\frac{aN}{10}+\frac{1}{2}=16.1 \geq F_{c_i}$ باشد، دسته $x=4$ است؛ بنابراین:

$$\text{محل دهک ششم}=\frac{aN}{10}+\frac{1}{2}=16.1 \xrightarrow[\text{عدد 4است.}]{\text{داده 12ام تا 18ام}} D_6=4$$

مثال۳ صدک 36ام در جدول داده‌های زیر کدام است؟

x_i	4	1	0	7
F_i	16	5	9	10

۱) 0 ۲) 0.36

۳) 2.5 ۴) 3.7

حل: گزینه ۴ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) محل صدک سی و ششم را تعیین می‌کنیم:

$$N=40 \quad \to \quad \frac{aN}{100}+\frac{1}{2}=\frac{36\times 40}{100}+\frac{1}{2}=14.9$$

	(داده اول تا 9ام)	(داده 10ام تا 14ام)	(داده 15ام تا 30ام)		
x_i	0	1	4	7	
F_i	9	5	16	10	$\sum F_i = N = 40$
F_{c_i}	9	14	30	40	

ج) اولین مرکز دسته‌ای که در آن $\frac{aN}{100}+\frac{1}{2}=14.9 \geq F_{c_i}$ باشد، دسته $x=4$ است؛ بنابراین:

$$\text{محل صدک سی و ششم}=\frac{aN}{100}+\frac{1}{2}=14.9 \xrightarrow[\text{داده 15ام عدد 4است.}]{\text{داده 14ام عدد 1است.}} P_{36}=\underset{\text{چهاردهم}}{1}+0.9(\underset{\text{پانزدهم}}{4}-\underset{\text{چهاردهم}}{1})=3.7$$

تبصره: اگر در جدول داده‌های دسته‌بندی شده با مرکز دسته، فراوانی نسبی مشاهدات داده شود، برای محاسبه چندک‌ها به صورت زیر عمل می‌کنیم:

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) فراوانی تجمعی نسبی دسته‌ها را به دست می‌آوریم.

ج) اولین مرکز دسته‌ای که فراوانی تجمعی نسبی آن به ترتیب $f_{c_i} \geq \frac{a}{4}$، $f_{c_i} \geq \frac{a}{10}$ و $f_{c_i} \geq \frac{a}{100}$ باشد، چارک، دهک و صدک aام داده‌هاست.

مثال چارک سوم داده‌های زیر کدام است؟

x_i	4	1	0	7
f_i	0.3	0.1	0.4	0.2

۱) 0 ۲) 4

۳) 2.5 ۴) 1

حل: گزینه ۲ درست است.

الف) مرکز دسته‌ها را به صورت صعودی مرتب می‌کنیم.

ب) فراوانی تجمعی نسبی دسته‌ها را به دست می‌آوریم:

x_i	0	1	4	7	
f_i	0.4	0.1	0.3	0.2	$\sum f_i = 1$
f_{c_i}	0.4	0.5	0.8	1	

ج) اولین مرکز دسته‌ای که در آن $f_{c_i} \geq \frac{3}{4} = 0.75$ باشد، دسته $x = 4$ است؛ بنابراین $Q_3 = 4$ خواهد بود.

چندک‌ها در داده‌های طبقه‌بندی شده

الف) فراوانی تجمعی $\left(F_{c_i}\right)$ جدول را محاسبه می‌کنیم.

ب) اولین طبقه‌ای (از چپ به راست) که فراوانی تجمعی‌اش بیشتر یا مساوی $\frac{aN}{4}$ یا $\frac{aN}{10}$ یا $\frac{aN}{100}$ باشد $\left(F_{c_i} \geq \frac{aN}{4}, \frac{aN}{10}, \frac{aN}{100}\right)$، طبقه چارک یا دهک یا صدک‌دار است $\left(L_i - U_i\right)$.

ج) اگر طبقه مورد نظر پیوسته نبود، آن را پیوسته می‌کنیم.

د) مقدار **چارک a**ام برابر است با:

$$Q_a = L_i + \frac{\frac{aN}{4} - F_{c_{i-1}}}{F_i} \times I \qquad a = 1,2,3$$

L_i: حد پایین طبقه چارک‌دار

$F_{c_{i-1}}$: فراوانی تجمعی طبقه ماقبل چارک‌دار

F_i: فراوانی مطلق طبقه چارک‌دار

I: طول طبقات (فاصله طبقات)

مقدار **دهک a**ام برابر است با:

$$D_a = L_i + \frac{\frac{aN}{10} - F_{c_{i-1}}}{F_i} \times I \qquad a = 1,2,\ldots,9$$

L_i: حد پایین طبقه دهک‌دار

$F_{c_{i-1}}$: فراوانی تجمعی طبقه ماقبل دهک‌دار

F_i: فراوانی مطلق طبقه دهک‌دار

I: طول طبقات (فاصله طبقات)

مقدار **صدک a**ام برابر است با:

$$P_a = L_i + \frac{\frac{aN}{100} - F_{c_{i-1}}}{F_i} \times I \qquad a = 1,2,\ldots,99$$

L_i: حد پایین طبقه صدک‌دار

$F_{c_{i-1}}$: فراوانی تجمعی طبقه ماقبل صدک‌دار

F_i: فراوانی مطلق طبقه صدک‌دار

I: طول طبقات (فاصله طبقات)

مثال ۱ چارک اول داده‌های زیر کدام است؟

$C-L$	$0-6$	$6-12$	$12-18$	$18-24$
F_i	3	10	7	4

۱) 7.8 ۲) 12
۳) 1.8 ۴) 6.3

حل: گزینه ۱ درست است.

الف) F_{c_i} را در جدول محاسبه می‌کنیم:

$C-L$	$0-6$	$6-12$	$12-18$	$18-24$	
F_i	3	10	7	4	$\sum F_i = N = 24$
F_{c_i}	3	13	20	24	

ب) چارک در اولین طبقه‌ای است که $\frac{aN}{4} = \frac{1\times 24}{4} = 6$ $F_{c_i} \geq$ باشد؛ بنابراین طبقه $(6-12)$ طبقه چارک‌دار است.

ج) طبقه پیوسته است.

د) مقدار چارک اول برابر است با:

$$Q_a = L_i + \frac{\frac{aN}{4} - F_{c_{i-1}}}{F_i} \times I \xrightarrow[N=24]{a=1} Q_1 = 6 + \frac{\frac{1\times 24}{4} - 3}{10} \times 6 = 7.8$$

مثال ۲ چارک سوم جدول زیر کدام است؟ (حسابداری ـ ۷۸)

CL	$2-5$	$6-9$	$10-13$
F_i	10	30	20

۱) 6.8 ۲) 9.5
۳) 10 ۴) 10.5

حل: گزینه ۴ درست است.

الف) F_{c_i} را در جدول محاسبه می‌کنیم:

CL	$2-5$	$6-9$	$10-13$	
F_i	10	30	20	$N = \sum F_i = 60$
F_{c_i}	10	40	60	

ب) چارک در اولین طبقه‌ای است که $\frac{aN}{4} = \frac{3\times 60}{4} = \frac{180}{4} = 45$ $F_{c_i} \geq$ باشد؛ بنابراین طبقه $(10-13)$ طبقه چارک‌دار است.

ج) با توجه به گسسته بودن طبقات، ابتدا طبقه سوم را پیوسته می‌کنیم. $\leftarrow$ $(9.5-13.5)$

د) مقدار چارک سوم برابر است با:

$$Q_a = L_i + \frac{\frac{aN}{4} - F_{c_{i-1}}}{F_i} \times I \xrightarrow[N=60]{a=3} Q_3 = 9.5 + \frac{\frac{3\times 60}{4} - 40}{20} \times 4 = 10.5$$

دقت کنید که فاصله طبقات، وقتی حدود طبقات پیوسته باشد، مورد نظر است؛ بنابراین در این مثال $I = 13.5 - 9.5 = 4$ است.

مثال ۳ با توجه به جدول طبقه‌بندی شده زیر، مقدار دهک ششم کدام است؟

$C-L$	$0-6$	$6-12$	$12-18$	$18-24$
f_i	0.3	0.1	0.4	0.2

۱) 12.5 ۲) 15
۳) 11 ۴) 6

حل: گزینه ۲ درست است.

الف) فراوانی تجمعی نسبی $\left(f_{c_i}\right)$ را در جدول محاسبه می‌کنیم:

C − L	0 − 6	6 − 12	12 − 18	18 − 24	
f_i	0.3	0.1	0.4	0.2	$\sum f_i = N = 1$
f_{c_i}	0.3	0.4	0.8	1	

ب) دهک در اولین طبقه‌ای است که $0.6 = \frac{6 \times 1}{10} = \frac{aN}{10} \leq f_{c_i}$ باشد؛ بنابراین طبقه $(12 - 18)$ طبقه دهک‌دار است.

ج) طبقه پیوسته است.

د) مقدار دهک ششم برابر است با:

$$D_a = L_i + \frac{\frac{aN}{10} - f_{c_{i-1}}}{f_i} \times I \xrightarrow[N=1]{a=6} D_6 = 12 + \frac{\frac{6 \times 1}{10} - 0.4}{0.4} \times 6 = 15$$

دقت کنید درصورتی‌که مشاهدات جدول، همانند این مثال با فراوانی نسبی $\left(f_i\right)$ داده شوند، برای محاسبه چندک‌ها به همان روش اصلی عمل می‌کنیم، فقط به‌جای N عدد 1 را قرار می‌دهیم.

مثال ۴ متغیرهای پیوسته در جدول زیر گروه‌بندی شده‌اند. متغیر 80 درصدی داده‌ها کدام است؟ (مدیریت و حسابداری ـ ۸۹)

حدود دسته	12 − 15	15 − 18	18 − 21	21 − 24	24 − 27	27 − 30
فراوانی	5	10	9	11	8	7

۱) 25.875 ۲) 25.625 ۳) 26.125 ۴) 26.225

حل: گزینه ۱ درست است.

منظور از متغیر 80 درصدی همان صدک 80ام است؛ بنابراین:

الف) F_{c_i} را در جدول محاسبه می‌کنیم:

C − L	12 − 15	15 − 18	18 − 21	21 − 24	24 − 27	27 − 30	
F_i	5	10	9	11	8	7	$N = \sum F_i = 50$
F_{c_i}	5	15	24	35	43	50	

ب) صدک در اولین طبقه‌ای است که $40 = \frac{80 \times 50}{100} = \frac{aN}{100} \leq F_{c_i}$ باشد؛ بنابراین طبقه $(24 - 27)$ طبقه صدک‌دار است.

ج) طبقه پیوسته است.

د) مقدار صدک هشتادم برابر است با:

$$P_a = L_i + \frac{\frac{aN}{100} - F_{c_{i-1}}}{F_i} \times I \xrightarrow[N=50]{a=80} P_{80} = 24 + \frac{\frac{80 \times 50}{100} - 35}{8} \times 3 = 24 + \frac{5}{8} \times 3 = 25.875$$

مثال ۵ داده‌های آماری پیوسته یک پژوهش در جدول زیر گروه‌بندی شده‌اند. چند درصد داده‌ها کمتر از 36.5 است؟

حدود دسته	20 − 25	25 − 30	30 − 35	35 − 40	40 − 45	45 − 50
فراوانی	12	17	19	20	15	7

(۸۹ ـ GIS)

۱) 56 ۲) 60 ۳) 63 ۴) 64

حل: گزینه ۲ درست است.
ازآنجاکه درصد داده‌های کمتر از 36.5 مورد نظر است؛ 36.5 یک چندک است (میانه، چارک، دهک یا صدک) و چون صدک شامل تمام چندک‌هاست، 36.5 را صدک aام فرض می‌کنیم. درواقع، a درصد مشاهدات کمتر یا مساوی 36.5 هستند. برای به دست آوردن مقدار a، ابتدا F_{c_i} را در جدول محاسبه می‌کنیم:

$C-L$	$20-25$	$25-30$	$30-35$	$35-40$	$40-45$	$45-50$	
F_i	12	17	19	20	15	7	$N=\sum F_i=90$
F_{c_i}	12	29	48	68	83	90	

با توجه به آنکه مقدار صدک برابر با 36.5 است، طبقه صدک‌دار طبقه $(35-40)$ خواهد بود؛ بنابراین:

$$P_a = L_i + \frac{\frac{aN}{100} - F_{c_{i-1}}}{F_i} \times I \rightarrow 36.5 = 35 + \frac{\frac{a\times 90}{100} - 48}{20} \times 5 \rightarrow 1.5 = \frac{0.9a - 48}{4} \rightarrow a = 60$$

✓ **دقت کنید!**

خاصیت سوم میانه برای تمام چندک‌ها نیز صادق است؛ یعنی چندک‌ها با هرگونه تغییر در کل داده‌ها به همان میزان تغییر می‌کنند.

میانگین (Mean)

اصلی‌ترین و مهم‌ترین شاخص مرکزی که نشان‌دهنده نقطه تعادل و مرکز ثقل جامعه است، میانگین یک جامعه است. درواقع اگر مشاهدات جامعه (x_i) را در یک ردیف به صورت مرتب قرار دهیم، آن‌گاه میانگین، نقطه ثقل جامعه خواهد بود به‌گونه‌ای که جمع جبری گشتاور (تفاوت داده‌ها) نسبت به میانگین صفر خواهد شد.

$$\text{داده‌ها}: (1,2,3) \xrightarrow{\text{میانگین} = \frac{1+2+3}{3} = 2} \text{تفاوت داده‌ها از میانگین} = (1-2)+(2-2)+(3-2)=0$$

انواع میانگین

میانگین با توجه به مقیاس اندازه‌گیری مشاهدات، دارای انواع مختلفی به شرح زیر است:

۱ـ میانگین هارمونیک $(\bar{x}_H)$
۲ـ میانگین هندسی $(\bar{x}_G)$
۳ـ میانگین حسابی $(\mu , \bar{x})$

✓ **دقت کنید!**

۱ـ با توجه به مقدار حاصل از میانگین حسابی، هندسی و هارمونیک، صرف‌نظر از واحد اندازه‌گیری رابطه زیر بین آن‌ها همیشه وجود دارد:

$$\boxed{\bar{x}_H \leq \bar{x}_G \leq \bar{x}(\mu)}$$

به عبارت دیگر، از بین سه میانگین، میانگین حسابی از نظر کمیت همیشه بیشترین مقدار و میانگین هارمونیک همیشه کمترین مقدار را دارد و اگر تمام داده‌ها یکسان (برابر) باشند، هر سه میانگین با هم برابر می‌شوند.

۲ـ میانگین‌های «پیراسته»، «وینزوری» و «وزنی» جزیی از میانگین حسابی هستند و در آن بررسی می‌شوند.

مثال کدام‌یک از روابط زیر بین میانگین حسابی $(\bar{x})$ و میانگین هندسی $(\bar{x}_G)$ و میانگین هارمونیک $(\bar{x}_H)$ برقرار است؟
(مدیریت ـ ۷۱ ، اقتصاد ـ ۷۴)

۱) $\bar{x}_H < \bar{x}_G < \bar{x}$ ۲) $\bar{x}_G < \bar{x}_H < \bar{x}$ ۳) $\bar{x}_G < \bar{x} < \bar{x}_H$ ۴) $\bar{x} < \bar{x}_G < \bar{x}_H$

حل: گزینه ۱ درست است.

میانگین هارمونیک (Harmonic Mean)

برای محاسبه میانگین هارمونیک $\left(\bar{x}_H\right)$ در مشاهدات جامعه یا نمونه به صورت زیر عمل می‌کنیم:

الف) مشاهدات بدون فراوانی

اگر $x_1, x_2, \ldots, x_N$ $\left(x_i \neq 0\right)$ N مشاهده دلخواه باشند، میانگین هارمونیک $\left(\bar{x}_H\right)$ داده‌ها برابر است با:

$$\bar{x}_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

مثال ۱ میانگین هارمونیک داده‌های 3 , 4 , 4 , 6 کدام است؟

۱) 2 ۲) 4 ۳) 6 ۴) 8

حل: گزینه ۲ درست است.

$$\bar{x}_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4}} = \frac{4}{\frac{1}{3} + \frac{1}{4} + \frac{1}{4} + \frac{1}{6}} = 4$$

محاسبه میانگین هارمونیک با استفاده از میانگین حسابی

اگر مشاهدات $x_1, x_2, \ldots, x_N$ مفروض باشد، عکس این مقادیر به ترتیب $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_N}$ است. میانگین حسابی این مقادیر برابر است با:

$$\text{میانگین حسابی معکوس مقادیر} = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}}{N}$$

و میانگین هارمونیک برابر است با:

$$\bar{x}_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

همان‌طورکه مشاهده می‌شود، میانگین هارمونیک چند داده برابر است با عکس میانگین حسابی معکوس داده‌ها.

ب) مشاهدات با فراوانی

اگر مشاهدات $\left(x_i\right)$ دارای فراوانی مطلق $\left(F_i\right)$ یا فراوانی نسبی $\left(f_i\right)$ باشند، میانگین هارمونیک داده‌ها برابر است با:

$$\bar{x}_H = \frac{\sum F_i}{\frac{F_1}{x_1} + \frac{F_2}{x_2} + \ldots} = \frac{\sum f_i}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots}$$

درواقع اگر f_i یا F_i به عنوان وزن داده‌ها (x_i) در نظر گرفته شوند، می‌توانیم به‌جای هر دو از w_i استفاده کرده و به رابطه نهایی زیر برسیم:

$$\bar{x}_H=\frac{\sum w_i}{\frac{w_1}{x_1}+\frac{w_2}{x_2}+\dots}=\frac{\sum w_i}{\sum\frac{w_i}{x_i}}\rightarrow\begin{cases}w_i=F_i\ ;\ \bar{x}_H=\dfrac{N}{\sum\frac{F_i}{x_i}}\\ w_i=f_i\ ;\ \bar{x}_H=\dfrac{1}{\sum\frac{f_i}{x_i}}\end{cases}$$

مثال ۲ با توجه به جدول فراوانی زیر میانگین هارمونیک داده‌ها کدام است؟

x_i	1	2	3
$w_i=F_i$	4	6	9

۱) 2.5 ۲) 1.9
۳) 1.2 ۴) 0.7

حل: گزینه ۲ درست است.

$$\begin{cases}\bar{x}_H=\dfrac{\sum w_i}{\frac{w_1}{x_1}+\frac{w_2}{x_2}+\frac{w_3}{x_3}}=\dfrac{19}{\frac{4}{1}+\frac{6}{2}+\frac{9}{3}}=\dfrac{19}{10}=1.9\\ \sum w_i=\sum F_i=19\end{cases}$$

مثال۳ با توجه به جدول فراوانی (وزنی) زیر، میانگین هارمونیک داده‌ها کدام است؟

x_i	1	2	4
$w_i=f_i$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

۱) 4 ۲) 6
۳) 2 ۴) 10

حل: گزینه ۳ درست است.

$$\begin{cases}\bar{x}_H=\dfrac{\sum w_i}{\frac{w_1}{x_1}+\frac{w_2}{x_2}+\frac{w_3}{x_3}}=\dfrac{1}{\frac{\frac{1}{4}}{1}+\frac{\frac{1}{4}}{2}+\frac{\frac{1}{2}}{4}}=\dfrac{1}{\frac{1}{2}}=2\\ \sum w_i=\sum f_i=1\end{cases}$$

کاربرد میانگین هارمونیک

برای محاسبه میانگین مشاهداتی که دارای **مقیاس با واحد ترکیبی** هستند (مانند متر بر ثانیه، کیلومتر بر ساعت، دور در دقیقه، تعداد در دقیقه، ریال در لیتر)، از میانگین هارمونیک استفاده می‌شود.
نام‌های دیگر میانگین هارمونیک، همساز و توافقی معکوس است.

مثال ۱ اگر یک اتومبیل مسافتی را با سرعت 100 کیلومتر بر ساعت طی کرده باشد و با سرعت 80 کیلومتر بر ساعت همان مسیر را برگشته باشد، برای محاسبه سرعت متوسط اتومبیل در رفت و برگشت از چه میانگینی استفاده می‌شود؟

۱) هندسی ۲) هارمونیک ۳) حسابی ۴) پیراسته

حل: گزینه ۲ درست است.

مثال۲ اتومبیلی $\frac{1}{3}$ مسافتی را با سرعت 60 کیلومتر بر ساعت و بقیه مسافت را با سرعت 120 کیلومتر بر ساعت طی کرده است. سرعت متوسط اتومبیل چقدر بوده است؟

۱) 90 ۲) 84 ۳) 120 ۴) 100

حل: گزینه ۱ درست است.

اولاً، واحد مشاهدات $x_1 = 60$ و $x_2 = 120$، واحد ترکیبی کیلومتر بر ساعت است.

ثانیاً، به مشاهده $x_1 = 60$، وزنی معادل $w_1 = \frac{1}{3}$ و به مشاهده $x_2 = 120$، وزنی معادل $w_2 = \frac{2}{3}$ نسبت داده شده است؛ بنابراین:

$$\bar{x}_H = \frac{\sum_{i=1}^{2} w_i}{\frac{w_1}{x_1} + \frac{w_2}{x_2}} = \frac{\frac{1}{3} + \frac{2}{3}}{\frac{\frac{1}{3}}{60} + \frac{\frac{2}{3}}{120}} = 90$$

مثال۳ 2 اتومبیل مسیر 100 کیلومتری را با سرعت 120 کیلومتر بر ساعت طی کرده‌اند و هنگام برگشت، اتومبیل اول با سرعت 80 کیلومتر بر ساعت و اتومبیل دوم $\frac{1}{4}$ مسافت را با سرعت 60 و بقیه مسافت را با سرعت 90 کیلومتر بر ساعت برگشته است. سرعت متوسط اتومبیل‌ها چقدر بوده است؟

۱) 82 ۲) 80 ۳) 96 ۴) 100

حل: گزینه ۳ درست است.

اولاً، واحد مشاهدات $x_1 = 120$، $x_2 = 80$، $x_3 = 60$ و $x_4 = 90$، واحد ترکیبی کیلومتر بر ساعت است.

ثانیاً، به مشاهده $x_1 = 120$ وزنی معادل $w_1 = 2$، به مشاهده $x_2 = 80$ وزنی معادل $w_2 = 1$، به مشاهده $x_3 = 60$ وزنی معادل $w_3 = \frac{1}{4}$ و به مشاهده $x_4 = 90$ وزنی معادل $w_4 = \frac{3}{4}$، نسبت داده شده است؛ بنابراین:

$$\bar{x}_H = \frac{\sum_{i=1}^{4} w_i}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \frac{w_3}{x_3} + \frac{w_4}{x_4}} = \frac{2 + 1 + \frac{1}{4} + \frac{3}{4}}{\frac{2}{120} + \frac{1}{80} + \frac{\frac{1}{4}}{60} + \frac{\frac{3}{4}}{90}} = \frac{4}{\frac{10}{240}} = 96$$

دقت کنید در این سؤال طول مسافت (100 کیلومتر) تأثیری در حل مسئله ندارد.

مثال۴ مقدار کار 3 کارگر کارخانه آجرسازی در روزهای مختلف اندازه‌گیری شده است. کارگر اولی در دو دقیقه، دومی در سه دقیقه و سومی در چهار دقیقه یک متر خشت را آماده می‌کنند. اگر این سه کارگر با هم کار کنند، به طور متوسط یک متر خشت را در چه مدتی آماده می‌کنند؟

۱) 3 ۲) 39 ۳) $\frac{1}{3}$ ۴) $\frac{36}{13}$

حل: گزینه ۴ درست است.

واحد مشاهدات $x_1 = 2$، $x_2 = 3$ و $x_3 = 4$، واحد ترکیبی مترخشت است؛ بنابراین با توجه به رابطه میانگین هارمونیک داریم:

$$\bar{x}_H = \frac{N}{\sum_{i=1}^{3} \frac{1}{x_i}} = \frac{3}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}} = \frac{3}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}} = \frac{36}{13}$$

مثال ۵ اتومبیلی مسیری را با سرعت 30 کیلومتر بر ساعت رفته و هنگام برگشت $\frac{1}{3}$ مسیر را با سرعت 60 کیلومتر بر ساعت و باقی‌مانده را با سرعت 40 کیلومتر بر ساعت برگشته است، سرعت متوسط اتومبیل چقدر بوده است؟

۱) 36 ۲) 40 ۳) 48 ۴) 54

حل: گزینه ۱ درست است.

اولاً، واحد مشاهدات $x_1 = 30$ و $x_2 = 60$ و $x_3 = 40$، واحد ترکیبی کیلومتر بر ساعت است.

ثانیاً، به مشاهده $x_1 = 30$ وزنی معادل $w_1 = 1$، به مشاهده $x_2 = 60$ وزنی معادل $w_2 = \frac{1}{3}$ و به مشاهده $x_3 = 40$ وزنی معادل $w_3 = \frac{2}{3}$ نسبت داده شده است؛ بنابراین:

$$\bar{x}_H = \frac{\sum_{i=1}^{3} w_i}{\frac{w_1}{x_1}+\frac{w_2}{x_2}+\frac{w_3}{x_3}} = \frac{1+\frac{1}{3}+\frac{2}{3}}{\frac{1}{30}+\frac{\frac{1}{3}}{60}+\frac{\frac{2}{3}}{40}} = \frac{2}{\frac{10}{180}} = 36$$

مثال ۶ در یک کارخانه 4 ماشین با سرعت 2 دور در ثانیه و 6 ماشین با سرعت 5 دور در ثانیه کار می‌کنند. سرعت متوسط این ماشین‌ها چند دور در ثانیه است؟

۱) 4.5 ۲) 3.125 ۳) 5 ۴) 2.5

حل: گزینه ۲ درست است.

اولاً، واحد مشاهدات $x_1 = 2$ و $x_2 = 5$، واحد ترکیبی دور در ثانیه است.

ثانیاً، به مشاهده $x_1 = 2$، وزنی معادل $w_1 = 4$ ماشین و به مشاهده $x_2 = 5$، وزنی معادل $w_2 = 6$ ماشین نسبت داده شده است؛ بنابراین:

$$\bar{x}_H = \frac{\sum_{i=1}^{2} w_i}{\frac{w_1}{x_1}+\frac{w_2}{x_2}} = \frac{4+6}{\frac{4}{2}+\frac{6}{5}} = \frac{10}{\frac{32}{10}} = \frac{100}{32} = 3.125$$

مثال ۷ یک اتومبیل فاصله 3 هزار کیلومتری را با سرعت 480 کیلومتر بر ساعت و فاصله 5 هزار کیلومتری را با سرعت 240 کیلومتر بر ساعت و فاصله 2 هزار کیلومتری را با سرعت 320 کیلومتر بر ساعت طی کرده است. سرعت متوسط این اتومبیل چند کیلومتر بر ساعت است؟

۱) 300 ۲) 150 ۳) 200 ۴) 600

حل: گزینه ۱ درست است.

با توجه به آنکه واحد مشاهدات، واحد ترکیبی کیلومتر بر ساعت است، از میانگین هارمونیک استفاده می‌کنیم. در این مثال کل مسافت $2+3+5=10$ هزار کیلومتر است که:

$$x_1 = 480 \quad w_1 = \frac{3}{10} = 0.3 \quad , \quad x_2 = 240 \quad w_2 = \frac{5}{10} = 0.5 \quad , \quad x_3 = 320 \quad w_3 = \frac{2}{10} = 0.2$$

$$\bar{x}_H = \frac{w_1+w_2+w_3}{\frac{w_1}{x_1}+\frac{w_2}{x_2}+\frac{w_3}{x_3}} = \frac{0.3+0.5+0.2}{\frac{0.3}{480}+\frac{0.5}{240}+\frac{0.2}{320}} = \frac{1}{\frac{0.3+1+0.3}{480}} = 300$$

میانگین هندسی (Geometric Mean)

میانگین هندسی $(\bar{x}_G)$ داده‌های $x_1, x_2, \ldots, x_N$ برابر است با:

$$\bar{x}_G = \sqrt[N]{x_1 \times x_2 \times \ldots \times x_n} = (x_1 \times x_2 \times \ldots \times x_N)^{\frac{1}{N}}$$

اگر داده‌ها (x_i ها) دارای فراوانی مطلق (F_i) یا فراوانی نسبی (f_i) باشند، این فراوانی‌ها به عنوان وزن داده‌ها در نظر گرفته شده و به جای هر دو (F_i یا f_i)، می‌توانیم از w_i به شکل زیر استفاده کنیم:

$$\bar{x}_G = \sqrt[\sum w_i]{x_1^{w_1} \times x_2^{w_2} \times \ldots \times x_k^{w_k}} = \left(\prod_{i=1}^{k} x_i^{w_i}\right)^{\frac{1}{N}}$$

برای F_i (فراوانی مطلق) $\sum w_i = N$

برای f_i (فراوانی نسبی) $\sum w_i = 1$

مثال ۱ میانگین هندسی داده‌های 4 , 8 , 2 کدام است؟

۱) $\frac{14}{3}$ ۲) 4 ۳) 8 ۴) $\frac{64}{3}$

حل: گزینه ۲ درست است.

$$\bar{x}_G = \sqrt[3]{x_1 \times x_2 \times x_3} = \sqrt[3]{4 \times 8 \times 2} = \sqrt[3]{2^2 \times 2^3 \times 2} = \sqrt[3]{2^6} = \sqrt[3]{\left(2^2\right)^3} = 2^2 = 4$$

مثال ۲ با توجه به جدول فراوانی زیر، میانگین هندسی داده‌ها کدام است؟

x_i	1	2	4
$w_i = F_i$	2	6	2

۱) 2 ۲) $\frac{24}{3}$ ۳) 4 ۴) $\frac{1024}{6}$

حل: گزینه ۱ درست است.

$$\begin{cases} \bar{x}_G = \sqrt[\sum w_i]{x_1^{w_1} \times x_2^{w_2} \times x_3^{w_3}} = \sqrt[10]{1^2 \times 2^6 \times \underbrace{4^2}_{\left(2^2\right)^2}} = \sqrt[10]{2^6 \times \left(2^2\right)^2} = \sqrt[10]{2^{10}} = 2 \\ \sum w_i = \sum F_i = N = 10 \end{cases}$$

مثال ۳ با توجه به جدول فراوانی زیر میانگین هندسی داده‌ها کدام است؟

x_i	1	2	4
$w_i = f_i$	0.2	0.6	0.2

۱) 0.024 ۲) 0.33 ۳) 2 ۴) 1

حل: گزینه ۳ درست است.

$$\begin{cases} \bar{x}_G = \sqrt[1]{1^{0.2} \times 2^{0.6} \times 4^{0.2}} = 2^{0.6} \times \left(2^2\right)^{0.2} = 2^{0.6} \times 2^{0.4} = 2^1 = 2 \\ \sum w_i = n = 1 \end{cases}$$

کاربرد میانگین هندسی

میانگین هندسی در مواردی که داده‌ها، نسبی (بدون واحد) باشند (مانند نسبت، درصد، نرخ رشد، نرخ تورم) استفاده می‌شود.
برای حل مسایل میانگین هندسی می‌توانیم دسته‌بندی زیر را در نظر بگیریم:

۱ـ مشاهدات $x_1, x_2, ..., x_N$ **به صورت نسبت داده شوند.**

در این حالت، میانگین هندسی از رابطه زیر محاسبه می‌شود:

$$\bar{x}_G = \sqrt[N]{x_1 \times x_2 \times ... \times x_N} = (x_1 \times x_2 \times ... \times x_N)^{\frac{1}{N}}$$

که در آن $x_1, x_2, ..., x_N$ نسبت دو مقدار هستند.

برای مثال، اگر قیمت یک کالا در سال دوم تولید آن 2 برابر سال اول و در سال سوم، 3 برابر سال دوم باشد به این معناست که نسبت قیمت این کالا در سال دوم به سال اول برابر با 2 و نسبت قیمت آن در سال سوم به سال دوم برابر با 3 است؛ به عبارت دیگر:

$$\frac{\text{قیمت کالا در سال دوم}}{\text{قیمت کالا در سال اول}} = 2 \qquad \text{و} \qquad \frac{\text{قیمت کالا در سال سوم}}{\text{قیمت کالا در سال دوم}} = 3$$

مشاهده می‌شود که مقادیر $x_1 = 2$ و $x_2 = 3$ نسبت دو قیمت هستند و بنا بر رابطه بالا میانگین هندسی آن‌ها که عبارت است از متوسط افزایش قیمت در این سه سال، به صورت زیر محاسبه می‌شود:

$$\bar{x}_G = \sqrt[2]{x_1 \times x_2} = \sqrt[2]{2 \times 3} \simeq 2.5$$

به عبارت دیگر، قیمت این کالا در این سه سال به طور متوسط 2.5 برابر شده است.

مثال ۱ با تغییر مدیریت در یک فروشگاه، فروش در سال اول 2 برابر سال قبل و در سال دوم 8 برابر سال اول شده است. به طور متوسط فروش از آغاز مدیریت جدید چند برابر شده است؟ (اقتصاد ـ ۷۰)

۱) 4 برابر ۲) 5 برابر ۳) 8 برابر ۴) 6 برابر

حل: گزینه ۱ درست است.

ازآنجاکه داده‌ها نسبی هستند، برای محاسبه متوسط از میانگین هندسی استفاده می‌کنیم:

$$\left.\begin{aligned} x_1 &= \frac{\text{فروش سال اول}}{\text{فروش سال قبل}} = 2 \\ x_2 &= \frac{\text{فروش سال دوم}}{\text{فروش سال اول}} = 8 \end{aligned}\right. \Rightarrow \bar{x}_G = \sqrt[2]{x_1 \times x_2} = \sqrt[2]{2 \times 8} = 4$$

بنابراین، از آغاز مدیریت جدید، فروش به طور متوسط 4 برابر شده است.

مثال ۲ نسبت قیمت گندم به ذرت طی سه سال گذشته 4,8,2 بوده است. میانگین این نسبت‌ها کدام است؟

۱) 4 ۲) 2 ۳) 8 ۴) 3

حل: گزینه ۱ درست است.

ازآنجاکه داده‌ها نسبی هستند، از میانگین هندسی استفاده می‌کنیم:

$$\left.\begin{aligned} \text{سال اول}: x_1 &= \frac{\text{گندم}}{\text{ذرت}} = 2 \\ \text{سال دوم}: x_2 &= \frac{\text{گندم}}{\text{ذرت}} = 8 \\ \text{سال سوم}: x_3 &= \frac{\text{گندم}}{\text{ذرت}} = 4 \end{aligned}\right. \Rightarrow \bar{x}_G = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = 4$$

بنابراین، میانگین قیمت گندم به ذرت طی سه سال گذشته 4 است.

مثال۳ سود شرکتی در سال ۱۳۶۶ نسبت به ۱۳۶۵ هشت برابر، در سال ۱۳۶۷ نسبت به ۱۳۶۶ دو برابر و در سال ۱۳۶۸ نسبت به ۱۳۶۷ چهار برابر شده است. سود شرکت در پایان سال ۱۳۶۸ نسبت به سال ۱۳۶۵، چند برابر شده است؟

۱) 4 ۲) 2 ۳) 27 ۴) 8

حل: گزینه ۱ درست است.

ازآنجاکه مشاهدات، اندازه‌های نسبی هستند، برای محاسبه سود شرکت در سال ۱۳۶۸ نسبت به ۱۳۶۵ که همان متوسط سود در این سه سال است، از میانگین هندسی استفاده می‌کنیم:

$$\text{سال اول}: x_1 = \frac{\text{سود سال ۱۳۶۶}}{\text{سود سال ۱۳۶۵}} = 8$$

$$\text{سال دوم}: x_2 = \frac{\text{سود سال ۱۳۶۷}}{\text{سود سال ۱۳۶۶}} = 2 \quad \Rightarrow \quad \bar{x}_G = \sqrt[3]{8 \times 2 \times 4} = \sqrt[3]{64} = 4$$

$$\text{سال سوم}: x_3 = \frac{\text{سود سال ۱۳۶۸}}{\text{سود سال ۱۳۶۷}} = 4$$

بنابراین، سود شرکت در پایان سال ۱۳۶۸ نسبت به سال ۱۳۶۵، 4 برابر شده است.

مثال۴ سه نفر کارشناس، اولویت اتومبیل بنز را از نظر سرعت نسبت به پژو تعیین کرده‌اند. مقیاس اولویت‌گذاری از $\frac{1}{8}$ تا 8 است. نفر اول اولویت بنز را 4، نفر دوم $\frac{1}{4}$ و نفر سوم 8 تعیین کرده است. متوسط اولویت اتومبیل بنز به پژو از نظر سرعت چقدر است؟

۱) 2 ۲) 4 ۳) 6 ۴) 8

حل: گزینه ۱ درست است.

ازآنجاکه داده‌ها به صورت نسبی هستند، برای محاسبه متوسط از میانگین هندسی استفاده می‌کنیم:

$$\text{نفر اول}: x_1 = \frac{\text{اولویت بنز}}{\text{اولویت پژو}} = 4$$

$$\text{نفر دوم}: x_2 = \frac{\text{اولویت بنز}}{\text{اولویت پژو}} = \frac{1}{4} \quad \Rightarrow \quad \bar{x}_G = \sqrt[3]{x_1 \times x_2 \times x_3} = \sqrt[3]{4 \times \frac{1}{4} \times 8} = \sqrt[3]{8} = \sqrt[3]{2^3} = 2$$

$$\text{نفر سوم}: x_3 = \frac{\text{اولویت بنز}}{\text{اولویت پژو}} = 8$$

بنابراین، متوسط اولویت اتومبیل بنز به پژو از نظر سرعت برابر 2 است.

مثال۵ میزان سود یک شرکت سهامی در دو سال گذشته برحسب درصد فروش به ترتیب 20 و 80 بوده است، کدام‌یک از کمیت‌های زیر به عنوان شاخص مرکزی، وضع سودآوری شرکت را بهتر نشان می‌دهد؟

۱) 20 ۲) 40 ۳) 60 ۴) 50

حل: گزینه ۲ درست است.

ازآنجاکه میزان سود برحسب درصد فروش مطرح شده است، داریم:

$$\text{سود سال اول} = 20\%\,(\text{فروش سال اول}) \to \text{سود سال اول} = 0.20\,(\text{فروش سال اول}) \to x_1 = \frac{\text{سود سال اول}}{\text{فروش سال اول}} = 0.20$$

$$\text{سود سال دوم} = 80\%\,(\text{فروش سال دوم}) \to \text{سود سال دوم} = 0.80\,(\text{فروش سال دوم}) \to x_2 = \frac{\text{سود سال دوم}}{\text{فروش سال دوم}} = 0.80$$

بنابراین، بهترین شاخص مرکزی برای تعیین وضعیت سودآوری، میانگین هندسی است:

$$\bar{x}_G = \sqrt[2]{x_1 \times x_2} = \sqrt[2]{0.20 \times 0.80} = 0.40$$

به عبارت دیگر متوسط سودآوری در دو سال گذشته 40% فروش بوده است.

۲ـ مشاهدات $x_1, x_2, \dots, x_N$ به صورت درصد (نرخ رشد یا تورم) داده شوند.

در این حالت با استفاده از رابطه $\left(\frac{x_i}{100}+1\right)$ درصدها را برحسب واحد «برابر» به دست آورده و مقدار متوسط را برحسب «برابر» از همان رابطه (۱) به صورت زیر محاسبه می‌کنیم:

$$\bar{x}_G = \sqrt[N]{\left(\frac{x_1}{100}+1\right)\left(\frac{x_2}{100}+1\right)\dots\left(\frac{x_N}{100}+1\right)}$$

برای مثال، اگر نرخ رشد قیمت یک کالا در سال دوم تولید آن 20 درصد سال اول و در سال سوم، 30 درصد سال دوم باشد به این معناست که قیمت این کالا در سال دوم نسبت به سال اول 20 درصد افزایش و قیمت آن در سال سوم نسبت به سال دوم 30 درصد افزایش داشته است؛ به عبارت دیگر:

p_1 : قیمت در سال اول

$p_2 = p_1 + 0.2\,p_1 = (1.2)p_1$: قیمت در سال دوم

$p_3 = p_2 + 0.3\,p_2 = (1.3)p_2$: قیمت در سال سوم

در این صورت داریم:

$$\frac{\text{قیمت در سال دوم}}{\text{قیمت در سال اول}} = \frac{p_2}{p_1} = \frac{(1.2)p_1}{p_1} = \text{برابر}\,1.2 = \frac{20}{100}+1 = \frac{x_1}{100}+1$$

$$\frac{\text{قیمت در سال سوم}}{\text{قیمت در سال دوم}} = \frac{p_3}{p_2} = \frac{(1.3)p_2}{p_2} = \text{برابر}\,1.3 = \frac{30}{100}+1 = \frac{x_2}{100}+1$$

$$\Rightarrow \quad \bar{x}_G = \sqrt[2]{(1.2)\times(1.3)} \simeq \text{برابر}\,1.25$$

بنابراین، متوسط نرخ رشد قیمت این کالا در این سه سال برابر است با:

$$\frac{x}{100}+1 = 1.25 \quad \rightarrow \quad x = (1.25-1)\times 100 = 25\%$$

مثال ۶ نرخ رشد تولیدات یک کارخانه تولیدی طی دو سال گذشته به ترتیب 80% و -20% بوده است. متوسط نرخ رشد تولیدات سالانه در این کارخانه چقدر است؟ (اقتصاد ـ ۸۵)

۱) 20% ۲) 30% ۳) 50% ۴) 60%

حل: گزینه ۱ درست است.

با توجه به آنکه داده‌ها با عنوان نرخ رشد برحسب «درصد» مطرح شده است، برای محاسبه متوسط نرخ رشد به صورت زیر عمل می‌کنیم:

$$\bar{x}_G = \sqrt[2]{\left(\frac{x_1}{100}+1\right)\left(\frac{x_2}{100}+1\right)} = \sqrt[2]{\left(\frac{80}{100}+1\right)\left(\frac{-20}{100}+1\right)} = \sqrt[2]{1.8 \times 0.8} = \sqrt[2]{1.44} = 1.2$$

متوسط نرخ رشد در هر سال 1.2 برابر معادل $(1.2-1)\times 100 = 20\%$ است.

۳ـ مشاهدات $x_1, x_2, \dots, x_N$ با اندازه‌های غیرنسبی (واحددار مانند تومان، ریال، تعداد) داده شوند.

در این حالت برای محاسبه متوسط نرخ رشد و یا نرخ تورم با استفاده از کسر ($\frac{x_i}{x_{i-1}}$ برای $i\geq 2$) داده‌ها را به نسبت تبدیل می‌کنیم؛ به عبارت دیگر:

$$\bar{x}_G = \sqrt[N-1]{\frac{x_2}{x_1}\times\frac{x_3}{x_2}\times\ldots\times\frac{x_N}{x_{N-1}}} = \sqrt[N-1]{\frac{x_N}{x_1}}$$

بنابراین:

$$\bar{x}_G = \sqrt[N-1]{\frac{x_N}{x_1}}$$

✓ دقت کنید!

N داده، $N-1$ نسبت ایجاد می‌کند.

برای مثال، فرض کنید قیمت یک کالا در سال اول تولید آن 400 تومان، در سال دوم، 600 تومان و در سال سوم 900 تومان باشد. در این صورت داریم:

$$x_1 = 400\ ,\ x_2 = 600\ ,\ x_3 = 900 \quad\Rightarrow\quad \bar{x}_G = \sqrt[3-1]{\frac{\cancel{x_2}}{x_1}\times\frac{x_3}{\cancel{x_2}}} = \sqrt[2]{\frac{x_3}{x_1}} = \sqrt[2]{\frac{900}{400}} = \frac{3}{2} = 1.5$$

بنابراین، قیمت این کالا در این سه سال به طور متوسط 1.5 برابر شده است که معادل $(1.5-1)\times 100 = \%50$ رشد است.

مثال ۷ تعداد کارکنان یک شرکت طی یک دوره پنجساله به ترتیب 40، 80، 130، 200 و 640 نفر شده است. متوسط نرخ رشد کارکنان چند درصد بوده است؟

۱) %100 ۲) %200 ۳) %150 ۴) %50

حل: گزینه ۱ درست است.

ازآنجاکه داده‌ها با اندازه‌های غیرنسبی و برحسب تعداد مطرح شده‌اند، متوسط نرخ رشد طی دوره $n=5$ ساله با استفاده از میانگین هندسی به صورت زیر محاسبه می‌شود:

$$x_1 = 40\ ,\ x_N = 640\ ,\ N = 5 \quad\Rightarrow\quad \bar{x}_G = \sqrt[N-1]{\frac{x_N}{x_1}} = \sqrt[4]{\frac{640}{40}} = \sqrt[4]{16} = 2$$

بنابراین، تعداد کارمندان در طول پنج سال، به طور متوسط 2 برابر شده که معادل $(2-1)\times 100 = \%100$ رشد است.

مثال ۸ فرض کنید شاخص قیمت خرده‌فروشی از 80 در سال ۶۹ به 180 در سال ۷۱ رسیده باشد، متوسط نرخ تورم در این فاصله زمانی چقدر بوده است؟

۱) %50 ۲) %150 ۳) %100 ۴) %200

حل: گزینه ۱ درست است.

ازآنجاکه داده‌ها با اندازه‌های غیرنسبی و برحسب قیمت مطرح شده‌اند، متوسط نرخ تورم طی دوره $n=3$ ساله (۶۹ تا ۷۱)، با استفاده از میانگین هندسی به صورت زیر محاسبه می‌شود:

$$x_1 = 80\ ,\ x_N = 180\ ,\ N = 3 \quad\Rightarrow\quad \bar{x}_G = \sqrt[N-1]{\frac{x_N}{x_1}} = \sqrt[2]{\frac{180}{80}} = \sqrt[2]{2.25} = 1.5$$

بنابراین، قیمت خرده‌فروشی در طول 3 سال، به طور متوسط 1.5 برابر شده که معادل $(1.5-1)\times 100 = \%50$ تورم است.

میانگین حسابی (Mathematical Mean)

میانگین حسابی $(\bar{x})$ داده‌های $x_1, x_2, ..., x_n$ برابر است با حاصل تقسیم مجموع مشاهدات بر تعداد مشاهدات.

$$\text{میانگین حسابی} = \frac{\text{مجموع مشاهدات}}{\text{تعداد مشاهدات}}$$

میانگین حسابی، «**معدل**» یا «**متوسط**» نیز نامیده می‌شود.

میانگین حسابی نمونه و جامعه

اگر $x_1, x_2, ..., x_N$ مشاهدات جامعه باشند، میانگین حسابی جامعه برابر است با:

$$\text{میانگین حسابی جامعه}: \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

اگر $x_1, x_2, ..., x_n$ نمونه‌ای n تایی از جامعه باشد، میانگین حسابی آن برابر است با:

$$\text{میانگین حسابی نمونه}: \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

✓ **دقت کنید!**

۱ـ مفاهیم میانگین نمونه $(\bar{x})$ و میانگین جامعه (μ) کاملاً با هم متفاوت است (در فصل ۵ بررسی می‌شود) اما روش محاسبه آن‌ها یکسان است (مجموع مشاهدات تقسیم بر تعداد مشاهدات).

۲ـ بیان دیگر میانگین حسابی، امید ریاضی $(E(X))$ است (در فصل ۳ بررسی می‌شود)؛ به عبارت دیگر:

$$E(X) \equiv \mu \equiv \bar{x}$$

مثال ۱ میانگین حسابی داده‌های $30, 40, 30, 20$ کدام است؟

۱) 20 ۲) 30 ۳) 25 ۴) 35

حل: گزینه ۲ درست است.

$$\mu = \frac{\sum x_i}{n} = \frac{20+30+40+30}{4} = \frac{120}{4} = 30$$

میانگین حسابی و تصاعد حسابی

اگر $x_1, x_2, ..., x_n$ یک تصاعد حسابی با قدرنسبت d باشد، آن‌گاه میانگین حسابی مشاهدات به صورت زیر محاسبه می‌شود:

$$x_1, x_2, ..., x_n \xrightarrow{\text{ها تصاعد حسابی } x_i} x_1, x_1+d, ..., x_1+(n-1)d \rightarrow \sum_{i=1}^{n} x_i = \frac{n(x_1+x_n)}{2}$$

$$\mu=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{\frac{n(x_1+x_n)}{2}}{n}=\frac{x_1+x_n}{2}$$

به عبارت دیگر در هر تصاعد حسابی، میانگین حسابی مشاهدات برابر است با:

$$\mu=\frac{x_1+x_n}{2}=\frac{\text{مشاهده اول}+\text{مشاهده }n\text{ام}}{2}$$

مثال ۲ میانگین حسابی داده‌های $1,2,3,\ldots,99$ کدام است؟

۱) 50 ۲) 100 ۳) 51 ۴) 45

حل: گزینه ۱ درست است.

این مشاهدات یک تصاعد حسابی با قدرنسبت $d=1$ هستند؛ بنابراین میانگین حسابی برابر است با:

$$\mu=\frac{x_1+x_n}{2}=\frac{1+99}{2}=50$$

✓ دقت کنید!

در داده‌هایی که تشکیل تصاعد حسابی می‌دهند، اگر تعداد داده‌ها فرد باشد، میانگین حسابی داده‌ها برابر با داده وسط است و اگر تعداد داده‌ها زوج باشد، میانگین حسابی داده‌ها برابر با میانگین دو داده وسط است.

مثال ۳ میانگین حسابی داده‌های زیر را محاسبه کنید.

حل:

$$232,235,\underbrace{238}_{\mu},241,244 \xrightarrow[n\text{ فرد}]{\text{تصاعد حسابی}} \mu=238$$

$$435,439,\underbrace{443,447},451,455 \xrightarrow[n\text{ زوج}]{\text{تصاعد حسابی}} \mu=\frac{443+447}{2}=445$$

میانگین حسابی وزنی (Weighted Mathematical Mean)

گاهی مشاهدات (x_i) به صورت زیر با جدول فراوانی (مطلق یا نسبی) داده می‌شوند:

x_i	x_1	x_2	$\cdots$	
F_i	F_1	F_2	$\cdots$	$\sum F_i=N$
$\frac{F_i}{N}=f_i=p_i$	f_1	f_2	$\cdots$	$\sum f_i=1$

در این حالت میانگین حسابی برابر است با:

$$\boxed{\bar{x}=\mu=\frac{\sum F_i x_i}{N}=\sum f_i x_i}$$

مثال ۱ با توجه به جدول فراوانی زیر مقدار میانگین کدام است؟

x_i	2	1	3
F_i (فراوانی مطلق)	5	3	2

۱) 1.9 ۲) 3.3
۳) 1.8 ۴) 6.3

حل: گزینه ۱ درست است.

$$\begin{cases} \mu = \dfrac{\sum F_i x_i}{N} = \dfrac{5 \times 2 + 3 \times 1 + 2 \times 3}{10} = \dfrac{19}{10} = 1.9 \\ \sum F_i = N = 10 \end{cases}$$

مثال ۲ با توجه به جدول فراوانی زیر مقدار میانگین کدام است؟

x_i	0	1	2	3
f_i (فراوانی نسبی)	0.2	0.3	0.1	0.4

۱) 1.5 ۲) 1.7
۳) 1.3 ۴) 1

حل: گزینه ۲ درست است.

$$\mu = \sum f_i x_i = 0.2 \times 0 + 0.3 \times 1 + 0.1 \times 2 + 0.4 \times 3 = 1.7$$

نکته: اگر f_i (فراوانی نسبی) یا F_i (فراوانی مطلق) را وزن هر داده (x_i) در نظر بگیریم، می‌توانیم به‌جای هر دو از w_i استفاده کرده و میانگین حاصل را «**میانگین حسابی وزنی**» با نماد $\bar{x}_w$ نام‌گذاری کنیم.

x_i	x_1	x_2	$\cdots$	
$f_i = w_i$	w_1	w_2	$\cdots$	$\sum w_i = N$
$F_i = w_i$	w_1	w_2	$\cdots$	$\sum w_i = 1$

$$\boxed{\bar{x}_w = \mu_w = \frac{\sum w_i x_i}{\sum w_i} \rightarrow \begin{cases} w_i = f_i \ : \ \bar{x}_w = \sum w_i x_i \\ w_i = F_i \ : \ \bar{x}_w = \dfrac{\sum w_i x_i}{N} \end{cases}}$$

مثال ۳ معدل یک دانشجو در 5 واحد 14 شده است. اگر نمره 12 را از نمرات او حذف کنیم، معدل این دانشجو چقدر می‌شود؟

۱) 14.5 ۲) تغییر نمی‌کند.
۳) 11 ۴) به ضریب نمره 12 بستگی دارد.

حل: گزینه ۴ درست است.

اگر هر نمره را با x_i و واحد آن را با w_i نشان دهیم، آن‌گاه $\sum w_i = 5$ تعداد واحدهاست و معدل به صورت زیر محاسبه می‌شود:

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{w_1 \times x_1 + \ldots + w_k \times 12 + \ldots}{5} = 14$$

ازآنجاکه در این مسئله ضریب نمره 12 (تعداد واحد $= w_k$) مشخص نیست، نمی‌توانیم مشخص کنیم چه ضریبی از 12 را باید کم کنیم و همین‌طور تعداد واحدهای باقی‌مانده نیز مشخص نمی‌شود؛ درنتیجه نمی‌توانیم معدل واحدهای باقی‌مانده را محاسبه کنیم.

دقت کنید اگر در سؤال معدل برای 5 درس (1 واحدی) مطرح می‌شد، معدل دروس باقی‌مانده 14.5 می‌شد:

$$\text{معدل‌قدیم} = \frac{\text{مجموع نمرات}}{5} = 14 \rightarrow \text{مجموع نمرات} = 70 \longrightarrow 70 - 12 = 58\ (\text{مجموع نمرات باقی‌مانده})$$

$$\text{معدل جدید} = \frac{58}{4} = 14.5$$

مثال ۴ میانگین ده عدد مساوی 12 است. اگر یک عدد را کنار بگذاریم، میانگین 9 عدد باقی‌مانده مساوی 11 می‌شود. عدد کنار گذاشته‌شده، کدام است؟

۱) 11 ۲) 12 ۳) 20 ۴) 21

حل: گزینه ۴ درست است.

با توجه به تعریف میانگین حسابی برای 10 عدد داریم:

$$\mu = \frac{\sum_{i=1}^{10} x_i}{10} = 12 \rightarrow \sum_{i=1}^{10} x_i = 120$$

اگر a همان عدد کنار گذاشته‌شده باشد، آن‌گاه میانگین 9 عدد باقی‌مانده به صورت زیر محاسبه می‌شود:

$$\frac{\sum_{i=1}^{10} x_i - a}{9} = 11 \rightarrow 120 - a = 99 \rightarrow a = 21$$

مثال ۵ میانگین 10 عدد برابر 12 است. دو عدد را اشتباهی به‌جای 8 و 4 برابر 18 و 14 گرفته‌ایم. میانگین درست چقدر می‌شود؟

۱) 8 ۲) 14 ۳) 10 ۴) 12

حل: گزینه ۳ درست است.

با توجه به تعریف میانگین حسابی از رابطه زیر مجموع 10 عدد برابر 120 است:

$$\mu = \frac{\sum_{i=1}^{10} x_i}{10} = 12 \rightarrow \sum_{i=1}^{10} x_i = 120$$

به ازای انتخاب عدد 18 به‌جای 8 و انتخاب عدد 14 به‌جای 4، باید از مجموع 10 عدد، اعداد 18 و 14 که به اشتباه وارد شده‌اند، کسر شود و به‌جای آن دو، اعداد 8 و 4 به مجموع 10 عدد اضافه شود؛ درنتیجه:

$$\frac{\sum_{i=1}^{10} x_i = 120 - 18 - 14 + 8 + 4}{10} = 10\ (\text{میانگین جدید})$$

میانگین پیراسته (Truncated Mean / Trimmed Mean)

اگر تعداد کمی از مشاهدات با بقیه داده‌ها اختلاف زیادی داشته باشند (همخوانی یا تجانس نداشته باشند)، ازآنجاکه می‌دانیم هر داده کوچک یا بزرگ در میانگین اثر می‌گذارد، اگر تعداد کمی از داده‌ها به طور غیرعادی کوچک یا بزرگ باشند، باعث تغییرات فاحش در میانگین می‌شوند؛ به همین دلیل برای این دسته از داده‌ها می‌توان از میانگین پیراسته استفاده کرد.

محاسبه میانگین پیراسته به صورت زیر است:

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم.

ب) از ابتدا و انتهای داده‌ها به تعداد $(a\times n-1)$ داده را حذف می‌کنیم.

ج) میانگین حسابی مشاهدات باقی‌مانده را محاسبه می‌کنیم.

توجه: انتخاب a در این میانگین، شرایط خاصی دارد؛ اگر $(a\times n-1)$ عددی اعشاری باشد، حد پایین آن را در نظر می گیریم (به سمت پایین گرد می‌کنیم).

مثال هزینه ماهانه یک خانواده تهرانی برحسب صد هزار تومان به صورت زیر به دست آمده است:

$$10 , 8 , 15 , 20 , 9 , 16 , 17 , 18 , 25 , 30 , 15 , 14.5$$

اگر $a = \%25$ باشد، میانگین پیراسته را محاسبه کنید.

حل:

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم:

$$8 , 9, 10, 14.5 , 15 , 15 , 16 , 17 , 18 , 20 ,25 , 30$$

ب) از ابتدا و انتهای داده‌ها تعداد $\left(a\times n-1=\frac{25}{100}\times 12-1=2\right)$ داده را حذف می‌کنیم.

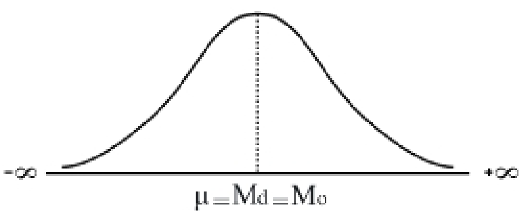
داده‌های پیراسته : $\cancel{8,9}$, 10, 14.5 , 15 , 15 , 16 , 17 , 18 , 20 , $\cancel{25,30}$

ج) میانگین حسابی داده‌های باقی‌مانده را محاسبه می‌کنیم:

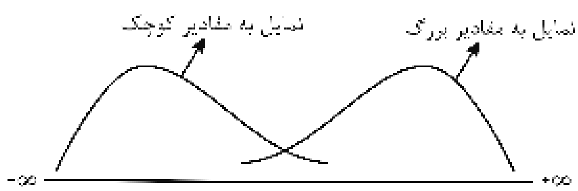
$$\mu_{\text{پیراسته}}=\frac{10+14.5+15+15+16+17+18+20}{8}=\frac{125.5}{8}=15.6875$$

نکته:

۱ـ در توزیع‌های متقارن، میانگین پیراسته به خوبی میانگین حسابی است.



۲ـ در توزیع‌های نامتقارن، یعنی زمانی‌که مقادیر جامعه ناهمگون است و منحنی توزیع به راست یا به چپ کشیده می‌شود، میانگین پیراسته بهتر از میانگین حسابی است و به‌جای آن به عنوان شاخص مرکزی به کار می‌رود.



میانگین وینزوری (Winsorized Mean)

میانگین وینزوری نوع دیگری از میانگین پیراسته است که برای محاسبه آن به صورت زیر عمل می‌کنیم:

الف) داده‌ها را به صورت صعودی مرتب می‌کنیم.

ب) از ابتدای داده‌ها به موقعیت $(a\times n)$ رفته و مقدار آن را در داده‌های قبل از آن جایگزین می‌کنیم. سپس از انتهای داده‌ها به سمت عقب به موقعیت $(a\times n)$ رفته و مقدار آن را در داده‌های بعد جایگزین می‌کنیم.

ج) میانگین حسابی مشاهدات جدید را محاسبه می‌کنیم.

توجه: انتخاب a در این میانگین شرایط خاصی دارد؛ اگر $(a\times n-1)$ عددی اعشاری باشد، حد پایین آن را در نظر می‌گیریم (به سمت پایین گرد می‌کنیم).

مثال هزینه ماهانه یک خانواده تهرانی برحسب صد هزار تومان به دست آمده است:

$10 , 8 , 15 , 20 , 9 , 16 , 17 , 18 , 25 , 30 , 15 , 14.5$

اگر $a = \%25$ باشد، میانگین وینزوری را محاسبه کنید.

حل:

الف) ابتدا داده‌ها را به صورت صعودی مرتب می‌کنیم:

$8 , 9, 10, 14.5 , 15 , 15 , 16 , 17 , 18 , 20 ,25 , 30$

ب) از ابتدای داده‌ها به موقعیت $(a\times n = 0.25\times 12 = 3)$ می‌رویم و مقدار آن را در داده‌های قبل از آن جایگزین می‌کنیم، همچنین از انتهای داده‌ها به سمت عقب به موقعیت $(a\times n)$ می‌رویم و مقدار آن را در داده‌های بعد جایگزین می‌کنیم:

داده‌های وینزوری : $\overset{10\,,\,10}{\cancel{8 , 9}} , \underset{\uparrow}{10} , 14.5 , 15 , 15 , 16 , 17 , 18 , \underset{\uparrow}{20} , \overset{20\,,\,20}{\cancel{25 , 30}}$

ج) میانگین حسابی مشاهدات جدید را محاسبه می‌کنیم:

$$\mu_{\text{وینزوری}} = \frac{10+10+10+14.5+15+15+16+17+18+20+20+20}{12} = \frac{185.5}{12} = 15.46$$

تفاوت میانگین پیراسته و وینزوری

میانگین پیراسته حالت خاصی از میانگین حسابی است؛ اگر از سری تغییرات و یا توزیع‌های فراوانی، مشاهدات کوچک‌تر از a درصد پایین و بزرگ‌تر از a درصد بالا را حذف کنیم، سپس میانگین حسابی را محاسبه کنیم، به آن میانگین پیراسته می‌گویند.

میانگین وینزوری نوعی از میانگین پیراسته است که در آن به‌جای مقادیر کوچک‌تر از a درصد پایین و بزرگ‌تر از a درصد بالا، مقدار عددی a درصد پایین و بالا را قرار می‌دهیم، سپس میانگین کل مشاهدات را محاسبه می‌کنیم.

خواص میانگین حسابی

۱ـ مجموع انحرافات (تفاضلات) همه نقاط از میانگین برابر صفر است.

$$\sum_{i=1}^{n}(x_i-\mu)=\sum_{i=1}^{n}(x_i-\bar{x})=0$$

به عبارت دیگر:

$$\sum_{i=1}^{n}(x_i-a)=0 \iff a=\mu=\bar{x}$$

از نظر فیزیکی (مکانیکی) میانگین، نقطه ثقل جامعه است که مجموع گشتاور نقاط نسبت به آن $\left(\sum(x_i-\mu)\right)$ همیشه صفر است.

۲ـ مجموع مجذور (توان دو) انحرافات (تفاضلات) همه نقاط از میانگین همیشه حداقل (مینیمم) است.

$$\sum(x_i-\mu)^2=\sum(x_i-\bar{x})^2=\min\ (\text{حداقل})$$

یعنی اگر b داده‌ای دلخواه از $x_1,\dots,x_n$ باشد، آن‌گاه:

$$\sum(x_i-\mu)^2<\sum(x_i-b)^2$$

به عبارت دیگر مجموع مجذور انحرافات از میانگین، از هر مجموع مجذور انحرافات نسبت به داده دلخواه، کوچک‌تر است:

$$\sum(x_i-a)^2=\min \ \rightarrow\ a=\mu=\bar{x}$$

۳ـ اگر a و b مقادیر ثابتی باشند (مثبت یا منفی):

الف) میانگین داده‌های برابر (مساوی) $a,a,a,a,\dots,a$، برابر a است.

$$a,a,\dots,a \ \rightarrow\ \mu(a)=a$$

ب) هرگاه به تمام مشاهدات (x_i) با میانگین μ، ثابت a را اضافه یا از آن کم کنیم، آن‌گاه میانگین مشاهدات جدید $(x_i \pm a)$ برابر با $\mu \pm a$ خواهد بود:

$$\begin{cases} x_1, x_2, \dots, x_n \\ \mu \end{cases} \xrightarrow{x_i \pm a} \begin{cases} x_1 \pm a,\ x_2 \pm a,\ \dots, x_n \pm a \\ \mu(x \pm a) = \mu \pm a \end{cases}$$

ج) هرگاه تمام مشاهدات (x_i) با میانگین μ، را در ثابت b ضرب یا بر آن تقسیم کنیم، آن‌گاه میانگین مشاهدات جدید (bx_i) یا $(\frac{1}{b}x_i)$ برابر با $b\mu$ یا $\frac{1}{b}\mu$ خواهد بود.

$$\begin{cases} x_1, x_2, \dots, x_n \\ \mu \end{cases} \begin{array}{l} \xrightarrow{bx_i} \begin{cases} bx_1,\ bx_2, \dots, bx_n \\ \mu(bx) = b\mu \end{cases} \\ \xrightarrow{\frac{1}{b}x_i} \begin{cases} \frac{1}{b}x_1,\ \frac{1}{b}x_2,\ \dots, \frac{1}{b}x_n \\ \mu\left(\frac{x}{b}\right) = \frac{1}{b}\mu \end{cases} \end{array}$$

د) با توجه به حالات (ب) و (ج) داریم:

$$\begin{cases} x_1, x_2, \dots, x_n \\ \mu \end{cases} \begin{array}{l} \xrightarrow{bx_i \pm a} \begin{cases} bx_1 \pm a,\ bx_2 \pm a,\ \dots, bx_n \pm a \\ \mu(bx \pm a) = b\mu \pm a \end{cases} \\ \xrightarrow{\frac{1}{b}x_i \pm a} \begin{cases} \frac{1}{b}x_1 \pm a,\ \frac{1}{b}x_2 \pm a,\ \dots, \frac{1}{b}x_n \pm a \\ \mu\left(\frac{x}{b} \pm a\right) = \frac{1}{b}\mu \pm a \end{cases} \end{array}$$

۴ـ میانگین تنها پارامتری است که اگر به جای تمام داده‌ها قرار گیرد، مجموع داده‌ها تغییری نمی‌کند.

$$\bar{x} = \frac{\sum x_i}{n} \rightarrow \sum x_i = n\bar{x} \rightarrow x_1 + x_2 + \dots + x_n = \underbrace{\bar{x} + \bar{x} + \dots + \bar{x}}_{n}$$

برای مثال،

$$\underbrace{1,2,3}_{\text{مجموع} = 6} \rightarrow \bar{x} = \frac{1+2+3}{3} = 2 \rightarrow \underbrace{2,2,2}_{\text{مجموع} = 6}$$

مثال اگر میانگین اعداد $x_1, x_2, \dots, x_n$ مساوی 20 باشد، مقدار $\sum_{i=1}^{n}(x_i - 20)$ کدام است؟

۱) صفر ۲) n ۳) $n^2 - 20$ ۴) 20

حل: گزینه ۱ درست است.

یکی از خواص مهم میانگین (μ) به صورت $\sum(x_i - \mu) = 0$ است که بیان می‌کند:

«مجموع تفاضلات (انحرافات) حول میانگین برابر 0 است.»

میانگین حسابی در داده‌های طبقه‌بندی شده

برای محاسبه میانگین حسابی در داده‌های طبقه‌بندی شده، به صورت زیر عمل می‌کنیم:

الف) مرکز طبقات را محاسبه می‌کنیم:

$$\text{مرکز طبقه}: x_i = \frac{\text{حد بالای طبقه} + \text{حد پایین طبقه}}{2} = \frac{L_i + U_i}{2}$$

ب) مقدار میانگین حسابی را از رابطه زیر به دست می‌آوریم:

$$\mu = \frac{\sum F_i x_i}{N} = \sum f_i x_i$$

جداول با حدود طبقات بزرگ

برای ساده کردن محاسبات در جداولی که حدود دسته‌های آن‌ها اعداد دورقمی یا بزرگ‌تر هستند، از روش زیر برای محاسبه میانگین حسابی استفاده می‌شود:

الف) طبقه‌ای را که دارای بیشترین فراوانی است $(L_k - U_k)$ انتخاب می‌کنیم و جدول فراوانی را به صورت زیر در نظر می‌گیریم:

$$\begin{cases} \text{مرکز طبقات جدید}: x'_i = \dfrac{x_i - a}{I} \\ \text{مرکز طبقه دارای بیشترین فراوانی}: a = \dfrac{L_k + U_k}{2} \\ \text{فاصله طبقات}: I \end{cases}$$

x'	$\cdots$	-2	-1	$\boxed{0}$	1	2	$\cdots$
$C-L$	$\cdots$	$\cdots$	$\cdots$	$\boxed{L_k - U_k}$	$\cdots$	$\cdots$	$\cdots$
F_i	$\cdots$	$\cdots$	$\cdots$	$\boxed{F_k}$	$\cdots$	$\cdots$	$\cdots$

ب) میانگین را با مرکز طبقات جدید محاسبه می‌کنیم:

$$\mu_{X'} = \frac{\sum F_i x'_i}{N}$$

ج) با توجه به خواص میانگین، میانگین داده‌های اصلی (μ_X) برابر است با:

$$\mu_{X'} = \mu\left(\frac{x_i - a}{I}\right) = \frac{\mu_X - a}{I} \longrightarrow \mu_X = \mu_{X'} \times I + a$$

مثال ۶ میانگین داده‌های جدول زیر کدام است؟

$C-L$	$20-30$	$30-40$	$40-50$
F_i	10	25	15

۱) 35　　۲) 36
۳) 37　　۴) 38

حل: گزینه ۲ درست است.

الف) طبقه $(30-40)$ دارای بیشترین فراوانی است؛ بنابراین جدول به صورت زیر خواهد بود:

x'	-1	$\boxed{0}$	1
$C-L$	$20-30$	$\boxed{30-40}$	$40-50$
F_i	10	$\boxed{25}$	15

ب) میانگین را با مرکز طبقات جدید محاسبه می‌کنیم:

x'	-1	0	1	
F_i	10	25	15	$N = \sum F_i = 50$
$F_i x'_i$	-10	0	15	$\sum F_i x'_i = 5$

$$\longrightarrow \mu_{X'} = \frac{\sum F_i x'_i}{N} = \frac{5}{50} = \frac{1}{10}$$

ج) میانگین داده‌های اصلی (μ_X) برابر است با:

$$\begin{cases} \mu_X = \mu_{X'} \times I + a \quad \rightarrow \quad \mu_X = \frac{1}{10} \times 10 + 35 = 1 + 35 = 36 \\ a = \frac{L_k + U_k}{2} = \frac{30+40}{2} = 35 \\ \mu_{X'} = \frac{1}{10} \ , \ I = 10 \end{cases}$$

مقایسه معیارهای مرکزی

سه ویژگی مهم پارامترهای مرکزی عبارت‌اند از:

ـ حداقل تعداد تخمین‌های اشتباه $(N_e = \min)$ در مد

ـ حداقل مجموع قدرمطلق انحرافات از میانه $\left(\sum |x_i - Md| = \min\right)$ در میانه

ـ حداقل مجذور خطا $\left(\sum (x_i - \mu)^2 = \sum (x_i - \bar{x})^2 = \min\right)$ در میانگین

از میان این سه ویژگی، حداقل مجذور خطا پایه بسیاری از تکنیک‌های آمار استنباطی بوده و از اهمیت خاصی برخوردار است؛ بنابراین، **میانگین** که دارای این ویژگی است **بهترین معیار مرکزی** محسوب می‌شود.

معیارهای پراکندگی (Measures of Dispersion)

هرگاه بخواهیم از بین جوامع آماری که می‌توانند ما را به هدفی مشخص برسانند، یکی را انتخاب کنیم، باید بتوانیم مشاهدات جوامع را به طور دقیق ارزیابی کنیم تا میزان تفاوت آن‌ها از یکدیگر مشخص شود.

میانگین، مهم‌ترین شاخص مرکزی است که می‌توان توزیع مشاهدات جامعه را اطراف آن بررسی کرد، اما هنگام مقایسه دو جامعه، ممکن است میانگین آن‌ها با هم برابر باشد؛ در این صورت نمی‌توان از این شاخص برای مقایسه تفاوت‌های دو جامعه استفاده کرد. در چنین شرایطی اگر **پراکندگی مشاهدات را حول مرکز بررسی و مقایسه کنیم**، به معیار جدیدی برای تصمیم‌گیری و انتخاب یک جامعه از بین جوامع می‌رسیم. معیاری که به این طریق به دست می‌آید به «**شاخص پراکندگی**» نیز معروف است.

برای مثال، فرض کنید یک شرکت عمرانی می‌خواهد از میان دو پیمانکار با نام‌های A و B یکی را برای واگذاری پروژه آینده‌اش انتخاب کند. داده‌های مربوط به مدت‌زمانی که این دو پیمانکار توانسته‌اند 4 پروژه قبلی خود را به اتمام برسانند (برحسب ماه)، به شرح زیر به دست آمده است:

پیمانکار A : 10,16,12,18

پیمانکار B : 15,14,13,14

در این مثال اگر شرکت بخواهد میانگین مدت‌زمان تحویل پروژه را ملاک انتخاب قرار دهد، دچار مشکل می‌شود زیرا میانگین مدت زمان تحویل پروژه هر دو پیمانکار برابر 14 ماه است؛ اما با کمی دقت می‌تواند متوجه شود که توزیع آماری این دو جامعه یکسان نیست، بنابراین برای انتخاب باید از پارامترهای پراکندگی کمک بگیرد. واضح است که پراکندگی داده‌های پیمانکار B که حول میانگین متمرکزترند، کمتر است.

تعریف: پارامترهای پراکندگی، معیارهایی برای تعیین میزان پراکندگی داده‌ها از یکدیگر یا میزان پراکندگی آن‌ها نسبت به بهترین مرکز (میانگین) است.

انواع معیارهای پراکندگی

۱) دامنه تغییرات
۲) دامنه میان‌چارکی (نیم‌دامنه)
۳) انحراف چارکی

← میزان پراکندگی داده‌ها را نسبت به یکدیگر تعیین می‌کنند.

۴) انحراف متوسط از میانگین
۵) واریانس و انحراف معیار
۶) نیمه‌واریانس

← میزان پراکندگی داده‌ها را نسبت به میانگین تعیین می‌کنند.

دامنه تغییرات (Range)

تعریف: دامنه تغییرات (R) عبارت است از تفاضل مقدار کوچک‌ترین مشاهده از مقدار بزرگ‌ترین مشاهده.

به عبارت دیگر، درصورتی‌که مشاهدات $x_1, x_2, ..., x_n$ از جامعه‌ای مفروض باشند، آن‌گاه دامنه تغییرات (R) به عنوان یکی از ساده‌ترین پارامترهای پراکندگی به صورت زیر محاسبه می‌شود:

$$\begin{cases} R = Max(x_i) - Min(x_i) \\ i=1,2,...,n \end{cases}$$

مثال ۱ دامنه تغییرات مربوط به مشاهدات $6, 7, 3, 2, 4$ کدام است؟

۱) 5 ۲) 2 ۳) 4 ۴) 7

حل: گزینه ۱ درست است.

$$\begin{cases} R = 7-2 = 5 \\ max(x_i) = 7, min(x_i) = 2 \end{cases}$$

مثال ۲ با توجه به جدول زیر دامنه تغییرات کدام است؟

مرکز دسته	-1	0	3	5
فراوانی	10	7	15	4

۱) 5 ۲) 6

۳) 4 ۴) قابل محاسبه نیست.

حل: گزینه ۲ درست است.

$$\begin{cases} R = 5-(-1) = 6 \\ max(x_i) = 5 \quad , \quad min(x_i) = -1 \end{cases}$$

مثال ۳ در جدول طبقه‌بندی شده زیر دامنه تغییرات کدام است؟

حدود طبقات	$1-4$	$4-7$	$7-10$	$10-13$
فراوانی	7	13	5	6

۱) 12 ۲) 13

۳) 10 ۴) 3

حل: گزینه ۱ درست است.

$$\begin{cases} R = 13-1 = 12 \\ max(x_i) = 13 \quad , \quad min(x_i) = 1 \end{cases}$$

یادآوری: در طبقه‌بندی داده‌ها، فاصله طبقات با استفاده از دامنه تغییرات (R)، محاسبه و داده‌ها طبقه‌بندی می‌شدند.

برای مثال، در این مسئله داریم:

$$\text{(فاصله طبقات)}\ I = \frac{R\ \text{(دامنه تغییرات)}}{K\ \text{(تعداد طبقات)}} = \frac{12}{4} = 3$$

موارد استفاده از دامنه تغییرات

الف) طبقه‌بندی مشاهدات جامعه

ب) محاسبه میزان پراکندگی در نمونه‌های کوچک (4 یا 5 تایی) که به آسانی قابل محاسبه است.

ج) کنترل کیفیت آماری

معایب دامنه تغییرات

هرچند دامنه تغییرات به عنوان یکی از ساده‌ترین معیارهای پراکندگی به آسانی قابل محاسبه است، اما در نمایش پراکندگی مشاهدات، پایداری (ثبات) لازم را ندارد، زیرا:

اولاً، از بین تمام داده‌ها $\left(x_n,...,x_2,x_1\right)$ تنها دو مشاهده (ماکزیمم و مینیمم) در نظر گرفته می‌شود.

ثانیاً، وضعیت توزیع سایر مشاهدات در نظر گرفته نمی‌شود؛ به عبارت دیگر، درصورتی‌که سایر مشاهدات به داده ماکزیمم یا مینیمم نزدیک باشند، دامنه تغییرات قادر به نمایش آن‌ها نیست.

برای مثال، به دو گروه از مشاهدات زیر دقت کنید:

$$(1)\quad 14\,,17\,,15\,,20\,,100\,,21\,,19 \qquad ; \qquad R=100-14=86$$

در این مشاهدات، علی‌رغم اینکه بیشتر مشاهدات به داده مینیمم (14) نزدیک‌ترند، دامنه تغییرات (86) بیشتر از پراکندگی واقعی مشاهدات است.

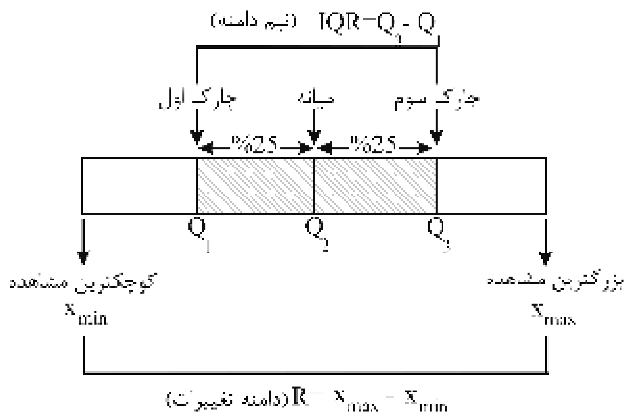
$$(2)\quad 14\,,\ 80\,,75\,,90\,,100\,,91\,,41 \qquad ; \qquad R=100-14=86$$

در این مشاهدات، علی‌رغم این‌که بیشتر مشاهدات به داده ماکزیمم (100) نزدیک‌ترند، باز هم دامنه تغییرات (86) شده است.

دامنه میان‌چارکی (نیم‌دامنه) (Interquantile Range)

تعریف: اگر پارامتر پراکندگی دامنه تغییرات (R) را برای 50 درصد از مشاهدات در فاصله Q_1 (چارک اول) تا Q_3 (چارک سوم) محاسبه کنیم، به شاخص پراکندگی جدیدی می‌رسیم که آن را دامنه میان‌چارکی (نیم‌دامنه) می‌نامند و با IQR نشان می‌دهند. $IQR=Q_3-Q_1$

در دامنه میان‌چارکی (نیم‌دامنه)، پارامتر مرکزی میانه را که 50 درصد داده‌هاست در نظر می‌گیریم و از پایین تا Q_1 (چارک اول) و از بالا تا Q_3 (چارک سوم) توسعه می‌دهیم.



مثال فرض کنید $Q_1=100$ و $Md=140$ و $Q_3=180$ باشد، نیم‌دامنه کدام است؟

۱) 40 ۲) 60 ۳) 30 ۴) 80

حل: گزینه ۴ درست است.

$$\text{نیم‌دامنه}: IQR=Q_3-Q_1=180-100=80$$

انحراف چارکی (Quantile Dispersion)

تعریف: انحراف چارکی برابر است با نصف دامنه میان‌چارکی (نیم‌دامنه).

$$SIQR = \frac{IQR}{2} = \frac{Q_3 - Q_1}{2}$$

در بعضی منابع، از انحراف چارکی $(SIQR)$ به جای دامنه میان‌چارکی (IQR) به عنوان شاخص پراکندگی استفاده می‌شود. اصولاً انحراف چارکی $(SIQR)$ برای بررسی پراکندگی جوامع از دامنه تغییرات (R) و دامنه میان‌چارکی (نیم‌دامنه) باثبات‌تر بوده و در نمایش پراکندگی داده‌ها بهتر از شاخص‌های مذکور است.

نکته: اگر تمام داده‌ها از Q_1 (چارک اول) تا Q_3 (چارک سوم) با هم برابر باشند، آن‌گاه انحراف چارکی $\left(\frac{Q_3 - Q_1}{2}\right)$ و دامنه میان‌چارکی (نیم‌دامنه) $(Q_3 - Q_1)$ برابر صفر هستند و برعکس.

$$Q_1 = Q_3 \Longleftrightarrow \begin{cases} \text{انحراف چارکی} = \dfrac{Q_3 - Q_1}{2} = 0 \\ \text{نیم دامنه} = Q_3 - Q_1 = 0 \end{cases}$$

مثال ۱ در $N = 10$ مشاهده زیر، چارک اول و سوم را به‌دست آورید.

2,3,6,6,6,6,6,6,9,15

حل:

$$\begin{cases} N = 10 \\ \text{محل چارک اول} = \dfrac{1N}{4} + \dfrac{1}{2} = 3 \rightarrow Q_1 = 6 \\ \text{محل چارک سوم} = \dfrac{3N}{4} + \dfrac{1}{2} = 8 \rightarrow Q_3 = 6 \end{cases} \rightarrow \begin{cases} \text{نیم‌دامنه} = 6 - 6 = 0 \\ \text{انحراف چارکی} = \dfrac{6-6}{2} = 0 \end{cases}$$

در این مشاهدات، تمام داده‌ها از $Q_1 = 6$ تا $Q_3 = 6$ با هم برابرند؛ بنابراین هیچ پراکندگی بین 50% میانی داده‌ها وجود ندارد و درنتیجه انحراف چارکی و نیم‌دامنه برابر صفر می‌شود.

نتیجه: اگر نیم‌دامنه (IQR) یا انحراف چارکی $(SIQR)$ برابر صفر شود، آن‌گاه 50 درصد مشاهدات از چارک اول (Q_1) تا چارک سوم (Q_3) با هم برابر هستند.

مثال ۲ در جدول داده‌های زیر، انحراف چارکی کدام است؟

C−L	10−13	13−16	16−19	19−22
F_i	5	8	4	3

۱) 3.75 ۲) 2.25
۳) 4 ۴) 2

حل: گزینه ۲ درست است.

ابتدا چارک اول (Q_1) و چارک سوم (Q_3) را محاسبه می‌کنیم:

C−L	10−13	13−16	16−19	19−22	
F_i	5	8	4	3	N = 20
F_{c_i}	5	13	17	20	

بعد از محاسبه فراوانی تجمعی $\left(F_{c_i}\right)$:

۱) اولین طبقه‌ای که در آن $F_{c_i} \ge \frac{N}{4} = 5$ باشد (طبقه $10-13$)، محل چارک اول $\left(Q_1\right)$ است:

$$Q_1 = L_i + \frac{\frac{1N}{4} - F_{c_{i-1}}}{F_i} \times I = 10 + \frac{5-0}{5} \times 3 = 13$$

۲) اولین طبقه‌ای که در آن $F_{c_i} \ge \frac{3N}{4} = 15$ باشد (طبقه $16-19$)، محل چارک سوم $\left(Q_3\right)$ است:

$$Q_3 = L_i + \frac{\frac{3N}{4} - F_{c_{i-1}}}{F_i} \times I = 16 + \frac{15-13}{4} \times 3 = 17.5$$

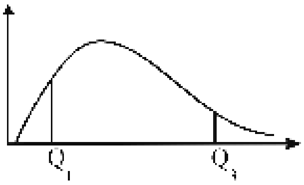
انحراف چارکی $\left(SIQR\right)$ بعد از محاسبه Q_1 و Q_3 به صورت زیر به دست می‌آید:

$$SIQR = \frac{Q_3 - Q_1}{2} = \frac{17.5-13}{2} = 2.25$$

موارد استفاده از انحراف چارکی

در صورت بروز یکی از شرایط زیر، استفاده از انحراف چارکی $\left(SIQR\right)$ به عنوان شاخص پراکندگی جامعه مناسب به نظر می‌رسد:

الف) در جوامع نامتقارن



ب) در توزیع‌هایی که تعداد اندکی مشاهده با اختلاف زیاد از سایر مشاهدات، وجود دارد.

برای مثال، در $N = 18$ مشاهده زیر:

$$\boxed{-200}, 2, 2, 2, 2, \underset{\uparrow Q_1}{3}, 3, 4, 4, 4, 5, 5, 5, 6, \underset{\uparrow Q_3}{6}, 6, 6, \boxed{100}, \boxed{400}$$

داده‌های -200 (در ابتدا) و 100 و 400 (در انتها) فاصله زیادی با سایر مشاهدات دارند و تعدادشان نیز اندک است (3مشاهده)؛ در این شرایط پارامتر پراکندگی انحراف چارکی برای نمایش میزان پراکندگی داده‌ها مناسب‌تر است، چرا که این داده‌ها تأثیری در $SIQR = \frac{Q_3 - Q_1}{2}$ ندارند و نمی‌توانند پراکندگی را بیش از حد نمایش دهند.

$$\begin{cases} N = 18 \\ \text{محل چارک اول } Q_1 = \frac{1N}{4} + \frac{1}{2} = 5 \quad \to \quad Q_1 = 3 \\ \text{محل چارک سوم } Q_3 = \frac{3N}{4} + \frac{1}{2} = 14 \quad \to \quad Q_3 = 6 \end{cases} \quad \to \quad SIQR = \frac{Q_3 - Q_1}{2} = 1.5$$

ج) در جداول طبقه‌بندی شده، زمانی که حدود ابتدا یا انتها، باز (نامشخص) باشند.

برای مثال:

$C-L$	<2	$2-5$	$5-8$	≥ 8	
F_i (فراوانی)	$\boxed{1}$	8	7	$\boxed{2}$	$N = \sum F_i = 18$

با توجه به داده‌های مثال قبل مشاهده می‌شود که تنها 1 مشاهده قبل از $x = 2$ وجود دارد (-200) و به همین دلیل حدود ابتدا نامشخص است. همین‌طور تنها 2 مشاهده بعد از $x = 8$ قرار دارد $(400,100)$؛ بنابراین حدود انتها نیز نامشخص است.

یادآوری: همان‌طور که در مبحث مربوط به موارد استفاده از شاخص مرکزی میانه مطرح شد، در تمام موارد (الف)، (ب) و (ج) مناسب‌ترین شاخص مرکزی میانه (Md) است.

انحراف متوسط از میانگین (Average Deviation of Mean)

هیچ‌یک از شاخص‌هایی که تا به حال در مورد آن‌ها صحبت شد ، قادر به بیان تمام تغییرات نیستند. تغییر زمانی مفهوم پیدا می‌کند که تمام داده‌ها نسبت به یک مبدأ مقایسه شوند و بهترین مبدأ یا مرکز داده‌ها، میانگین است؛ اما ازآنجاکه مجموع انحرافات از میانگین همیشه صفر است، $\sum(x_i - \mu)$، می‌توانیم از متوسط قدرمطلق انحرافات حول میانگین استفاده کنیم تا به پارامتر پراکندگی جدیدی به نام انحراف متوسط از میانگین $(A.D)$ برسیم. برای محاسبه انحراف متوسط از میانگین از رابطه زیر استفاده می‌کنیم:

داده‌های طبقه‌بندی نشده : $A.D_\mu = \dfrac{\sum_{i=1}^{n} |x_i - \mu_x|}{N}$

داده‌های طبقه‌بندی شده : $A.D_\mu = \dfrac{\sum F_i |x_i - \mu_x|}{N}$ یا $A.D_\mu = \sum f_i |x_i - \mu_x|$

مثال ۱ انحراف متوسط از میانگین داده‌های $2,4,5,3,4,6$ در جامعه‌ای برابر است با:

۱) 1 ۲) 2 ۳) 4 ۴) 8

حل: گزینه ۱ درست است.

$$\begin{cases} A.D_\mu = \dfrac{\sum|x_i - \mu_x|}{N} = \dfrac{|2-4|+|4-4|+|5-4|+|3-4|+|4-4|+|6-4|}{6} = \dfrac{6}{6} = 1 \\ \mu_X = \dfrac{\sum x_i}{N} = \dfrac{2+4+5+3+4+6}{6} = \dfrac{24}{6} = 4 \end{cases}$$

مثال ۲ انحراف متوسط از میانگین در داده‌های جدول فراوانی زیر کدام است؟

x_i	2	5	6
F_i	4	4	2

۱) 4 ۲) 1.8 ۳) 0.5 ۴) 1.6

حل: گزینه ۴ درست است.

$$\begin{cases} A.D_\mu = \dfrac{\sum F_i|x_i - \mu_x|}{N} = \dfrac{4|2-4|+4|5-4|+2|6-4|}{10} = \dfrac{16}{10} = 1.6 \\ \mu_X = \dfrac{\sum F_i x_i}{N} = \dfrac{4\times2+4\times5+2\times6}{4+4+2} = \dfrac{40}{10} = 4 \end{cases}$$

✓ **دقت کنید!**

در شرایطی که داده‌ها طبقه‌بندی شده باشند، از مرکز طبقه به‌جای x_i در محاسبات استفاده می‌کنیم.

خواص انحراف متوسط از میانگین

1) $x_1 = x_2 = \dots = x_n \longleftrightarrow A.D_\mu = 0$ (اگر تمام داده‌ها با هم برابر باشند، انحراف متوسط از میانگین صفر است.)

2) $A.D(x \pm a) = A.D_\mu$ (اگر به متغیرها عدد ثابتی اضافه یا از آن کم شود، انحراف متوسط تغییری نمی‌کند.)

نتیجه: هرگاه $\frac{c}{100}$ از یک مقدار به هر داده اضافه یا از آن کم شود، انحراف متوسط تغییری نمی‌کند.

$$3)\ A.D_{(ax)} = |a| \times A.D_x \ ,\ A.D_{\left(\frac{x}{a}\right)} = \frac{A.D_x}{|a|}$$

(اگر متغیرها را در عدد ثابتی (مثبت یا منفی) ضرب یا بر آن تقسیم کنیم، انحراف متوسط در قدرمطلق آن عدد ضرب یا بر قدرمطلق آن عدد تقسیم می‌شود.)

(انحراف متوسط اعداد ثابت صفر است.) $4)\ A.D(a) = 0$

معایب انحراف متوسط از میانگین

۱- از علامت جبری داده‌ها صرف‌نظر می‌کند.

۲- نقص اساسی این پارامتر این است که تأثیر انحرافات بزرگ را در شرایطی که تعداد زیادی انحرافات کوچک در برابر تعداد کمی انحرافات بزرگ وجود داشته باشد، نشان نمی‌دهد (برای رفع این اشکال، از واریانس استفاده می‌شود).

واریانس (Variance)

یکی از مهم‌ترین شاخص‌های اندازه‌گیری پراکندگی مشاهدات $(x_1,...,x_N)$ حول میانگین، واریانس است. هنگام محاسبه انحراف متوسط از میانگین ازآنجاکه $\sum(x_i-\mu)/N$ برابر با صفر می‌شد از قدرمطلق انحرافات $\left(\sum|x_i-\mu|/N\right)$ استفاده کردیم. مهم‌ترین مشکل این روش آن بود که انحرافات بزرگ را به خوبی نمایش نمی‌داد. برای رفع این مشکل از مجذور (توان دو) انحرافات $\left(\sigma^2=\sum(x_i-\mu)^2/N\right)$ استفاده می‌کنیم که شاخص جدیدی برای اندازه‌گیری میزان پراکندگی به‌ویژه انحرافات بزرگ محسوب می‌شود و به آن **واریانس** (σ^2) گفته می‌شود.

مثال کدام یک از پارامترهای زیر بیشتر تحت تأثیرات انحرافات بزرگ است؟ (حسابداری ـ ۸۲)

۱) واریانس
۲) نیم‌دامنه
۳) انحرافات چارکی
۴) انحراف متوسط از میانگین

حل: گزینه ۱ درست است.

واریانس بیش از سایر پارامترها تحت تأثیر انحرافات بزرگ قرار دارد.

انحراف معیار (Standard Deviation)

درصورتی‌که واحد اصلی مشاهدات $(x_1,...,x_N)$ برابر با m باشد، آن‌گاه واحد اندازه‌گیری واریانس مجذور واحد اصلی یعنی m^2 است. برای مثال، اگر واحد اصلی مشاهدات متر باشد، واحد واریانس ^{2}(متر) یا همان مترمربع است. ازآنجاکه واریانس به صورت مجذور واحد اولیه محاسبه می‌شود، نمی‌توانیم میزان پراکندگی مشاهدات را از روی آن به راحتی بررسی کنیم. برای حل این مشکل از جذر مثبت واریانس $\left(\sqrt{\sigma^2}\right)$ استفاده می‌کنیم و آن را **انحراف معیار** (σ) می‌نامیم. به این ترتیب به معیاری از پراکندگی می‌رسیم که با واحد اصلی مشاهدات $(x_1,...,x_N)$ بیان می‌شود.

نکته: در مفهوم مالی، «انحراف معیار» همان «ریسک» است.

ملاک مقایسه

هرگاه برای مقایسه دو جامعه تنها ملاک، واریانس $\left(\sigma^2\right)$ یا انحراف معیار (σ) باشد، می‌توان نتیجه گرفت که سایر عوامل تأثیرگذار در مقایسه با هم برابرند؛ بنابراین، جامعه‌ای که واریانس $\left(\sigma^2\right)$ یا انحراف معیار (σ) کمتری داشته باشد، پراکندگی کمتری دارد؛ درنتیجه متمرکزتر بوده و از دقت بیشتری برخوردار است.

مثال دستگاه A در اندازه‌گیری مکرر از شیء واحدی دارای واریانس $\sigma^2 = 9$ بوده است. دستگاه B در اندازه‌گیری مکرر از همان شیء دارای واریانس $\sigma^2 = 25$ بوده است. (مدیریت ـ ۷۴)

۱) دستگاه A دقیق‌تر است.
۲) دستگاه B دقیق‌تر است.
۳) دستگاه A اندازه‌گیری‌های بزرگ‌تری از دستگاه B به دست می‌دهد.
۴) دستگاه B اندازه‌گیری‌های بزرگ‌تری از دستگاه A به دست می‌دهد.

حل: گزینه ۱ درست است.
دستگاه A واریانس کمتری دارد بنابراین پراکندگی‌اش کمتر و دقتش بیشتر است.

$$\sigma_A^2 = 9 < \sigma_B^2 = 25$$

محاسبه واریانس

برای محاسبه واریانس داده‌های یک جامعه $\left(x_1,\ldots,x_N\right)$ می‌توانیم یکی از دو رابطه زیر را به کار ببریم:

$$(1)\quad \sigma^2 = \frac{\sum\left(x_i-\mu\right)^2}{N} = \text{میانگین مجذور تفاضل داده‌ها از میانگین}$$

$$(2)\quad \sigma^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2 = (\text{میانگین مجذور } x) - (\text{میانگین } x)^2$$

در محاسبه واریانس $\left(\sigma^2\right)$ و انحراف معیار (σ) داده‌های خام $\left(x_N,\ldots,x_2,x_1\right)$، استفاده از رابطه (1) برای سهولت در محاسبه توصیه می‌شود.

برای **محاسبه انحراف معیار** جامعه (σ)، کافی است بعد از محاسبه واریانس $\left(\sigma^2\right)$ از آن جذر مثبت بگیریم:

$$\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}} = \sqrt{\frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2}$$

واریانس و تصاعد حسابی

اگر $x_1, x_2, \ldots, x_N$ یک تصاعد حسابی با قدرنسبت d باشد، واریانس مشاهدات به صورت زیر به‌دست می‌آید:

$$x_1,x_2,\ldots,x_N \xrightarrow{\text{ها تصاعد حسابی } x_i} x_1, x_1+d,\ldots,x_1+(N-1)d$$

$$\boxed{\sigma^2 = \frac{d^2\left(N^2-1\right)}{12}}$$

مثال ۱ واریانس مشاهدات $2, 4, 1, 4, 5, 2$ در یک جامعه برابر است با:

۱) 2 ۲) $\sqrt{2}$ ۳) $\sqrt{3}$ ۴) $\frac{3}{2}$

حل: گزینه ۱ درست است.

با استفاده از روش ۱ داریم:

$$\begin{cases} \sigma^2 = \dfrac{\sum(x_i-\mu)^2}{N} = \dfrac{(2-3)^2+(5-3)^2+(4-3)^2+(1-3)^2+(4-3)^2+(2-3)^2}{6} = 2 \\ \mu = \dfrac{\sum x_i}{N} = \dfrac{2+5+4+1+4+2}{6} = 3 \end{cases}$$

البته روش ۲ نیز باید به همین نتیجه برسد:

$$\sigma^2 = \frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2 = \frac{2^2+5^2+4^2+1^2+4^2+2^2}{6} - \left(\frac{2+5+4+1+4+2}{6}\right)^2 = \frac{66}{6} - 3^2 = 2$$

در این مثال واریانس $\sigma^2 = 2$ و انحراف معیار $\sigma = \sqrt{2}$ است.

مثال ۲ انحراف معیار داده‌های $8, 10, 12, 14, 16, 18, 20$ کدام است؟

۱) 2 ۲) 4 ۳) 1 ۴) 16

حل: گزینه ۲ درست است.

این مشاهدات تصاعد حسابی با قدرنسبت $d = 2$ هستند؛ بنابراین واریانس آن‌ها برابر است با:

$$\begin{cases} \sigma^2 = \dfrac{d^2(N^2-1)}{12} = \dfrac{2^2(7^2-1)}{12} = \dfrac{4(49-1)}{12} = \dfrac{4\times 48}{12} = 16 \quad \to \quad \sigma = 4 \\ N = 7 \,,\, d = 2 \end{cases}$$

مثال ۳ در جامعه‌ای به حجم $N = 10$، درصورتی‌که مجموع مشاهدات 40 و مجموع مجذور مشاهدات 250 باشد، انحراف معیار جامعه کدام است؟

۱) $\sqrt{3}$ ۲) 3 ۳) 9 ۴) 4

حل: گزینه ۲ درست است.

$$\begin{cases} \sigma^2 = \dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N}\right)^2 = \dfrac{250}{10} - \left(\dfrac{40}{10}\right)^2 = 9 \quad \to \quad \sigma^2 = 9, \sigma = 3 \\ \text{مجموع مجذور مشاهدات} = \sum x^2 = 250 \\ \text{مجموع مشاهدات} = \sum x = 40 \;\text{،}\; N = 10 \end{cases}$$

در این مثال واریانس $\sigma^2 = 9$ و انحراف معیار $\sigma = 3$ است.

مثال ۴ اگر میانگین مقادیر x برابر 10 و میانگین مقادیر مجذورات x برابر 1000 باشد، واریانس مقادیر x چقدر است؟

(اقتصاد ـ ۸۴)

۱) 600 ۲) 700 ۳) 800 ۴) 900

حل: گزینه ۴ درست است.

$$\begin{cases} \sigma^2 = \dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N}\right)^2 = 1000 - 10^2 = 900 \\ \text{میانگین مجذورات} = \dfrac{\sum x^2}{N} = 1000 \\ \text{میانگین } X = \dfrac{\sum x}{N} = 10 \end{cases}$$

مثال ۵ اگر میانگین و انحرافات معیار X برابر 2 باشد، میانگین X^2 چقدر است؟

۱) 4 ۲) 8 ۳) 19 ۴) 32

حل: گزینه ۲ درست است.

$$\begin{cases} \sigma^2 = \dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N}\right)^2 \rightarrow 4 = \dfrac{\sum x^2}{N} - 2^2 \rightarrow \dfrac{\sum x^2}{N} = 8 \\ \text{میانگین } X = \dfrac{\sum x}{N} = 2 \\ \sigma = 2 \rightarrow \sigma^2 = 4 \end{cases}$$

خواص واریانس و انحراف معیار

۱ــ واریانس و انحراف معیار داده‌های مساوی، برابر با صفر است و برعکس.

$$x_1 = x_2 = \cdots = x_N \rightleftarrows \sigma^2 = 0 \ , \ \sigma = 0$$

یادآوری: در داده‌های مساوی، میانگین برابر با همان داده‌های مساوی است.

مثال ۱ داده‌های زیر در دست است. انحراف معیار این داده‌ها برابر است با: (مدیریت ـ ۷۴)

98750 , 98750 , 98750

۱) 98750 ۲) 9875 ۳) 1 ۴) 0

حل: گزینه ۴ درست است.

داده‌های 98750 , 98750 , 98750 برابر هستند، بنابراین:

$$\begin{cases} \text{واریانس} : \sigma^2 = 0 \\ \text{انحراف معیار} : \sigma = 0 \\ \text{میانگین} : \mu = 98750 \end{cases}$$

مثال ۲ واریانس داده‌های آماری صفر است، کدام نتیجه‌گیری در این مورد درست است؟

۱) میانگین داده‌ها صفر است. ۲) تمام داده‌ها برابرند.

۳) داده‌ها متقارن هستند. ۴) میانگین بیشتر از هر داده است.

حل: گزینه ۲ درست است.

در داده‌های مساوی، واریانس و انحراف معیار برابر صفر و میانگین برابر داده‌هاست.

۲ــ واریانس و انحراف معیار هر مقدار ثابت مانند a برابر با صفر است.

$$\sigma^2(a) = 0 \ , \ \sigma(a) = 0$$

یادآوری: میانگین هر مقدار ثابت مانند a برابر با a است $(\mu(a)=a)$.

مثال ۳ میانگین و واریانس عدد 4 کدام است؟ (راست به چپ)

۱) $4\,,\,4$ ۲) $0\,,\,0$ ۳) $0\,,\,4$ ۴) $4\,,\,0$

حل: گزینه ۳ درست است.

$$\mu(4)=4 \quad , \quad \sigma^2(4)=0 \quad , \quad \sigma(4)=0$$

۳ـ هرگاه مقدار ثابت a به تمام داده‌ها اضافه یا از آن‌ها کم شود، واریانس و انحراف معیار هیچ تغییری نمی‌کنند.

$$\begin{cases} x_1, x_2, \dots, x_n \\ \sigma_x^2 \\ \sigma_x \end{cases} \to \begin{cases} x_1 \pm a\,,\, x_2 \pm a\,,\, \dots\,,\, x_n \pm a \\ \sigma^2(x \pm a) = \sigma_x^2 \\ \sigma(x \pm a) = \sigma_x \end{cases}$$

تبصره: با توجه به این خاصیت، اگر تمام داده‌های دارای قسمت مشترکی باشند، می‌توان اعداد قسمت مشترک را حذف و واریانس باقی‌مانده اعداد را محاسبه کرد؛ درواقع داده‌ها را از عدد قسمت مشترک کم می‌کنیم.

۴ـ هرگاه تمام داده‌ها در مقدار ثابت b ضرب یا بر آن تقسیم شوند، واریانس در مجذور آن (b^2) و انحراف معیار در قدرمطلق آن $(|b|)$ ضرب یا بر آن تقسیم می‌شود.

$$\begin{cases} x_1, x_2, \dots, x_n \\ \sigma_x^2 \\ \sigma_x \end{cases} \to \begin{cases} bx_1,\ bx_2\,,\, \dots\,,\, bx_n \\ \sigma^2(bx) = b^2\sigma_x^2 \\ \sigma(bx) = |b|\sigma_x \end{cases}$$

$$\begin{cases} x_1, x_2, \dots, x_n \\ \sigma_x^2 \\ \sigma_x \end{cases} \to \begin{cases} \dfrac{x_1}{b}, \dfrac{x_2}{b}, \dots, \dfrac{x_n}{b} \\ \sigma^2\left(\dfrac{x}{b}\right) = \dfrac{1}{b^2}\sigma_x^2 \\ \sigma\left(\dfrac{x}{b}\right) = \dfrac{1}{|b|}\sigma_x \end{cases}$$

۵ـ هرگاه به تمام داده‌ها c درصد از یک مقدار (a) اضافه یا از آن‌ها کم شود، حالت ۳ و اگر c درصد از هر داده به داده‌ها اضافه یا از آن‌ها کم شود، حالت ۴ رخ می‌دهد.

c درصد از هر داده به آن اضافه یا از آن کم شود:

$$\begin{cases} \sigma^2\left(x \pm \dfrac{c}{100}x\right) = \sigma^2\left(\left(1 \pm \dfrac{c}{100}\right)x\right) = \left(1 \pm \dfrac{c}{100}\right)^2 \sigma_x^2 \\ \sigma\left(x \pm \dfrac{c}{100}x\right) = \sigma\left(\left(1 \pm \dfrac{c}{100}\right)x\right) = \left|1 \pm \dfrac{c}{100}\right|\sigma_x \end{cases}$$

(مشابه حالت ۴)

c درصد از یک مقدار (a) به هر داده اضافه یا از آن کم شود:

$$\begin{cases} \sigma^2\left(x \pm \overset{0}{\cancel{\dfrac{c}{100}a}}\right) = \sigma_x^2 \\ \sigma\left(x \pm \overset{0}{\cancel{\dfrac{c}{100}a}}\right) = \sigma_x \end{cases}$$

(مشابه حالت۳)

مثال ۴ در یک کارخانه میانگین حقوق کارمندان 120 هزار تومان با انحراف معیار 10 هزار تومان است. مقدار میانگین و واریانس و انحراف معیار با هر یک از تغییرات زیر چگونه خواهد شد؟

الف) به حقوق تمام کارمندان 5 هزار تومان اضافه یا از آن‌ها کم کنیم.

ب) حقوق همه کارمندان را 2 برابر کنیم.

ج) از حقوق هر کارمند 20% حقوقش را کم کنیم.

د) از حقوق هر کارمند 20% میانگین حقوق را کم کنیم.

حل:

الف)

$$\begin{cases} x_i \pm 5 \\ \mu(x \pm 5) = \mu_x \pm 5 \begin{matrix} \nearrow 125 \\ \searrow 115 \end{matrix} \\ \sigma^2\left(x_i \overset{0}{\cancel{\pm 5}}\right) = \sigma_x^2 = 100 \\ \sigma\left(x_i \overset{0}{\cancel{\pm 5}}\right) = \sigma_x = 10 \end{cases}$$

ب)

$$\begin{cases} 2x_i \\ \mu(2x_i) = 2\mu_x = 2 \times 120 = 240 \\ \sigma^2(2x_i) = 4\sigma_x^2 = 4 \times 100 = 400 \\ \sigma(2x_i) = |2|\sigma_x = 2 \times 10 = 20 \end{cases}$$

ج)

$$\begin{cases} x_i - 0.20x_i = 0.8x_i \\ \mu(0.8x_i) = 0.8 \times 120 = 96 \\ \sigma^2(0.8x_i) = (0.8)^2 \times \sigma_x^2 = 0.64 \times 100 = 64 \\ \sigma(0.8x_i) = |0.8| \times \sigma_x = 0.8 \times 10 = 8 \end{cases}$$

د)

$$\begin{cases} x_i - 0.2\mu_x = x_i - 0.2 \times 120 = x_i - 24 \\ \mu(x_i - 24) = \mu_x - 24 = 120 - 24 = 96 \\ \sigma^2\left(x_i \overset{0}{\cancel{-24}}\right) = \sigma_x^2 = 100 \\ \sigma\left(x_i \overset{0}{\cancel{-24}}\right) = \sigma_x = 10 \end{cases}$$

مثال ۵ 10 عدد معین را در 4 ضرب کرده و 20 را از آن‌ها کم می‌کنیم. میانگین و انحراف معیار اعداد جدید به ترتیب 4 و 2 است. میانگین و انحراف معیار اولیه کدام هستند؟

۱) $\mu_x = 4 \,,\, \sigma_x = 1$ ۲) $\mu_x = 5 \,,\, \sigma_x = 1$ ۳) $\mu_x = 6 \,,\, \sigma_x = \frac{1}{2}$ ۴) $\mu_x = 4 \,,\, \sigma_x = 1$

حل: گزینه ۳ درست است.

با توجه به صورت مسئله، میانگین و انحراف معیار داده‌های $4x_1-20,\ 4x_2-20, \dots, 4x_{10}-20$ به ترتیب برابر 4 و 2 است؛ بنابراین:

$$\begin{cases}\mu(4x-20)=4 \ \rightarrow\ 4\mu_x-20=4 \ \rightarrow\ \mu_x=6\\ \sigma(4x-20)=2 \ \rightarrow\ |4|\sigma_x=2 \ \rightarrow\ \sigma_x=\dfrac{1}{2}\end{cases}$$

دقت کنید که تعداد داده‌ها $N=10$ در صورت مسئله تأثیری بر حل مسئله ندارد، زیرا فقط خواص میانگین و انحراف معیار بررسی می‌شوند.

مثال ۶ اگر واریانس $1, 2, 3, 4, 5$ برابر a^2 باشد، واریانس اعداد $202, 204, 206, 208, 210$ کدام است؟

۱) $4a^2$ ۲) $2a^2$ ۳) $2a^2+10$ ۴) $4a^2+10$

حل: گزینه ۱ درست است.

با توجه به صورت مسئله، داده‌های $x=1,2,3,4,5$ با تبدیل $2x+200$ به صورت $202,204,206,208,210$ ظاهر می‌شوند؛ بنابراین داریم:

$$\begin{cases}1,2,3,4,5\\ \sigma_x^2=a^2\end{cases} \xrightarrow{2x+200} \begin{cases}202\ ,\ 204\ ,\ 206\ ,\ 208\ ,\ 210\\ \sigma^2(2x+200)=2^2\sigma_x^2=4a^2\end{cases}$$

مثال ۷ انحراف معیار داده‌های $x_1, x_2, \dots, x_n$ برابر a است. انحراف معیار داده‌های $-2x_1+3, \dots, -2x_n+3$ چقدر است؟

۱) $-2a+3$ ۲) $-2a$ ۳) $2a$ ۴) $4a$

حل: گزینه ۳ درست است.

$$\begin{cases}x_1, x_2, \dots, x_n\\ \sigma_x=a\end{cases} \rightarrow \begin{cases}-2x_1+3, -2x_2+3, \dots, -2x_n+3\\ \sigma(-2x\underbrace{\cancel{+3}}_{0})=|-2|\sigma_x=2\sigma_x=2a\end{cases}$$

مثال ۸ اگر واریانس چند عدد مساوی 10 باشد و به هر یک از داده‌ها 20% اضافه شود، واریانس چقدر خواهد شد؟

۱) 11.2 ۲) 11.44 ۳) 14.4 ۴) 10

حل: گزینه ۳ درست است.

$$\begin{cases}\sigma^2(x+0.20x)=\sigma^2(1.2x)=(1.2)^2\sigma_x^2=1.44\times 10=14.4\\ \sigma_x^2=10\end{cases}$$

دقت کنید اگر در صورت سؤال مطرح می‌شد 0.20 یک عدد ثابت مانند (μ) به داده‌ها اضافه شود، آن‌گاه σ^2 تغییری نمی‌کرد و گزینه ۴ درست بود.

$$\sigma^2\left(x+\underbrace{\cancel{0.20\mu}}^{0}\right)=\sigma_x^2=10$$

مثال ۹ اگر واریانس اعداد $x_1, x_2, \dots, x_n$ مساوی a باشد، انحراف معیار اعداد $2x_1-1$، $2x_2-1$، … و $2x_n-1$ کدام است؟

۱) $\sqrt{a}$ ۲) $2\sqrt{a-1}$ ۳) $2\sqrt{a}$ ۴) $4\sqrt{a}$

حل: گزینه ۳ درست است.

$$\begin{cases} x_1, x_2, \dots, x_n \\ \sigma_x^2 = a\ ,\ \sigma_x = \sqrt{a} \end{cases} \rightarrow \begin{cases} 2x_1 - 1\ ,\ 2x_2 - 1\ , \dots, 2x_n - 1 \\ \sigma(2x-1) = |2|\sigma_x = 2\sqrt{a} \end{cases}$$

مثال ۱۰ اگر انحراف معیار داده‌های آماری X_1، X_2، ...، X_n، برابر 4 باشد، واریانس داده‌های زیر کدام است؟

$2X_1+1,\ 2X_2+1,\ \dots,\ 2X_n+1$

۱) 8 ۲) 16 ۳) 32 ۴) 64

حل: گزینه ۴ درست است.

$$\begin{cases} x_1, x_2, \dots, x_n \\ \sigma_x = 4, \sigma_x^2 = 16 \end{cases} \rightarrow \begin{cases} 2x_1 + 1\ ,\ 2x_2 + 1, \dots, 2x_n + 1 \\ \sigma^2(2x+1) = 2^2\sigma_x^2 = 4 \times 16 = 64 \end{cases}$$

محاسبه واریانس در داده‌های دارای فراوانی

اگر مشاهدات جامعه دارای فراوانی مطلق (F_i) یا فراوانی نسبی $\left(f_i = \frac{F_i}{N}\right)$ باشند، یکی از دو رابطه زیر را به کار می‌بریم:

$$\begin{cases} (1)\ \ \sigma^2 = \dfrac{\sum F_i (x_i - \mu)^2}{N} = \sum f_i (x_i - \mu)^2 \\ (2)\ \ \sigma^2 = \dfrac{\sum F_i x_i^2}{N} - \left(\dfrac{\sum F_i x_i}{N}\right)^2 = \sum f_i x_i^2 - \left(\sum f_i x_i\right)^2 \end{cases}$$

تبصره ۱: برای محاسبه واریانس (σ^2) و انحراف معیار (σ) در داده‌های دارای فراوانی (f_i یا F_i)، استفاده از رابطه (2) برای سهولت در محاسبه توصیه می‌شود.

تبصره ۲: در جداول طبقه‌بندی شده، از مرکز طبقه به‌جای x_i استفاده می‌شود.

تبصره ۳: برای محاسبه انحراف معیار (σ)، کافی است پس از محاسبه σ^2 از آن جذر مثبت بگیریم.

مثال ۱ واریانس داده‌ها با جدول فراوانی زیر کدام است؟

x_i	-2	-1	0	1
F_i	2	3	1	4

۱) 1.5 ۲) 1.41
۳) 2.5 ۴) 0.8

حل: گزینه ۲ درست است.

همان‌طورکه توصیه شد، برای محاسبه واریانس در داده‌های دارای فراوانی از رابطه (2) استفاده می‌کنیم:

x_i^2	4	1	0	1	
x_i	-2	-1	0	1	
F_i	2	3	1	4	$N = \sum F_i = 10$

$$\sigma^2 = \frac{\sum F_i x_i^2}{N} - \left(\frac{\sum F_i x_i}{N}\right)^2 = \frac{2\times4+3\times1+1\times0+4\times1}{10} - \left(\frac{2\times(-2)+3\times(-1)+1\times0+4\times1}{10}\right)^2 = \frac{15}{10} - \frac{9}{100} = 1.41$$

مثال ۲ انحراف معیار داده‌ها با جدول فراوانی زیر کدام است؟

x_i	-1	0	1	2
f_i	0.1	0.2	0.3	0.4

۱) 1 ۲) 2

۳) $\sqrt{2}$ ۴) 4

حل: گزینه ۱ درست است.

$$\sigma^2=\sum f_i x_i^2-\left(\sum f_i x_i\right)^2$$

$$=[\underbrace{0.1\times(-1)^2}_{0.1}+\underbrace{0.2\times 0^2}_{0}+\underbrace{0.3\times 1^2}_{0.3}+\underbrace{0.4\times 2^2}_{1.6}]-[\underbrace{0.1\times(-1)}_{-0.1}+\underbrace{0.2\times 0}_{0}+\underbrace{0.3\times 1}_{0.3}+\underbrace{0.4\times 2}_{0.8}]^2=2-1^2=1$$

یادآوری: در بعضی از منابع به‌جای استفاده از نماد F_i برای فراوانی مطلق از f_i استفاده می‌شود. با کمی دقت می‌توانیم از اشتباه در مسایل جلوگیری کنیم، به یاد دارید که همیشه $0\le f_i\le 1$ و $F_i\ge 0$ است.

مثال ۳ در یک جدول توزیع فراوانی که شامل 100 مشاهده است، میانگین 6 و $\sum_{i=1}^{k} f_i x_i^2=5000$ است. واریانس این مشاهدات چقدر است؟ (اقتصاد ـ ۸۵)

۱) 1 ۲) 14 ۳) 44 ۴) 49

حل: گزینه ۲ درست است.

در این سؤال با توجه به مقدار $\sum f_i x_i^2=5000$، منظور از f_i همان F_i است.

$$\begin{cases}\sigma^2=\dfrac{\sum F_i x_i^2}{N}-\left(\dfrac{\sum F_i x_i}{N}\right)^2=\dfrac{5000}{100}-(6)^2=50-36=14\\ \mu=\dfrac{\sum F_i x_i}{N}=6\\ N=100\end{cases}$$

مثال ۴ واریانس کدام‌یک از چهار مجموعه زیر بیشتر است؟ (اقتصاد ـ ۷۰)

۱) $1, 1, 2, 4, 5, 7, 8, 8$ ۲) $1, 1, 1, 4, 5, 8, 8, 8$

۳) $1, 1, 1, 1, 8, 8, 8, 8$ ۴) $1, 2, 3, 4, 5, 6, 7, 8$

حل: گزینه ۳ درست است.

راه حل اول:

در این سؤال نیازی به محاسبه واریانس نیست. اولاً، در تمام گزینه‌ها مجموع داده‌ها $\sum_{i=1}^{8} x_i=36$ و میانگین $\mu=\frac{36}{8}=4.5$ است؛

ثانیاً، چون برای واریانس باید $\sigma^2=\dfrac{\sum(x_i-\mu)^2}{N}$ محاسبه شود و در گزینه ۳ عبارت $(x_i-\mu)^2$ برای تمام مشاهدات یا $(1-4.5)^2$ است یا $(8-4.5)^2$ که نسبت به مشاهدات گزینه‌های دیگر بیشتر است؛ درنتیجه واریانس بیشتری دارد. در گزینه۴ عبارت $(x_i-\mu)^2$ فقط در دو مشاهده $(1-4.5)^2$ و $(8-4.5)^2$ است و در سایر مشاهدات کمتر از این دو مقدار است؛ بنابراین کمترین واریانس را در بین گزینه‌ها دارد. پس گزینه‌ها به ترتیب بیشترین واریانس تا کمترین واریانس عبارت‌اند از: ۳، ۲، ۱ و ۴.

راه حل دوم:
در گزینه ۳ انحراف داده‌ها از یکدیگر 0 یا 7 است، اما در گزینه‌های دیگر انحراف داده‌ها از یکدیگر بین 0 و 7 است؛ بنابراین گزینه ۳ پراکندگی بیشتر و درنتیجه واریانس بیشتری دارد. درعین‌حال در گزینه ۴ چون داده‌ها به هم نزدیک هستند، کمترین پراکندگی و درنتیجه کمترین واریانس را دارند.

جداول با حدود طبقات بزرگ

برای ساده کردن محاسبات در جداولی که حدود دسته‌های آن‌ها اعداد دورقمی یا بزرگ‌تر هستند، از روش زیر برای محاسبه واریانس استفاده می‌شود:

الف) طبقه‌ای را که دارای بیشترین فراوانی است $(L_k - U_k)$ انتخاب می‌کنیم و جدول فراوانی را به صورت زیر در نظر می‌گیریم:

x'	$\cdots$	-2	-1	$\boxed{0}$	1	2	$\cdots$
$C-L$	$\cdots$	$\cdots$	$\cdots$	$\boxed{L_k - U_k}$	$\cdots$	$\cdots$	$\cdots$
F_i	$\cdots$	$\cdots$	$\cdots$	$\boxed{F_k}$	$\cdots$	$\cdots$	$\cdots$

مرکز طبقات جدید : $x'_i = \dfrac{x_i - a}{I}$

مرکز طبقه دارای بیشترین فراوانی : $a = \dfrac{L_k + U_k}{2}$

فاصله طبقات : I

ب) واریانس را با مرکز طبقات جدید محاسبه می‌کنیم:

$$\sigma^2_{X'} = \frac{\sum F_i (x'_i)^2}{N} - (\mu_{X'})^2 = \frac{\sum F_i (x'_i)^2}{N} - \left(\frac{\sum F_i x'_i}{N}\right)^2$$

ج) با توجه به خواص واریانس، واریانس داده‌های اصلی (σ^2_X) برابر است با:

$$\sigma^2_{X'} = \sigma^2\left(\frac{x_i - a}{I}\right) = \frac{1}{I^2}\sigma^2_X \longrightarrow \sigma^2_X = \sigma^2_{X'} \times I^2$$

مثال ۵ انحراف معیار داده‌های جدول زیر کدام است؟

$C-L$	$20-30$	$30-40$	$40-50$
F_i	10	25	15

۱) 0.7
۲) 7
۳) 0.4
۴) 4

حل: گزینه ۲ درست است.
برای محاسبه انحراف معیار، ابتدا باید واریانس داده‌ها را به دست آوریم.
الف) طبقه $(30-40)$ دارای بیشترین فراوانی است؛ بنابراین جدول به صورت زیر خواهد بود:

x'	-1	$\boxed{0}$	1
$C-L$	$20-30$	$\boxed{30-40}$	$40-50$
F_i	10	$\boxed{25}$	15

ب) واریانس را با مرکز طبقات جدید محاسبه می‌کنیم:

x'	-1	0	1	
F_i	10	25	15	$N = \sum F_i = 50$
$F_i x'_i$	-10	0	15	$\sum F_i x'_i = 5$
$F_i (x'_i)^2$	10	0	15	$\sum F_i (x'_i)^2 = 25$

$$\sigma_{X'}^2=\frac{\sum F_i(x_i')^2}{N}-\left(\frac{\sum F_i x_i'}{N}\right)^2=\frac{25}{50}-\left(\frac{5}{50}\right)^2=0.5-0.01=0.49$$

ج) واریانس داده‌های اصلی $\left(\sigma_X^2\right)$ برابر است با:

$$\begin{cases}\sigma_X^2=\sigma_{X'}^2\times I^2 \quad\to\quad \sigma_X^2=(0.49)\times(10)^2=49\\ \sigma_{X'}^2=0.49 \ \ ,\ \ I=10\end{cases}$$

بنابراین انحراف معیار داده‌ها برابر است با:

$$\sigma_X=\sqrt{\sigma_X^2}=\sqrt{49}=7$$

محاسبه واریانس نمونه

با توجه به رابطه واریانس جامعه به صورت زیر:

$$\sigma^2=\frac{\sum(x_i-\mu)^2}{N}=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2=\frac{1}{N}\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{N}\right]$$

اگر برای یک نمونه n تایی، $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$ باشد، واریانس $x_1,\dots,x_n$ برابر است با:

$$\boxed{S^2=\frac{\sum(x_i-\bar{x})^2}{n-1}=\frac{1}{n-1}\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right]}$$

نکته:

۱- اگر S_x^2 واریانس یک نمونه n تایی باشد، آن‌گاه S_x انحراف معیار نمونه خواهد بود.

۲- تمام خواص واریانس جامعه $\left(\sigma^2\right)$ برای واریانس نمونه $\left(S^2\right)$ نیز صادق است.

مثال ۱ واریانس نمونه متشکل از سه عدد 567921120 , 567921124 , 567921122 کدام است؟ (اقتصاد ـ ۷۹)

۱) $\frac{8}{3}$ ۲) 4 ۳) $\frac{25124}{3}$ ۴) 25112

حل: گزینه ۲ درست است.

اولاً، در صورت سؤال کلمه «نمونه» ذکر شده، بنابراین منظور، واریانس نمونه S^2 است. ثانیاً، همان‌طور که در خاصیت سوم واریانس مطرح شد، اگر تمام داده‌ها دارای قسمت مشترکی باشند، برای محاسبه واریانس می‌توان بخش مشترک را حذف (داده‌ها را از عدد مشترک کم می‌کنیم) و واریانس بخش غیر مشترک را محاسبه کرد.

$$S^2(x-567921120)=S^2(x) \quad\to\quad x=0\ ,\ 4\ ,\ 2$$

بنابراین کافی است با حذف بخش مشترک (567921120) واریانس نمونه $0\,,\,4\,,\,2$ را به صورت زیر محاسبه کنیم:

$$\begin{cases}S^2=\frac{\sum(x_i-\bar{x})^2}{n-1}=\frac{(0-2)^2+(4-2)^2+(2-2)^2}{2}=4\\ \bar{x}=\frac{\sum x_i}{n}=\frac{0+2+4}{3}=2\end{cases}$$

اگر در صورت سؤال واریانس جامعه $\left(\sigma^2\right)$ خواسته شده بود، آن‌گاه:

$$\sigma^2 = \frac{\sum\left(x_i - \mu\right)^2}{N} = \frac{8}{3}$$

و در این صورت گزینه ۱ درست بود.

مثال ۲ در نمونه $x_1, \dots, x_{10}$ اگر $\sum_{i=1}^{10} x_i = 60$ و $\sum_{i=1}^{10} x_i^2 = 396$ باشد، واریانس کدام است؟

۱) 2   ۲) 3   ۳) 4   ۴) 5

حل: گزینه ۳ درست است.

با توجه به ذکر کلمه «نمونه» در صورت سؤال منظور واریانس نمونه $\left(S^2\right)$ است، بنابراین:

$$\begin{cases} S^2 = \dfrac{\sum\left(x_i - \bar{x}\right)^2}{n-1} = \dfrac{1}{n-1}\left[\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}\right] = \dfrac{1}{9}\left[396 - \dfrac{(60)^2}{10}\right] = 4 \\ \sum x_i^2 = 396 \,,\, \sum x_i = 60 \,,\, n = 10 \end{cases}$$

مثال ۳ اگر $y = 4 + 2u$ و $S_u^2 = 1$ باشد، مقدار S_y^2 کدام است؟

۱) 1   ۲) 2   ۳) 3   ۴) 4

حل: گزینه ۴ درست است.

همان‌طورکه می‌دانیم $\sigma^2 = \dfrac{\sum\left(x_i - \mu\right)^2}{N}$ (واریانس جامعه) و $S^2 = \dfrac{\sum(x_i - \bar{x})^2}{n-1}$ (واریانس نمونه) از نظر محاسبه با هم تفاوت دارند، اما خواص σ^2 و S^2 یکسان است؛ بنابراین:

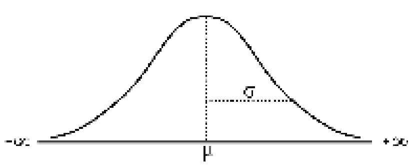
$$S^2\left(y\right) = S^2\left(4 + 2u\right) = 2^2 S_u^2 = 4 \times 1 = 4$$

کاربردهای انحراف معیار

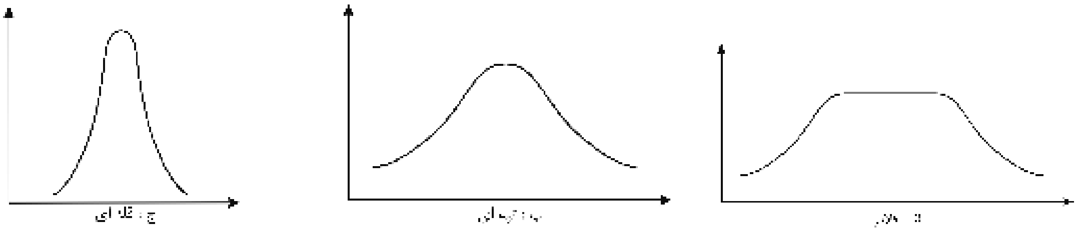
۱ـ درصدهای منحنی نرمال

اطلاعات آماری دنیای واقعی هرگز نمی‌توانند دقیقاً با یک منحنی فراوانی ریاضی نمایش داده شوند، ولی در هر حال انواع مختلف اطلاعات آماری چنان الگویی دارند که تقریباً دارای شکل نرمال (متقارن) است؛ بنابراین منحنی نرمال وسیله مهمی برای توصیف بسیاری از مجموعه‌های آماری است.

چنانکه می‌دانیم میانگین (μ) یک شاخص تمایل مرکزی و انحراف معیار (σ) شاخصی برای تعیین میزان پراکندگی در اطراف میانگین است. با توجه به شکل، هر دو شاخص μ و σ برای توصیف منحنی نرمال حیاتی هستند.

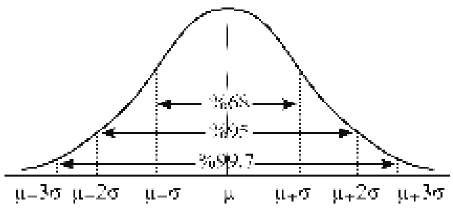


هرچند به طور عمومی یک شکل نرمال وجود دارد، اما این شکل به صورت‌های مختلف ظاهر می‌شود؛ به این معنی که منحنی‌های نرمال در میانگین (نقطه توازن) و پراکندگی (تغییرات) متفاوت‌اند. در شکل زیر سه توزیع با کشیدگی‌های متفاوت رسم شده است؛ منحنی (الف) را اصطلاحاً «فلاتی»، (ب) را «تپه‌ای» و (ج) را «قله‌ای» گویند.



اگر منحنی نرمال پراکندگی زیادی داشته باشد، شبیه منحنی (الف) خواهد بود و هرچه پراکندگی کمتری داشته باشد، به شکل (ج) نزدیک‌تر خواهد شد.

در هر منحنی نرمال حدود 68% سطح زیر منحنی بین فاصله یک انحراف معیار به طرف چپ میانگین تا یک انحراف معیار به طرف راست میانگین قرار دارد؛ یعنی حدود 68% از مشاهدات در فاصله $\mu\pm\sigma$ قرار می‌گیرند (از $\mu-\sigma$ تا $\mu+\sigma$).

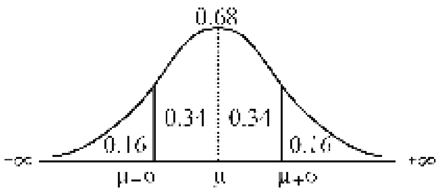


اگر این فاصله به دو انحراف معیار در اطراف میانگین گسترش یابد (یعنی $\mu\pm2\sigma$)، آن‌گاه نزدیک به 95% سطح زیر منحنی در این فاصله قرار می‌گیرد و اگر به اندازه سه انحراف معیار به طرفین میانگین برویم، سطح زیر منحنی نرمال در این فاصله 99.7% خواهد بود. به عبارت دیگر تقریباً تمام مشاهدات مربوط به یک جامعه نرمال در فاصله $\mu\pm3\sigma$ قرار دارند.

مثال اگر توزیع نمرات دانشجویان در یک کلاس 50 نفری تقریباً نرمال باشد و آنانی که نمره‌ای کمتر از $\mu-\sigma$ گرفته‌اند مردود اعلام شوند، تقریباً چند نفر در این کلاس مردود اعلام می‌شوند؟

۱) 5 ۲) 8 ۳) 12 ۴) 16

حل: گزینه ۲ درست است.



مطابق شکل مقابل، 68% سطح زیر منحنی نرمال در فاصله ($\mu-\sigma$ تا $\mu+\sigma$) قرار می‌گیرد و خارج از این فاصله، 32% سطح زیر منحنی را تشکیل می‌دهد که با توجه به تقارن منحنی، 16% از آن مربوط به $x<\mu-\sigma$ و 16% دیگر مربوط به $x>\mu+\sigma$ است.

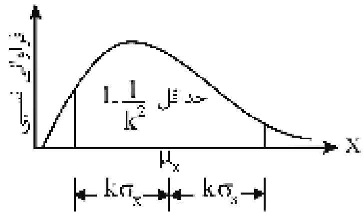
بنابراین داریم:

$$\text{تعداد دانشجویان مردودی} = 50\times \underbrace{P\left(x<\mu-\sigma\right)}_{\approx 0.16}\approx 8$$

۲ـ قضیه چی‌بی‌شف (Chebysheff)

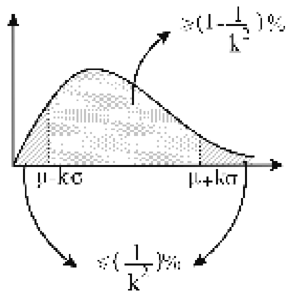
درصدهای تقریبی منحنی نرمال برای همه مجموعه‌های آماری صادق نیستند. چی‌بی‌شف برای یافتن قاعده‌ای که همواره بتواند مورد استفاده قرار گیرد، قضیه‌ای بیان کرد که برای سطوح جوامع نامعلوم (غیرنرمال) نیز حدودی به دست می‌آید.

قضیه: اگر $x_1, x_2, \dots, x_N$، $\left(x_i\geq 0\right)$، N مشاهده از **جامعه‌ای پیوسته و غیرنرمال (نامعلوم)** با میانگین μ_x و واریانس σ_x^2 باشند، آن‌گاه حداقل $\left(1-\frac{1}{k^2}\right)$ درصد از مشاهدات در دامنه (فاصله) k $\left(k\geq1\right)$ انحراف معیار از میانگین قرار دارد.



$$\boxed{P\left(\mu-k\sigma\le x\le\mu+k\sigma\right)\ge 1-\frac{1}{k^2}}$$

نتیجه: با توجه به درستی قضیه بالا، بدیهی است دامنه مشاهدات خارج از فاصله مطرح‌شده «**حداکثر** $\frac{1}{k^2}$ **درصد**» خواهد بود تا مجموع کل مشاهدات $1=(\%100)$ شود.



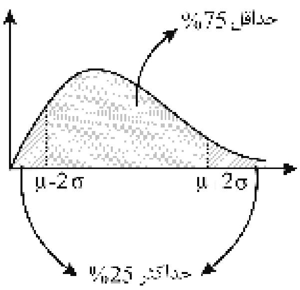
$$\begin{cases} P(\ \mu-k\sigma\le\ x\ \le\ \ \mu+k\sigma)\ge 1-\dfrac{1}{k^2} & \textbf{(حداقل)}\\ P\left(x\le\mu-k\sigma\quad \text{or}\quad x\ge\mu+k\sigma\right)\le\ \dfrac{1}{k^2} & \textbf{(حداکثر)}\end{cases}$$

مثال ۱ در یک جامعه غیرنرمال، چند درصد از مشاهدات در دامنه $\mu\pm2\sigma$ قرار دارد؟ (مدیریت ـ ۷۲)

۱) $\%95.44$ ۲) حداقل $\%75$ ۳) حداکثر $\%25$ ۴) حداقل $\frac{8}{9}$

حل: گزینه ۲ درست است.

از آنجاکه جامعه، غیرنرمال است:



$$\mu\pm2\sigma\ \xrightarrow{k=2}\ 1-\frac{1}{k^2}=1-\frac{1}{2^2}=\frac{3}{4}=\%75\ \ \text{حداقل}$$

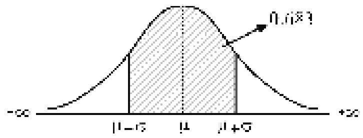
اگر در سؤال درصد مشاهدات خارج دامنه مطرح می‌شد، گزینه ۳ درست بود و اگر جامعه نرمال ذکر می‌شد، گزینه ۱ درست بود.

مقایسه درصد انحرافات جامعه غیرنرمال و نرمال

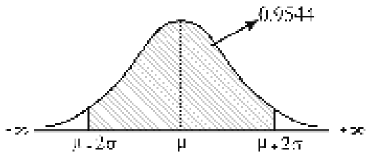
جامعه غیرنرمال: بنا بر قضیه چی‌بی‌شف در یک جامعه غیرنرمال، **حداقل** $\left(1-\frac{1}{k^2}\right)\%$ از مشاهدات در دامنه (فاصله) $k\ (k\ge1)$ انحراف معیار از میانگین قرار دارد. در جدول زیر دیده می‌شود که با افزایش مقدار k، درصد مشاهدات به سمت $1=\%100$ میل می‌کند.

k	1	2	3	4	…
$\mu \pm k\sigma$	$\mu \pm \sigma$	$\mu \pm 2\sigma$	$\mu \pm 3\sigma$	$\mu \pm 4\sigma$	…
حداقل $\left(1-\frac{1}{k^2}\right)$	$0=\%0$	$\frac{3}{4}=\%75$	$\frac{8}{9}\simeq\%89$	$\frac{15}{16}\simeq\%94$	…

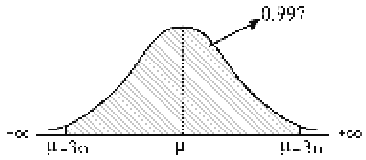
جامعه نرمال: در یک جامعه نرمال (زنگی، متقارن)، همواره درصد مشاهدات ثابتی در دامنه k انحراف معیار حول میانگین وجود دارد، که با افزایش مقدار k، به سمت $1=\%100$ میل می‌کند.



$\mu \pm \sigma \simeq \%68$ درصد مشاهدات



$\mu \pm 2\sigma \simeq \%95.44$ درصد مشاهدات



$\mu \pm 3\sigma \simeq \%99.7$ درصد مشاهدات

نتیجه: همان‌طورکه دیده می‌شود، درصد مشاهدات انحرافات حول میانگین در جامعه نرمال به ازای هر k، مقداری بیش از حداقل مقدار جامعه غیرنرمال در ناحیه مشابه است. این نشان می‌دهد که ذکر کلمه «حداقل» در جامعه غیرنرمال به این دلیل است که قابل تعمیم به همه توزیع‌ها از جمله نرمال باشد.

روش حل مسایل کاربردی چی‌بی‌شف

هرگاه به یک مسئله کاربردی چی‌بی‌شف برخورد کردیم، با توجه به نوع سؤال مطرح شده به یکی از سه روش زیر عمل می‌کنیم :

الف) اگر در مسئله محاسبه احتمال (حداقل یا حداکثر درصد مشاهدات) مورد نظر باشد، ابتدا با برقراری رابطه زیر مقدار k (تعداد انحراف از میانگین) را محاسبه می‌کنیم:

$$\begin{cases} \underbrace{\mu - k\sigma}_{a} \\ \underbrace{\mu + k\sigma}_{b} \end{cases} \longrightarrow \frac{b-a}{2} = k\sigma$$

سپس با توجه به سؤال، مقدار احتمال حداقل $1-\frac{1}{k^2}$ یا حداکثر $\frac{1}{k^2}$ خواهد بود.

ب) اگر در مسئله محاسبه حدودی مورد نظر باشد که مقدار احتمال آن‌را به صورت حداقل $1-\frac{1}{k^2}$ یا حداکثر $\frac{1}{k^2}$ بیان کرده است، ابتدا مقدار k را محاسبه می‌کنیم، سپس به کمک μ, σ داده شده در مسئله، حدود خواسته شده را به دست می‌آوریم:

درصد داده‌شده $\geq 1-\dfrac{1}{k^2} \xrightarrow{k=?}$ حدود خواسته‌شده : $(\mu-k\sigma \leq x \leq \mu+k\sigma)$

درصد داده‌شده $\leq \dfrac{1}{k^2} \xrightarrow{k=?}$ حدود خواسته‌شده : $(x \leq \mu-k\sigma \quad or \quad x \geq \mu+k\sigma)$

ج) گاهی هم محدوده و هم احتمال (حداقل یا حداکثر درصد مشاهدات) داده شده و یکی از مقادیر μ یا σ خواسته می‌شود؛ در این حالت برای محاسبه هر یک به صورت زیر عمل می‌کنیم:

محاسبه میانگین (μ):

$$\begin{cases} \underbrace{\mu-k\sigma}_{a} \\ \underbrace{\mu+k\sigma}_{b} \end{cases} \longrightarrow \mu=\frac{a+b}{2}$$

محاسبه انحراف معیار (σ): ابتدا به کمک احتمال داده‌شده به صورت حداقل $1-\dfrac{1}{k^2}$ یا حداکثر $\dfrac{1}{k^2}$ مقدار k را محاسبه کرده، سپس به کمک حدود داده‌شده به شکل زیر مقدار σ را تعیین می‌کنیم:

$$\begin{cases} \underbrace{\mu-k\sigma}_{a} \\ \underbrace{\mu+k\sigma}_{b} \end{cases} \longrightarrow \frac{b-a}{2}=k\sigma \longrightarrow \sigma=?$$

مثال ۲ تعداد مشتریانی که روزانه به یک فروشگاه مراجعه می‌کنند دارای میانگین 24 و انحراف معیار 4 نفر است. در یک روز خاص، احتمال اینکه حداقل بین 16 تا 32 نفر مشتری به فروشگاه مراجعه کنند، چقدر است؟ (اقتصاد ـ ۸۶)

۱) 0.25 ۲) 0.68 ۳) 0.75 ۴) 0.95

حل: گزینه ۳ درست است.

با توجه به حالت الف داریم:

اولاً:

$$\frac{b-a}{2}=k\sigma \rightarrow \frac{32-16}{2}=4k \rightarrow k=2$$

ثانیاً:

$$P(\mu-k\sigma<x<\mu+k\sigma)\geq 1-\frac{1}{k^2} \rightarrow 1-\frac{1}{k^2}=1-\frac{1}{2^2}=\frac{3}{4}=\%75$$ حداقل

مثال ۳ متوسط دستمزد روزانه کارگران یک کارخانه 10 هزار تومان با انحراف معیار یک هزار تومان است. چه نسبتی از کارگران دارای دستمزد روزانه‌ای بیشتر از 12 هزار تومان یا کمتر از 8 هزار تومان هستند؟ (اقتصاد ـ ۸۸)

۱) حداکثر %25 ۲) حداقل %25 ۳) حداکثر %75 ۴) حداقل %75

حل: گزینه ۱ درست است.

با توجه به نامعلوم بودن توزیع جامعه و همچنین با مشاهده گزینه‌ها که احتمال حداقل یا حداکثر را بیان می‌کنند از قضیه چی‌بی‌شف استفاده می‌کنیم:

$$P\left(x>\underbrace{\mu+k\sigma}_{b} \quad or \quad x<\underbrace{\mu-k\sigma}_{a}\right)\leq\left(\frac{1}{k^2}\right)$$ حداکثر

در این سؤال نیز:

$$P(x>12 \quad \text{یا} \quad x<8) \le \frac{1}{k^2} \quad \text{حداکثر}$$

حال با توجه به حالت الف داریم:

اولاً:

$$k\sigma=\frac{b-a}{2} \xrightarrow[\sigma^2=1]{b=12,a=8} k\times 1=\frac{12-8}{2} \to k=2$$

ثانیاً:

$$\frac{1}{k^2}=\frac{1}{2^2}=\frac{1}{4} \to P(x>12 \quad \text{یا} \quad x<8) \le \frac{1}{4}=\%25 \quad \text{حداکثر}$$

مثال ۴ در یک کارخانه کمپوت‌سازی میانگین وزن محصولات 250 گرم و واریانس 2.25 محاسبه شده است. طبق قانون چی‌بی‌شف وزن حداقل 64 درصد از این نوع محصولات در کدام بازه قرار دارد؟ (مدیریت ـ ۸۴)

۱) $(245\ ,255)$ ۲) $(246,254)$ ۳) $(247.5,252.5)$ ۴) $(247.75,\ 252.5)$

حل: گزینه ۳ درست است.

با توجه به روش ب به صورت زیر عمل می‌کنیم:

اولاً:

$$P(\mu-k\sigma\le x\le \mu+k\sigma)\ge 1-\frac{1}{k^2} \to 1-\frac{1}{k^2}=\frac{64}{100} \to \frac{1}{k^2}=\frac{36}{100} \to k=\frac{10}{6}=\frac{5}{3}$$

ثانیاً:

$$x\in(\mu-k\sigma\ ,\ \mu+k\sigma) \xrightarrow[k=\frac{5}{3}]{\mu=250\ ,\sigma=\frac{3}{2}} \left(250-\frac{5}{3}\times\frac{3}{2}\ ,\ 250+\frac{5}{3}\times\frac{3}{2}\right) \to (247.5\ ,\ 252.5)$$

مثال ۵ طبق قانون چی‌بی‌شف انتظار می‌رود 84 درصد مشاهدات در دامنه $(72,88)$ قرار گیرند. مقدار انحراف معیار این مشاهدات کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) 2.4 ۲) 3.2 ۳) 3.6 ۴) 4.2

حل: گزینه ۲ درست است.

با توجه به روش ج به صورت زیر عمل می‌کنیم:

اولاً:

$$1-\frac{1}{k^2}=\frac{84}{100} \to \frac{1}{k^2}=\frac{16}{100} \to k=\frac{10}{4}=\frac{5}{2}$$

ثانیاً:

$$\frac{b-a}{2}=k\sigma \to \frac{88-72}{2}=\frac{5}{2}\sigma \to 8=\frac{5}{2}\sigma \to \sigma=\frac{16}{5}=3.2$$

نمایش‌های دیگر از قضیه چی‌بی‌شف

در بعضی منابع و سؤالات، قضیه چی‌بی‌شف به صورت‌های دیگری نمایش داده می‌شوند که در ادامه آن‌ها را بررسی می‌کنیم:

نمایش اول

با توجه به آنکه همواره رابطه زیر برقرار است:

$$\mu-k\sigma\le x\le\mu+k\sigma \to -k\sigma\le x-\mu\le k\sigma \to |x-\mu|\le k$$

$$x\le\mu-k\sigma \quad \text{or} \quad x\ge\mu+k\sigma \to x-\mu\le -k\sigma \quad \text{or} \quad x-\mu\ge +k\sigma \to |x-\mu|\ge k\sigma$$

قضیه چی‌بی‌شف را می‌توان به صورت زیر نیز در نظر گرفت:

$$\boxed{\begin{array}{l} P(|x-\mu| \le k\sigma) \ge 1-\frac{1}{k^2} \\ P(|x-\mu| \ge k\sigma) \le \frac{1}{k^2} \end{array}}$$

چی‌بی‌شف در جامعه استاندارد

در یک جامعه استاندارد $\mu = 0$ و $\sigma^2 = 1$ است. حال اگر قضیه چی‌بی‌شف را برای یک متغیر استاندارد با استفاده از نمایش اول به کار ببریم، خواهیم داشت:

$$P(|x-\mu| \le k\sigma) \ge 1-\frac{1}{k^2} \xrightarrow{\mu=0,\sigma=1} \boxed{P(|x| \le k) \ge 1-\frac{1}{k^2}}$$

احتمال آنکه قدرمطلق یک متغیر استاندارد حداکثر k باشد، حداقل $1-\frac{1}{k^2}$ است.

$$P(|x-\mu| \ge k\sigma) \le \frac{1}{k^2} \xrightarrow{\mu=0,\sigma=1} \boxed{P(|x| \ge k) \le \frac{1}{k^2}}$$

احتمال آنکه قدرمطلق یک متغیر استاندارد حداقل k باشد، حداکثر $\frac{1}{k^2}$ است.

مثال ۶ بنا به نابرابری چه‌بی‌شف (Chebysheff) احتمال اینکه قدرمطلق یک متغیر استاندارد شده حداقل k باشد:(اقتصاد ـ ۷۹)

۱) برابر $\frac{1}{k^2}$ است.
۲) برابر k^2 است به شرط آنکه توزیع متغیر نرمال باشد.
۳) حداکثر $\frac{1}{k^2}$ است.
۴) حداکثر $\frac{1}{k^2}$ است به شرط آنکه توزیع متغیر نرمال باشد.

حل: گزینه ۳ درست است.

مثال ۷ اگر متغیر تصادفی x دارای میانگین 8 و واریانس 9 باشد برای $P(|x-8| \ge 6)$ داریم: (محیط زیست ـ ۸۴)

۱) $p \le \frac{1}{4}$
۲) $p \le \frac{3}{4}$
۳) $p \ge \frac{1}{4}$
۴) $p \ge \frac{3}{4}$

حل: گزینه ۱ درست است.

طبق نمایش اول قضیه چی‌بی‌شف داریم:

$$\begin{cases} P(|x-\mu| \ge k\sigma) \le \frac{1}{k^2} \xrightarrow{\mu=8\ ,\ k\sigma=6} P(|x-8| \ge 6) \le \frac{1}{4} \\ k\sigma = 6 \xrightarrow{\sigma=3} k = 2 \end{cases}$$

مثال ۸ چنانچه $E(x) = 2$ و $E(x^2) = 8$ باشد، بر اساس نابرابری چه‌بی‌شف، $P(|x| \ge 8)$ برابر است با: (اقتصاد ـ ۷۵)

۱) حداکثر $\frac{1}{9}$
۲) حداکثر $\frac{1}{4}$
۳) برابر $\frac{1}{9}$
۴) حداقل $\frac{1}{9}$

حل: گزینه ۱ درست است.

با فرض آنکه:

$$\begin{cases} \mu = E(x) = 2 \\ \sigma^2 = E(x^2) - E(x)^2 = 8 - 4 = 4 \quad \to \quad \sigma = 2 \end{cases}$$

با توجه به نمایش اول قضیه چی‌بی‌شف، داریم:

$$P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

درنتیجه برای این مسئله داریم:

$$\begin{cases} P(|x|>8) = P(|x-\mu|>8-\mu) = P(|x-\mu|>6) \leq \frac{1}{3^2} = \frac{1}{9} \quad \text{حداکثر} \\ k\sigma = 6 \xrightarrow{\sigma=2} k = 3 \end{cases}$$

نمایش دوم

درصورتی‌که در نمایش اول از قضیه چی‌بی‌شف $k\sigma = \varepsilon$ در نظر بگیریم، آن‌گاه $k = \frac{\varepsilon}{\sigma}$ و $\frac{1}{k^2} = \frac{\sigma^2}{\varepsilon^2}$ است و داریم:

$$\begin{matrix} P(|x-\mu| \leq k\sigma) \geq 1 - \frac{1}{k^2} \\ P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2} \end{matrix} \longrightarrow \boxed{\begin{matrix} P(|x-\mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2} \\ P(|x-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \end{matrix}}$$

مثال ۹ متغیر تصادفی X دارای میانگین صفر و واریانس 4 است. کران بالایی برای $P(|X| \geq 3)$ کدام است؟

۱) $\frac{1}{3}$ ۲) $\frac{2}{3}$ ۳) $\frac{4}{9}$ ۴) $\frac{5}{9}$

حل: گزینه ۳ درست است.

با توجه به نمایش دوم قضیه چی‌بی‌شف داریم:

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \xrightarrow[\sigma^2=4]{\varepsilon=3\,,\,\mu=0} P(|X-0| \geq 3) \leq \frac{4}{3^2} = \frac{4}{9}$$

نقاط قوت و ضعف نامساوی چی‌بی‌شف

این نامساوی قادر است درصد مشاهدات قرارگرفته در فاصله k انحراف معیار از میانگین را بدون توجه به توزیع مشاهدات (نرمال یا غیرنرمال) بیان کند اما درصد تعیین‌شده توسط این نامساوی دقت کافی را ندارد، چون با کلمه حداقل همراه است؛ «حداقل 75% از مشاهدات در فاصله دو انحراف معیار از میانگین قرار دارند». کلمه حداقل، از 75% تا 100% مشاهدات را شامل می‌شود و یک احتمال دقیق را بیان نمی‌کند.

۳ـ متغیر استاندارد Z

تعریف: در هر جامعه آماری هرگاه مقدار ثابت میانگین (μ) را از تک‌تک مشاهدات (x) کم کنیم و سپس حاصل را بر مقدار ثابت انحراف معیار (σ) تقسیم کنیم به متغیر جدیدی می‌رسیم که میانگین آن (0) و انحراف معیار آن 1 است. این متغیر را اصطلاحاً متغیر استاندارد می‌نامند و با Z نشان می‌دهند.

$$\text{متغیر استاندارد}: Z=\frac{x-\mu}{\sigma}$$

$$\begin{cases}\mu_Z=\mu\left(\frac{x}{\sigma}-\frac{\mu}{\sigma}\right)=\frac{1}{\sigma}\mu-\frac{\mu}{\sigma}=0\\ \sigma_Z=\sigma\left(\frac{x}{\sigma}-\cancel{\frac{\overset{0}{\mu}}{\sigma}}\right)=\frac{1}{\sigma}\sigma=1\end{cases}$$

✓ **دقت کنید!**

هر متغیر استاندارد (Z) نشان می‌دهد که مشاهده (x) در چند انحراف معیار از میانگین بیشتر است، زیرا:

$$Z=\frac{x-\mu}{\sigma}\quad\rightarrow\quad x=\mu+Z\sigma$$

کاربرد متغیر استاندارد

هر متغیر استاندارد یک معیار نسبی و درنتیجه بدون واحد است. به همین دلیل از آن برای مقایسه دو جامعه با توزیع مختلف استفاده می‌کنیم.

مثال ۱۰ دانشجویی در درس اقتصاد نمره 17 گرفته و میانگین و انحراف معیار نمرات کلاس به ترتیب 15 و 4 است. وی در امتحان درس آمار که دارای میانگین 16 و انحراف معیار 6 بوده، نمره 18 گرفته است. در کدام درس نمره وی به نسبت بهتر بوده است؟

حل:

ابتدا مقدار استاندارد هر یک از نمرات دانشجو را در دروس اقتصاد و آمار به دست می‌آوریم:

$$\text{اقتصاد}: Z=\frac{17-15}{4}=0.5$$

$$\text{آمار}: Z=\frac{18-16}{6}=0.33$$

همان‌طورکه مشاهده می‌شود، نمره این دانشجو در درس اقتصاد 0.5 انحراف معیار بالای میانگین است درحالی‌که در درس آمار 0.33 انحراف معیار بالای میانگین است؛ بنابراین نمره او در درس اقتصاد به نسبت بهتر بوده است.

تأثیر تغییرات جامعه بر متغیر استاندارد

هرگاه Z، اندازه استانداردشده متغیر x از یک جامعه با استفاده از رابطه $Z_x=\frac{x-\mu}{\sigma}$ باشد، آن‌گاه هرگونه **تغییر در** x **هیچ تأثیری بر** Z **(استاندارد) ندارد**، تنها در بعضی شرایط علامت آن را عوض می‌کند.

به تمام داده‌ها ثابت a اضافه یا از آن کم شود.	$1)\ Z_{x\pm a}=Z_x$
تمام داده‌ها در ثابت a ضرب یا بر آن تقسیم شود.	$2)\ Z_{\frac{x}{a}}=Z_{ax}=\begin{cases}Z_x & (\text{مثبت } a)\\ -Z_x & (\text{منفی } a)\end{cases}$
a درصد هر داده به آن اضافه یا از آن کم شود.	$3)\ Z_{x\pm\frac{a}{100}x}=Z_x$

مثال ۱۱ اگر اندازه x از جامعه‌ای برحسب متغیر استاندارد $z_x = 10$ باشد، درصورتی‌که هر یک از تغییرات زیر در جامعه اعمال شود، مقدار استاندارد شده Z چه تغییری می‌کند؟

الف) به تمام داده‌ها 2 واحد اضافه یا از آن کم کنیم.

ب) به تمام داده‌ها 0.20 هر داده را اضافه یا از آن کم کنیم.

ج) تمام داده‌ها را در 2 ضرب یا برآن تقسیم کنیم.

د) تمام داده‌ها را در -2 ضرب یا بر آن تقسیم کنیم.

حل:

الف)

$Z_{x\pm2} = Z_x = 10$ (تغییری نمی‌کند)

زیرا:

$$\begin{cases} Z_{x\pm2} = \dfrac{(x\pm2)-\mu(x\pm2)}{\sigma(x\pm2)} = \dfrac{(x\pm2)-(\mu\pm2)}{\sigma} = \dfrac{x-\mu}{\sigma} = Z_x = 10 \\ \mu(x\pm2) = \mu\pm2 \\ \sigma(x\cancelto{0}{\pm2}) = \sigma \end{cases}$$

ب)

$Z_{x\pm0.20x} = Z_x = 10$ (تغییری نمی‌کند)

زیرا:

$$Z_{x+0.2x} = Z_{1.2x} = \frac{1.2x-\mu(1.2x)}{\sigma(1.2x)} = \frac{1.2x-1.2\mu}{1.2\sigma} = \frac{x-\mu}{\sigma} = Z_x = 10$$

$$Z_{x-0.2x} = Z_{0.8x} = \frac{0.8x-\mu(0.8x)}{\sigma(0.8x)} = \frac{0.8x-0.8\mu}{0.8\sigma} = \frac{x-\mu}{\sigma} = Z_x = 10$$

ج)

$Z_{2x} = Z_{\frac{x}{2}} = Z_x = 10$ (تغییری نمی‌کند)

د)

$Z_{-2x} = Z_{\frac{x}{-2}} = -Z_x = -10$ (بدون تغییر منفی می‌شود)

زیرا:

$$Z_{2x} = \frac{2x-\mu(2x)}{\sigma(2x)} = \frac{2x-2\mu}{2\sigma} = \frac{x-\mu}{\sigma} = Z_x = 10$$

$$Z_{-2x} = \frac{-2x-\mu(-2x)}{\sigma(-2x)} = \frac{-2x-(-2\mu)}{|-2|\sigma} = -\frac{x-\mu}{\sigma} = -Z_x = -10$$

واریانس و تصحیح شپارد (Sheppard Correction)

هنگام محاسبه واریانس در داده‌های طبقه‌بندی شده، از مرکز طبقات به جای x_i ها (داده‌های واقعی) استفاده می‌کنیم؛ در این حالت واریانس به‌دست‌آمده ممکن است با واریانس مربوط به داده‌های واقعی (x_i ها) اختلاف داشته باشد. تأثیر این خطا روی میانگین و واریانس به شکل زیر بررسی می‌شود:

الف) در میانگین، خطای ناشی از میانگین به‌دست‌آمده به علت مثبت و منفی بودن خطاها جبران می‌شود. به همین دلیل از مجموع خطاها صرف‌نظر می‌کنیم.

ب) در واریانس، چون مجذور (توان دو) خطاهای مثبت و منفی در نظر گرفته می‌شود، خطاها یکدیگر را خنثی نمی‌کنند؛ بنابراین واریانس به‌دست‌آمده، از واریانس واقعی بزرگ‌تر است و پراکندگی را بیشتر نشان می‌دهد.

تصحیح شپارد

شپارد پیشنهاد کرده است برای تصحیح خطای ناشی از محاسبه واریانس در شرایط ذکرشده، مقدار $\frac{I^2}{12}$ ($I^2 =$ مجذور فاصله طبقات) را از واریانس به‌دست‌آمده $\left(\sigma_x^2\right)$ کم کنیم تا به مقدار دقیق‌تری برای واریانس برسیم؛ این واریانس را **واریانس تصحیح شده** می‌نامند و آن را با σ_c^2 نشان می‌دهند:

$$\text{واریانس شپارد}: \sigma_c^2 = \sigma_x^2 - \frac{I^2}{12}$$

شرایط استفاده از تصحیح شپارد

از تصحیح شپارد برای محاسبه واریانس N داده طبقه‌بندی شده در شرایط زیر استفاده می‌شود:

۱ـ متغیر x پیوسته باشد.

۲ـ تعداد داده‌ها (N) حداقل 1000 باشد $(N \geq 1000)$.

۳ـ توزیع فراوانی جامعه متقارن یا اندکی متقارن باشد.

مثال ۱ تصحیح شپارد در مورد واریانس N داده از متغییر پیوسته در مواردی به کار می‌رود که $N > \cdots$ و تابع توزیع فراوانی .... باشد. (برنامه ریزی شهری ـ ۸۵)

۱) 100 ـ اندکی متقارن
۲) 100 ـ غیر متقارن
۳) 1000 ـ اندکی متقارن
۴) 1000 ـ غیر متقارن

حل: گزینه ۳ درست است.

مثال ۲ در 1500 داده آماری پیوسته طبقه‌بندی شده با فاصله طبقات 3، مقدار واریانس برابر 7 محاسبه شده است. نمودار توزیع فراوانی آن متقارن است. انحراف معیار تصحیح‌شده طبق پیشنهاد شپارد، کدام است؟ (سنجش از دور ـ ۸۶)

۱) 2.2
۲) 2.4
۳) 2.5
۴) 2.6

حل: گزینه ۳ درست است.

$$\begin{cases} \sigma^2_{\text{شپارد}} = \sigma^2 - \frac{I^2}{12} = 7 - \frac{3^2}{12} = 7 - 0.75 = 6.25 \rightarrow \sigma_{\text{شپارد}} = \sqrt{6.25} = 2.5 \\ \text{فاصله طبقات } I = 3 \end{cases}$$

میانگین و واریانس کل چند جامعه آماری

اگر k جامعه با تعداد مشاهدات $N_1, \ldots, N_k$، با میانگین‌های $\mu_1, \ldots, \mu_k$ و واریانس‌های $\sigma_1^2, \ldots, \sigma_k^2$ مفروض باشند، آن‌گاه میانگین کل (μ) و واریانس کل $\left(\sigma^2\right)$ داده‌های آماری به صورت زیر محاسبه می‌شوند:

میانگین کل

$$\mu = \frac{N_1\mu_1 + \ldots + N_k\mu_k}{N_1 + N_2 + \ldots + N_k} = \frac{\sum N_i\mu_i}{N}$$

$$\sum N_i = N_1 + \ldots + N_k = N$$

تحلیل حالات

۱ـ اگر میانگین k جامعه با هم برابر باشد، آن‌گاه میانگین کل نیز با آن‌ها برابر می‌شود.

$$\mu=\mu_1=\mu_2=\cdots=\mu_k$$

۲ـ اگر تعداد نمونه‌ها در k جامعه برابر باشد، آن‌گاه میانگین کل برابر با میانگین حسابی k جامعه می‌شود.

$$\mu=\frac{\sum_{i=1}^{k}\mu_i}{k}$$

واریانس کل

$$\boxed{\begin{array}{l}\sigma^2=\dfrac{\sum N_i\sigma_i^2}{N}+\dfrac{\sum N_i(\mu_i-\mu)^2}{N}\\ \sum N_i=N_1+\ldots+N_k=N\end{array}}$$

$\dfrac{\sum N_i(\mu_i-\mu)^2}{N}$: میانگین مجذور انحرافات میانگین‌ها از میانگین کل

$\dfrac{\sum N_i\sigma_i^2}{N}$: میانگین واریانس جوامع

✓ دقت کنید!

همان‌طور که مشاهده می‌شود، قبل از محاسبه واریانس کل (σ^2) ابتدا باید میانگین کل (μ) محاسبه شود.

تحلیل حالات

۱- اگر میانگین k جامعه با هم برابر باشد، آن‌گاه $\dfrac{\sum N_i(\mu_i-\mu)^2}{N}=0$ و واریانس کل همان میانگین واریانس k جامعه است.

$$\sigma^2=\frac{\sum N_i\sigma_i^2}{N}$$

۲- اگر حداقل دو میانگین در جوامع با هم متفاوت باشند، آن‌گاه $\dfrac{\sum N_i(\mu_i-\mu)^2}{N}>0$ و واریانس کل از میانگین واریانس جوامع بزرگ‌تر است.

$$\sigma^2=\frac{\sum N_i\sigma_i^2}{N}+\underbrace{\frac{\sum N_i(\mu_i-\mu)^2}{N}}_{>0}\rightarrow\ \sigma^2>\frac{\sum N_i\sigma_i^2}{N}$$

نتیجه: هرگاه چند جامعه به صورت یک جامعه کل ترکیب شوند، واریانس این جامعه بزرگ‌تر از میانگین واریانس جوامع تشکیل‌دهنده آن خواهد بود و فقط درصورتی‌که میانگین جوامع با هم برابر باشند، واریانس جامعه کل با میانگین واریانس جوامع تشکیل‌دهنده برابر می‌شود؛ یعنی:

$$\sigma^2_{\text{کل}}\geq\frac{\sum_{i=1}^{k}N_i\sigma_i^2}{N}\qquad;\qquad N=N_1+N_2+\ldots+N_k$$

مثال ۱ در دو نمونه 10 تایی و 30 تایی، میانگین‌ها به ترتیب برابر 1 و 5 است. میانگین کل برای نمونه 40 تایی چند است؟

۱) 2.5 ۲) 3 ۳) 3.5 ۴) 4

حل: گزینه ۴ درست است.

$$\begin{cases} \mu = \dfrac{n_1\mu_1 + n_2\mu_2}{n_1+n_2} = \dfrac{10\times1+30\times5}{10+30} = 4 \\ \mu_1 = 1\ ,\ \mu_2 = 5\ ;\ n_1 = 10\ ,\ n_2 = 30 \end{cases}$$

مثال ۲ میانگین نمرات آمار و واریانس آن‌ها در دو کلاس به صورت زیر است:

کلاس	1	2
تعداد دانشجو (N_i)	20	30
میانگین نمرات (μ_i)	15	10
واریانس نمرات (σ_i^2)	17	12

میانگین و واریانس نمرات کل دانشجویان دو کلاس چقدر است؟ (اقتصاد ـ ۸۱)

۱) 12.5 و 20 ۲) 12 و 20 ۳) 12.5 و 35 ۴) 12 و 35

حل: گزینه ۲ درست است.

$$\mu_{\text{کل}} = \frac{N_1\mu_1 + N_2\mu_2}{N_1+N_2} = \frac{20\times15+30\times10}{20+30} = 12$$

$$\sigma^2_{\text{کل}} = \frac{\sum_{i=1}^{k} N_i\sigma_i^2}{N} + \frac{\sum_{i=1}^{k} N_i(\mu_i-\mu)^2}{N} = \frac{20\times17+30\times12}{20+30} + \frac{20(15-12)^2+30(10-12)^2}{20+30}$$

$$= \frac{340+360}{50} + \frac{180+120}{50} = 14+6 = 20$$

نیمه‌واریانس (Semivariance)

نیمه‌واریانس یکی از شاخص‌های پراکندگی است که برای استخراج انحرافات نامناسب و نامطلوب به کار می‌رود. برای مثال در داده‌های مربوط به سود، مقادیر کوچک‌تر از میانگین و در داده‌های مربوط به زیان، مقادیر بزرگ‌تر از میانگین نامطلوب هستند.

تعریف: متوسط مجذور مقادیر نامطلوب را نیمه‌واریانس می‌نامند و آن را با S.V نشان می‌دهند.

$$S.V = \frac{\sum_{i=1}^{k}(x_i-\mu)^2}{K}$$

در این رابطه فرض بر آن است که مشاهدات $x_1,\dots,x_N$ در جامعه موجود هستند و فقط k تای آن‌ها $(k<N)$ نامطلوب هستند.

(البته برای محاسبه میانگین داده‌ها از رابطه $\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$ استفاده می‌کنیم.)

مثال مشاهدات $7, 5, 3, 1$ را در نظر بگیرید. اگر انحراف بالای میانگین نامطلوب باشد، نیمه‌واریانس کدام است؟

۱) 4 ۲) 5 ۳) 3 ۴) صفر

حل: گزینه ۲ درست است.

$$\mu = \frac{\sum x_i}{N} = \frac{1+3+5+7}{4} = 4$$

با توجه به مقدار میانگین $(\mu = 4)$، مشاهده می‌شود که داده‌های 5 و 7 بالای میانگین هستند و داده‌های نامطلوب محسوب می‌شوند؛ درنتیجه واریانس این داده‌ها همان نیمه‌واریانس است.

$$S.V = \frac{\sum_{k=1}^{2}\left(x_i - \mu\right)^2}{k} = \frac{(5-4)^2 + (7-4)^2}{2} = 5$$

معیارهای پراکندگی نسبی

گاهی معیارهای پراکندگی (که تا به حال بررسی شدند) نمی‌توانند برای بررسی و مقایسه جوامع به کار روند. در چنین شرایطی از معیارهای پراکندگی نسبی استفاده می‌شود که عبارت‌اند از:

۱ـ ضریب پراکندگی (ضریب تغییرات) (CV)

۲ـ ضریب چولگی (انحراف از تقارن) (Sk)

۳ـ ضریب کشیدگی (E)

شرایطی که هر یک از این معیارها در آن به‌کار می‌روند به صورت خلاصه عبارتند از:

CV $\rightarrow$ واحدهای اندازه‌گیری متفاوت یا $\mu_1 \neq \mu_2$
$\begin{cases} \sigma_1 = \sigma_2 \rightarrow \mathbf{Sk} \\ \sigma_1 \neq \sigma_2 \rightarrow \mathbf{E} \end{cases}$ $\rightarrow$ واحدهای اندازه‌گیری یکسان و $\mu_1 = \mu_2$

ضریب پراکندگی (ضریب تغییرات) (Coefficient of Variation)

فرض کنید می‌خواهیم در یکی از دو شرکت تجاری A و B سرمایه‌گذاری کنیم. درصورتی‌که میانگین سود آن‌ها در سال‌های قبل برابر باشد، سرمایه‌گذاری در شرکتی پرمخاطره‌تر است که انحراف معیار بیشتری دارد، اما اگر میانگین سود یا واحد اندازه‌گیری آن‌ها یکسان نباشد مقایسه ریسک سرمایه‌گذاری در دو شرکت با ضریب پراکندگی انجام می‌شود.

تعریف: حاصل تقسیم انحراف معیار به میانگین را ضریب تغییرات می‌نامند و آن را با CV نشان می‌دهند.

$$CV = \frac{\sigma}{\mu}$$

ضریب تغییرات، یک نسبت بدون واحد و مقدار آن یک عدد حقیقی است $(-\infty < CV < +\infty)$.

برای محاسبه **درصد ضریب تغییرات** کافی است نسبت به‌دست‌آمده از محاسبه $CV = \frac{\sigma}{\mu}$ را در 100 ضرب کنیم؛ به عبارت دیگر:

$$\text{درصد ضریب تغییرات} = CV \times 100 = \frac{\sigma}{\mu} \times 100$$

مثال ۱ با فرض اینکه داشته باشیم $\sum_{i=1}^{3} x_i = 3$ و $\sum_{i=1}^{3} x_i^2 = 6$، ضریب تغییرات کدام است؟ (حسابداری ـ ۸۰)

۱) 0.5　　۲) 1　　۳) 1.5　　۴) 2

حل: گزینه ۲ درست است.

$$\begin{cases} CV = \frac{\sigma}{\mu} = \frac{1}{1} = 1 \\ \mu = \frac{\sum x_i}{N} = \frac{3}{3} = 1 \\ \sigma^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2 = \frac{6}{3} - 1^2 = 1 \quad \rightarrow \quad \sigma = 1 \end{cases}$$

مثال ۲ در 40 داده آماری مجموع داده‌ها برابر 100 و مجموع مجذورات آن‌ها 340 است. ضریب پراکندگی کدام است؟
(حسابداری و مدیریت ـ ۸۵)

۱) 0.4 ۲) 0.6 ۳) 0.8 ۴) 0.9

حل: گزینه ۲ درست است.

$$\begin{cases} CV_x = \dfrac{\sigma_x}{\mu_x} = \dfrac{3/2}{5/2} = \dfrac{3}{5} = 0.6 \\ \mu_x = \dfrac{\sum x_i}{N} = \dfrac{100}{40} = \dfrac{5}{2} \\ \sigma^2 = \dfrac{\sum x_i^2}{N} - \left(\dfrac{\sum x_i}{N}\right)^2 = \dfrac{340}{40} - \left(\dfrac{5}{2}\right)^2 = \dfrac{34}{4} - \dfrac{25}{4} = \dfrac{9}{4} \quad \rightarrow \quad \sigma = \dfrac{3}{2} \end{cases}$$

مثال ۳ میانگین سن یک گروه 12 سال و ضریب تغییرات سن آنان 20 درصد است. انحراف معیار سن آنان چقدر است؟
(مدیریت ـ ۷۹)

۱) 0.6 ۲) 2.4 ۳) 60 ۴) 240

حل: گزینه ۲ درست است.

با توجه به رابطه مربوط به درصد ضریب تغییرات داریم:

$$\text{درصد ضریب تغییرات}: CV \times 100 = \frac{\sigma}{\mu} \times 100 \quad \rightarrow \quad 20 = \frac{\sigma}{12} \times 100 \quad \rightarrow \quad \sigma = 2.4$$

خواص ضریب پراکندگی (تغییرات)

۱ـ ضریب تغییرات هر عدد ثابت (مانند a) برابر صفر است. $CV_a = 0$

مثال ۱ ضریب تغییرات (Coefficient of Variation) عدد 5 برابر است با: (اقتصاد ـ ۷۴)

۱) 5 ۲) 1 ۳) ∞ ۴) 0

حل: گزینه ۴ درست است.

واریانس هر عدد ثابت برابر صفر است؛ درنتیجه انحراف معیار آن نیز صفر می‌شود. میانگین یا امید ریاضی هر عدد نیز مساوی خود آن عدد است؛ بنابراین:

$$CV_5 = \frac{\cancelto{0}{\sigma(5)}}{\mu(5)} = \frac{0}{5} = 0$$

۲- درصورتی‌که تمام مشاهدات با هم برابر باشند، ضریب تغییرات برابر صفر می‌شود و برعکس.

$$x_1 = x_2 = \cdots = x_N = a \iff \begin{cases} CV = \dfrac{\sigma}{\mu} = 0 \\ \mu = a \quad , \quad \sigma = 0 \end{cases}$$

یادآوری: واریانس (σ^2) و انحراف معیار (σ) داده‌های مساوی، برابر صفر و میانگین آن‌ها برابر با همان مقدار مساوی است.

۳- اگر به تمام داده‌ها مقدار ثابتی (a) را اضافه یا از آن‌ها کم کنیم، ضریب تغییرات به صورت زیر محاسبه می‌شود:

$$CV_{x \pm a} = \frac{\sigma(x \cancelto{0}{\pm a})}{\mu(x \pm a)} = \frac{\sigma_x}{\mu_x \pm a}$$

برای محاسبه حتماً به مقادیر σ_x و μ_x نیاز داریم و داشتن مقدار CV_x کافی نیست.

مثال ۲ در جامعه‌ای ضریب تغییرات 0.6 است. اگر از تمام داده‌ها 1 واحد کم کنیم، ضریب تغییرات چقدر خواهد شد؟

۱) 0.6 ۲) 0.7 ۳) 0.5 ۴) قابل محاسبه نیست.

حل: گزینه ۴ درست است.

$$\begin{cases} CV_{x-1} = \dfrac{\sigma(x \overset{0}{\cancel{-1}})}{\mu(x-1)} = \dfrac{\sigma_x}{\mu_x - 1} = ? \\ CV_x = \dfrac{\sigma_x}{\mu_x} = 0.6 \end{cases}$$

با توجه به آنکه $CV_x = 0.6$ است و مقادیر σ_x و μ_x قابل محاسبه نیستند، امکان محاسبه CV_{x-1} در این وضعیت وجود ندارد.

مثال ۳ اگر میانگین و انحراف معیار متغیر تصادفی x به ترتیب 1 و 2 باشد، درصد ضریب تغییرات $y = x + 3$ چقدر است؟

۱) %25 ۲) %50 ۳) %75 ۴) %100

حل: گزینه ۲ درست است.

$$\begin{cases} CV_{x+3} = \dfrac{\sigma(x \overset{0}{\cancel{+3}})}{\mu(x+3)} = \dfrac{\sigma_x}{\mu_x + 3} = \dfrac{2}{1+3} = 0.5 \\ CV_{x+3} \times 100 = 50 \end{cases}$$

۴ـ اگر تمام داده‌ها را در عدد ثابتی مانند b ضرب یا بر آن تقسیم کنیم، مقدار ضریب تغییرات هیچ تغییری نمی‌کند و فقط علامت b را اختیار می‌کند.

$$\begin{cases} CV_{bx} = (\text{علامت } b)\,CV_x \\ CV_{\frac{x}{b}} = (\text{علامت } b)\,CV_x \end{cases}$$

مثال ۴ اگر در جامعه‌ای که مقدار ضریب تغییرات آن 73.1 درصد به دست آمده است، تمام متغیرها را در عدد ثابت 10 ضرب کنیم، ضریب تغییرات چقدر خواهد شد؟

۱) %7.31 ۲) %73.1

۳) %731 ۴) نمی‌توان تعیین کرد.

حل: گزینه ۲ درست است،

تغییری در ضریب تغییرات ایجاد نمی‌شود، زیرا:

$$\begin{cases} CV_{10x} = CV_x = \%73.1 \\ CV_x = \%73.1 \end{cases}$$

مثال ۵ ضریب تغییرات (CV) متغیر X برابر 0.2 است، اگر هر یک از مقادیر X را بر $a = 10$ تقسیم کنیم، ضریب تغییرات چه مقدار خواهد شد؟ (اقتصاد ـ ۸۳)

۱) 0.02 ۲) 0.2 ۳) 0.04 ۴) 2

حل: گزینه ۲ درست است.

تغییری در ضریب تغییرات ایجاد نمی‌شود، زیرا:

$$\begin{cases} CV_{\frac{x}{10}} = CV_x = 0.2 \\ CV_x = 0.2 \end{cases}$$

۵ـ هرگاه a درصد از هر داده به آن اضافه یا از آن کم شود، وضعیتی مانند خاصیت چهارم به وجود می‌آید و مقدار ضریب تغییرات هیچ تغییری نمی‌کند.

$$CV_{x \pm \frac{a}{100}x} = CV_{\left(1 \pm \frac{a}{100}\right)x} = CV_x$$

مثال ۶ در جامعه‌ای ضریب پراکندگی (تغییرات) برابر 70 درصد است. در صورت انجام هر یک از تغییرات زیر، ضریب پراکندگی چه تغییری می‌کند؟

الف) تمام داده‌ها را در عدد ثابت 10 ضرب یا بر آن تقسیم کنیم.

ب) به هر یک از داده‌ها عدد ثابت 10 را اضافه یا از آن کم کنیم.

ج) به هر یک از داده‌ها 0.20 خود داده را اضافه یا از آن کم کنیم.

حل: با توجه به آنکه $CV_x = 0.70$ است، داریم:

الف) بنا بر به خاصیت ۴ ضریب پراکندگی، مقدار ضریب پراکندگی تغییری نمی‌کند؛ به عبارت دیگر:

$$\begin{cases} CV_{10x} = CV_x = 0.70 \\ CV_{\frac{x}{10}} = CV_x = 0.70 \end{cases}$$

ب) بنا بر به خاصیت ۳ ضریب پراکندگی، به مقادیر μ_x و σ_x نیاز داریم؛ بنابراین با توجه به اطلاعات مسئله، مقدار ضریب پراکندگی قابل محاسبه نیست:

$$\begin{cases} CV_{x+10} = \dfrac{\sigma(x \overset{0}{\cancel{+10}})}{\mu(x+10)} = \dfrac{\sigma_x}{\mu_x + 10} = ? \\[2ex] CV_{x-10} = \dfrac{\sigma(x \overset{0}{\cancel{-10}})}{\mu(x-10)} = \dfrac{\sigma_x}{\mu_x - 10} = ? \end{cases}$$

ج) بنا بر به خاصیت ۵ ضریب پراکندگی، مقدار ضریب پراکندگی تغییری نمی‌کند؛ یعنی:

$$\begin{cases} CV_{(x+0.20x)} = CV_{(1.2x)} = CV_x = 0.70 \\ CV_{(x-0.20x)} = CV_{(0.8x)} = CV_x = 0.70 \end{cases}$$

کاربرد و موارد استفاده از ضریب پراکندگی

هنگام مقایسه دو جامعه از نظر ثبات و پراکندگی در شرایط زیر باید از ضریب پراکندگی (CV) استفاده کنیم، بدیهی است در این وضعیت واریانس و انحراف معیار به تنهایی کاربردی ندارند.

۱ـ دو یا چند جامعه هنگام مقایسه با هم، دارای مشاهدات ناهمگون (نامتجانس) از نظر واحد اندازه‌گیری باشند.

مثال ۱ برای تعیین آنکه در 30 روز گذشته، به نسبت، قیمت دلار از ثبات بیشتری برخوردار بوده است یا یورو، استفاده از کدام شاخص آماری مناسب‌تر است؟ (اقتصاد ـ ۸۲)

۱) انحراف متوسط ۲) ضریب پراکندگی ۳) ضریب چولگی ۴) واریانس

حل: گزینه ۲ درست است.
ازآنجاکه دلار و یورو دو واحد متفاوت هستند، از ضریب پراکندگی برای مقایسه آن‌ها با هم استفاده می‌شود.
در این حالت با استفاده از $CV_{\text{دلار}}$ و $CV_{\text{یورو}}$ دو نسبت بدون واحد را با هم مقایسه می‌کنیم.

تبصره: در شرایطی ممکن است مقیاس اندازه‌گیری در دو جامعه یکسان باشد، ولی بزرگی مشاهدات به طور قابل ملاحظه‌ای با هم اختلاف داشته باشد. برای مثال، در مقایسه پراکندگی میزان سود صنعت نفت و کشاورزی، با آنکه هر دو واحد یکسانی دارند اما درآمد حاصل از نفت با درآمد حاصل از کشاورزی قابل مقایسه نیست. در این وضعیت نیز باید از ضریب پراکندگی (CV) برای مقایسه استفاده کنیم.

۲- هرگاه دو یا چند جامعه میانگین‌های متفاوتی داشته باشند، در شرایطی که واریانس آن‌ها هم ممکن است یکسان باشد.
در این وضعیت برای مقایسه دو جامعه از نظر پراکندگی، استفاده از انحراف معیار بدون در نظر گرفتن میانگین امکان‌پذیر نیست. پراکندگی زمانی مفهوم دارد که نسبت به میانگین در نظر گرفته شود یعنی $CV = \frac{\sigma}{\mu}$ که معیار مناسبی برای بررسی پراکندگی دو جامعه در این شرایط است.

مثال ۲ ضریب تغییرات برای مقایسه تغییرات دو جامعه در حالت زیر به کار می‌رود: (مدیریت ـ ۷۳ ، حسابداری ـ ۷۸)
۱) میانگین‌های مختلف داشته باشد.
۲) دارای مقادیر منفی باشند.
۳) دارای دامنه تغییرات یکسان باشند.
۴) دارای میانگین‌های مختلف باشند و یا با واحدهای اندازه‌گیری مختلف اندازه‌گیری شده باشند.

حل: گزینه ۴ درست است.

معیار مقایسه دو جامعه با ضریب پراکندگی

درصورتی‌که دو یا چند جامعه، شرایط مطرح‌شده در قسمت قبل را داشته باشند، باید از ضریب پراکندگی برای مقایسه آن‌ها استفاده کرد. جامعه‌ای که ضریب پراکندگی (CV) کمتری دارد، از ثبات بیشتری برخوردار است و پراکندگی کمتری دارد.

مثال ۳ متوسط درآمد ماهانه کارگران کارخانه A، 17 هزار تومان با واریانس 4 است. در کارخانه B متوسط درآمد ماهانه 250 هزار ریال با واریانس 900 است. (مدیریت ـ ۷۴)
۱) اختلاف درآمد در کارخانه A بیش از کارخانه B است.
۲) اختلاف درآمد در کارخانه B بیش از کارخانه A است.
۳) درآمدهای اکثر افراد کارخانه A کم‌تر از اکثر افراد کارخانه B است.
۴) کمترین درآمد در کارخانه A بیش از کارخانه B است.

حل: گزینه ۲ درست است.
واحدهای اندازه‌گیری و میانگین‌ها برابر نیستند، بنابراین برای مقایسه بین دو جامعه از ضریب تغییرات استفاده می‌کنیم:

$$CV_A = \frac{\sigma_A}{\mu_A} = \frac{2}{17} \times 100 = 11.76$$

$$CV_B = \frac{\sigma_B}{\mu_B} = \frac{30}{250} \times 100 = 12$$

چون $CV_A < CV_B$ است، پراکندگی در کارخانه B بیشتر و اختلاف درآمد در آن بیش از کارخانه A است.

مثال ۴ میانگین حقوق کارکنان بخش فنی و شهرسازی شهرداری یک شهر 180000 تومان با انحراف معیار 4300 تومان و میانگین حقوق کارکنان بخش مالی آن شهرداری 160000 تومان با انحراف معیار 3200 تومان است. ثبات حقوق کدام بخش بیشتر است؟ و اختلاف درصد ضریب تغییر چیست؟ (برنامه‌ریزی شهری ـ ۸۴)

۱) فنی و 0.21 ۲) مالی و 0.21 ۳) فنی و 0.39 ۴) مالی و 0.39

حل: گزینه ۴ درست است.

واحدهای هر دو جامعه یکسان (تومان) اما میانگین‌های آن‌ها با هم متفاوت است؛ بنابراین برای مقایسه دو جامعه از ضریب تغییرات استفاده می‌کنیم؛ جامعه‌ای که مقدار CV در آن کمتر باشد، ثبات بیشتری خواهد داشت.

$$\begin{cases} CV_{\text{فنی}} \times 100 = \dfrac{\sigma}{\mu} \times 100 = \dfrac{4300}{180000} \times 100 = 2.39 \\ CV_{\text{مالی}} \times 100 = \dfrac{\sigma}{\mu} \times 100 = \dfrac{3200}{160000} \times 100 = 2 \end{cases}$$

$$CV_{\text{مالی}} < CV_{\text{فنی}} \rightarrow \text{مالی باثبات‌تر}$$

$$CV_{\text{فنی}} - CV_{\text{مالی}} = 0.39 \ (\text{درصد})$$

گشتاورها (Moments)

برای توضیح شاخص ضریب چولگی که یکی از معیارهای تعیین شکل توزیع از لحاظ قرینگی است، ابتدا باید مفهوم گشتاورها بیان شود.

گشتاورها به سه دسته تقسیم می‌شوند:

۱ـ گشتاور حول نقطه دلخواه a (گشتاورهای عمومی)

۲ـ گشتاور حول مبدأ صفر (گشتاورهای اولیه)

۳ـ گشتاور حول میانگین (گشتاورهای مرکزی)

۱ـ گشتاور حول نقطه دلخواه a (گشتاورهای عمومی)

گشتاور مرتبه nام نسبت به عدد دلخواه a $\left(M_n(a)\right)$ به صورت‌های زیر تعریف می‌شود:

$$\text{داده‌های طبقه‌بندی نشده}: M_n(a) = \frac{\sum (x_i - a)^n}{N}$$

$$\text{داده‌های طبقه‌بندی شده}: M_n(a) = \frac{\sum F_i (x_i - a)^n}{N} = \sum f_i (x_i - a)^n$$

اگر در گشتاورهای عمومی $n = 1, 2, 3, 4$ را جایگذاری کنیم، گشتاورهای مراتب اول، دوم، سوم و چهارم به دست می‌آیند:

$$n = 1 \quad ; \quad M_1(a) = \frac{\sum (x_i - a)}{N} \quad , \quad n = 2 \quad ; \quad M_2(a) = \frac{\sum (x_i - a)^2}{N}$$

$$n = 3 \quad ; \quad M_3(a) = \frac{\sum (x_i - a)^3}{N} \quad , \quad n = 4 \quad ; \quad M_4(a) = \frac{\sum (x_i - a)^4}{N}$$

۲- گشتاور حول مبدأ صفر (گشتاورهای اولیه)

گشتاور مرتبه nام نسبت به مبدأ (صفر) (m_n) به صورت زیر تعریف می‌شود:

$$\text{داده‌های طبقه‌بندی نشده}: \ m_n = \frac{\sum x_i^n}{N}$$

$$\text{داده‌های طبقه‌بندی شده}: \ m_n = \sum \frac{F_i x_i^n}{N} = \sum f_i x_i^n$$

با جایگذاری $a = 0$ در رابطه گشتاورهای عمومی، گشتاورهای اولیه به دست می‌آیند.

اگر در گشتاورهای اولیه $n = 1, 2, 3, 4$ را جایگذاری کنیم، گشتاورهای اولیه مراتب اول، دوم، سوم و چهارم به دست می‌آیند:

$$n = 1 \ ; \ m_1 = \frac{\sum x_i}{N} = \mu_x \quad , \quad n = 2 \ ; \ m_2 = \frac{\sum x_i^2}{N}$$

$$n = 3 \ ; \ m_3 = \frac{\sum x_i^3}{N} \quad , \quad n = 4 \ ; \ m_4 = \frac{\sum x_i^4}{N}$$

در حالت کلی می‌توان به این نتیجه رسید که:

$$m_n = E\left(x^n\right)$$

گشتاور مرتبه اول نسبت به مبدأ (صفر) برابر میانگین حسابی است، یعنی:

$$m_1 = \mu_x = \bar{x}$$

۳- گشتاور حول میانگین (گشتاورهای مرکزی)

گشتاور مرتبه nام نسبت به میانگین (μ_n) به صورت زیر تعریف می‌شود:

$$\text{داده‌های طبقه‌بندی نشده}: \mu_n = \frac{\sum (x_i - \mu)^n}{N}$$

$$\text{داده‌های طبقه‌بندی شده}: \mu_n = \frac{\sum F_i (x_i - \mu)^n}{N} = \sum f_i (x_i - \mu)^n$$

با جایگذاری $a = \mu$ در رابطه گشتاورهای عمومی، گشتاورهای مرکزی به دست می‌آیند.

گشتاور مرتبه اول نسبت به میانگین برابر صفر و گشتاور مرتبه دوم نسبت به میانگین برابر واریانس است، یعنی:

$$\mu_1 = 0$$

$$\mu_2 = \sigma_x^2 = Var(x)$$

بنابراین داریم:

$$\sigma_x^2 = \mu_2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2 = \frac{\sum x_i^2}{N} - \mu^2$$

مثال ۱ گشتاورهای مرتبه اول و دوم نسبت به مبدأ صفر به صورت $m_1 = 10$, $m_2 = 125$ به دست آمده‌اند. ضریب تغییرات صفت متغیر x چند درصد است؟

۱) 25　　۲) 50　　۳) 30　　۴) 60

حل: گزینه ۲ درست است.

$$CV = \frac{\sigma}{\mu} = \frac{5}{10} = \%50$$

$$\sigma^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2 = m_2 - m_1^2 = 125 - 10^2 = 25 \rightarrow \sigma = 5$$

$$\mu = \frac{\sum x_i}{N} = m_1 = 10$$

نکات مهم گشتاورها

۱- گشتاورهای آماری نسبت به مبدأ (صفر) و نسبت به میانگین حسابی، حالات خاصی از گشتاورهای آماری نسبت به نقطه دلخواه a هستند.

۲- به کمک گشتاورهای آماری می‌توان معیارها، پارامترها و روابط آماری را به طور ساده و خلاصه بیان کرد.

۳- گشتاور مرتبه اول نسبت به مبدأ (صفر) برابر میانگین حسابی است، یعنی $\boxed{m_1 = \bar{x}}$

۴- گشتاور مرتبه اول نسبت به میانگین برابر صفر است، یعنی $\boxed{\mu_1 = 0}$

۵- گشتاور مرتبه دوم نسبت به میانگین برابر واریانس است، یعنی $\boxed{\mu_2 = \sigma^2}$

۶- گشتاورهای آماری را می‌توان طبق روابطی به هم تبدیل کرد.

تبدیل گشتاورها به یکدیگر

الف) تبدیل گشتاورهای عمومی به گشتاورهای مرکزی

$$\boxed{\mu_n = (M - M_1)^n}$$

$$\mu_1 = (M - M_1)^1 = M_1 - M_1^1 = M_1 - M_1 = 0$$

$$\mu_2 = (M - M_1)^2 = M_2 - 2M_1M_1^1 + M_1^2 = M_2 - 2M_1^2 + M_1^2 = M_2 - M_1^2$$

$$\mu_3 = (M - M_1)^3 = M_3 - 3M_2M_1^1 + 3M_1M_1^2 - M_1^3 = M_3 - 3M_1M_2 + 2M_1^3$$

$$\mu_4 = (M - M_1)^4 = M_4 - 4M_3M_1^1 + 6M_2M_1^2 - 4M_1M_1^3 + M_1^4 = M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4$$

✓ **دقت کنید!**

همیشه در به توان رساندن گشتاورها، Mهایی که اندیس ندارند وقتی به توان می‌رسند، عدد توان به جای اندیسشان قرار می‌گیرد.

ب) تبدیل گشتاورهای عمومی به گشتاورهای اولیه

$$\boxed{m_n = (M + a)^n}$$

$$m_1 = (M + a) = M_1 + a$$

$$m_2 = (M + a)^2 = M_2 + 2aM_1 + a^2$$

$$m_3 = (M + a)^3 = M_3 + 3aM_2 + 3a^2M_1 + a^3$$

$$m_4 = (M + a)^4 = M_4 + 4aM_3 + 6a^2M_2 + 4a^3M_1 + a^4$$

ج) تبدیل گشتاورهای اولیه به گشتاورهای مرکزی

$$\boxed{\mu_n = (m - m_1)^n}$$

$$\mu_1 = (m - m_1)^1 = m_1 - m_1 = 0$$

$$\mu_2 = (m - m_1)^2 = m_2 - 2m_1m_1^1 + m_1^2 = m_2 - m_1^2 = \frac{\sum x_i^2}{N} - \bar{x}^2 = \sigma^2$$

$$\mu_3 = (m - m_1)^3 = m_3 - 3m_2m_1^1 + 3m_1m_1^2 - m_1^3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = (m - m_1)^4 = m_4 - 4m_3m_1^1 + 6m_2m_1^2 - 4m_1m_1^3 + m_1^4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

د) تبدیل گشتاورهای اولیه به گشتاورهای عمومی

$$\boxed{M_n = (m - a)^n}$$

$$M_1 = (m - a)^1 = m_1 - a$$

$$M_2 = (m - a)^2 = m_2 - 2am_1 + a^2$$

$$M_3 = (m - a)^3 = m_3 - 3am_2 + 3a^2m_1 - a^3$$

$$M_4 = (m - a)^4 = m_4 - 4am_3 + 6a^2m_2 - 4a^3m_1 + a^4$$

مثال ۲ در جامعه‌ای به حجم $n = 10$ کمیت‌های زیر محاسبه شده است:

$$\sum F_i x_i = 40 \ , \ \sum F_i x_i^2 = 400 \ , \ \sum F_i x_i^3 = 600$$

گشتاوری مرکزی مرتبه سوم (μ_3) کدام است؟ (مدیریت ـ ۷۱)

۱) -292 ۲) 292 ۳) -548 ۴) 485

حل: گزینه ۱ درست است.

با توجه به آنکه بین گشتاور مرکزی و گشتاور حول مبدأ، رابطه $\mu_n = (m - m_1)^n$ برقرار است، داریم:

$$m_1 = \frac{\sum F_i x_i}{N} = \frac{40}{10} = 4 \ \text{و} \ m_2 = \frac{\sum F_i x_i^2}{N} = \frac{400}{10} = 40 \ \text{و} \ m_3 = \frac{\sum F_i x_i^3}{N} = \frac{600}{10} = 60$$

بنابراین:

$$\mu_3 = m_3 + 2m_1^3 - 3m_2m_1 = 60 + 2(4)^3 - 3(40)(4) = 60 + 128 - 480 = -292$$

مثال ۳ گشتاورهای مرتبه اول و دوم نسبت به مبدأ $a = 40$ به صورت $M_1 = 2 \ , \ M_2 = 85$ به دست آمده‌اند. میانگین و واریانس جامعه کدام است؟

۱) 42 و 83 ۲) 45 و 81 ۳) 42 و 81 ۴) 45 و 83

حل: گزینه ۳ درست است.

برای به دست آوردن میانگین $\left(\mu = \frac{\sum x_i}{N}\right)$، به گشتاور مرتبه اول حول مبدأ صفر نیاز داریم، یعنی:

$$m_1 = \frac{\sum x_i}{N}$$

$$m_n = (M + a)^n \rightarrow m_1 = M_1 + a \rightarrow \mu = m_1 = \frac{\sum x_i}{N} = 2 + 40 = 42$$

برای به دست آوردن واریانس جامعه $\left(\sigma^2=\frac{\sum(x_i-\mu)^2}{N}\right)$، به گشتاور مرکزی مرتبه دوم نیاز داریم، یعنی:

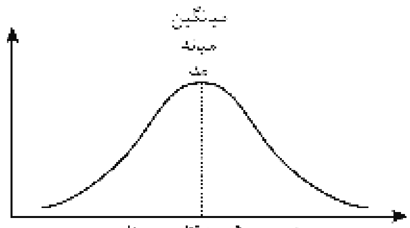
$$\mu_2=\frac{\sum(x_i-\mu)^2}{N}$$

$$\mu_n=(M-M_1)^n \rightarrow \mu_2=(M-M_1)^2=M_2-M_1^2 \rightarrow \sigma^2=\mu_2=\frac{\sum(x_i-\mu)^2}{N}=85-2^2=81$$

چولگی (انحراف از قرینگی) (Skewness)

در مقایسه دو یا چند جامعه با یکدیگر، ابتدا از پارامترهای مرکزی استفاده می‌شود، اما در صورت تساوی برخی از پارامترهای مرکزی (مانند میانگین)، اختلاف جوامع آماری به کمک شاخص‌های پراکندگی (مانند انحراف معیار) مشخص می‌شود. گاهی پارامترهای پراکندگی نیز به علت مساوی بودن، مناسب نیستند. برای رفع این مشکل از معیاری به نام **چولگی (انحراف از قرینگی)** استفاده می‌شود.

توزیع متقارن: هر توزیعی مانند **توزیع نرمال** که در آن پارامترهای مرکزی (مد، میانگین و میانه) با یکدیگر مساوی باشند، توزیع متقارن نامیده می‌شود.

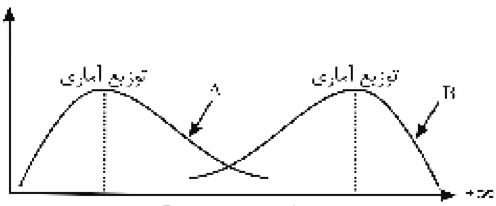


توزیع زنگی شکل و متقارن

هرچه یک توزیع با توزیع متقارن تفاوت بیشتری داشته باشد، معیار چولگی (انحراف از قرینگی) در آن بیشتر است.

هنگام مقایسه دو یا چند جامعه هرگاه **میانگین و واریانس آن‌ها باهم برابر** باشد، از معیاری به نام **چولگی (انحراف از قرینگی)** استفاده می‌کنیم. این معیار، **انحراف جوامع را از توزیع متقارن (نرمال)** نشان می‌دهد.

برای مثال، در شکل زیر دو توزیع دارای میانگین و واریانس مساوی هستند، اما توزیع یکسانی ندارند.



توزیع های نامتقارن A و B

در توزیع A تمایل جامعه به مقادیر کوچک‌تر است زیرا مد جامعه سمت چپ قرار دارد، درحالی‌که در توزیع B تمایل جامعه به مقادیر بزرگ‌تر است و درنتیجه مد جامعه در نقطه مقابل قرار دارد. این تفاوت را چولگی یا انحراف از قرینگی می‌گویند.

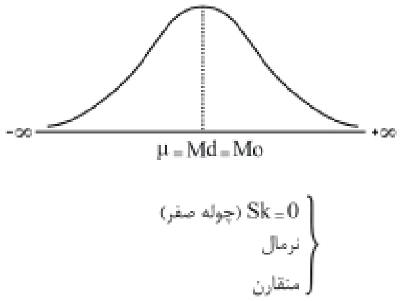
ضریب چولگی

تعریف: میزان انحراف هر توزیع از توزیع نرمال (متقارن) را ضریب چولگی می‌نامند و آن را با Sk یا α_3 نشان می‌دهند.

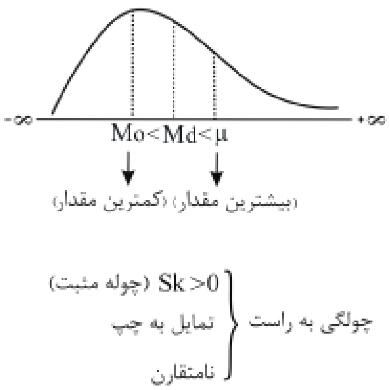
ضریب چولگی در توزیع نرمال (متقارن) برابر با صفر است $(\alpha_3=Sk=0)$، درنتیجه اگر توزیعی با آن تفاوت داشته باشد، آن‌گاه ضریب چولگی مخالف صفر و $Sk>0$ یا $Sk<0$ است.

انواع چولگی جوامع

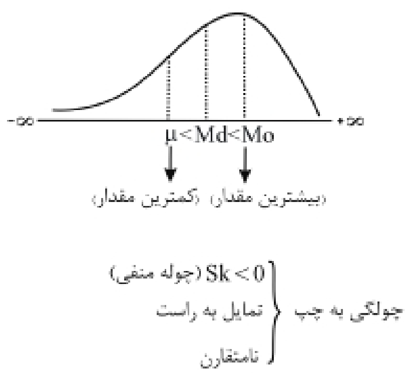
ازآنجاکه ضریب چولگی (Sk) به‌دست‌آمده از هر توزیع، می‌تواند « صفر، مثبت یا منفی» باشد ، انواع چولگی در جامعه به صورت زیر بررسی می‌شود:



الف) چوله صفر (متقارن)
$Mo = Md = \mu$
میانگین = میانه = مد (نماد)



ب) چولگی مثبت (چولگی به راست)
تمایل به چپ (تمایل منحنی به مقادیر کوچکتر)
(بیشترین) $Mo < Md < \mu$ (کمترین)
میانگین > میانه > مد (نما)
راست وسط چپ



ج) چولگی منفی (چولگی به چپ)
تمایل به راست (تمایل منحنی به مقادیر بزرگتر)
(بیشترین) $\mu < Md < Mo$ (کمترین)
مد (نما) > میانه > میانگین
راست وسط چپ

مثال ۱ در یک توزیع متمایل به راست، کدام گزینه صحیح است؟ (حسابداری ـ ۷۸)

۱) $\mu_x > Md > Mo$ ۲) $\mu_x < Mo < Md$ ۳) $Mo > Md > \mu_x$ ۴) $Md < Mo < \mu_x$

حل: گزینه ۳ درست است.

زمانی که تمایل توزیع به سمت راست است، بدین معناست که توزیع چوله به چپ است.

مثال ۲ در توزیعی با چولگی منفی انتظار می‌رود که .... کمترین مقدار را داشته باشد. (اقتصاد ـ ۸۱)

۱) دامنه تغییرات ۲) میانه ۳) میانگین ۴) نما

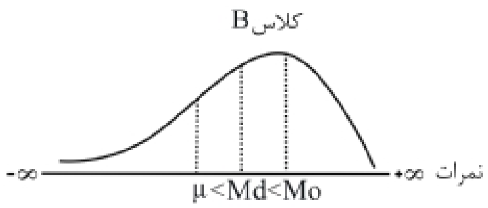
حل: گزینه ۳ درست است.

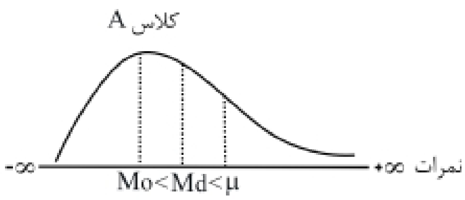
در جامعه‌ای با چولگی منفی همواره رابطه $\mu < Md < Mo$ برقرار است، بنابراین کمترین مقدار را «میانگین» و بیشترین مقدار را «مد (نما)» دارد.

مفهوم چولگی

برای درک بهتر مفهوم چولگی به دو مثال زیر توجه کنید.

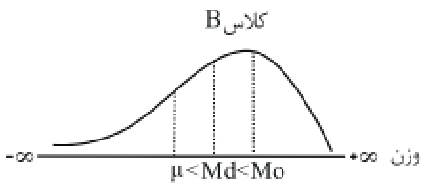
مثال هنگام مقایسه نمرات پایان ترم دو کلاس، با فرض آنکه میانگین و واریانس آن‌ها با هم برابر باشد، کلاسی بهتر است که در آن نمرات تمایل به مقدار بیشتر دارند.

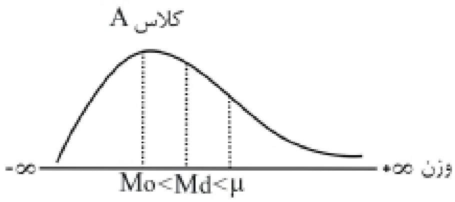




در شکل بالا کلاس B بهتر از کلاس A است، زیرا تمایل (سنگینی منحنی نمرات) به سمت نمرات بیشتر است، اما در کلاس A تمایل به سمت نمرات کمتر است.

مثال هنگام مقایسه لاغری دانش‌آموزان دو کلاس، با فرض آنکه میانگین و واریانس آن‌ها با هم برابر باشد، کلاسی بهتر است که در آن وزن‌ها تمایل به مقدار کمتر دارند.





در شکل بالا کلاس A بهتر از کلاس B است، زیرا تمایل (سنگینی منحنی وزن) به سمت وزن‌های کمتر است، اما در کلاس B تمایل به سمت وزن‌های بیشتر است.

مثال ۱ در توزیع درآمد خانوارهای ایران موقعیت میانگین و میانه چگونه است؟ (برنامه ریزی شهری ـ ۸۴)

۱) میانه از میانگین بزرگ‌تر است.

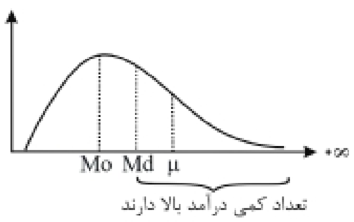
۲) میانگین بزرگ‌تر از میانه است.

۳) میانه و میانگین تقریباً بر روی هم قرار می‌گیرند.

۴) میانه و میانگین دقیقاً بر روی هم قرار می‌گیرند.

حل: گزینه ۲ درست است.

با توجه به اینکه در ایران تعداد خانوارهای پردرآمد کم و تعداد خانوارهای کم‌درآمد زیاد است، توزیع درآمد خانوارها تمایل به مقادیر کمتر دارد و درنتیجه چوله به راست (تمایل به چپ) و $Mo < Me < \mu$ است که نشان می‌دهد میانگین بزرگ‌تر از میانه است.



نتیجه: همان‌طورکه دیده شد نمی‌توان گفت کدام‌یک از چولگی‌ها (راست یا چپ) بهتر است زیرا صفت مورد اندازه‌گیری مشخص می‌کند کدام چولگی بهتر است.

چنانکه مشاهده شد:

در مثال اول، صفت مورد نظر «نمرات» بود که نمرات بیشتر ایده‌آل جامعه است.

در مثال دوم، صفت مورد نظر «رژیم لاغری» بود که وزن کمتر ایده‌آل جامعه است.

در مثال سوم، صفت مورد نظر «درآمد خانوارها» بود که درآمد بیشتر ایده‌آل جامعه است.

تفسیر قدرمطلق ضریب چولگی

علاوه بر علامت ضریب چولگی که نشان‌دهنده نوع چولگی هر توزیع است:

$(Sk > 0)$	$(Sk = 0)$	$(Sk < 0)$
چوله به راست	متقارن	چوله به چپ

قدرمطلق ضریب چولگی $|Sk|$ نشان‌دهنده میزان اختلاف قرینگی توزیع جامعه با توزیع نرمال است. بدیهی است هرچه $|Sk|$ بزرگ‌تر شود، میزان چولگی (عدم تقارن) بیشتر شده و اختلاف توزیع جامعه با نرمال زیاد خواهد بود.

برای بررسی $|Sk|$ استانداردهایی وجود دارد:

الف) $|Sk| \le 0.1$:

جامعه از نظر تقارن تقریباً نرمال (متقارن) است، تقریباً چولگی ندارد (چولگی قابل اغماض).

ب) $0.1 < |Sk| \le 0.5$:

جامعه از نظر تقارن تفاوت اندکی با توزیع نرمال دارد، چولگی اندک است (چولگی غیرقابل اغماض).

ج) $|Sk| > 0.5$:

جامعه از نظر تقارن تفاوت فاحشی با توزیع نرمال دارد، چولگی زیاد است (چولگی غیرقابل اغماض).

مثال ۲ اگر ضریب چولگی توزیع یک جامعه -0.66 باشد، آن‌گاه جامعه مورد مطالعه ... (حسابداری ـ ۸۳)

۱) نرمال است.
۲) با جامعه نرمال تفاوت مختصری دارد.
۳) با جامعه نرمال تفاوت فاحش دارد.
۴) با اطلاعات داده شده نمی‌توان قضاوت کرد.

حل: گزینه ۳ درست است.

با توجه به آنکه $Sk = -0.66 < 0$، چولگی منفی (چوله به چپ) است ولی از نظر قدرمطلق به صورت زیر تحلیل می‌شود:

$|Sk| = 0.66 \rightarrow |Sk| > 0.5 \rightarrow$ پس چولگی جامعه با جامعه نرمال تفاوت فاحش دارد.

محاسبه ضریب چولگی

همان‌طورکه گفته شد، شاخص اندازه‌گیری برای تعیین میزان انحراف توزیع‌ها از تقارن (نرمال)، ضریب چولگی (Sk) نامیده می‌شود. این معیار یک نسبت بدون واحد است و به صورت زیر محاسبه و تحلیل می‌شود.

برای محاسبه ضریب چولگی، روابط بسیاری وجود دارد که آن‌ها را بررسی می‌کنیم:

الف) ضریب چولگی گشتاوری	$Sk=\alpha_3=\frac{\mu_3}{\sigma^3}=\frac{\frac{\sum(x_i-\mu)^3}{N}}{\sigma^3}$
ب) ضریب چولگی پیرسون	$Sk_1=\frac{(\mu-Mo)}{\sigma}\ ,\ Sk_2=\frac{3(\mu-Md)}{\sigma}$
ج) ضریب چولگی چندکی (چارکی ـ صدکی)	$Sk_Q=\frac{Q_3-2Q_2+Q_1}{Q_3-Q_1}$

با توجه به داده‌های مسئله تشخیص می‌دهیم که از کدام‌یک از روابط بالا استفاده کنیم.

الف) ضریب چولگی گشتاوری

ضریب چولگی گشتاوری مهم‌ترین رابطه برای محاسبه ضریب چولگی است.

$$Sk=\alpha_3=\frac{\mu_3}{\sigma^3}=\frac{\frac{\sum(x_i-\mu)^3}{N}}{\sigma^3}=\frac{\frac{\sum F_i(x_i-\mu)^3}{N}}{\sigma^3}$$

در این رابطه داریم:

$$\begin{cases}\mu_3=\text{گشتاور مرکزی مرتبه سوم حول میانگین}=\frac{\sum(x_i-\mu)^3}{N}=\frac{\sum F_i(x_i-\mu)^3}{N}\\ \text{واریانس}:\sigma^2=\mu_2=\text{گشتاور مرکزی مرتبه دوم حول میانگین}=\frac{\sum(x_i-\mu)^2}{N}=\frac{\sum F_i(x_i-\mu)^2}{N}\\ \sigma^2=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2=\frac{\sum F_i x_i^2}{N}-\left(\frac{\sum F_i x_i}{N}\right)^2\\ \text{انحراف معیار}:\sigma=\sqrt{\sigma^2}\end{cases}$$

مثال ۱ اگر $n=50$, $\sum x_i^2=3250$ و $\mu_x=7$ و $\sum(x_i-\mu_x)^3=96$ باشد، ضریب چولگی جامعه کدام است؟

(مدیریت ـ ۷۳)

۱) $\%3$ ۲) $\%6$ ۳) $\%1.96$ ۴) $\%2.3$

حل: گزینه ۱ درست است.

$$\begin{cases}Sk=\frac{\frac{\sum(x_i-\mu_x)^3}{N}}{\sigma^3}=\frac{\frac{96}{50}}{4^3}=0.03=\%3\\ \sigma^2=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2=\frac{\sum x_i^2}{N}-\mu_x^2=\frac{3250}{50}-(7)^2=65-49=16\ \rightarrow\ \sigma=4\end{cases}$$

مثال ۲ گشتاورهای مراتب اول، دوم و سوم نسبت به مبدأ $a=10$ به صورت $M_1=2$ و $M_2=20$ و $M_3=88$ به دست آمده است. ضریب چولگی توزیع کدام است؟

۱) -0.25 ۲) 0.25 ۳) 0.5 ۴) -0.5

حل: گزینه ۱ درست است.

با توجه به اینکه برای محاسبه ضریب چولگی به گشتاور مرکزی مرتبه سوم $\left(\mu_3=\frac{\sum(x_i-\mu)^3}{N}\right)$ و گشتاور مرکزی مرتبه دوم $\left(\sigma^2=\mu_2=\frac{\sum(x_i-\mu)^2}{N}\right)$ نیاز داریم بنا برابطه تبدیل گشتاورهای عمومی به مرکزی، داریم:

$$\mu_n=\left(M-M_1\right)^n$$

$$\mu_1=\left(M-M_1\right)^1=M_1-M_1=0$$

$$\mu_2=\left(M-M_1\right)^2=M_2-M_1^2=20-2^2=16\rightarrow\sigma^2=16\rightarrow\sigma=4$$

$$\mu_3=\left(M-M_1\right)^3=M_3-3M_1M_2+2M_1^3=88-3\times2\times20+2\times2^3=-16$$

$$Sk=\frac{\mu_3}{\sigma^3}=\frac{-16}{4^3}=\frac{-1}{4}=-0.25$$

ب) ضریب چولگی پیرسون

پیرسون، دو رابطه تجربی برای محاسبه ضریب چولگی ارائه کرد:

$$Sk_1=\frac{(\mu-Mo)}{\sigma}\qquad \text{(رابطه اول)}$$

$$Sk_2=\frac{3(\mu-Md)}{\sigma}\qquad \text{(رابطه دوم)}$$

رابطه سه معیار مرکزی

تنها رابطه موجود بین سه پارامتر مرکزی میانگین (μ) و میانه (Md) و مد (Mo) بویژه زمانی که چولگی «خفیف، متعادل یا ضعیف» است، از تساوی دو رابطه پیرسون به‌دست می‌آید:

$$\boxed{\mu-Mo=3\left(\mu-Md\right)}$$

مثال ۱ در یک جامعه با میانگین 13 و واریانس 6.25 و مد 15 ، ضریب چولگی پیرسون کدام است؟

۱) -0.8 ۲) -0.4 ۳) 0.4 ۴) 0.8

حل: گزینه ۱ درست است.

$$Sk=\frac{(\mu-Mo)}{\sigma}=\frac{13-15}{2.5}=\frac{-2}{2.5}=-0.8$$

مثال ۲ در یک جامعه با میانگین و انحراف معیار 8 و 4، مقدار میانه برابر 10.5 است ضریب چولگی جامعه کدام است؟

۱) 2.25 ۲) -1.875 ۳) -2.25 ۴) 1.875

حل: گزینه ۲ درست است.

$$Sk=\frac{3(\mu-me)}{\sigma}=\frac{3(8-10.5)}{4}=-1.875$$

یادآوری: در بعضی از منابع به‌جای μ در تمام روابط بالا از $\bar{x}$ استفاده می‌کنند.

مثال ۳ در یک توزیع با چولگی خفیف، میانگین حسابی $\bar{X}=52.4$ و میانه $Me=51.8$ به دست آمده است. مد توزیع کدام است؟ (مدیریت ـ ۷۱ ، اقتصاد ـ ۷۴)

۱) 53.6 ۲) 50.6 ۳) 54.2 ۴) 51.6

حل: گزینه ۲ درست است.

همواره رابطه $\bar{x}-Mo=3(\bar{x}-Md)$ بین سه شاخص مرکزی برقرار است؛ به‌ویژه زمانی که توزیع دارای چولگی خفیف باشد؛ بنابراین:

$$(52.4-Mo)=3(52.4-51.8)\ \rightarrow\ Mo=50.6$$

ج) ضریب چولگی چندکی

همان‌طورکه در پارامتر پراکندگی «انحراف چارکی» توضیح داده شد، در بعضی توزیع‌ها امکان محاسبه میانگین و واریانس وجود ندارد یا محاسبه آن منطقی به نظر نمی‌رسد؛ در این وضعیت برای محاسبه چولگی، از پارامترهای چندکی مانند ضریب چولگی چارکی، دهکی، صدکی به صورت زیر استفاده می‌شود:

ضریب چولگی چارکی: $Sk_Q=\frac{Q_3-2Q_2+Q_1}{Q_3-Q_1}$

ضریب چولگی صدکی: $Sk_P=\frac{P_{90}-2P_{50}+P_{10}}{P_{90}-P_{10}}$

که البته در این روابط داریم:

$$\begin{cases} \text{(میانه) } Md = \text{(چارک دوم) } Q_2 = \text{(دهک پنجم) } D_5 = \text{(صدک پنجاهم) } P_{50} \\ \text{(دهک نهم) } D_9 = \text{(صدک نودم) } P_{90} \\ \text{(دهک اول) } D_1 = \text{(صدک دهم) } P_{10} \end{cases}$$

مثال ۱ فرض کنید میانه 50، دهک اول 10 و صدک نودم 90 باشد. ضریب چولگی کدام است؟

۱) $\frac{1}{2}$ ۲) 0 ۳) 0.44 ۴) 2

حل: گزینه ۲ درست است.

$$\begin{cases} Sk_P=\frac{P_{90}-2P_{50}+P_{10}}{P_{90}-P_{10}}=\frac{90-2\times 50+10}{90-10}=0 \\ \text{(صدک نودم) } P_{90}=90 \text{ ، (صدک دهم) } P_{10}=10 = \text{(دهک اول) } D_1 \\ \text{(صدک پنجاهم) } P_{50}=50 = \text{(میانه) } Md \end{cases}$$

مثال ۲ در یک توزیع فراوانی داده‌ها، چارک‌های اول، دوم و سوم به ترتیب 36، 61 و 76 است. نوع توزیع از نظر تقارن نسبت به نرمال چگونه است؟ (حسابداری و مدیریت ـ ۸۶)

۱) چوله به راست ـ تقریباً نرمال ۲) چوله به راست ـ تفاوت اندک

۳) چوله به چپ ـ تقریباً نرمال ۴) چوله به چپ ـ تفاوت اندک

حل: گزینه ۴ درست است.

با توجه به اینکه مقادیر چارک توزیع در مسئله داده شده است، برای محاسبه ضریب چولگی، از رابطه ضریب چولگی چندکی استفاده می‌کنیم:

$$\begin{cases} \text{(پیرسون)}\ Sk_Q = \dfrac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \dfrac{76 - 2\times 61 + 36}{76 - 36} = \dfrac{-10}{40} = -0.25 \\ Q_1 = 36 \quad , \quad Q_2 = 61 \quad , \quad Q_3 = 76 \end{cases}$$

با توجه به منفی بودن علامت ضریب چولگی، توزیع داده‌ها چوله به چپ است و چون $0.1 < |Sk| \le 0.5$ است، چولگی توزیع از نظر تقارن تفاوت اندکی با توزیع نرمال دارد.

تأثیر تغییر مشاهدات در چولگی

تغییر در مشاهدات، هیچ‌گونه تغییری در ضریب چولگی ایجاد نمی‌کند، فقط ممکن است علامت آن را عوض کند:

مشاهدات در a ضرب و با b جمع یا از آن کم شوند.	$Sk_{ax\pm b} = \begin{cases} Sk_x & a>0 \\ -Sk_x & a<0 \end{cases}$
a درصد به هر مشاهده اضافه یا از آن کم شود.	$Sk_{x\pm(\%a)x} = Sk_{(1\pm\%a)x} = Sk_x$

مثال ۱ اگر تمام داده‌ها با عدد 10 جمع و سپس بر -2 تقسیم شوند، ضریب چولگی داده‌ها چه تغییری می‌کند؟

۱) نصف می‌شود.
۲) قرینه می‌شود.
۳) به اندازه $\frac{1}{10}$ افزایش می‌یابد.
۴) به اندازه 10 واحد افزایش می‌یابد.

حل: گزینه ۲ درست است.

$$Sk_{\frac{x+10}{-2}} = Sk_{-\frac{x}{2}-5} = -Sk_x$$

مثال ۲ ضریب چولگی حقوق کارمندان یک شرکت -0.6 است. اگر 25% به حقوق کارکنان اضافه شود، ضریب چولگی حقوق‌های جدید کدام است؟

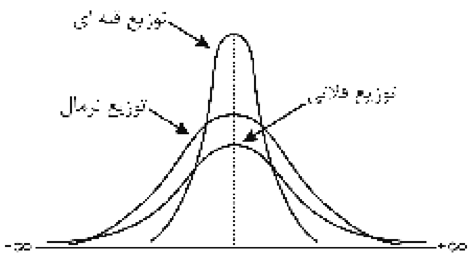
۱) 0.45
۲) -0.15
۳) -0.75
۴) -0.6

حل: گزینه ۴ درست است.

هرگاه a% به مشاهدات اضافه یا از آن‌ها کم شود، ضریب چولگی تغییری نمی‌کند.

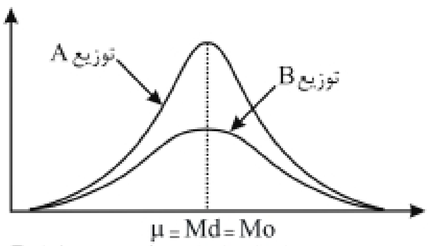
کشیدگی (Kurtosis)

هنگام مقایسه دو جامعه که توزیع آن‌ها متقارن و میانگین آن‌ها برابر است، توزیعی که منحنی کوتاه‌تری نسبت به نرمال دارد (فلاتی)، دارای پراکندگی بیشتر و توزیعی که منحنی بلندتری نسبت به نرمال دارد (قله‌ای)، دارای پراکندگی کمتری است. دو جامعه با این شرایط را می‌توان با استفاده از شاخص کشیدگی (نقطه اوج) با هم مقایسه کرد.



تعریف: اندازه بلندی (اوج) مربوط به منحنی هر توزیع، را شاخص کشیدگی (نقطه اوج) می‌نامند و آن را با α_4 نشان می‌دهند.

در مقایسه دو توزیع A و B با توجه به شکل روبرو:



پارامترهای مرکزی هر دو توزیع یکسان ولی کشیدگی توزیع A از B بیشتر است

مشاهده می‌شود که در هر دو توزیع A و B پارامترهای مرکزی (میانگین، میانه، مد) برابر هستند و چولگی آن‌ها نیز یکسان است (هر دو متقارن $Sk_A = Sk_B = 0$).

تنها تفاوت بین دو توزیع A و B میزان کشیدگی آن‌هاست؛ توزیع A کشیدگی بیشتری نسبت به توزیع B دارد، درنتیجه پراکندگی توزیع A کمتر از توزیع B و تمرکز آن حول میانگین، بیشتر از توزیع B است.

محاسبه شاخص کشیدگی

برای محاسبه شاخص کشیدگی (نقطه اوج) از رابطه زیر استفاده می‌کنیم:

$$\text{کشیدگی (گشتاوری)} : \alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{\frac{\sum (x_i-\mu)^4}{N}}{(\sigma^2)^2} = \frac{\frac{\sum F_i(x_i-\mu)^4}{N}}{(\sigma^2)^2}$$

که در آن:

$\mu_4 =$ گشتاور مرکزی مرتبه چهارم حول میانگین

$$\text{واریانس} : \sigma^2 = \mu_2 = \frac{\sum (x_i-\mu)^2}{N} = \frac{\sum F_i(x_i-\mu)^2}{N} = \sum f_i(x_i-\mu)^2$$

✓ **دقت کنید!**

۱- علت استفاده از اصطلاح «کشیدگی گشتاوری» آن است که در محاسبه آن از گشتاور μ_4 استفاده می‌شود.

۲- شاخص کشیدگی $\left(\frac{\mu_4}{\sigma^4}\right)$ یک عدد بدون واحد است.

کشیدگی منحنی نرمال

مقدار شاخص کشیدگی (نقطه اوج) برای منحنی متقارن نرمال (زنگی) برابر $\alpha_4 = 3$ است.

مثال ۱ کشیدگی گشتاور یک توزیع آماری، 3 است. کدام گزینه صحیح است؟ (مدیریت ـ ۷۵)

۱) پراکندگی آن در حد توزیع نرمال است.

۲) پراکندگی آن زیاد است.

۳) پراکندگی آن پایین است.

۴) جهت اظهار نظر به اطلاعات بیشتری نیاز است.

حل: گزینه ۱ درست است.

ضریب کشیدگی

تعریف: شاخص سنجش میزان پراکندگی جامعه را نسبت به توزیع نرمال که با استفاده از تفاضل کشیدگی آن‌ها به‌دست می‌آید، ضریب کشیدگی می‌نامند و با E نشان می‌دهند.

$$E = \text{کشیدگی توزیع جامعه} - \text{کشیدگی توزیع نرمال} = \frac{\mu_4}{\sigma^4} - 3$$

مثال ۲ اگر $N = 1000$, $\sum F_i \left(x_i - \mu_x\right)^4 = 5000$ و انحراف معیار جامعه 2 باشد، مقدار ضریب کشیدگی کدام است؟ (مدیریت ـ ۷۸)

۱) -2.69 ۲) -2.53 ۳) 0.31 ۴) 2.53

حل: گزینه ۱ درست است.

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{\frac{\sum F_i \left(x_i - \mu\right)^4}{N}}{\sigma^4} - 3 = \frac{\frac{5000}{1000}}{2^4} - 3 = -2.69$$

مثال ۳ گشتاورهای مرتبه اول تا چهارم نسبت به مبدأ $a = 3$ به صورت زیر استخراج شده‌اند. ضریب کشیدگی توزیع چقدر است؟

$M_1 = 0 \quad M_2 = 1.2 \quad M_3 = 0 \quad M_4 = 3.6$

۱) -1.5 ۲) -0.5 ۳) 0.5 ۴) 2.1

حل: گزینه ۲ درست است.

ازآنجاکه برای محاسبه ضریب کشیدگی به گشتاور مرکزی مرتبه چهارم $\left(\mu_4 = \frac{\sum (x_i - \mu)^4}{N}\right)$ و گشتاوری مرکزی مرتبه دوم $\left(\sigma^2 = \mu_2 = \frac{\sum (x_i - \mu)^2}{N}\right)$ نیاز داریم، از رابطه تبدیل گشتاورهای عمومی به مرکزی استفاده می‌کنیم:

$$\mu_n = \left(M - M_1\right)^n$$

$$\mu_1 = (M - M_1)^1 = M_1 - M_1 = 0$$

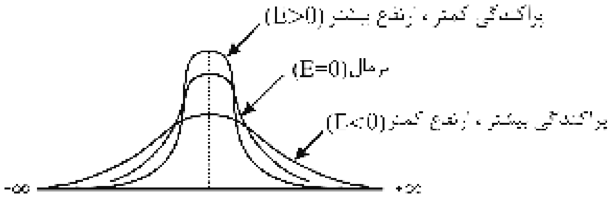
$$\mu_2 = (M - M_1)^2 = M_2 - M_1^2 = 1.2 - 0^2 = 1.2 \rightarrow \sigma^2 = 1.2$$

$$\mu_4 = \left(M - M_1\right)^4 = M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4 = 3.6$$

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{3.6}{(1.2)^2} - 3 = -0.5$$

تفسیر ضریب کشیدگی

با توجه به شکل زیر می‌توانیم مقدار ضریب کشیدگی (E) را تحلیل و بررسی کنیم:



الف) $E < 0$ (ضریب کشیدگی منفی)

۱- کشیدگی توزیع جامعه کوتاه‌تر و کمتر از کشیدگی توزیع نرمال است $(\alpha_4 < 3)$.

۲- پراکندگی توزیع جامعه از توزیع نرمال بیشتر است.

۳- تمرکز جامعه حول میانگین از توزیع نرمال کمتر است.

ب) $E = 0$ (ضریب کشیدگی صفر)

کشیدگی توزیع جامعه هم‌اندازه و برابر با کشیدگی توزیع نرمال است $(\alpha_4 = 3)$. این توزیع‌ها را توزیع‌هایی با کشیدگی متوسط می‌نامند.

ج) $E > 0$ (ضریب کشیدگی مثبت)

۱- کشیدگی توزیع جامعه بلندتر و بیشتر از توزیع نرمال است $(\alpha_4 > 3)$.

۲- پراکندگی توزیع جامعه از توزیع نرمال کمتر است.

۳- تمرکز جامعه حول میانگین از توزیع نرمال بیشتر است.

تفسیر قدرمطلق ضریب کشیدگی

بدون در نظر گرفتن علامت E که کوتاه‌تر یا بلندتر بودن توزیع جامعه را نسبت به توزیع نرمال نشان می‌دهد، قدرمطلق ضریب کشیدگی $|E|$، میزان اختلاف پراکندگی توزیع جامعه را با توزیع نرمال مورد بررسی قرار می‌دهد. هرچه $|E|$ بیشتر شود، تفاوت پراکندگی توزیع جامعه با توزیع نرمال نیز بیشتر می‌شود.

تفسیر قدرمطلق ضریب کشیدگی $|E|$ با توجه به استانداردهای مطرح‌شده به شرح زیر است:

الف) $|E| \leq 0.1$

توزیع جامعه از نظر پراکندگی و کشیدگی، تقریباً نرمال است (تفاوت قابل اغماض).

ب) $0.1 < |E| \leq 0.5$

توزیع جامعه از نظر پراکندگی و کشیدگی، تفاوت اندکی با توزیع نرمال دارد (تفاوت غیرقابل اغماض).

ج) $|E| > 0.5$

توزیع جامعه از نظر پراکندگی و کشیدگی، اختلاف فاحشی با توزیع نرمال دارد (تفاوت غیرقابل اغماض).

مثال ۴ در یک توزیع آماری ضریب کشیدگی برابر -0.08 محاسبه شده است. پراکندگی این توزیع و منحنی آن چگونه است؟ (حسابداری و مدیریت ـ ۸۵)

۱) تقریباً نرمال ـ بلندتر از نرمال
۲) تقریباً نرمال ـ کوتاه‌تر از نرمال
۳) تفاوت اندکی با نرمال ـ کوتاه‌تر از نرمال
۴) تفاوت اندکی با نرمال ـ بلندتر از نرمال

حل: گزینه ۲ درست است.

با توجه به علامت منفی ضریب کشیدگی $(E=-0.08<0)$ متوجه می‌شویم که منحنی توزیع کوتاه‌تر از نرمال بوده و چون $|E|\le 0.1$ است، پراکندگی توزیع تقریباً نرمال است.

کشیدگی چندکی

آن دسته از توزیع‌هایی که امکان محاسبه گشتاور برای آن‌ها وجود ندارد، با استفاده از چندک‌ها توصیف می‌شوند. (منظور جوامع نامتقارن است که از شاخص مرکزی $Q_2 = Md$ و شاخص پراکندگی انحراف چارکی $\left(SIQR=\frac{Q_3-Q_1}{2}\right)$ استفاده می‌شود.)

برای محاسبه شاخص کشیدگی و ضریب کشیدگی در این توزیع‌ها به صورت زیر عمل می‌کنیم.

$$\begin{cases} \text{کشیدگی چندکی} = \dfrac{SIQR}{P_{90}-P_{10}} \\ \text{کشیدگی چندکی توزیع نرمال} = 0.263 \\ \text{ضریب کشیدگی چندکی}: E_P = \dfrac{SIQR}{P_{90}-P_{10}} - 0.263 \end{cases}$$

در رابطه بالا:

$$SIQR=\frac{Q_3-Q_1}{2} = \text{انحراف چارکی} \qquad P_{90} = \text{صدک نودم} \qquad P_{10} = \text{صدک دهم}$$

نتیجه:

جامعه متقارن

$$\begin{cases} \text{کشیدگی گشتاوری} = \dfrac{\mu_4}{\sigma^4} = \dfrac{\dfrac{\sum F_i (x_i-\mu)^4}{N}}{(\sigma^2)^2} \\ \text{کشیدگی گشتاوری توزیع نرمال} = 3 \\ \text{ضریب کشیدگی گشتاور)}: E = \dfrac{\mu_4}{\sigma^4} - 3 \end{cases}$$

جامعه نامتقارن

$$\begin{cases} \text{کشیدگی چندکی} = \dfrac{SIQR}{P_{90}-P_{10}} = \dfrac{\dfrac{Q_3-Q_1}{2}}{P_{90}-P_{10}} \\ \text{کشیدگی چندکی نرمال} = 0.263 \\ \text{ضریب کشیدگی چندکی}: E_P = \dfrac{SIQR}{P_{90}-P_{10}} - 0.263 \end{cases}$$

مثال کشیدگی (Kurtosis) چندکی و گشتاوری توزیع نرمال به ترتیب (از چپ به راست) کدام است؟ (مدیریت ـ ۸۰)

۱) $(0\,,\,0)$ ۲) $(0.263\,,\,3)$ ۳) $(0.263\,,\,0.263)$ ۴) $(3\,,\,3)$

حل: گزینه ۲ درست است.

کشیدگی چندکی توزیع نرمال همواره مساوی 0.263 و کشیدگی گشتاوری توزیع نرمال همیشه برابر 3 است.

تأثیر تغییر مشاهدات درکشیدگی

تغییر در مشاهدات، هیچ‌گونه تغییری در کشیدگی و درنتیجه ضریب کشیدگی ایجاد نمی‌کند:

مشاهدات در a ضرب و با b جمع یا از آن کم شوند.	$E_{ax\pm b} = E_x$
aدرصد به هر مشاهده اضافه یا از آن کم شود.	$E_{x\pm(\%a)x} = E_{(1\pm\%a)x} = E_x$

مثال ۱ اگر تمام داده‌ها با عدد 10 جمع و سپس بر 2− تقسیم شوند، ضریب کشیدگی داده چه تغییری می‌کند؟

۱) نصف می‌شود. ۲) تغییری نمی‌کند.

۳) به اندازه $\frac{1}{10}$ افزایش می‌یابد. ۴) به اندازه 10 واحد افزایش می‌یابد.

حل: گزینه ۲ درست است.

$$E_{\frac{x+10}{-2}} = E_{-\frac{x}{2}-5} = E_x$$

مثال ۲ درصورتی‌که داده‌های آماری یک جامعه دو برابر شوند، ضریب کشیدگی آن:

۱) نصف می‌شود. ۲) دو برابر می‌شود. ۳) بدون تغییر می‌ماند. ۴) قرینه می‌شود.

حل: گزینه ۳ درست است.

تغییر در داده‌ها، هیچ تغییری در مقدار ضریب کشیدگی ایجاد نمی‌کند.

محاسبه شاخص‌ها در جداول با حدود باز

همواره با داشتن یک جدول طبقه‌بندی شده به صورت زیر که حدود تمام طبقات در آن مشخص است، می‌توان تمام شاخص‌های مرکزی و پراکندگی را محاسبه کرد.

C − L	1 − 2	2 − 5	5 − 8	8 − 11	11 − 14
(فراوانی) F_i	1	10	17	20	3

حال جدول زیر را در نظر بگیرید که حدود طبقات آن در ابتدا یا انتها باز (نامشخص) باشد:

C − L	<2	2 − 5	5 − 8	8 − 11	≥ 11
(فراوانی) F_i	1	10	17	20	3

شاخص‌های غیرقابل محاسبه در جداول با حدود باز

ازآنجاکه امکان محاسبه مرکز طبقه (x_i) در طبقات ابتدا و انتها وجود ندارد، امکان محاسبه میانگین و تمام شاخص‌هایی که برای محاسبه نیاز به میانگین دارند وجود ندارد. این شاخص‌ها عبارت‌اند از:

۱- گشتاور مرکزی مرتبه nام

۲- واریانس و انحراف معیار

۳- ضریب چولگی گشتاوری

۴- ضریب کشیدگی گشتاوری

شاخص‌های قابل محاسبه در جدول با حدود باز

با توجه به آنکه در جداولی مانند جدول بالا مناسب‌ترین پارامتر مرکزی، میانه (Md) و مناسب‌ترین پارامتر پراکندگی، انحراف چارکی $\left(\frac{Q_3-Q_1}{2}\right)$ است، پارامترهای قابل محاسبه عبارت‌اند از:

۱- میانه

۲- چارک اول، دوم و سوم

۳- نیم‌دامنه

۴- انحراف چارکی

۵- تمام چندک‌های (دهک و صدک) بین چارک اول و چارک سوم

۶- ضریب چولگی چندکی

۷- ضریب کشیدگی چندکی

مثال کدام‌یک از شاخص‌های آماری برای توزیع زیر قابل محاسبه نیست؟

x_i	<5	$5-8$	$8-11$	$11-14$	≥ 14	
F_i	3	10	17	6	4	$\sum F_i = 40$

۱) میانه

۲) انحراف چارکی

۳) ضریب چولگی گشتاوری

۴) چارک اول

حل: گزینه ۳ درست است.

نمایش هندسی مشاهدات

برای نمایش توزیع‌های فراوانی، اغلب از نمودارها استفاده می‌شود، اما هر داده با توجه به نوع مقیاس آن، به شیوه‌ای متفاوت ترسیم می‌شود.

انواع نمودارهای آماری با توجه به مقیاس

۱ـ نمودارهای کمّی (برای مقیاس‌های نسبی و فاصله‌ای)
۲ـ نمودارهای کیفی (برای مقیاس‌های اسمی و رتبه‌ای)

نکته: مهم‌ترین نمودارهای کمّی و کیفی عبارت‌اند از:

نمودارهای کمّی
- ۱ـ نمودار بافت‌نگار (مستطیلی) (Histogram Chart)
- ۲ـ نمودار چندضلعی (پلی‌گون) (Polygon Chart)
- ۳ـ نمودار فراوانی تجمعی (اجایو) (Cumulative Frequency Chart) (Ogive)
- ۴ـ نمودارهای تحلیل اکتشافی داده‌ها
  - ۱- شاخه و برگ (Stem and Leaf Plot)
  - ۲- جعبه‌ای (Box Plot)

نمودارهای کیفی (وصفی)
- ۱- نمودار ستونی (میله‌ای) (Bar Chart)
- ۲- نمودار دایره‌ای (Pie Chart)
- ۳- نمودار پارتو (Pareto Chart)

مثال ۱ کدام‌یک از نمودارهای زیر برای توصیف داده‌ها با مقیاس اسمی مناسب‌تر است؟ (حسابداری ـ ۷۹)

۱) جعبه‌ای ۲) ریشه و برگ ۳) پارتو ۴) چندضلعی

حل: گزینه ۳ درست است.

مثال ۲ بهترین نمایش تصویری برای مقایسه دو مجموعه داده اسمی، کدام است؟ (حسابداری ـ ۸۱)

۱) جعبه‌ای ۲) بافت‌نگار ۳) نمودار میله‌ای ۴) چندضلعی

حل: گزینه ۳ درست است.

مثال ۳ کدام‌یک از نمودارهای زیر برای توصیف داده‌های فاصله‌ای مناسب‌تر است؟ (مدیریت ـ ۷۶)

۱) مستطیلی ۲) پارتو ۳) دایره‌ای ۴) ستونی

حل: گزینه ۱ درست است.

نمودارهای کمّی (Numerical Charts)

از طریق نمودارهای کمّی که برای داده‌هایی با مقیاس فاصله‌ای و نسبتی به کار می‌روند، می‌توان به آسانی بعضی از شاخص‌های جامعه مانند تقارن، مقدار چندک‌ها، مد و میانه را به دست آورد.

نمودار بافت‌نگار (مستطیلی، هیستوگرام) (Histogram Chart)

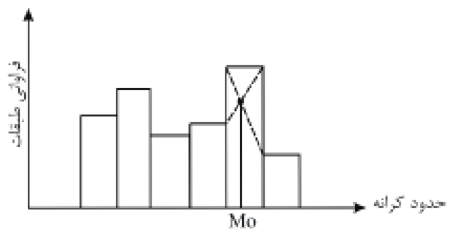
۱- برای رسم این نمودار از فراوانی مطلق یا فراوانی نسبی و حدود واقعی طبقات (کرانه‌های طبقات) استفاده می‌شود.
۲- در نمودار بافت‌نگار محور افقی دستگاه مختصات با حدود واقعی طبقات (کرانه‌های طبقات) و محور عمودی آن با فراوانی نسبی یا مطلق مدرج می‌شود.
۳- روی کرانه هر طبقه، مستطیلی عمودی رسم می‌کنیم به طوری که مساحت آن برابر با فراوانی نسبی آن طبقه باشد؛ به عبارت دیگر:

$$\text{ارتفاع مستطیل} = \frac{\text{فراوانی نسبی طبقه}}{\text{طول طبقه}}$$

۴- مساحت هر یک از مستطیل‌های بافت‌نگار نشان‌دهنده نسبت مشاهدات موجود در هر طبقه است.

محاسبه مد از طریق نمودار بافت‌نگار

بلندترین مستطیل در نمودار بافت‌نگار، شامل نقطه مد داده‌هاست.
دو خط مورب از دو گوشه بلندترین مستطیل رسم می‌کنیم؛ محل تقاطع دو خط، نقطه مد داده‌هاست.
نحوه ترسیم دو خط مورب: یک خط مورب از گوشه راست بلندترین مستطیل به گوشه راست مستطیل سمت چپ آن می‌کشیم و خط مورب دیگر را از گوشه چپ بلندترین مستطیل به گوشه چپ مستطیل سمت راست آن رسم می‌کنیم.



مثال ۴ برای رسم هیستوگرام (نمودار مستطیلی)، محورهای x و y بر اساس کدام اندازه‌ها مدرج می‌شوند؟ (مدیریت ـ ۷۰)

۱) کرانه‌های طبقات و فراوانی طبقات
۲) حدود طبقات و چگالی
۳) حد وسط طبقات و فراوانی مطلق
۴) مقادیر متغیر x و فراوانی‌های تجمعی

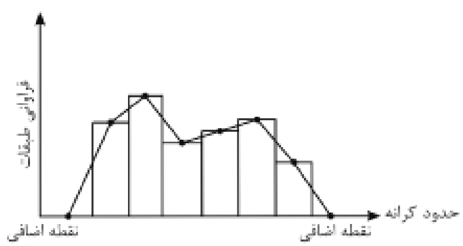
حل: گزینه ۱ درست است.

نمودار چندضلعی (پلی‌گون) (Polygon Chart)

۱ـ در نمودار چندضلعی، مرکز هر دسته (نقطه میانی) در محور افقی و فراوانی نسبی یا مطلق دسته‌ها روی محور عمودی قرار می‌گیرند.
۲ـ این نمودار شامل نقطه‌هایی در صفحه مختصات است که طول آن نقطه‌ها مرکز دسته‌ها و عرض نقطه‌ها فراوانی دسته‌ها را نشان می‌دهد.
۳ـ برای رسم نمودار چندضلعی، دو نقطه دیگر به نقطه‌های موجود در صفحه مختصات اضافه کرده، سپس نقطه‌ها را از ابتدا به هم متصل می‌کنیم.

«طول طبقه + مرکز طبقه آخر» = نقطه دوم
«طول طبقه − مرکز طبقه اول» = نقطه اول

۴ـ اگر در نمودار بافت‌نگار مرکز مستطیل‌ها را به هم متصل کنیم، به نمودار چندضلعی می‌رسیم. در نمودار چندضلعی به‌جای رسم ستون‌های متصل به هم، از خطوط مستقیم استفاده شده است.
۵ـ در تحقیقاتی که هدف، مقایسه دو یا چند جامعه آماری باشد، نمودار چندضلعی ترجیح داده می‌شود.



مثال ۵ برای مقایسه دو توزیع فراوانی مربوط به حقوق پرداختی به کارگران مرد و زن در یک کارخانه، کدام‌یک از نمودارهای زیر مناسب‌تر است؟ (اقتصاد ـ ۷۷)

۱) پلی‌گون (چندگوش) فراوانی نسبی
۲) نمودار میله‌ای فراوانی مطلق
۳) نمودار تجمعی (ogive) فراوانی مطلق
۴) هیستوگرام (بافت نگار) فراوانی نسبی

حل: گزینه ۱ درست است.

نمودار فراوانی تجمعی (اجایو) (Cumulative Frequency Chart) (Ogive Chart)

۱ـ نمودار فراوانی تجمعی بر اساس توزیع فراوانی تجمعی مشاهدات رسم می‌شود.
۲ـ دو روش برای رسم آن وجود دارد:
الف) روی محور عمودی دستگاه مختصات، فراوانی تجمعی داده‌ها و روی محور افقی آن، مرکز دسته‌ها قرار می‌گیرند.
ب) روی محور عمودی دستگاه مختصات، فراوانی تجمعی داده‌ها و روی محور افقی آن، حد بالای دسته‌ها قرار می‌گیرند.
در روش (الف)، ارزش همه مقادیر داخل هر طبقه یکسان فرض شده است، اما در روش (ب) فرض بر یکنواخت نبودن ارزش مقادیر هر طبقه است.
محور عمودی در هر دو روش، فراوانی تجمعی داده‌هاست اما درصورتی‌که با فراوانی نسبی تجمعی مدرج شده باشد، عمل نقطه‌یابی مفیدتر خواهد بود.
در هر دو روش پس از به دست آوردن نقاط در صفحه مختصات از ابتدا نقاط را به هم متصل می‌کنیم.
۳ـ با نمودار فراوانی تجمعی به راحتی می‌توان هر یک از چندک‌ها را محاسبه کرد.
در این نمودار به سؤال‌هایی نظیر «چند درصد از مشاهدات پایین‌تر از x (یک نقطه دلخواه) قرار دارد؟» به سادگی پاسخ داده می‌شود.
۴ـ این نمودار برای مقایسه توزیع‌های فراوانی دو یا چند جامعه که از نظر تعداد با هم مساوی هستند، مفید است، مانند مقایسه میزان رشد تورم در دو یا چند کشور.

تحلیل اکتشافی داده‌ها (Exploratory Data Analysis)

۱ـ چون این نمودارها اغلب در مراحل اولیه تحلیل داده‌ها مفید هستند، به روش‌های تحلیل اکتشافی داده‌ها معروفند.
۲ـ نمودارهای شاخه برگ (ریشه برگ) و جعبه‌ای از نوع تحلیل اکتشافی داده‌ها هستند.

مثال ۶ کدام‌یک از نمودارهای زیر از نوع تحلیل اکتشافی است؟ (حسابداری ـ ۸۲)

۱) میله‌ای
۲) اجایو
۳) بافتنگار
۴) ریشه و برگ

حل: گزینه ۴ درست است.

نمودار شاخه و برگ (ریشه و برگ) (Stem and Leaf Plot)

۱ـ در این نمودار از داده‌های خام دسته‌بندی نشده استفاده می‌شود.
۲ـ در این نمودار ارقام مشاهدات را به دو بخش شاخه و برگ تقسیم می‌کنیم؛ شاخه شامل یک یا چند رقم اولیه و برگ شامل ارقام باقی‌مانده است. مثلاً داده 63 به شاخه 6 و برگ 3 تقسیم می‌شود.
۳ـ تعیین تعداد شاخه‌ها به تمایل تصمیم‌گیرنده بستگی دارد. پس از تعیین شاخه‌ها، آن‌ها را در یک ستون در سمت چپ قرار می‌دهیم و سپس برگ‌های مربوط به هر شاخه را به ترتیب صعودی جلوی شاخه آن می‌نویسیم.

۴ـ این نمودار همانند نمودار بافت‌نگار به ما امکان می‌دهد که بعضی از خصوصیات (مد و میانه) را به سرعت از روی نمودار متوجه شویم. اما در این نمودار برخلاف بافت‌نگار اعداد اصلی از بین نمی‌روند.
۵ـ در نمودار شاخه برگ محاسبه چندک‌ها به راحتی امکان‌پذیر است.

مثال ۷ کدام نمودار برای نمایش مشاهدات کمّی طبقه‌بندی نشده به کار می‌رود؟ (مدیریت ـ ۸۱)

۱) پارتو ۲) چندضلعی ۳) ریشه و برگ ۴) بافت‌نگار

حل: گزینه ۳ درست است.

مثال ۸ داده‌های زیر مربوط به نمرات آمار 12 دانشجوی حسابداری است:

15 , 11 , 16 , 4 , 10 , 14 , 13 , 7 , 20 , 9 , 11 , 15

نمودار شاخه و برگ مربوط به نمرات آمار را رسم کنید.

حل:

داده	دهگان	یکان
4	0	4
7	0	7
9	0	9
10	1	0
11	1	1
11	1	1
13	1	3
14	1	4
15	1	5
15	1	5
16	1	6
20	2	0

شاخه	برگ
0	4 7 9
1	0 1 1 3 4 5 5 6
2	0

بنابراین، نمودار شاخه برگ این نمرات به صورت زیر است:

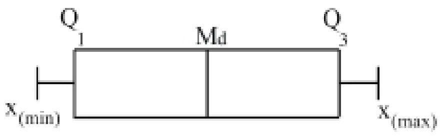
شاخه	برگ
0	4 7 9
1	0 1 1 3 4 5 5 6
2	0

نمودار جعبه‌ای (Box Plot)

۱ـ یکی از مفیدترین نمودارهای اکتشافی برای مقایسه دو یا چند جامعه آماری، نمودار جعبه‌ای است.
۲ـ نحوه رسم نمودار جعبه‌ای به صورت زیر است:
الف) داده حداقل و حداکثر را پیدا کنید.
ب) چارک اول، میانه و چارک سوم را پیدا کنید.
ج) از چارک اول تا چارک سوم یک جعبه رسم کنید به طوری که میانه در وسط جعبه قرار گیرد و سپس دو طرف جعبه را با خط راست به داده حداقل و حداکثر رسم کنید.

۳ـ داده حداقل و حداکثر گاهی ریشه نیز نامیده می‌شوند.

۴ـ جعبه شامل اختلاف چارک اول و سوم است.



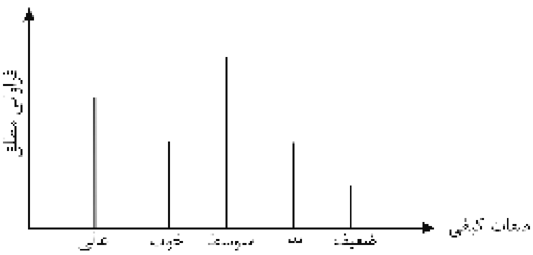
نمودارهای کیفی (Descriptive Charts)

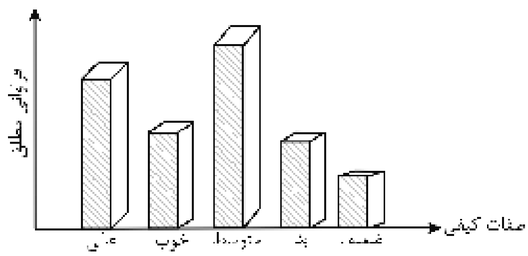
این دسته از نمودارها برای نمایش هندسی داده‌های کیفی به کار می‌روند (خوب، بد، متوسط، گروه A و B و C و ...). در این نمودارها برخلاف نمودار بافت‌نگار، فاصله مفهومی ندارد؛ به عبارت دیگر در این نمودارها هر مقدار را یک طبقه به حساب می‌آوریم.

نمودار ستونی (میله‌ای) (Bar Chart)

۱ـ در این نمودار در محور افقی دستگاه مختصات، کیفیت مشاهدات (صفات کیفی مشاهدات، خوب، بد، ...) و در محور عمودی دستگاه مختصات، فراوانی مطلق یا نسبی هر گروه قرار می‌گیرد.

۲ـ در نمودار ستونی خطهایی ضخیم (مکعب مستطیل) از صفات کیفی به اندازه فراوانیشان رسم می‌کنیم. توجه کنید که در شکل میله‌ای آن، خطوط جایگزین مستطیل‌ها می‌شوند تا بر این موضوع تأکید شود که فراوانی‌ها واقعاً روی فاصله‌ها پخش نشده‌اند.





نمودار دایره‌ای (Pie Chart)

۱ـ نمودار دایره‌ای ابزار مناسبی برای تجسم مشاهدات کیفی است.

۲ـ نمودار دایره‌ای برحسب درصد بیان می‌شود و موارد استفاده فراوانی دارد.

۳ـ مراحل رسم این نمودار به صورت زیر است:

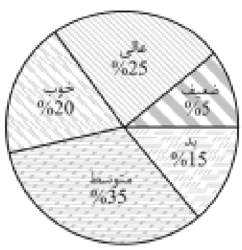
الف) فراوانی‌های مطلق را به فراوانی‌های نسبی تبدیل کنید.

ب) با استفاده از رابطه $\theta_i = 360° \times f_i$ زاویه هر قطعه را تعیین کنید.

ج) با استفاده از زاویه هر قطعه، مساحت دایره را تقسیم کنید.

د) روی هر قطعه، درصد مربوط به آن و نوع مشاهده را ذکر کنید.

۴ـ به این نمودار، نمودار کلوچه‌ای نیز گفته می‌شود.



مثال ۹ در دانشکده‌ای 300 نفر دوره روزانه، 200 نفر دوره شبانه دوره کارشناسی و 100 نفر در تحصیلات تکمیلی تحصیل می‌کنند. در نمودار دایره‌ای این اطلاعات، زاویه متناظر با 100 نفر تحصیلات تکمیلی کدام است؟ (برنامه ریزی شهری ـ ۸۱)

۱) $60°$ ۲) $80°$ ۳) $90°$ ۴) $120°$

حل: گزینه ۱ درست است.

x	روزانه کارشناسی	شبانه کارشناسی	تحصیلات تکمیلی	جمع
فراوانی مطلق F_i	300	200	100	$600 = N$
فراوانی نسبی f_i	$\frac{300}{600}=\frac{1}{2}$	$\frac{200}{600}=\frac{1}{3}$	$\frac{100}{600}=\frac{1}{6}$	1
زاویه هر قطعه	$\frac{1}{2}\times 360° = 180°$	$\frac{1}{3}\times 360° = 120°$	$\boxed{\frac{1}{6}\times 360° = 60°}$	

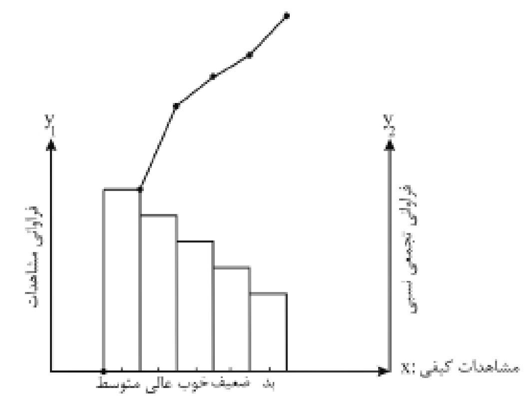
نمودار پارتو (Pareto Chart)

۱ـ نمودار پارتو نوعی نمودار ستونی است که محور افقی آن نوع مشاهدات کیفی و محور عمودی آن فراوانی مطلق مشاهدات است. در این نمودار ستون‌ها مستطیل‌شکل هستند.

۲ـ نمودار پارتو همیشه به ترتیب نزولی فراوانی‌ها رسم می‌شود، یعنی در یک نگاه مستطیل‌های آن به ترتیب ارتفاع (از بزرگ به کوچک) چیده شده‌اند؛ به عبارت دیگر، پروقوع‌ترین موضوع در سمت چپ قرار می‌گیرد.

۳ـ این نمودار سه محور دارد که محور سوم آن یک محور عمودی دیگر در طرف راست محور افقی است و بر اساس فراوانی تجمعی نسبی مدرج شده است. درواقع نمودار پارتو شامل یک منحنی (حاصل اتصال درصدهای تجمعی طبقات) و چندین مستطیل است.

۴ـ نمودار پارتو در تحلیل موجودی انبار کالاها، نواقص سیستم‌ها، توزیع درآمد و توزیع کارمندان سازمان‌ها کاربرد فراوانی دارد. این نمودار بخش مهمی از برنامه کنترل کیفیت محسوب می‌شود، زیرا توجه را به بحرانی‌ترین نواقص معطوف می‌کند.



تست‌های طبقه‌بندی شده

تعاریف

۱. کدام مقیاس برای اندازه‌گیری از ویژگی‌های بهتری برخوردار است؟ **(حسابداری ــ ۸۳ ، ۸۴ ، ۸۸)**

۱) نسبتی ۲) اسمی ۳) رتبه‌ای ۴) فاصله‌ای

۲. اولین مرحله در یک تحقیق علمی کدام است؟ **(حسابداری و مدیریت ــ ۸۶)**

۱) فرضیه‌سازی ۲) جمع‌آوری داده‌ها ۳) هدف‌گذاری ۴) تحلیل یافته

۳. متغیرهای تصادفی کمّی به کدام دو دسته تقسیم می‌شوند؟ **(حسابداری و مدیریت ــ ۸۷)**

۱) اسمی ـ ترتیبی ۲) پیوسته ـ ترتیبی ۳) گسسته ـ پیوسته ۴) گسسته ـ اسمی

۴. فراوانی مطلق چیست؟ **(برنامه‌ریزی شهری ــ ۸۶)**

۱) تکرار هر داده ۲) دسته‌بندی داده‌ها
۳) نسبت فراوانی به تعداد نمونه ۴) ضریبی از فراوانی تجمعی

۵. جامعه آماری چیست؟ **(برنامه‌ریزی شهری ــ ۸۶)**

۱) سرشماری جمعیت
۲) تعداد کل اعضای جامعه
۳) نمونه‌گیری تصادفی که در سرشماری به کار می‌رود.
۴) مجموعه‌ای از افراد یا اشیا که موضوع مورد مطالعه هستند.

۶. متغیرها را از این نظر که قابل اندازه‌گیری باشند یا نه، به کدام گروه‌ها تقسیم می‌کنند؟ **(برنامه‌ریزی شهری ــ ۸۶)**

۱) پیوسته ـ کمّی ۲) پیوسته ـ گسسته ۳) کمّی ـ کیفی ۴) گسسته ـ کیفی

۷. کدام‌یک از متغیرهای زیر کیفی است؟ **(برنامه‌ریزی شهری ــ ۸۶)**

۱) متوسط درجه سالانه شهر ۲) تعداد مسافرین حمل و نقل عمومی شهر
۳) میزان آلودگی هوای شهر ۴) مراقبت از فضای سبز شهر

۸. متغیرهای تصادفی کمّی به کدام دو دسته تقسیم می‌شوند؟ **(برنامه‌ریزی شهری ـ ۸۶)**

۱) اسمی ـ ترتیبی ۲) ترتیبی ـ اسمی ۳) گسسته ـ پیوسته ۴) گسسته ـ اسمی

۹. مزیت اطلاعات کمّی بر کیفی چیست؟ **(برنامه‌ریزی شهری ـ ۸۶)**

۱) اندازه‌گیری دقیق از موضوع و قابلیت تعمیم نتایج ۲) توصیف دقیق از اطلاعات و قابلیت تفسیر آن

۳) قابلیت تفسیر اطلاعات و نتیجه‌گیری تجریدی ۴) قابلیت دستیابی به برداشت‌های مختلف

۱۰. کدام‌یک از متغیرهای زیر در سطح اسمی اندازه‌گیری می‌شود؟ **(برنامه‌ریزی شهری ـ ۸۷)**

۱) دما ۲) تعداد پل‌های منطقه ۳

۳) مناطق بیست و دوگانه تهران ۴) میزان آلودگی هوا

۱۱. «مجموعه‌ای از افراد یا اشیای مورد مطالعه» و «موضوع مورد مطالعه» به ترتیب چه نامیده می‌شود؟ **(برنامه‌ریزی شهری ـ ۸۷)**

۱) جامعهٔ آماری ـ متغیر تصادفی ۲) جامعهٔ آماری ـ پارامتر

۳) نمونهٔ آماری ـ پارامتر ۴) نمونهٔ آماری ـ متغیر تصادفی

۱۲. در کدام مرحله از آمار، آماره‌ها با نمونه‌های تصادفی محاسبه می‌شوند؟ **(GIS ـ ۸۶)**

۱) آزاد از توزیع ۲) استنباطی ۳) توصیفی ۴) ناپارامتریک

۱۳. وزن محصولات تولیدشده در یک شرکت، دارای کدام نوع مقیاس است؟ **(GIS ـ ۸۶)**

۱) اسمی ۲) رتبه‌ای ۳) فاصله‌ای ۴) نسبی

۱۴. در اندازه قطر درختان یک باغ، جامعهٔ آماری کدام است ؟ **(GIS ـ ۸۷)**

۱) قطر درختان ۲) عمر درختان ۳) درختان باغ ۴) درختان باغات مجاور

۱۵. کدام عبارت تعریف صفت مشخصه است؟ **(GIS ـ ۸۸)**

۱) صفتی که از فردی به فرد دیگر تغییر کند. ۲) عنصر مشترک جوامع آماری مختلف

۳) متمایزکنندهٔ عناصر جامعه از یکدیگر ۴) صفت مشترک بین کلیهٔ افراد جامعه

۱۶. در یک آزمون سنجش معلومات، دانش‌آموزان تیزهوش با کد 1، دانش‌آموزان خوب با کد 2 و دانش‌آموزان متوسط با کد 3 نشان داده شده‌اند. در این آزمون از چه نوع مقیاسی برای این انتساب استفاده شده است؟ **(محیط زیست ـ ۸۷)**

۱) اسمی ۲) ترتیبی ۳) فاصله‌ای ۴) نسبتی

فراوانی

۱۷. در جدول فراوانی تجمعی داده‌های دسته‌بندی شده اگر درصد فراوانی نسبی دسته وسط 16 باشد، فراوانی مطلق در دسته چهارم کدام است؟ **(حسابداری و مدیریت ـ ۸۷)**

مرکز دسته	9	12	15	18	21
فراوانی تجمعی	8	25	a	58	75

۱) 24 ۲) 21

۳) 19 ۴) 17

۱۸. در جدول فراوانی تجمعی داده‌های دسته‌بندی شده زیر اگر درصد فراوانی نسبی دسته وسط 24 باشد، فراوانی مطلق دسته چهارم کدام است؟ (برنامه‌ریزی شهری ــ ۸۶)

مرکز دسته	21	19	17	15	13
فراوانی تجمعی	50	41	A	14	5

۱) 14 ۲) 15 ۳) 16 ۴) 17

معیارهای تمرکز

مد (نما)

۱۹. از 50 کارمند بانک، سؤال شده است که «مناسب‌ترین زمان شروع به کار در صبح را چه ساعتی می‌دانید؟» نتایج به‌دست‌آمده دارای میانگین 8.45، میانه 8.15 و نمای 7.15 بوده است. بر اساس این اطلاعات مناسب‌ترین ساعت برای شروع کار بانک‌ها کدام است؟ (اقتصاد ــ ۸۸)

۱) 7.15 ۲) 8.0 ۳) 8.15 ۴) 8.45

۲۰. در جدول توزیع فراوانی دسته‌بندی شده اگر مد جامعه 16 باشد، فراوانی مطلق دسته چهارم کدام است؟

حدود دسته	$18-21$	$15-18$	$12-15$	$9-12$
فراوانی مطلق	a	15	12	8

(حسابداری و مدیریت ــ ۸۷)

۱) 7 ۲) 9 ۳) 10 ۴) 11

۲۱. در جدول داده‌های طبقه‌بندی شده زیر اگر مد جامعه 25.5 باشد، درصد فروانی نسبی دسته سوم کدام است؟

حدود دسته	$27-31$	$23-27$	$19-23$	$15-19$
فراوانی مطلق	14	a	12	7

(برنامه‌ریزی شهری ــ ۸۷)

۱) 32 ۲) 34 ۳) 35 ۴) 36

میانه

۲۲. مقدار میانه در جدول توزیع فراوانی زیر کدام است؟ (اقتصاد ــ ۸۸)

x_i	10	11	12	13	14
f_i	5	10	14	26	55

۱) 12 ۲) 13 ۳) 13.5 ۴) 14

۲۳. میانه در توزیع آماری 50 مشاهده دسته‌بندی شده برابر 41 می‌باشد. اگر طول دسته‌ها 5، فراوانی طبقه میانه‌دار 10 و مجموع فراوانی‌های ماقبل طبقه میانه‌دار برابر 18 باشد، حدود دسته میانه‌دار کدام است؟

(حسابداری و مدیریت ــ ۸۶)

۱) $(36.5,41.5)$ ۲) $(37,42)$ ۳) $(37.5,42.5)$ ۴) $(38,43)$

۲۴. در 60 مشاهده دسته‌بندی شده، میانه 23، فاصله طبقات 3، فراوانی طبقه میانه‌دار 12 و فراوانی تجمعی طبقه میانه‌دار 38 می‌باشد. حدود دسته میانه‌دار کدام است؟ (برنامه‌ریزی شهری ـ ۸۷)

۱) $(21\,,\,24)$ ۲) $(22\,,\,25)$ ۳) $(21.5\,,\,24.5)$ ۴) $(22.5\,,\,25.5)$

۲۵. با توجه به جدول روبه‌رو، میانه مقادیر کدام است؟ (محیط زیست ـ ۸۶)

فراوانی	رده
2	$4-6$
7	$6-8$
4	$8-10$
22	$10-12$
15	$12-14$

۱) 11
۲) 12
۳) $\frac{122}{11}$
۴) $\frac{150}{11}$

خواص میانه

۲۶. مشاهدات $x_1, x_2, 15, x_4, x_5$ دارای انحراف معیار صفر هستند. آن‌گاه میانه مشاهدات $(2x_1+1)$، $(2x_2+1)$، ...، $(2x_5+1)$ برابر است با: (اقتصاد ـ ۸۷)

۱) 18 ۲) 21 ۳) 24 ۴) 31

چندک

۲۷. چارک سوم داده‌های جدول توزیع فراوانی زیر کدام است؟ (اقتصاد ـ ۸۶)

X	$2-4$	$4-6$	$6-8$	$8-10$
f	5	3	8	4

۱) $\frac{31}{4}$ ۲) $\frac{32}{3}$
۳) $\frac{62}{4}$ ۴) $\frac{32}{6}$

۲۸. در جدول توزیع فراوانی زیر، چارک سوم کدام است؟ (GIS ـ ۸۶)

X_i	$15-20$	$20-25$	$25-30$	$30-35$	$35-40$
f_i	22	28	16	12	14

۱) 30.75 ۲) 31.25
۳) 31.6 ۴) 32.3

۲۹. در 120 داده آماری، کوچک‌ترین و بزرگ‌ترین آن‌ها به ترتیب 12 و 54 می‌باشد. این داده‌ها در 7 طبقه دسته‌بندی شده‌اند به طوری که مقدار دهک ششم برابر 32 و در دسته وسط واقع است. اگر فراوانی مطلق این طبقه 9 باشد، درصد فراوانی نسبی تجمعی آن کدام است ؟ (GIS ـ ۸۷)

۱) 63 ۲) 65 ۳) 68 ۴) 72

میانگین هارمونیک

۳۰. یک هواپیما فاصله 3 هزار کیلومتری را با سرعت 600 کیلومتر در ساعت، فاصله 5 هزار کیلومتری را با سرعت 750 کیلومتر در ساعت و فاصله 4 هزار کیلومتری را با سرعت 800 کیلومتر در ساعت طی می‌کند. سرعت متوسط آن کدام است؟ (حسابداری و مدیریت ـ ۸۸)

۱) 715 ۲) 717 ۳) 720 ۴) 725

۳۱. راننده اتومبیلی $\frac{1}{3}$ مسافت بین دو شهر را با سرعت 120 کیلومتر در ساعت، $\frac{1}{4}$ این مسافت را با سرعت 80 کیلومتر در ساعت و بقیه مسافت را با سرعت 100 کیلومتر در ساعت طی کرده است. سرعت متوسط این راننده در مسیر بین دو شهر کدام است؟ (برنامه‌ریزی شهری ـ ۸۸)

۱) 98.7 ۲) 99.3 ۳) 101.6 ۴) 102.3

۳۲. راننده اتومبیلی $\frac{2}{3}$ مسافتی را با سرعت 80 کیلومتر در ساعت و $\frac{1}{4}$ این مسافت را با سرعت 90 کیلومتر در ساعت و بقیه مسافت را با سرعت 60 کیلومتر در ساعت طی کرده است. سرعت متوسط او در این مسافت کدام است؟ (GIS ـ ۸۶)

۱) 78.6 ۲) 80 ۳) 80.8 ۳) 82

میانگین هندسی

۳۳. فزونی میانگین حسابی از میانگین هندسی داده‌های

x_i	9	12	16
f_i	2	3	2

کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) 0.24 ۲) 0.25 ۳) 0.27 ۴) 0.28

۳۴. تعداد دانشجویان پذیرفته‌شده در یک دانشکده در پنج سال متوالی به شرح زیر است:

سال‌ها	1380	1381	1382	1383	1384
تعداد	100	280	310	350	400

متوسط درصد (نرخ) رشد سالانه دانشجویان پذیرفته‌شده در این دانشکده کدام است؟ (اقتصاد ـ ۸۶)

۱) %38 ۲) %41 ۳) %50 ۴) %56

۳۵. فروش یک فروشگاه در سال گذشته 80% افزایش یافته و امسال 80%کاهش داشته است. متوسط نرخ رشد فروش سالانه این فروشگاه در این دو سال چند درصد بوده است؟ (اقتصاد ـ ۸۸)

۱) 40% کاهش ۲) 60% کاهش ۳) 20% افزایش ۴) صفر

۳۶. میانگین هندسی رشته اعداد $6, 24, 8, 72, 96$ کدام است؟ (GIS ـ ۸۷)

۱) 12 ۲) 18 ۳) 24 ۴) 36

۳۷. نسبت میزان بارندگی طی سه سال گذشته به ترتیب $\frac{20}{21}$، $\frac{15}{14}$ و $\frac{245}{432}$ بوده است. میانگین این نسبت‌ها کدام است؟ (GIS ـ ۸۸)

۱) $\frac{5}{6}$ ۲) $\frac{15}{16}$ ۳) $\frac{25}{28}$ ۴) $\frac{65}{63}$

میانگین حسابی

۳۸. توزیع فراوانی هزینه متوسط ماهانه خانوارها در جدول زیر داده شده است:

هزینه متوسط	15	35	60	85	210	380
تعداد خانوار	32	48	12	5	2	1

میانگین هزینه 20% از پرخرج‌ترین خانوارها چقدر است؟ (اقتصاد ـ ۸۶)

۱) 19.45 ۲) 41.05 ۳) 97.25 ۴) 108.15

۳۹. اگر میانگین حسابی 50 داده آماری متقارن برابر 12 باشد، اختلاف میانگین پیراسته $LN = 20$ این داده‌ها از میانگین کل کدام است؟ (GIS ـ ۸۶)

۱) -1 ۲) صفر ۳) 0.5 ۴) 1

۴۰. میانگین داده‌های پیوسته از جدول زیر کدام است؟ (حسابداری و مدیریت ـ ۸۸)

حدود دسته	$216-232$	$232-248$	$248-264$	$264-280$
فراوانی	16	29	13	12

۱) 242.8 ۲) 243.2 ۳) 243.6 ۴) 244.8

۴۱. میانگین داده‌های آماری در جدول زیر به صورت $25+A$ است. A کدام است؟ (برنامه‌ریزی شهری ـ ۸۷)

حدود دسته	$15-19$	$19-23$	$23-27$	$27-31$	$31-35$
فراوانی مطلق	6	12	15	9	8

۱) 0.02 ۲) 0.06 ۳) 0.06 ۴) 0.08

۴۲. کمترین و بیشترین داده‌های آماری 12.5 و 32.5 می‌باشد. این داده‌ها در 4 طبقه دسته‌بندی شده‌اند. فراوانی تجمعی آنها به ترتیب $10, 29, 45$ و 60 می‌باشد. میانگین این داده‌ها کدام است؟ (GIS ـ ۸۶)

۱) 23 ۲) 23.5 ۳) 24 ۴) 24.5

معیارهای پراکندگی

انحراف چارکی

۴۳. نمرات مسئولیت‌پذیری کارمندان یک شرکت از صفر تا 30 طبقه‌بندی شده است. انحراف چارکی کدام است؟

حدود دسته	<10	$10-14$	$14-18$	$18-22$	$22-26$	≥ 26
فراوانی	4	8	10	12	9	7

(حسابداری و مدیریت ـ ۸۸)

۱) 2.9 ۲) 4.6 ۳) 6.4 ۴) 9.2

۴۴. در یک جامعه آماری چارک اول، دوم، سوم به ترتیب $52, 70$ و 84 شده است. مقدار انحراف چارکی کدام است؟ (برنامه‌ریزی شهری ـ ۸۶)

۱) 32 ۲) 18 ۳) 16 ۴) 14

۴۵. در جدول فراوانی داده‌های آماری زیر انحراف چارکی کدام است؟ (برنامه‌ریزی شهری ـ ۸۸)

فاصله‌دسته	<8	$8-11$	$11-14$	$14-17$	$17-20$	≥ 20
فراوانی	5	12	15	13	8	7

۱) 3.10 ۲) 3.15 ۳) 3.20 ۴) 3.25

۴۶.چارک اول ،دوم و سوم یک جامعه آماری به ترتیب 31 , 72 و 95 محاسبه شده است. مقدار انحراف چارکی کدام است؟ (GIS ـ ۸۶)

۱) 32 ۲) 33 ۳) 34 ۴) 64

۴۷.توزیع نمرات مسئولیت پذیری کارکنان یک شرکت در جدول زیر تنظیم شده است. انحراف چارکی کدام است؟ (GIS ـ ۸۸)

نمرات	<7	$7-10$	$10-13$	$13-16$	$16-19$	≥ 19
فراوانی	7	9	16	19	8	5

۱) 2.4 ۲) 2.7 ۳) 4.8 ۴) 5.4

واریانس

۴۸.انحراف متوسط از میانگین در 12 داده آماری صفر و میانگین آن‌ها 15 می‌باشد. اگر داده‌های 20، 16 و 24 به آن‌ها اضافه شود، واریانس 15 داده جدید کدام است؟ (GIS ـ ۸۷)

۱) 5.24 ۲) 5.42 ۳) 6.13 ۴) 6.31

واریانس نمونه

۴۹.هرگاه در یک نمونه‌گیری حجم نمونه افزایش یابد و مجموع مربعات تفاضل از میانگین ثابت بماند، واریانس چه تغییری می‌کند؟ (محیط زیست ـ ۸۷)

۱) افزایش می‌یابد. ۲) کاهش می‌یابد. ۳) دو برابر می‌شود. ۴) تغییری نمی‌کند.

خواص واریانس

۵۰. ارتفاع چهار منطقه کوهستانی برحسب متر نسبت به دریا عبارت است از $141, 155, 169, 135$. واریانس مقادیر مذکور چقدر است؟ (محیط زیست ـ ۸۶)

۱) 170 ۲) 173 ۳) 176 ۴) 180

کاربرد واریانس

۵۱. کدام پارامتر، بیشتر تحت تأثیر انحراف بزرگ است؟ (GIS ـ ۸۵ ، ۸۷)

۱) انحراف چارکی ۲) انحراف متوسط از میانگین

۳) واریانس ۴) نیم‌دامنه

قضیه چی‌بی‌شف

۵۲.طبق قانون چی‌بی‌شف انتظار می‌رود 84 درصد مشاهدات در دامنه $(72 , 88)$ قرار گیرند. مقدار انحراف معیار این مشاهدات کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) 2.4 ۲) 3.2 ۳) 3.6 ۴) 4.2

۵۳. در یک آزمون مهارت میانگین و واریانس نمرات به ترتیب 75 و 64 بوده است. بنا بر قانون چی‌بی‌شف انتظار می‌رود حداقل چند درصد نمرات بین دو عدد 63 و 87 قرار گیرند؟ (حسابداری و مدیریت ـ ۸۸)

۱) 45 ۲) 50 ۳) 55 ۴) 65

۵۴.تعداد مشتریانی که روزانه به یک فروشگاه مراجعه می‌کنند دارای میانگین 24 و انحراف معیار 4 نفر می‌باشد. در یک روز خاص، احتمال اینکه حداقل بین 16 تا 32 نفر مشتری به فروشگاه مراجعه کنند، چقدر است؟ **(اقتصاد ــ ۸۶)**

۱) 0.25 ۲) 0.68 ۳) 0.75 ۴) 0.95

۵۵.جامعه‌ای با میانگین 30 و انحراف معیار 5 را در نظر بگیرید. حداقل چه درصدی از مشاهدات در فاصله 15 تا 45 قرار می‌گیرند؟ **(اقتصاد ــ ۸۷)**

۱) 0.61 ۲) 0.75 ۳) 0.83 ۴) 0.89

۵۶.متوسط دستمزد روزانه کارگران یک کارخانه 10 هزار تومان با انحراف معیار یک هزار تومان است. چه نسبتی از کارگران دارای دستمزد روزانه‌ای بیشتر از 12 هزار تومان یا کمتر از 8 هزار تومان هستند؟ **(اقتصاد ــ ۸۸)**

۱) حداکثر 25% ۲) حداقل 25% ۳) حداکثر 75% ۴) حداقل 75%

۵۷.در یک کارگاه، مطالعه بر روی عملکرد 20 کارگر نشان می‌دهد مدت‌زمان انتظار برای تکمیل عملیات تولیدی با میانگین 9.8 و انحراف معیار 1.2 دقیقه است. طبق قانون چی‌بی‌شف، حداقل 36 درصد مشاهدات در کدام فاصله زمانی قرار می‌گیرد؟ **(برنامه‌ریزی شهری ــ ۸۸)**

۱) $(7.4 , 12.2)$ ۲) $(8.2 , 11.4)$ ۳) $(8.3 , 11.3)$ ۴) $(8.6 , 11)$

تصحیح شپارد

۵۸. در داده‌های طبقه‌بندی شده با متغیرهای پیوسته و توزیع فراوانی اندکی متقارن، فاصله طبقات 6، تعداد جامعه 2400 و مقدار میانگین و واریانس به ترتیب 25 و 12 محاسبه شده است. واریانس تصحیح‌شده شپارد، کدام است؟ **(حسابداری و مدیریت ــ ۸۷)**

۱) 11.25 ۲) 10.5 ۳) 10 ۴) 9

۵۹.در داده‌های آماری دسته‌بندی شده با متغیر پیوسته، تعداد داده‌ها خیلی زیاد و تابع توزیع فراوانی اندکی متقارن است. واریانس محاسبه‌شده با مقایسه واریانس واقعی چگونه است ؟ **(GIS ــ ۸۶)**

۱) همواره بیشتر ۲) همواره کمتر ۳) دقیقاً برابر ۴) کمتر یا بیشتر

۶۰. در 1500 داده آماری پیوسته طبقه‌بندی شده با فاصله طبقات 3 مقدار واریانس برابر 7 محاسبه شده است. نمودار توزیع فراوانی آن متقارن است، انحراف معیار تصحیح‌شده طبق پیشنهاد شپارد، کدام است؟ **(GIS ــ ۸۶)**

۱) 2.25 ۲) 2.4 ۳) 2.5 ۴) 2.6

میانگین و واریانس کل

۶۱. واریانس کل داده‌ها، متشکل از سه گروه جدول روبه‌رو کدام است؟ **(حسابداری و مدیریت ــ ۸۶)**

۱) 48.4 ۲) 47.5 ۳) 45.8 ۴) 45

N	100	200	700
μ	14	18	20
σ^2	50	60	40

۶۲. 20 داده آماری با میانگین 12 و انحراف معیار 2 را با 10 داده آماری دیگر با میانگین 9 و انحراف معیار 3 در نظر می‌گیریم. واریانس 30 داده موجود کدام است؟ (حسابداری و مدیریت ــ ۸۷)

۱) 7.6 ۲) 7.9 ۳) 8.1 ۴) 8.4

۶۳. اگر مشاهداتی به تعداد 100, 200 و 450 به ترتیب با واریانس‌های 16, 25 و 20 به صورت جامعه‌ای واحد ترکیب شوند، درصورتی‌که میانگین این مشاهدات متفاوت باشند، کدام عدد برای واریانس جامعه کل مورد قبول است؟ (برنامه ریزی شهری ــ ۸۸)

۱) 19.8 ۲) 20.6 ۳) 20.8 ۴) 22.1

۶۴. تعداد کارکنان سه گروه متمایز از کارخانه‌ای 100، 150 و 50 نفر است که واریانس نمرات مسئولیت‌پذیر آنان به ترتیب 12, 14 و 9 محاسبه شده است. اگر میانگین‌ها متفاوت باشند، کدام عدد برای واریانس نمرات کل این کارکنان مورد قبول است؟ (GIS ــ ۸۶)

۱) 12.3 ۲) 12.4 ۳) 12.5 ۴) 12.6

۶۵. اگر 50 مشاهده با میانگین 40 و واریانس 25 و 100 مشاهده دیگر با میانگین 55 و واریانس 16 ترکیب شوند، واریانس جامعه کل کدام است؟ (GIS ــ ۸۷)

۱) 19 ۲) 24 ۳) 42 ۴) 69

۶۶. متوسط 4 عدد دو رقمی 12 و متوسط 5 عدد دو رقمی دیگر نیز 18 می‌باشد. متوسط این 9 عدد چقدر است؟ (محیط زیست ــ ۸۶)

۱) 13.5 ۲) 15 ۳) $\frac{138}{9}$ ۴) $\frac{150}{9}$

۶۷. میانگین طول عمر 20 دستگاه، 10 سال و میانگین طول عمر 15 دستگاه دیگر 12 سال می‌باشد، در این صورت میانگین کل نمونه‌ها چقدر است؟ (محیط زیست ــ ۸۶)

۱) 40 ۲) 18 ۳) 11 ۴) $\frac{380}{35}$

۶۸. در یک مطالعه آماری، میانگین وزن 30 نفر، 55 کیلوگرم و میانگین وزن 20 نفر دیگر، 65 کیلوگرم است. میانگین وزن این دو گروه روی هم چند کیلوگرم است؟ (محیط زیست ــ ۸۸)

۱) 58 ۲) 59 ۳) 60 ۴) 64

معیارهای پراکندگی نسبی

ضریب تغییرات

۶۹. برای تشخیص آنکه در دو هفته گذشته یورو باثبات‌تر بوده است یا ین ژاپن، کدام شاخص مناسب‌تر است؟ (اقتصاد ــ ۸۷)

۱) واریانس ۲) میانگین وزنی
۳) میانگین مجذور خطا ۴) ضریب پراکندگی (ضریب تغییرات)

۷۰. میانگین و واریانس 10 داده آماری به ترتیب 16 و 17 می‌باشد. اگر داده‌های 21, 13 و 14 به آن‌ها اضافه شود، ضریب پراکندگی 13 داده جدید کدام است؟ (GIS ـ ۸۷)

۱) 0.25 ۲) 0.28 ۳) 0.32 ۴) 0.35

چولگی

۷۱. اگر از داده‌ها عدد $\frac{1}{2}$ کم شود، ضریب چولگی چه تغییری خواهد داشت؟ (اقتصاد ـ ۸۷)

۱) بدون تغییر باقی می‌ماند.
۲) به اندازه $\frac{1}{8}$ چولگی آن افزایش می‌یابد.
۳) به اندازه $\frac{1}{2}$ چولگی آن افزایش می‌یابد.
۴) به اندازه 0.5 واحد از ضریب چولگی آن کم می‌شود.

۷۲. توزیع نمرات دانشجویان در دو کلاس A و B دارای میانگین و واریانس مساوی است. وقتی می‌توان نمرات دانشجویان کلاس A را به نسبت بهتر دانست که: (اقتصاد ـ ۸۷)

۱) ضریب چولگی آن مثبت باشد.
۲) میانه نمرات آن کمتر از میانگین باشد.
۳) میانگین نمرات آن کمتر از میانه باشد.
۴) میانگین نمرات آن بیشتر از نما باشد.

۷۳. میانگین، انحراف معیار و ضریب چولگی سود دو شرکت تجاری A و B در چند سال گذشته به صورت زیر بوده است. بر این اساس: (اقتصاد ـ ۸۸)

A	B
$\mu_A = 10$	$\mu_B = 10$
$\sigma_A = 2$	$\sigma_B = 2$
$\alpha_3 = -1$	$\alpha_3 = +1$

۱) احتمال سودآوری بیشتر از 10 در شرکت B بیشتر است.
۲) ریسک سرمایه‌گذاری در شرکت B کمتر است.
۳) احتمال سودآوری بیشتر از 10 در شرکت A بیشتر است.
۴) ضریب پراکندگی در شرکت A کمتر است.

۷۴. در یک توزیع چوله به راست کدام رابطه صحیح است؟ (GIS ـ ۸۶)

۱) $\mu < Md < Mo$ ۲) $Md < Mo < \mu$ ۳) $Mo < Md < \mu$ ۴) $\mu < Mo < Md$

۷۵. در توزیع نرمال میانگین، میانه و مد جامعه نسبت به یکدیگر چه رابطه‌ای دارند؟ (محیط زیست ـ ۸۷)

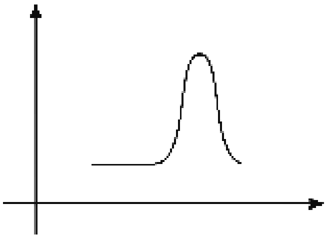
۱) میانه $\leq$ مد $=$ میانگین
۲) میانگین $\leq$ میانه $\leq$ مد
۳) مد $\leq$ میانه $\leq$ میانگین
۴) میانگین $=$ میانه $=$ مد

۷۶. نمرات در یک کلاس 40 نفری مطابق جدول زیر توزیع شده است. در این صورت .... (محیط زیست ـ ۸۸)

فراوانی	نمره
1	$0-5$
3	$5-10$
26	$10-15$
10	$15-20$

۱) تقارن نمرات مستقل است.
۲) نمرات چوله به چپ است.
۳) نمرات چوله به راست است.
۴) نمی‌توان به درستی تقارن یا عدم تقارن نمرات را تعیین نمود.

۷۷. فرض کنید نمودار داده‌ها در جامعه‌ای به صورت زیر است. در آن صورت در این جامعه همواره کدام رابطه برقرار است؟ **(محیط زیست ـ ۸۸)**



۱) مد جامعه کمتر از میانه جامعه و میانه جامعه از میانگین جامعه کمتر است.
۲) میانه جامعه کمتر از مد جامعه و مد جامعه از میانگین جامعه کمتر است.
۳) میانگین جامعه کمتر از میانه جامعه و میانه جامعه از مد جامعه کمتر است.
۴) میانگین جامعه کمتر از مد جامعه و مد جامعه از میانه جامعه کمتر است.

ضریب چولگی گشتاوری و تحلیل چولگی

۷۸. در 50 داده آماری با میانگین 15 و واریانس 4 داریم: $\sum_{i=1}^{50}(x_i-15)^3=24$ **، ضریب چولگی و تفاوت جامعه از نظر تقارن با توزیع نرمال چگونه است؟** **(حسابداری و مدیریت ـ ۸۷)**

۱) 0.12، تقریباً نرمال
۲) 0.06، تقریباً نرمال
۳) 0.12، تفاوت اندک با نرمال
۴) 0.06، تفاوت اندک با نرمال

۷۹. در جامعه‌ای با حجم $N=20$ کمیت‌های زیر محاسبه شده‌اند. ضریب چولگی توزیع چقدر است؟ **(اقتصاد ـ ۸۶)**

$$\sum(x_i-\mu)^2=180\ ,\sum(x_i-\mu)^3=-180$$

۱) -1
۲) -0.33
۳) 0.77
۴) 1

۸۰. در 100 داده آماری با میانگین 7 مجموع مربعات تمام داده‌ها 6500 , $\sum_{i=1}^{100}(x_i-7)^3=192$ ضریب چولگی جامعه چند درصد است؟ **(برنامه‌ریزی شهری ـ ۸۶)**

۱) 4.6
۲) 3
۳) 2.3
۴) 2

۸۱. جدول توزیع فراوانی انحراف از میانگین داده شده است. ضریب چولگی کدام است؟ **(برنامه‌ریزی شهری ـ ۸۸)**

$x-\bar{x}$	-3	-1	0	1	3
f	5	10	20	13	4

۱) $-\frac{3}{13\sqrt{2}}$
۲) $-\frac{2}{13\sqrt{2}}$
۳) $\frac{1}{13}$
۴) $\frac{1}{2\sqrt{2}}$

۸۲. اگر $N=40$ ، $\sum x_i^2=2440$ ، $\sum x_i=240$ و $\sum(x_i-6)^3=75$ باشد، ضریب چولگی جامعه چند درصد است؟ (GIS ـ ۸۶)

۱) 1.5
۲) 2
۳) 2.5
۴) 3

۸۳. در 50 داده آماری با میانگین 12 و واریانس 4 داریم $\sum f_i(x_i-12)^3=-24$ ، این توزیع چگونه است؟ (GIS ـ ۸۷)

۱) چوله به راست ـ تقریباً نرمال
۲) چوله به چپ ـ تقریباً نرمال
۳) چوله به راست ـ تفاوت اندک با نرمال
۴) چوله به چپ ـ تفاوت اندک با نرمال

۸۴.با توجه به جدول انحراف داده‌ها از میانگین، ضریب چولگی کدام است؟ (GIS ـ ۸۸)

$x-\bar{x}$	-5	-3	-1	1	3	5
f	2	5	4	8	2	3

۱) $\frac{2\sqrt{3}}{25}$ ۲) $\frac{5\sqrt{3}}{27}$ ۳) $\frac{3\sqrt{3}}{125}$ ۴) $\frac{6\sqrt{3}}{125}$

چولگی پیرسون، چندکی و تحلیل چولگی

۸۵. در یک توزیع فراوانی داده‌ها، چارک‌های اول، دوم و سوم به ترتیب 36، 61 و 76 می‌باشد، نوع توزیع از نظر تقارن نسبت به نرمال چگونه است؟ (حسابداری و مدیریت ـ ۸۶)

۱) چوله به راست ـ تقریباً نرمال ۲) چوله به راست ـ تفاوت اندک

۳) چوله به چپ ـ تقریباً نرمال ۴) چوله به چپ ـ تفاوت اندک

۸۶.در یک جدول طبقه‌بندی شده چارک اول، دوم و سوم به ترتیب 12 ,15 و 17 محاسبه شده است، نوع چولگی و از نظر قرینگی، با توزیع نرمال چگونه است؟ (حسابداری و مدیریت ـ ۸۸)

۱) چوله به راست ـ تفاوت اندک با نرمال ۲) چوله به راست ـ تفاوت فاحش با نرمال

۳) چوله به چپ ـ تفاوت اندک با نرمال ۴) چوله به چپ ـ تقریبا نرمال

۸۷. در یک جامعه آماری چارک اول و دوم و سوم به ترتیب 26 ,35 و 42 محاسبه شده است. ضریب چولگی و نوع آن چگونه است؟ (برنامه‌ریزی شهری ـ ۸۷)

۱) -0.125، چوله به چپ ۲) -0.125، چوله به راست

۳) 0.125، چوله به چپ ۴) 0.125، چوله به راست

رابطه سه معیار تمرکز (در چولگی ضعیف)

۸۸. میانه و مد یک جامعه آماری به ترتیب 54 و 72 می‌باشد، توزیع جامعه از نظر چولگی معقول است، میانگین کدام است؟ (GIS ـ ۸۶)

۱) 45 ۲) 48 ۳) 63 ۴) 81

۸۹.در توزیع‌های یک‌مدی (تک‌نما)، اگر $\bar{x}$، میانگین حسابی، Mo، مد و Md، میانه باشد، در این صورت کدام رابطه همواره به صورت تجربی برقرار است؟ (محیط زیست ـ ۸۷)

۱) $Mo-Md=3(Mo-\bar{x})$ ۲) $Md-\bar{x}=3(Mo-\bar{x})$

۳) $Md-\bar{x}=3(Md-Mo)$ ۴) $\bar{x}-Mo=3(\bar{x}-Md)$

کشیدگی

۹۰. در یک توزیع نرمال، کشیدگی گشتاوری و کشیدگی صدکی به ترتیب کدام است؟ (از راست به چپ) (GIS ـ ۸۶)

۱) 0.263 و 0.263 ۲) 0.263 و 3 ۳) 3 و 0.263 ۴) 3 و 3

ضریب کشیدگی

۹۱. اگر $N = 20$، $\sum_{i=1}^{20}(x_i - \mu)^4 = 8640$ و واریانس جامعه 12 باشد، ضریب کشیدگی کدام است؟

(حسابداری و مدیریت ــ ۸۶)

۱) صفر ۲) 1 ۳) 2 ۴) 3

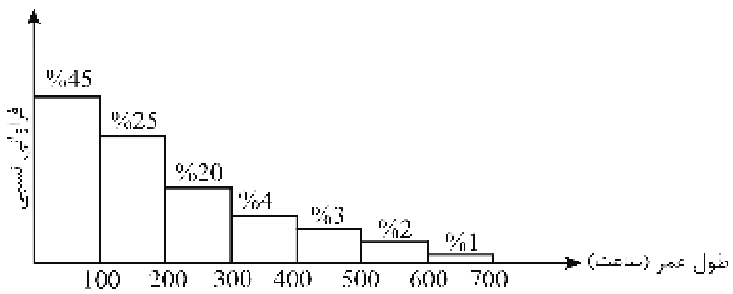
نمودارها

۹۲. در داده‌های جدول زیر، واریانس بین چارک اول و چارک سوم کدام است؟ (GIS ــ ۸۶)

شاخه	برگ					
0	1	1	1	2	2	4
1	0	0	3	4	6	9
2	0	1	2	3		

۱) 28
۲) 28.5
۳) 29
۴) 29.25

۹۳. توزیع طول عمر 1000 تولید نوعی قطعه ترانزیستوری که در یک کارخانه تولید شده است به شکل زیر است. اگر بدانیم 20% از قطعه‌ها از درجه کیفیت کمتری برخوردار می‌باشند و دارای طول عمر پایین هستند، پس این تعداد کمتر از چه مدت عمر می‌کنند؟ **(محیط زیست ــ ۸۷)**



۱) 20
۲) 44.4
۳) 50.5
۴) 70

پاسخ‌های تشریحی

تعاریف

۱. گزینه ۱ درست است.

ترتیب صعودی مقیاس‌ها از نظر قوت عبارت است از: اسمی ≪ ترتیبی ≪ فاصله‌ای ≪ نسبتی

۲. گزینه ۳ درست است.

مراحل تحقیق علمی در آمار عبارت‌اند از:

۱ـ مشخص کردن هدف
۲ـ جمع‌آوری داده‌ها
۳ـ تجزیه و تحلیل داده‌ها
۴ـ بیان یافته‌ها (فرضیه‌سازی)

۳. گزینه ۳ درست است.

- متغیرها
  - کمّی
    - گسسته
    - پیوسته
  - کیفی
    - اسمی
    - ترتیبی

۴. گزینه ۱ درست است.

فراوانی مطلق هر داده، یک عدد طبیعی است که میزان تکرار هر داده را در نمونه یا جامعه نشان می‌دهد. دقت کنید که گزینه ۳ (نسبت فراوانی به نمونه) نشان‌دهنده فراوانی نسبی است.

۵. گزینه ۴ درست است.

تعریف دیگر جامعه آماری: تعدادی از عناصر مطلوب مورد نظر که حداقل دارای یک صفت مشخصه باشند.

۶. گزینه ۳ درست است.

صفات متغیر از نظر قابلیت اندازه‌گیری به دو دسته کمی و کیفی تقسیم می‌شوند.
صفات متغیر کمّی: امکان اندازه‌گیری و بیان یک عدد واحددار وجود دارد و به دو دسته پیوسته (قد و وزن) و گسسته (تعداد اعضای خانواده) تقسیم می‌شوند.
صفات متغیر کیفی: امکان اندازه‌گیری با ابزارهای رایج وجود ندارد و نمی‌توان آن را به صورت عددی واحددار بیان کرد، مانند گروه خونی و رنگ پوست.

۷. گزینه ۴ درست است.

صفات متغیر کیفی: امکان اندازه‌گیری با ابزارهای رایج وجود ندارد و نمی‌توان آن را به صورت عددی واحددار بیان کرد مانند گروه خونی و رنگ پوست.

در این سؤال مراقبت از فضای شهر قابلیت اندازه‌گیری ندارد و نمی‌توان آن را به صورت عدد بیان کرد؛ بنابراین صفت کیفی است.

۸. گزینه ۳ درست است.

صفات متغیر کمّی: امکان اندازه‌گیری و بیان یک عدد واحددار وجود دارد و به دو دسته پیوسته (قد و وزن) و گسسته (تعداد اعضای خانواده) تقسیم می‌شوند.

۹. گزینه ۱ درست است.

۱۰. گزینه ۳ درست است.

صفات متغیر کیفی (اسمی): متغیرهایی که قابل اندازه‌گیری و شمارش نیستند و هیچ نوع ترتیبی و مقایسه‌ای بین متغیرها امکان‌پذیر نیست و تنها برای نام‌گذاری استفاده می‌شوند.

در این سؤال نیز مناطق ۲۲ گانه تهران قابل اندازه‌گیری نیستند و تنها برای نام‌گذاری است و ارزش دیگری ندارد.

۱۱. گزینه ۱ درست است.

جامعه آماری: مجموعه‌ای از افراد یا اشیای مورد مطالعه، جامعه آماری نامیده می‌شود.

نمونه آماری: تعداد محدودی از عناصر جامعه که خصوصیات جامعه اصلی را داشته باشند.

پارامتر: شاخصی که با اندازه‌گیری از تمام عناصر جامعه به دست می‌آید.

متغیر تصادفی: موضوع مورد مطالعه در تحقیقات آماری متغیر تصادفی است که بسته به نوع موضوع، کمّی یا کیفی خواهد بود.

۱۲. گزینه ۲ درست است.

یادآوری: در آمار استنباطی از روی داده‌های نمونه به محاسبه برآوردگرها (آماره‌ها) می‌پردازیم.

۱۳. گزینه ۴ درست است.

با توجه به اینکه وزن افراد یک متغیر کمّی و پیوسته است، مقیاس فاصله‌ای یا نسبتی دارد و چون نسبت را حفظ می‌کند، مقیاس نسبتی برای آن مناسب‌تر است.

۱۴. گزینه ۳ درست است.

جامعه آماری: درختان باغ (مجموعه‌ای از عناصر مورد نظر که حداقل دارای یک صفت مشترک باشند)

متغیر تصادفی: قطر درخت (از عنصری به عنصر دیگر متغیر است و همان موضوع مورد بررسی است)

دقت کنید که چون موضوع مورد مطالعه ضخامت درختان است، بنابراین قطر درختان متغیر تصادفی خواهد بود.

۱۵. گزینه ۴ درست است.

صفت مشخصه: صفتی مشترک بین همه عناصر جامعه آماری که آن‌ها را از سایر جوامع متمایز می‌کند.

۱۶. گزینه ۲ درست است.

با توجه به اینکه نام‌گذاری علاوه بر اسم بودن، ترتیب را نیز بین نام‌ها برقرار می‌کند، از مقیاس ترتیبی استفاده شده است.

متوسط > خوب > تیزهوش

$1 > 2 > 3$

فراوانی

۱۷. گزینه ۲ درست است.

بهتر است ابتدا با توجه به داده‌های مسئله جدول توزیع فراوانی را کامل کنیم. می‌دانیم که فراوانی تجمعی طبقه آخر برابر = N تعداد داده‌هاست. بنابراین با داشتن فراوانی نسبی دسته وسط می‌توانیم فراوانی مطلق دسته وسط را به دست آوریم:

دسته وسط

دسته	(1)	(2)	(3)	(4)	(5)
F_{c_i}	8	25	$\boxed{a=37}$	58	$75=N$
F_i			$\boxed{12}$	$\boxed{21}$	

$$\begin{cases} f_i=\dfrac{F_i}{N} \rightarrow f_3=\dfrac{F_3}{N} \rightarrow 0.16=\dfrac{F_3}{75} \rightarrow F_3=12 \\ F_i=F_{c_i}-F_{c_{i-1}} \rightarrow F_{c_i}=F_i+F_{c_{i-1}} \rightarrow F_{c_3}=F_3+F_{c_2}=12+25=37 \\ F_i=F_{c_i}-F_{c_{i-1}} \rightarrow F_4=F_{c_4}-F_{c_3} \rightarrow F_4=58-37=21 \\ N=F_c=75 \end{cases}$$

۱۸. گزینه ۲ درست است.

نکته:

۱- فراوانی تجمعی دسته آخر با تعداد داده‌ها (N) برابر است $\leftarrow F_{c_5}=N=50$

۲-

$$\begin{cases} f_i=\dfrac{F_i}{N} \quad (\text{فراوانی نسبی}) \\ f_i\times 100=\dfrac{F_i}{N}\times 100 \quad (\text{درصد فراوانی نسبی}) \end{cases}$$

درصد فراوانی نسبی دسته وسط $=f_3\times 100=24 \rightarrow \dfrac{F_3}{50}\times 100=24 \rightarrow F_3=12$ (فراوانی مطلق دسته سوم)

حال با داشتن F_3 می‌توان مقدار a (فراوانی تجمعی دسته سوم $=F_{c_3}$) را به دست آورد.

۳ـ فراوانی تجمعی هر دسته برابر است با فراوانی مطلق آن دسته به‌علاوه فراوانی تجمعی دسته قبل $\left(F_{c_i}=F_i+F_{c_{i-1}}\right)$:

$a=F_{c_3}=F_3+F_{c_2}=12+14=26$

حال با کامل شدن ردیف فراوانی تجمعی جدول به راحتی ردیف فراوانی مطلق به دست می‌آید، زیرا:

$F_i=F_{c_i}-F_{c_{i-1}}$

$F_4=F_{c_4}-F_{c_3}=41-26=15$

دسته	اول	دوم	سوم	چهارم	پنجم
F_{c_i}	5	14	$\boxed{a=26}$	41	$50=N$
F_i	5	9	12	$\boxed{15}$	9
$f_i\times 100$			24		

معیارهای تمرکز

مد (نما)

۱۹.گزینه ۱ درست است.

سؤال مطرح‌شده نوعی «نظرسنجی» است و درواقع هدف، دانستن نظر مردم درباره ساعات کار بانک‌هاست؛ بنابراین مهم این است که بدانیم بیشترین رأی (نظر) مردم بر چه ساعتی بوده است. به عبارت دیگر شاخص مد که تنها کاربرد آن در نظر سنجی است، بهترین معیار خواهد بود.

۲۰. گزینه ۲ درست است.

با توجه به اینکه مد جامعه 16 است، حتماً در دسته سوم $(15-18)$ است. حال با توجه به فرمول محاسبه مد داریم:

$$Mo = \text{حدپایین} + \frac{d_1}{d_1+d_2} \times \text{طول دسته}$$

$$16 = 15 + \frac{(15-12)}{(15-12)+(15-a)} \times 3 \rightarrow 1 = \frac{3\times 3}{18-a} \rightarrow 18-a=9 \rightarrow a=9$$

۲۱.گزینه ۲ درست است.

یادآوری:

$$\text{درصد فراوانی نسبی} = \frac{\text{فراوانی مطلق}}{\text{تعداد داده‌ها}} \times 100 \rightarrow f_i = \frac{F_i}{N} \times 100$$

برای به دست آوردن درصد فراوانی نسبی دسته سوم، به F_3 یعنی α و N تعداد داده‌ها نیاز داریم.

دقت کنید که مد جامعه 25.5 است، پس حتماً در دسته $(23-27)$ قرار دارد. حال با توجه به فرمول محاسبه مد داریم:

$$Mo = \text{حدپایین دسته} + \frac{d_1}{d_1+d_2} \times \text{طول دسته}$$

$$25.5 = 23 + \frac{(\alpha-12)}{(\alpha-12)+(\alpha-14)} \times 4 \rightarrow 4\alpha - 48 = 5\alpha - 65 \rightarrow \boxed{F_3 = \alpha = 17}$$

$$\begin{cases} f_3 = \dfrac{F_3}{N} \times 100 = \dfrac{17}{50} \times 100 = 34 \\ N = \sum F_i = 7+12+17+14 = 50 \end{cases}$$

میانه

۲۲.گزینه ۳ درست است.

ابتدا X_iها باید به ترتیب صعودی مرتب شوند که البته در این سؤال مرتب بوده‌اند.

X_i	10	11	12	13 (داده 30ام تا 55ام)	14 (داده 56ام تا 110ام)	
F_i	5	10	14	26	55	$\sum F_i = N = 110$
F_{c_i}	5	15	29	55	110	

$$\text{محل میانه}: \frac{N}{2} + \frac{1}{2} = \frac{110}{2} + \frac{1}{2} = 55.5$$

$$\text{مقدار میانه}: Me = X_{(55)} + 0.5\left(X_{(56)} - X_{(55)}\right) = 13 + 0.5(14-13) = 13.5$$

۲۳.گزینه ۳ درست است.

$$\begin{cases} Md = L_i + \dfrac{\frac{N}{2} - F_{c_{i-1}}}{F_i} \times I \rightarrow 41 = L_i + \dfrac{\frac{50}{2} - 18}{10} \times 5 \rightarrow L_i = 41 - 3.5 = 37.5 \\ Md = 41 \ , \ N = 50 \ , \ F_{c_{i-1}} = 18 \ , \ F_i = 10 \ , \ I = 5 \end{cases}$$

حد پایین دسته $: L_i = 37.5 \rightarrow$ حد بالای دسته $: U_i = 37.5 + 5 = 42.5 \rightarrow$ حدود دسته $: (37.5 \ , \ 42.5)$

۲۴.گزینه ۲ درست است.

برای محاسبه میانه، فراوانی تجمعی طبقه ماقبل نیاز است؛ اما در این مسئله فراوانی مطلق طبقه میانه داده شده است، بنابراین:

$$\begin{cases} F_{c_i} = F_{c_{i-1}} + F_i \rightarrow 35 = F_{c_{i-1}} + 12 \rightarrow F_{c_{i-1}} = 26 \\ F_{c_i} = 38 \ , \ F_i = 12 \end{cases}$$

$$\begin{cases} Md = L_i + \dfrac{\frac{N}{2} - F_{c_{i-1}}}{F_i} \times I \rightarrow 23 = L_i + \dfrac{\frac{60}{2} - 26}{12} \times 3 \rightarrow L_i = 23 - 1 = 22 \\ Md = 23 \ , \ N = 60 \ , \ F_{c_{i-1}} = 26 \ , \ F_i = 12 \ , \ I = 3 \end{cases}$$

حد پایین دسته $: L_i = 22 \rightarrow$ حد بالای دسته $: U_i = 22 + 3 = 25 \rightarrow$ حدود دسته $: (22 \ , \ 25)$

۲۵.گزینه ۳ درست است.

الف) جدول F_{c_i} را محاسبه می‌کنیم:

رده	$4-6$	$6-8$	$8-10$	$10-12$	$12-14$
F_i	2	7	4	22	15
F_{c_i}	2	9	13	35	$50 = N$

ب) میانه در اولین طبقه‌ای است که $F_{c_i} \geq \frac{N}{2} = \frac{50}{2} = 25$ باشد؛ بنابراین طبقه $(10-12)$ طبقه میانه‌دار است

ج) مقدار میانه برابر است با:

$$Md = L_i + \dfrac{\frac{N}{2} - F_{c_{i-1}}}{F_i} \times I \rightarrow Md = 10 + \dfrac{\frac{50}{2} - 13}{22} \times 2 = 10 + \frac{12}{22} \times 2 = 10 + \frac{12}{11} = \frac{110 + 12}{11} = \frac{122}{11}$$

خواص میانه

۲۶.گزینه ۴ درست است.

یادآوری:

(۱) خواص انحراف معیار: انحراف معیار داده‌های مساوی برابر صفر است: $x_1 = x_2 = \ldots = x_n \Leftrightarrow \sigma = 0$

(۲) خواص میانه: طبق خاصیت خطی بودن میانه داریم: $Md(ax + b) = aMd + b$

(۳) خواص میانه: میانه در داده‌های مساوی برابر داده‌هاست: $x_i = a, a, a, \ldots, a \rightarrow Md = a$

با توجه به یادآوری (۱) چون انحراف معیار داده‌ها برابر صفر است، 5 داده با هم برابرند و چون مقدار یکی 15 است، همه مشاهدات 15 هستند.

$$\sigma = 0 \xrightarrow{(1)} x_1 = x_2 = x_3 = x_4 = x_5 = 15 \xrightarrow{(3)} Md_x = 15$$

$$(2)\ Md(2x + 1) = 2Md_x + 1 = 2 \times 15 + 1 = 31$$

چندک

۲۷. گزینه ۱ درست است.

الف) جدول F_{c_i} را محاسبه می‌کنیم:

X	2 − 4	4 − 6	6 − 8	8 − 10	
F_i	5	3	8	4	$N=\sum F_i=20$
F_{c_i}	5	8	16	20	

ب) چارک در طبقه‌ای است که $F_{c_i} \geq \frac{aN}{4}=\frac{3\times 20}{4}=15$ باشد؛ بنابراین، طبقه $(6-8)$ طبقه چارک‌دار است.

ج) مقدار چارک سوم برابر است با:

$$Q_a=L_i+\frac{\frac{aN}{4}-F_{c_{i-1}}}{F_i}\times I \xrightarrow[N=20]{a=3} Q_3=6+\frac{\frac{3\times 20}{4}-8}{8}\times 2=6+\frac{7}{4}=\frac{31}{4}$$

توجه: منظور از f همان F (فراوانی مطلق) است. در بعضی کتاب‌ها فراوانی مطلق را با f و فراوانی تجمعی را با F نمایش می‌دهند. البته می‌دانیم که f فراوانی نسبی مقداری بین $(0,1)$ دارد، بنابراین تشخیص آن ساده است.

۲۸. گزینه ۲ درست است.

الف) جدول F_{c_i} را محاسبه می‌کنیم:

حدود دسته	15−20	20−25	25−30	30−35	35−40	
F_i	22	28	16	12	14	$N=\sum F_i=92$
F_{c_i}	22	50	66	78	92	

ب) چارک در دسته‌ای است که $F_{c_i} \geq \frac{aN}{4}=\frac{3\times 92}{4}=69$ باشد؛ بنابراین، دسته $(30-35)$ دسته چارک‌دار است.

ج) مقدار چارک سوم برابر است با:

$$Q_a=L_i+\frac{\frac{aN}{4}-F_{c_{i-1}}}{F_i}\times I \xrightarrow[N=92]{a=3} Q_3=30+\frac{\frac{3\times 92}{4}-66}{12}\times 5=31.25$$

توجه: منظور از f_i در جدول توزیع فراوانی صورت سؤال همان F_i (فراوانی مطلق) است که البته به سادگی متوجه می‌شویم، زیرا f_i (فراوانی نسبی) همواره مقداری بین صفر و یک است.

۲۹. گزینه ۲ درست است.

ابتدا باید فاصله طبقات را به دست آوریم تا بر اساس آن بتوانیم جدول توزیع فراوانی را رسم کنیم. توجه کنید که چون تعداد طبقات 7 تاست، طبقه وسط، طبقه چهارم خواهد بود.

$$\begin{cases} I=\frac{R}{k}=\frac{x_{(n)}-x_{(1)}}{7}=\frac{54-12}{7}=\frac{42}{7}=6 \\ I:\text{فاصله طبقات} \quad , \quad k:\text{تعداد طبقات} \quad , \quad R:\text{دامنه تغییرات} \end{cases}$$

حدود دسته‌ها	12−18	18−24	24−30	$\overbrace{30-36}^{\text{طبقه وسط}}$	…
F_i				9	

حال به دنبال فراوانی تجمعی نسبی دسته وسط هستیم، پس باید F_{c_i} را پیدا کنیم؛ می‌دانیم که دهک ششم در طبقه وسط یعنی $(30-36)$ قرار دارد، بنابراین داریم:

$$D_a = L_i + \frac{\frac{aN}{10} - F_{c_{i-1}}}{F_i} \times I \to 32 = 30 + \frac{\frac{6\times 120}{10} - F_{c_{i-1}}}{9} \times 6 \to F_{c_{i-1}} = 69$$

$$F_{c_i} = F_i + F_{c_{i-1}} = 9 + 69 = 78 \to f_{c_i} = \frac{F_{c_i}}{N} = \frac{78}{120} = 0.65 = \%65$$

میانگین هارمونیک

۳۰. گزینه ۳ درست است.

با توجه به واحد ترکیبی کیلومتر در ساعت، از میانگین هارمونیک استفاده می‌کنیم. دقت کنید که فاصله طی‌شده با سرعت معین درواقع وزن سرعت هواپیما محسوب می‌شود.

$$\bar{x}_H = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{3+5+4}{\frac{3}{600}+\frac{5}{750}+\frac{4}{800}} = \frac{12}{\frac{1}{200}+\frac{1}{150}+\frac{1}{200}} = \frac{12}{\frac{1}{100}+\frac{1}{150}} = \frac{12\times 1500}{15+10} = \frac{12\times 1500}{25} = 12\times 60 = 720$$

۳۱. گزینه ۲ درست است.

با توجه به واحد ترکیبی «کیلومتر در ساعت» از میانگین هارمونیک استفاده می‌کنیم:

$$\bar{x}_H = \frac{\sum w_i}{\sum \frac{w_i}{X_i}} = \frac{\frac{1}{3}+\frac{1}{4}+\left(1-\frac{1}{3}-\frac{1}{4}\right)}{\frac{\frac{1}{3}}{120}+\frac{\frac{1}{4}}{80}+\frac{1-\frac{1}{3}-\frac{1}{4}}{100}} = \frac{1}{\frac{1}{360}+\frac{1}{320}+\frac{1}{240}} = \frac{2880}{8+9+12} = \frac{2880}{29} = 99.3$$

دقت کنید که کل مسیر 1 واحد در نظر گرفته می‌شود که در این سؤال یک مسیر به سه قسمت شکسته شده $\frac{1}{3}$ و $\frac{1}{4}$ و بقیه یعنی $1-\frac{1}{3}-\frac{1}{4}=\frac{5}{12}$.

۳۲. گزینه ۲ درست است.

با توجه به واحد ترکیبی «کیلومتر در ساعت» از میانگین هارمونیک برای متوسط سرعت استفاده می‌کنیم:

$$\bar{x}_H = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{\frac{2}{3}+\frac{1}{4}+\frac{1}{12}}{\frac{\frac{2}{3}}{80}+\frac{\frac{1}{4}}{90}+\frac{\frac{1}{12}}{60}} = \frac{1}{\frac{\frac{9}{3}+\frac{4}{4}+\frac{6}{12}}{360}} = \frac{360\times 2}{9} = 80 \quad ، \quad \text{بقیه مسافت} = 1-\left(\frac{2}{3}+\frac{1}{4}\right) = \frac{1}{12}$$

میانگین هندسی

۳۳. گزینه ۴ درست است.

توجه: منظور از f_i همان F_i (فراوانی مطلق) است. در بعضی از کتاب‌ها فراوانی مطلق را با f_i و فراوانی تجمعی را با F_i نمایش می‌دهند. البته می‌دانیم که f_i (فراوانی نسبی) عددی بین $(0, 1)$ است، بنابراین تشخیص آن ساده است.

باید هر دو میانگین حسابی و هندسی را محاسبه کرد، سپس تفاضل میانگین حسابی از هندسی را به دست آورد.

x_i	9	12	16	
F_i	2	3	2	$N=7$

$$\bar{x}=\frac{\sum F_i x_i}{N}=\frac{2\times 9+3\times 12+2\times 16}{7}=\frac{86}{7}=12.28$$

$$\bar{x}_G=\sqrt[N]{\prod_{i=1}^{n} x_i^{F_i}}=\sqrt[7]{\underbrace{9^2}_{(3^2)^2}\times 12^3\times \underbrace{16^2}_{(4^2)^2}}=\sqrt[7]{\underbrace{3^4\times 4^4}_{12^4}\times 12^3}=\sqrt[7]{12^4\times 12^3}=\sqrt[7]{12^7}=12$$

فزونی میانگین حسابی از میانگین‌های هندسی : $\bar{x}-\bar{x}_G=12.28-12=0.28$

۳۴.گزینه ۲ درست است.

با توجه به رابطه سوم میانگین هندسی داریم:

$$\bar{x}_G=\sqrt[84-80]{\frac{N_{84}}{N_{80}}}=\sqrt[4]{\frac{400}{100}}=\sqrt[4]{4}=\sqrt[4]{2^2}=\sqrt[2]{2}=1.41\rightarrow(1.41-1)\times 100=\%41$$

بهتر است مقدار $\sqrt{2}=1.41$, $\sqrt{3}=1.73$ را به خاطر داشته باشیم.

۳۵. گزینه ۱ درست است.

با توجه به اینکه نرخ رشد فروش برحسب درصد داده شده است، بنا بر رابطه دوم میانگین هندسی داریم:

یادآوری: «درصد» هر سال را با جمع کردن با عدد یک به «برابر» تبدیل کرده و از رابطه اول میانگین هندسی استفاده می‌کنیم:

$$\bar{x}_G=\sqrt[n]{\prod_{i=1}^{n} x_i}=\sqrt[2]{(+0.8+1)(-0.8+1)}=\sqrt[2]{1.8\times 0.2}=\sqrt[2]{0.36}=0.6\text{«برابر»}\rightarrow(0.6-1)=-0.4=\%40\ \text{کاهش}$$

دقت کنید که درواقع رابطه بالا به صورت زیر به دست آمده است:

$$\begin{cases}\text{سال اول}: x\\ \text{سال دوم}: x+0.8x=1.8x\\ \text{سال سوم}: 1.8x-0.8(1.8x)=0.36x\end{cases}$$

$$\bar{x}_G=\sqrt[\text{تعداد سال‌ها}-1]{\frac{N_3}{N_1}}=\sqrt[3-1]{\frac{0.36x}{X}}=\sqrt[2]{0.36}=0.6\ \text{برابر}\rightarrow(0.6-1)=-0.4=\%40\ \text{کاهش}$$

۳۶.گزینه ۳ درست است.

$$\bar{x}_G=\sqrt[n]{\prod_{i=1}^{n} X_i}=\sqrt[5]{96\times 72\times 8\times 24\times 6}=\sqrt[5]{2^5\times 3\times 2^3\times 3^2\times 2^3\times 2^3\times 3\times 2\times 3}=\sqrt[5]{2^{15}\times 3^5}=2^3\times 3=24$$

۳۷. گزینه ۱ درست است.

با توجه به اینکه میانگین نسبت خواسته شده است، از میانگین هندسی استفاده می‌کنیم:

$$\bar{x}_G=\sqrt[n]{\prod_{i=1}^{n} x_i}=\sqrt[3]{\frac{20}{21}\times\frac{15}{14}\times\frac{245}{432}}=\sqrt[3]{\frac{2^2\times 5\times 3\times 5\times 7^2\times 5}{3\times 7\times 2\times 7\times 2^4\times 3^3}}=\sqrt[3]{\frac{5^3}{2^3\times 3^3}}=\frac{5}{2\times 3}=\frac{5}{6}$$

میانگین حسابی

۳۸. گزینه ۳ درست است.

با توجه به دو نکته

۱ـ هزینه متوسط در جدول فراوانی به ترتیب صعودی است.

۲ـ تعداد کل خانوارها برابر $N = 100$ است.

چون میانگین 20% از پرخرج‌ترین خانوارها خواسته شده ، از انتهای جدول که هزینه متوسط بیشتر است به میزان مجموع 20 خانوار جدا می‌کنیم و سپس میانگین جدول جدا شده را حساب می‌کنیم.

هزینه متوسط	15	35	60	85	210	380	
تعداد خانوار	32	48	12	5	2	1	$N=100$

20 خانوار

هزینه متوسط	60	85	210	380	
تعداد خانوار	12	5	2	1	$N=20$

$$\mu = \frac{\sum F_i x_i}{N} = \frac{12 \times 60 + 5 \times 85 + 2 \times 210 + 1 \times 380}{20} = \frac{1945}{20} = 97.25$$

۳۹.گزینه ۲ درست است.

یادآوری: درصورتی‌که در میانگین پیراسته $LN = 20$ باشد، از دو طرف داده‌ها به تعداد $N \times 20\%$ داده کم می‌شود و میانگین داده‌های باقی‌مانده میانگین پیراسته خواهد بود.

در این سؤال $N \times 20\% = 0.2 \times 50 = 10$ است، بنابراین از دو طرف داده‌های مرتب‌شده 10 داده حذف می‌شود.

توجه: وقتی داده‌ها متقارن هستند، یعنی فاصله آن‌ها از داده وسط به سمت کمترین و بیشترین داده یکسان است (درواقع تصاعد حسابی تشکیل خواهند داد)؛ برای مثال $1, 3, 5, 7, 9, 11, 13, 15$ داده‌هایی متقارن هستند. می‌دانیم که میانگین در داده‌هایی که تشکیل تصاعد حسابی می‌دهند اگر N فرد باشد، داده وسط و اگر N زوج باشد، میانگین دو داده وسط است و یا به عبارت دیگر: $\mu = \frac{x_{(1)} + x_{(n)}}{2}$

بنابراین میانگین داده‌های بالا برابر است با $\mu = \frac{1+15}{2} = 8$.

حال اگر میانگین پیراسته را با $LN = 20$ برای این داده‌ها به دست آوریم، ابتدا باید از دو طرف داده‌ها $N \times 20\% = 0.2 \times 8 = 3.2$ سه داده را حذف کنیم؛ داده‌های باقی‌مانده 9 , 7 هستند که میانگین آن‌ها همچنان 8 خواهد ماند.

در این سؤال تعداد داده‌ها زوج است $N = 50$ ، پس درصورتی‌که میانگین حسابی آن‌ها 12 باشد، میانگین دو داده وسط 12 بوده است. همچنین هنگامی که 10 داده نیز از ابتدا و انتهای داده‌ها حذف شود میانگین 30 داده باقی‌مانده (N زوج) باز هم 12 خواهد ماند، زیرا هنوز میانگین دو داده وسط آن 12 است.

۴۰. گزینه ۴ درست است.

حدود دسته‌های جدول اعداد بزرگی هستند؛ بنابراین برای محاسبه میانگین به صورت زیر عمل می‌کنیم:

الف) طبقه $(232-248)$ دارای بیشترین فراوانی است؛ بنابراین جدول به صورت زیر خواهد بود:

x'	-1	$\boxed{0}$	1	2
$C-L$	$216-232$	$\boxed{232-248}$	$248-264$	$264-280$
F_i	16	$\boxed{29}$	13	12

ب) میانگین را با مرکز طبقات جدید محاسبه می‌کنیم:

x'	-1	0	1	2	
F_i	16	29	13	12	$N=\sum F_i=70$
$F_i x'_i$	-16	0	13	24	$\sum F_i x'_i=21$

$$\longrightarrow \quad \mu_{X'}=\frac{\sum F_i x'_i}{N}=\frac{21}{70}=\frac{3}{10}$$

ج) میانگین داده‌های اصلی (μ_X) برابر است با:

$$\begin{cases}\mu_X=\mu_{X'}\times I+a \ \rightarrow\ \mu_X=\dfrac{3}{10}\times 16+240=\dfrac{48}{10}+240=4.8+240=244.8\\ a=\dfrac{L_k+U_k}{2}=\dfrac{232+248}{2}=240\\ \mu_{X'}=\dfrac{3}{10}\ ,\ I=16\end{cases}$$

۴۱. گزینه ۴ درست است.

$$\mu=25+A \ \rightarrow\ A=\mu-25 \ \rightarrow\ A=\mu_{(x_i-25)}$$

بهتر است ابتدا مرکز دسته‌ها را از عدد 25 کم کنیم، سپس مجدداً میانگین بگیریم. میانگین داده‌های جدید با عدد A برابر است:

حدود دسته	$15-19$	$19-23$	$23-27$	$27-31$	$31-35$	
$x_i=\dfrac{\text{حد پایین + حد بالا}}{2}$	17	21	25	29	33	
x_i-25	-8	-4	0	4	8	
F_i	6	12	15	9	8	$N=\sum F_i=50$

$$A=\mu_{(x_i-25)}=\frac{\sum F_i(x_i-25)}{N}=\frac{6\times(-8)+12\times(-4)+15\times 0+9\times 4+8\times 8}{50}=\frac{4}{50}=0.08$$

۴۲. گزینه ۱ درست است.

ابتدا باید فاصله طبقات را به دست آوریم تا بر اساس آن بتوانیم جدول توزیع فراوانی را تشکیل دهیم

$$\begin{cases} I=\dfrac{R}{k}=\dfrac{x_{(n)}-x_{(1)}}{k}=\dfrac{32.5-12.5}{4}=\dfrac{20}{4}=5\\ I:\ \text{فاصله طبقات}\ ,\ k:\ \text{تعداد طبقات}\ ,\ R:\ \text{دامنه تغییرات}\end{cases}$$

حال فراوانی مطلق هر دسته با توجه به رابطه $F_i=F_{c_i}-F_{c_{i-1}}$ به دست می‌آید.

حدود دسته	$12.5-17.5$	$17.5-22.5$	$22.5-27.5$	$27.5-32.5$
F_{c_i}	10	29	45	$60=N$
F_i	10	$29-10=19$	$45-29=16$	$60-45=15$
$x_i=\dfrac{\text{حد پایین + حد بالا}}{2}$	15	20	25	30

$$\mu=\frac{\sum F_i x_i}{N}=\frac{10\times 15+19\times 20+16\times 25+15\times 30}{60}=\frac{1380}{60}=23$$

معیارهای پراکندگی

انحراف چارکی

۴۳.گزینه ۲ درست است.

حدود دسته	<10	$10-14$	$14-18$ ($Q_1 \downarrow$)	$18-22$	$22-26$ ($Q_3 \downarrow$)	≥ 26
F_i (فراوانی)	4	8	10	12	9	7
F_{c_i} (فراوانی تجمعی)	4	12	22	34	43	$50=N$

$$\text{انحراف چارکی}: SIQR=\frac{Q_3-Q_1}{2}=\frac{23.5-14.2}{2}=\frac{9.3}{2}=4.65$$

اولین دسته‌ای که فراوانی تجمعی‌اش بزرگ‌تر یا مساوی $\frac{N}{4}$ باشد، محل چارک اول است.

$$\text{محل چارک اول}: F_{c_i}\geq\frac{N}{4}=\frac{50}{4}=12.5 \quad \rightarrow \quad \text{دسته سوم}: (14-18)$$

$$Q_1=\text{حد پایین دسته}+\frac{\frac{N}{4}-F_{c_{i-1}}}{F_i}\times\text{طول دسته}=14+\frac{\frac{50}{4}-12}{10}\times 4=14.2$$

اولین دسته‌ای که فراوانی تجمعی‌اش بزرگ‌تر یا مساوی $\frac{3N}{4}$ باشد، محل چارک سوم است.

$$\text{محل چارک سوم}: F_{c_i}\geq\frac{3N}{4}=\frac{3\times 50}{4}=37.5 \quad \rightarrow \quad \text{دسته پنجم}: (22-26)$$

$$Q_3=\text{حد پایین دسته}+\frac{\frac{3N}{4}-F_{c_{i-1}}}{F_i}\times\text{طول دسته}=22+\frac{\frac{3\times 50}{4}-34}{9}\times 4=23.5$$

۴۴.گزینه ۳ درست است.

$$Q_1=52 \quad , \quad Q_2=70 \quad , \quad Q_3=84$$

$$\text{انحراف چارکی}: SIQR=\frac{Q_3-Q_1}{2}=\frac{84-52}{2}=\frac{32}{2}=16$$

۴۵.گزینه ۴ درست است.

حدود دسته	<8	$8-11$ ($Q_1 \downarrow$)	$11-14$	$14-17$ ($Q_3 \downarrow$)	$17-20$	≥ 20
F_i (فراوانی)	5	12	15	13	8	7
F_{c_i} (فراوانی تجمعی)	5	17	32	45	53	$60=N$

$$\text{انحراف چارکی}: SIQR=\frac{Q_3-Q_1}{2}=\frac{17-10.5}{2}=\frac{6.5}{2}=3.25$$

اولین دسته‌ای که فراوانی تجمعی‌اش بزرگ‌تر یا مساوی $\frac{N}{4}$ باشد، محل چارک اول است.

$$F_{c_i} \geq \frac{N}{4} = \frac{60}{4} = 15 \quad \rightarrow \quad \text{دسته دوم}: (8-11)$$

$$Q_1 = \text{حد پایین دسته} + \frac{\frac{N}{4} - F_{c_{i-1}}}{F_i} \times \text{طول دسته} = 8 + \frac{\frac{60}{4} - 5}{12} \times 3 = 10.5$$

اولین دسته‌ای که فراوانی تجمعی‌اش بزرگ‌تر یا مساوی $\frac{3N}{4}$ باشد، محل چارک سوم است.

$$F_{c_i} \geq \frac{3N}{4} = \frac{3 \times 60}{4} = 45 \quad \rightarrow \quad \text{دسته چهارم } (14-17)$$

$$Q_3 = \text{حد پایین دسته} + \frac{\frac{3N}{4} - F_{c_{i-1}}}{F_i} \times \text{طول دسته} = 14 + \frac{\frac{3 \times 60}{4} - 32}{13} \times 3 = 17$$

۴۶. گزینه ۱ درست است.

$$\text{انحراف چارکی}: SIQR = \frac{Q_3 - Q_1}{2} = \frac{95-31}{2} = \frac{64}{2} = 32$$

۴۷. گزینه ۲ درست است.

نمرات	<7	$7-10$ ($Q_1 \downarrow$)	$10-13$	$13-16$ ($Q_3 \downarrow$)	$16-19$	≥ 19
F_i (فراوانی)	7	9	16	19	8	5
F_{c_i} (فراوانی تجمعی)	7	16	32	51	59	$64 = N$

$$\text{انحراف چارکی}: SIQR = \frac{Q_3 - Q_1}{2} = \frac{15.52 - 10}{2} = 2.76$$

اولین دسته‌ای که فراوانی تجمعی آن بزرگ‌تر یا مساوی $\frac{N}{4}$ باشد، محل چارک اول است.

$$F_{c_i} \geq \frac{N}{4} = \frac{64}{4} = 16 \quad \rightarrow \quad \text{دسته دوم } (7-10)$$

$$Q_1 = \text{حد پایین دسته} + \frac{\frac{N}{4} - F_{c_{i-1}}}{F_i} \times \text{طول دسته} = 7 + \frac{\frac{64}{4} - 7}{9} \times 3 = 10$$

اولین دسته‌ای که فراوانی تجمعی آن بزرگ‌تر یا مساوی $\frac{3N}{4}$ باشد، محل چارک سوم است.

$$F_{c_i} \geq \frac{3N}{4} = \frac{3 \times 64}{4} = 48 \quad \rightarrow \quad \text{دسته چهارم } (13-16)$$

$$Q_3 = \text{حد پایین دسته} + \frac{\frac{3N}{4} - F_{c_{i-1}}}{F_i} \times \text{طول دسته} = 13 + \frac{\frac{3 \times 64}{4} - 32}{19} \times 3 = 15.52$$

واریانس

۴۸.گزینه ۳ درست است.

$$A\cdot D_\mu=\frac{\sum|x_i-\mu|}{N}=0\rightarrow x_i=\mu\rightarrow x_1=x_2=\ldots=x_{12}=\mu=15$$

$$\mu_{\text{جدید}}=\frac{\sum_{i=1}^{15}x_i}{15}=\frac{\sum_{i=1}^{12}x_i+20+16+24}{12+3}=\frac{12\times15+60}{15}=\frac{240}{15}=16$$

$$\sigma^2_{\text{جدید}}=\frac{\sum_{i=1}^{12}(x_i-\mu)^2+(20-16)^2+(16-16)^2+(24-16)^2}{15}=\frac{12\times(15-16)^2+16+0+64}{15}=\frac{92}{15}=6.13$$

واریانس نمونه

۴۹.گزینه ۲ درست است.

یادآوری:

$$\text{واریانس نمونه}\ :S^2=\frac{\sum(x_i-\bar{x})^2\rightarrow\ \text{مجموع مربعات تفاضل از میانگین}}{n-1}$$

در این سؤال، صورت کسر واریانس را ثابت در نظر گرفته‌ایم و مخرج کسر یعنی حجم نمونه را افزایش داده‌ایم که با افزایش مخرج کسر، کل کسر یعنی واریانس کاهش می‌یابد.

خواص واریانس

۵۰. گزینه ۲ درست است.

$X_i:135, 141, 155, 169$

با توجه به خواص واریانس برای راحتی در محاسبه بهتر است داده‌ها را از یک عدد ثابت کم کنیم:

$$\sigma^2\left(x_i\cancel{-a}^{\,0}\right)=\sigma_x^2$$

زیرا واریانس به‌دست‌آمده برابر با واریانس اعداد اولیه خواهد بود. در میان این داده‌ها بهترین عدد برای کم کردن 150 است.

$(X_i-150):-15, -9, 5, 19$

$$\mu_{x_i-150}=\frac{\sum(x_i-150)}{N}=\frac{-15-9+5+19}{4}=0$$

$$\sigma^2_{(x_i-150)}=\sigma_x^2=\frac{\sum(x_i-\mu)^2}{N}\overset{\mu=0}{=\!=}\frac{\sum x_i^2}{N}=\frac{(-15)^2+(-9)^2+5^2+(19)^2}{4}=\frac{692}{4}=173$$

کاربرد واریانس

۵۱. گزینه ۳ درست است.

انحراف متوسط از میانگین در نمایش پراکندگی داده‌ها با ثبات‌تر از انحراف چارکی و نیم‌دامنه است، اما با توجه به مشکلات زیر نمی‌تواند انحرافات بزرگ را به‌خوبی نمایش دهد:

۱ـ علامت جبری داده‌ها در نظر گرفته نمی‌شود.

۲ـ تأثیر انحرافات بزرگ در مقابل تعداد زیادی انحراف کوچک قابل نمایش نیست.
در این وضعیت واریانس با توجه به آنکه مجذور انحرافات را در نظر می‌گیرد، تأثیر انحرافات بزرگ را به خوبی نمایش می‌دهد.

قضیه چی‌بی‌شف

۵۲.گزینه ۲ درست است.

با توجه به قضیه چی‌بی‌شف:

$$\begin{cases} P\left(\underbrace{\mu-k\sigma}_{a}\le x\le\underbrace{\mu+k\sigma}_{b}\right)\to \text{حداقل}\left(1-\frac{1}{k^2}\right)\text{درصد} \\ \frac{b-a}{2}=k\sigma \end{cases}$$

بنابراین:

$$\text{اولاً}: 1-\frac{1}{k^2}=\frac{84}{100}\to\frac{1}{k^2}=\frac{16}{100}\to k=\frac{10}{4}=\frac{5}{2}$$

$$\text{ثانیاً}: \frac{b-a}{2}=k\sigma\to\frac{88-72}{2}=\frac{5}{2}\sigma\to 8=\frac{5}{2}\sigma\to\sigma=\frac{16}{5}=3.2$$

۵۳. گزینه ۳ درست است.

با توجه به قضیه چی‌بی‌شف:

$$\begin{cases} P\left(\underbrace{\mu-k\sigma}_{a}< x<\underbrace{\mu+k\sigma}_{b}\right)\text{درصد}\to\left(1-\frac{1}{k^2}\right)\text{حداقل} \\ \frac{b-a}{2}=k\sigma \end{cases}$$

$$\text{اولاً}: \frac{b-a}{2}=k\sigma\xrightarrow[\sigma^2=64]{a=63\ ,\ b=87}\frac{87-63}{2}=\sqrt{64}k\to 12=8k\to k=\frac{12}{8}=\frac{3}{2}$$

$$\text{ثانیاً}: 1-\frac{1}{k^2}=1-\frac{1}{\left(\frac{3}{2}\right)^2}=1-\frac{4}{9}=\frac{5}{9}=0.55=\%55$$

۵۴.گزینه ۳ درست است.

با توجه به قضیه چی‌بی‌شف:

$$\begin{cases} P\left(\underbrace{\mu-k\sigma}_{a}\le x\le\underbrace{\mu+k\sigma}_{b}\right)\to \text{حداقل}\left(1-\frac{1}{k^2}\right)\text{درصد} \\ \frac{b-a}{2}=k\sigma \end{cases}$$

بنابراین:

$$\text{اولاً}: \frac{b-a}{2}=k\sigma\to\frac{32-16}{2}=4k\to k=2$$

$$\text{ثانیاً}: 1-\frac{1}{k^2}=1-\frac{1}{2^2}=\frac{3}{4}=\%75\ \text{حداقل}$$

۵۵.گزینه ۴ درست است.

با توجه به نامعلوم بودن توزیع جامعه و خواستن «حداقل احتمال»، از قضیه چی‌بی‌شف استفاده می‌کنیم:

$$\begin{cases} P(\underbrace{\mu-k\sigma}_{a}\le x\le\underbrace{\mu+k\sigma}_{b})\ge 1-\dfrac{1}{k^2} & \to \quad P(15<x<45)\ge 0.89 \\ \dfrac{b-a}{2}=k\sigma \end{cases}$$

$$\text{اولاً}: \frac{b-a}{2}=k\sigma \to \frac{45-15}{2}=k\sigma \xrightarrow{\sigma=5} 15=5k \to k=3$$

$$\text{ثانیاً}: 1-\frac{1}{k^2}=1-\frac{1}{3^2}=1-\frac{1}{9}=\frac{8}{9}\simeq 0.89$$

۵۶.گزینه ۱ درست است.

با توجه به نامعلوم بودن توزیع جامعه و همچنین مشاهده گزینه‌ها که احتمال حداقل یا حداکثر را بیان می‌کنند، از قضیه چی‌بی‌شف استفاده می‌کنیم:

$$\begin{cases} P(x>\underbrace{\mu+k\sigma}_{b} \quad \text{یا} \quad x<\underbrace{\mu-k\sigma}_{a}) \le \left(\dfrac{1}{k^2}\right) \text{حداکثر} \\ \dfrac{b-a}{2}=k\sigma \end{cases}$$

در این سؤال نیز:

$$P(X>12 \quad \text{یا} \quad X<8)\le\frac{1}{k^2} \quad \text{حداکثر}$$

$$\text{اولاً}: k\sigma=\frac{b-a}{2} \xrightarrow[\sigma^2=1]{b=12,a=8} k\times 1=\frac{12-8}{2} \to k=2$$

$$\text{ثانیاً}: \frac{1}{k^2}=\frac{1}{2^2}=\frac{1}{4} \quad \to \quad P(x>12 \quad \text{یا} \quad x<8)\le\frac{1}{4}=\%25 \quad \text{حداکثر}$$

۵۷.گزینه ۳ درست است.

با توجه به قضیه چی‌بی‌شف:

$$\begin{cases} P\left(\underbrace{\mu-k\sigma}_{a}<x<\underbrace{\mu+k\sigma}_{b}\right)\ge\left(1-\dfrac{1}{k^2}\right) \text{حداقل} \\ \dfrac{b-a}{2}=k\sigma \end{cases}$$

بنابراین:

$$\text{اولاً}: 1-\frac{1}{k^2}=\frac{36}{100} \quad \to \quad \frac{1}{k^2}=\frac{64}{100} \quad \to \quad k=\frac{10}{8}=\frac{5}{4}$$

$$\text{ثانیاً}: (\mu-k\sigma\ ,\ \mu+k\sigma) \xrightarrow[k=\frac{5}{4}]{\mu=9.8\ ,\ \sigma=1.2} \left(9.8-\frac{5}{4}\times1.2\ ,\ 9.8+\frac{5}{4}\times1.2\right)=(8.3\ ,\ 11.3)$$

تصحیح شپارد

۵۸. گزینه ۴ درست است.

با توجه به اینکه شرایط تصحیح شپارد برقرار است، یعنی:

۱ـ داده‌ها طبقه‌بندی شده و پیوسته هستند،

۲ـ توزیع فراوانی اندکی متقارن است،

۳ـ $N > 1000$،

واریانس تصحیح‌شده شپارد به صورت زیر محاسبه می‌شود:

$$\begin{cases} \sigma^2_{\text{شپارد}} = \sigma^2 - \frac{I^2}{12} = 12 - \frac{6^2}{12} = 12 - 3 = 9 \\ \text{فاصله طبقات} : I = 6 \ , \ \sigma^2 = 12 \end{cases}$$

۵۹. گزینه ۱ درست است.

یادآوری: هرگاه:

۱ـ داده‌های آماری دسته‌بندی شده با متغیر پیوسته باشند،

۲ـ تابع توزیع فراوانی متقارن یا اندکی متقارن باشد،

۳ـ $N \geq 1000$،

واریانس محاسبه‌شده به روش معمول از واریانس واقعی جامعه بیشتر می‌شود و در این شرایط از واریانس تصحیح‌شده شپارد استفاده می‌کنیم.

$$\sigma^2_{\text{شپارد}} = \sigma^2 - \frac{I^2}{12}$$

۶۰. گزینه ۳ درست است.

$$\begin{cases} \sigma^2_{\text{شپارد}} = \sigma^2 - \frac{I^2}{12} = 7 - \frac{3^2}{12} = 7 - 0.75 = 6.25 \rightarrow \sigma_{\text{شپارد}} = \sqrt{6.25} = 2.5 \\ \text{فاصله طبقات } I = 3 \end{cases}$$

میانگین و واریانس کل

۶۱. گزینه ۱ درست است.

$$\text{میانگین کل} : \mu = \frac{\sum N_i \mu_i}{\sum N_i} = \frac{100 \times 14 + 200 \times 18 + 700 \times 20}{100 + 200 + 700} = \frac{19000}{1000} = 19$$

$$\text{واریانس کل} : \sigma^2 = \frac{\sum N_i \sigma_i^2 + \sum N_i (\mu_i - \mu)^2}{\sum N_i}$$

$$= \frac{100 \times 50 + 200 \times 60 + 700 \times 40 + \overset{2500}{\cancel{100(14-19)^2}} + \overset{200}{\cancel{200(18-19)^2}} + \overset{700}{\cancel{700(20-19)^2}}}{1000} = \frac{48400}{1000} = 48.4$$

۶۲.گزینه ۱ درست است.

$$\begin{cases} \sigma^2_{\text{کل}} = \dfrac{\sum N_i\sigma_i^2}{\sum N_i} + \dfrac{\sum N_i(\mu_i-\mu)^2}{\sum N_i} = \dfrac{20\times 2^2+10\times 3^2}{20+10} + \dfrac{20(12-11)^2+10(9-11)^2}{20+10} = 7.6 \\ \mu_{\text{کل}} = \dfrac{\sum N_i\mu_i}{\sum N_i} = \dfrac{20\times 12+10\times 9}{20+10} = \dfrac{330}{30} = 11 \end{cases}$$

	N_i	μ_i	σ_i^2
جامعه اول	20	12	4
جامعه دوم	10	9	9

۶۳.گزینه ۴ درست است.

$$\sigma^2_{\text{کل}} = \frac{\sum N_i\sigma_i^2}{\sum N_i} + \underbrace{\frac{\sum N_i(\mu_i-\mu\quad)^2}{\sum N_i}}_{\text{مقداری بزرگ‌تر از صفر}} = \frac{100\times 16+200\times 25+450\times 20}{100+200+450} + k = \frac{15600}{750} + k = 20.8 + k$$

دقت کنید که وقتی میانگین‌های جوامع متفاوت هستند، میانگین کل نیز با آن‌ها متفاوت است و کسر دوم در محاسبه واریانس کل صفر نخواهد شد و حتماً مقداری بزرگ‌تر از صفر است؛ بنابراین در بین گزینه‌های این سؤال گزینه‌ای می‌تواند واریانس کل باشد که از 20.8 بزرگ‌تر باشد و تنها گزینه ۴ بزرگ‌تر $(22.1 > 20.8)$ و پاسخ سؤال است.

۶۴.گزینه ۴ درست است.

$$\text{واریانس کل}: \sigma^2 = \frac{\sum N_i\sigma_i^2}{\sum N_i} + \underbrace{\frac{\sum N_i(\mu_i-\mu\quad)^2}{\sum N_i}}_{>0} \rightarrow \sigma^2 > \frac{\sum N_i\sigma_i^2}{\sum N_i}$$

$$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_{\text{کل}} \rightarrow \sum(\mu_i - \mu_{\text{کل}})^2 > 0$$

$$\sigma^2_{\text{کل}} > \frac{\sum N_i\sigma_i^2}{\sum N_i} = \frac{100\times 12+150\times 14+50\times 9}{100+150+50} = \frac{3750}{300} = 12.5 \rightarrow \sigma^2_{\text{کل}} > 12.5$$

دقت کنید که وقتی میانگین‌های جوامع متفاوت هستند، میانگین کل نیز با آن‌ها متفاوت است و کسر دوم در محاسبه واریانس کل صفر نخواهد شد و حتماً مقداری بزرگ‌تر از صفر است؛ بنابراین در بین گزینه‌های این سؤال گزینه‌ای می‌تواند واریانس کل باشد که از 12.5 بزرگ‌تر باشد و تنها گزینه ۴ بزرگ‌تر $(12.6 > 12.5)$ و پاسخ سؤال است.

۶۵.گزینه ۴ درست است.

$$\sigma^2_{\text{کل}} = \frac{\sum N_i\sigma_i^2}{\sum N_i} + \frac{\sum N_i(\mu_i-\mu_{\text{کل}})^2}{\sum N_i} = \frac{50\times 25+100\times 16}{50+100} + \frac{50(40-50)^2+100(55-50)^2}{50+100} = 69$$

$$\mu_{\text{کل}} = \frac{\sum N_i\mu_i}{\sum N_i} = \frac{50\times 40+100\times 55}{50+100} = \frac{7500}{150} = 50$$

۶۶.گزینه ۳ درست است.

$$\mu = \frac{\sum N_i\mu_i}{\sum N_i} = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2} = \frac{4\times 12 + 5\times 18}{4+5} = \frac{48+90}{9} = \frac{138}{9}$$

$N_1 = 4 \ , \ \mu_1 = 12$

$N_2 = 5 \ , \ \mu_2 = 18$

۶۷.گزینه ۴ درست است.

$$\mu = \frac{\sum N_i\mu_i}{\sum N_i} = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2} = \frac{20\times 10 + 15\times 12}{20+15} = \frac{200+180}{35} = \frac{380}{35}$$

$N_1 = 20 \ , \ \mu_1 = 10$

$N_2 = 15 \ , \ \mu_2 = 12$

۶۸.گزینه ۲ درست است.

$$\mu = \frac{\sum N_i\mu_i}{\sum N_i} = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2} = \frac{30\times 55 + 20\times 65}{30+20} = \frac{1650+1300}{50} = \frac{2950}{50} = 59$$

$N_1 = 30 \ , \ \mu_1 = 55$

$N_2 = 20 \ , \ \mu_2 = 65$

معیارهای پراکندگی نسبی

ضریب تغییرات

۶۹.گزینه ۴ درست است.

یادآوری: ضریب تغییرات برای مقایسه دو جامعه وقتی استفاده می‌شود که یکی از شرایط زیر برقرار باشد:

۱- دو جامعه دارای واحد اندازه‌گیری متفاوت باشند.

۲- دو جامعه میانگین متفاوتی داشته باشند.

در این سؤال نیز واحد اندازه‌گیری جامعه اول «یورو» و واحد اندازه‌گیری جامعه دوم «ین» است، بنابراین طبق شرط اول ضریب پراکندگی مناسب‌ترین شاخص برای مقایسه است.

۷۰. گزینه ۱ درست است.

$$CV_{\text{جدید}} = \frac{\sigma}{\mu} = \frac{4}{16} = \frac{1}{4} = 0.25$$

$$\mu_{\text{جدید}} = \frac{\sum_{i=1}^{13} x_i}{13} = \frac{\sum_{i=1}^{10} x_i + 13 + 21 + 14}{10+3} = \frac{10\times 16 + 48}{13} = \frac{208}{13} = 16$$

$$\sigma^2_{\text{جدید}} = \frac{\sum_{i=1}^{10}(x_i - 16)^2 + (13-16)^2 + (21-16)^2 + (14-16)^2}{13} = \frac{10\times 17 + 9 + 25 + 4}{13} = 16 \rightarrow \sigma_{\text{جدید}} = 4$$

دقت کنید که چون میانگین جدید با میانگین قدیم برابر شد، در محاسبه واریانس توانستیم $\sum_{i=1}^{10} (x_i - \mu_{\text{جدید}})^2$ را همان $\sum_{i=1}^{10}(x_i - 16)^2 = N\sigma^2_{\text{قدیم}} = 10\times 17$ در نظر بگیریم؛ در غیر این صورت باید می‌دانستیم که داده‌های قبلی هر کدام چه عددی بوده است.

چولگی

۷۱. گزینه ۱ درست است.

نکته: تغییر در مبدأ و مقیاس داده‌ها تأثیری در ضریب چولگی ندارد؛ تنها درصورتی‌که داده‌ها در ضریبی منفی ضرب شوند، علامت منفی در ضریب چولگی ضرب می‌شود.

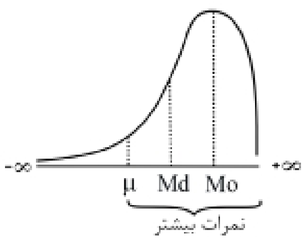
$$Sk(ax+b) \begin{cases} \xrightarrow{a>0} Sk \\ \xrightarrow{a<0} -Sk \end{cases} \quad \rightarrow \quad Sk\left(x-\frac{1}{2}\right)=Sk \quad \rightarrow$$ Sk بدون تغییر

۷۲. گزینه ۳ درست است.

یادآوری: ملاک مقایسه دو جامعه وقتی میانگین و واریانس آنها با هم برابر باشد، ضریب چولگی و یا ضریب کشیدگی است.

در این سؤال نیز میانگین و واریانس دو کلاس A و B با هم برابر است و با توجه به گزینه‌ها ضریب چولگی ملاک مقایسه قرار گرفته است.

دقت کنید می‌خواهیم نمرات دو کلاس را مقایسه کنیم، بنابراین نمرات کلاس A در صورتی بهتر است که نمرات بالاتری داشته باشد یعنی تمایل نمرات به راست باشد که در این صورت توزیع چوله به چپ (ضریب چولگی منفی) خواهد بود.



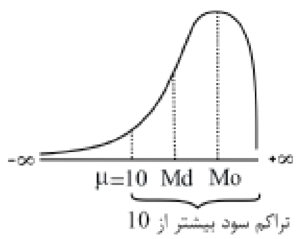
می‌دانیم که وقتی توزیع جامعه چوله به چپ $(Sk<0)$ است (تمایل به راست) رابطه زیر بین معیارهای مرکزی برقرار است:

مد > میانه > میانگین

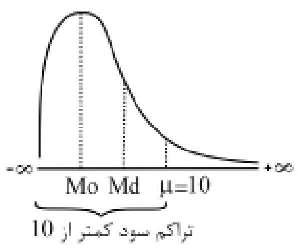
بنابراین باید میانگین نمرات کمتر از میانه باشد.

۷۳. گزینه ۳ درست است.

با توجه به اینکه میانگین و انحراف معیار سود هر دو شرکت یکسان است ($\mu_A=\mu_B=10$ و $\sigma_A=\sigma_B=2$)، ضریب پراکندگی آن‌ها نیز یکسان خواهد شد $\left(CV_A=CV_B=\frac{2}{10}\right)$ و در این وضعیت معیار مقایسه دو شرکت ضریب چولگی آن‌هاست.



توزیع سود شرکت A با ضریب چولگی $(\alpha_3=-1)$ منفی، چوله به چپ است، درنتیجه تراکم سود به راست و احتمال سود بیشتر از 10 (میانگین) در آن بیشتر است.



اما در شرکت B که ضریب چولگی $(\alpha_3=+1)$ مثبت است، توزیع سود شرکت چوله به راست و تراکم سود چوله به چپ است، درنتیجه احتمال سود کمتر از 10 (میانگین) در آن بیشتر است.

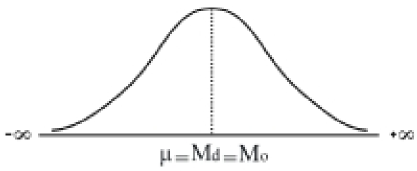
۷۴.گزینه ۳ درست است.

چوله به راست: $Mo < Md < \mu$

چوله به چپ: $\mu < Md < Mo$

۷۵.گزینه ۴ درست است.

توزیع نرمال متقارن است و هر سه معیار مرکزی آن بر هم منطبق هستند و محور تقارن را تشکیل می‌دهند (مد = میانه = میانگین).



۷۶.گزینه ۳ درست است.

برای تشخیص چولگی (انحراف از قرینگی) جامعه کافی است دو معیار از سه معیار تمرکز (میانگین، میانه، مد) را به دست آوریم، سپس با توجه به رابطه آن‌ها نوع چولگی را تشخیص دهیم.

نمره	$0-5$	$5-10$	$10-15$ ↓	$15-20$
F_i (فراوانی)	1	3	26	10
F_{C_i} (فراوانی تجمعی)	1	4	30	$40 = N$

دسته‌ای که دارای بیشترین فراوانی است، دسته مددار است $(10-15)$:

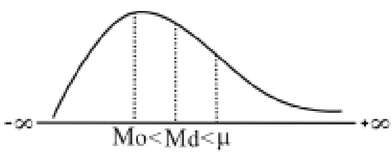
$$\text{مد}: Mo = \text{حد پایین دسته} + \frac{d_1}{d_1+d_2} \times \text{طول دسته} = 10 + \frac{(26-3)}{(26-3)+(26-10)} \times 5 = 12.95$$

اولین دسته‌ای که فراوانی تجمعی‌اش بیشتر و یا مساوی $\frac{N}{2}$ باشد، دسته میانه است.

$$F_{C_i} \geq \frac{N}{2} = \frac{40}{2} = 20 \rightarrow \text{دسته سوم} : (10-15)$$

$$\text{میانه}: Md = \text{حد پایین دسته} + \frac{\frac{N}{2} - F_{c_{i-1}}}{F_i} \times \text{طول دسته} = 10 + \frac{20-4}{26} \times 5 = 13.07$$

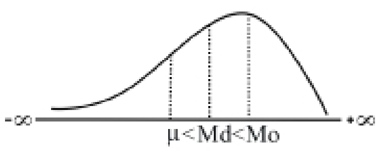
با توجه به اینکه مقدار مد از میانه کمتر است و میانه همواره بین مد و میانگین قرار دارد، رابطه زیر بین سه معیار مرکزی برقرار است و چولگی جامعه مثبت (به راست) است $(Mo < Md < \mu)$.



$$Mo = 12.95 < Md = 13.07 < \mu$$

۷۷. گزینه ۳ درست است.

با توجه به شکل، توزیع جامعه چوله به چپ بوده و رابطه زیر بین سه معیار تمرکز آن برقرار است.



مد > میانه > میانگین

ضریب چولگی گشتاوری و تحلیل چولگی

۷۸. گزینه ۲ درست است.

$$\begin{cases} Sk=\dfrac{\dfrac{\sum(x_i-\mu)^3}{N}}{\sigma^3}=\dfrac{\dfrac{\sum(x_i-15)^3}{50}}{2^3}=\dfrac{\dfrac{24}{50}}{8}=0.06 \\ \mu=15\ ,\ \sigma^2=4\to\sigma=2\ ,\ N=50 \end{cases}$$

با توجه به اینکه $|Sk|\le 0.1$ است، چولگی توزیع تقریباً نرمال است.

۷۹. گزینه ۲ درست است.

$$\begin{cases} Sk=\dfrac{\dfrac{\sum(x_i-\mu)^3}{N}}{\sigma^3}=\dfrac{\dfrac{-180}{20}}{3^3}=\dfrac{-9}{27}=\dfrac{-1}{3}=-0.33 \\ \sigma^2=\dfrac{\sum(x_i-\mu)^2}{N}=\dfrac{180}{20}=9\to\sigma=3 \end{cases}$$

۸۰. گزینه ۲ درست است.

$$\begin{cases} Sk=\dfrac{\dfrac{\sum(x_i-\mu)^3}{N}}{\sigma^3}=\dfrac{\dfrac{\sum(x_i-7)^3}{N}}{\sigma^3}=\dfrac{\dfrac{192}{100}}{4^3}=0.03=\%3 \\ \sigma^2=\dfrac{\sum x^2}{N}-\left(\dfrac{\sum x}{N}\right)^2=\dfrac{6500}{100}-(7)^2=16\to\sigma=4 \\ N=100\ ,\ \mu=7\ ,\ \sum x^2=6500 \end{cases}$$

۸۱. گزینه ۱ درست است.

$x-\bar{x}$	-3	-1	0	1	3	
$(x-\bar{x})^2$	9	1	0	1	9	
$(x-\bar{x})^3$	-27	-1	0	1	27	
F	5	10	20	13	4	$\sum F=N=52$

$$\text{ضریب چولگی}: Sk=\dfrac{\dfrac{\sum F_i(x_i-\bar{x})^3}{N}}{\sigma^3}=\dfrac{\dfrac{5(-27)+10(-1)+20\times0+13\times1+4\times27}{52}}{(\sqrt{2})^3}=\dfrac{\dfrac{-24}{52}}{2\sqrt{2}}=\dfrac{-3}{13\sqrt{2}}$$

$$\sigma^2=\dfrac{\sum F_i(x_i-\bar{x})^2}{N}=\dfrac{5\times9+10\times1+20\times0+13\times1+4\times9}{52}=\dfrac{104}{52}=2\ \to\ \sigma=\sqrt{2}$$

۸۲.گزینه ۱ درست است.

$$\begin{cases} Sk = \dfrac{\dfrac{\sum(x_i-\mu)^3}{N}}{\sigma^3} = \dfrac{\dfrac{75}{40}}{5^3} = 0.015 = \%1.5 \\ \sigma^2 = \dfrac{\sum x_i^2}{N} - \left(\dfrac{\sum x_i}{N}\right)^2 = \dfrac{2440}{40} - \left(\dfrac{240}{40}\right)^2 = 61-36=25 \quad \to \quad \sigma = 5 \end{cases}$$

۸۳. گزینه ۲ درست است.

$$Sk = \frac{\dfrac{\sum F_i(x_i-\mu)^3}{N}}{\sigma^3} = \frac{\dfrac{-24}{50}}{2^3} = -0.06$$

با توجه به اینکه علامت ضریب چولگی منفی است، توزیع جامعه چوله به چپ است. همچنین چون $|Sk| \le 0.1$ است، چولگی توزیع تقریباً نرمال است.

توجه: منظور از f_i در صورت سؤال همان F_i (فراوانی مطلق) است؛ درصورتی‌که آن را به اشتباه فراوانی نسبی فرض کنیم، $Sk = \dfrac{\sum f_i(x_i-\mu)^3}{\sigma^3} = \dfrac{-24}{8} = -3$ می‌شود که در این صورت چولگی جامعه با توزیع نرمال تفاوت فاحش دارد و در گزینه‌ها موجود نیست.

۸۴.گزینه ۴ درست است.

$X-\bar{X}$	-5	-3	-1	1	3	5	
$(X-\bar{X})^2$	25	9	1	1	9	25	
$(X-\bar{X})^3$	-125	-27	-1	1	27	125	
F	2	5	4	8	2	3	$\sum F = N = 24$

$$Sk = \frac{\dfrac{\sum F(X-\bar{X})^3}{N}}{\sigma^3} = \frac{\dfrac{2(-125)+5(-27)+4(-1)+8(1)+2(27)+3(125)}{24}}{\left(\dfrac{10}{2\sqrt{3}}\right)^3} = \frac{\dfrac{48}{24}}{\dfrac{1000}{8\times3\sqrt{3}}} = \frac{6\sqrt{3}}{125}$$

$$\sigma^2 = \frac{\sum F(X-\bar{X})^2}{N} = \frac{2(25)+5(9)+4(1)+8(1)+2(9)+3(25)}{24} = \frac{200}{24} = \frac{100}{12} \quad \to \quad \sigma = \sqrt{\frac{100}{12}} = \frac{10}{2\sqrt{3}}$$

چولگی پیرسون، چندکی و تحلیل چولگی

۸۵. گزینه ۴ درست است.

$Q_1 = 36 \quad , \quad Q_2 = 61 \quad , \quad Q_3 = 76$

تنها راه به دست آوردن چولگی از روی چارک‌ها، فرمول سوم چولگی پیرسون است:

چولگی پیرسون : $Sk_Q = \dfrac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \dfrac{76 - 2\times 61 + 36}{76-36} = \dfrac{-10}{40} = -0.25$

با توجه به منفی بودن علامت ضریب چولگی، توزیع داده‌ها چوله به چپ است و چون $0.1 < |Sk| \le 0.5$، از نظر تقارن تفاوت اندکی با توزیع نرمال دارد.

۸۶.گزینه ۳ درست است.

با توجه به اینکه مقادیر چارک‌ها داده شده است، از فرمول چولگی چندکی استفاده می‌کنیم:

$$Sk = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{17 - 2\times 15 + 12}{17 - 12} = \frac{-1}{5} = -0.2$$

چون ضریب چولگی منفی است، توزیع جامعه چوله به چپ بوده و چون $0.1 \le |Sk| = |-0.2| < 0.5$ است، توزیع جامعه تفاوت اندکی با توزیع نرمال دارد.

۸۷. گزینه ۱ درست است.

با توجه به داده‌های مسئله، از ضریب چولگی چندکی استفاده می‌کنیم:

$$Sk = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{42 - 2\times 35 + 26}{42 - 26} = \frac{-2}{16} = -0.125$$

چون علامت ضریب چولگی منفی است، توزیع جامعه چوله به چپ است.

رابطه سه معیار تمرکز (در چولگی ضعیف)

۸۸. گزینه ۱ درست است.

تنها رابطه میان سه معیار تمرکز وقتی که چولگی جامعه ضعیف (معقول) باشد، به صورت زیر است:

$$\begin{cases} \mu - Mo = 3(\mu - Md) & \rightarrow \quad \mu - 72 = 3(\mu - 54) \quad \rightarrow \quad 2\mu = 90 \quad \rightarrow \quad \mu = 45 \\ Md = 54\ ,\ Mo = 72 \end{cases}$$

۸۹.گزینه ۴ درست است.

تنها رابطه بین سه معیار تمرکز (میانگین، میانه، نما) در حالتی است که توزیع جامعه از چولگی خفیفی برخوردار باشد که در این صورت دو فرمول چولگی پیرسون با هم برابر می‌شود و رابطه زیر به دست می‌آید:

$$\mu - Mo = 3(\mu - Md)$$

کشیدگی

۹۰. گزینه ۳ درست است.

ضریب کشیدگی

۹۱.گزینه ۱ درست است.

$$\text{ضریب کشیدگی} : E = \frac{\dfrac{\sum (x_i - \mu)^4}{N}}{\sigma^4} - 3 = \frac{\dfrac{8640}{20}}{(12)^2} - 3 = \frac{432}{144} - 3 = 3 - 3 = 0$$

نمودارها

۹۲.گزینه ۴ درست است.

شاخه	برگ					
0	1	1	1	2	2	4
1	0	0	3	4	6	9
2	0	1	2	3		

توجه: در نمودار شاخه و برگ، شاخه رقم دهگان عدد و برگ رقم یکان آن است که به صورت مرتب در نمودار ظاهر می‌شوند؛ پس تعداد برگ‌ها، تعداد داده‌ها را مشخص می‌کند. برای مثال، در نمودار بالا داده‌ها عبارتند از:

$x_i : 1 , 1 , 1 , 2 , 2 , 4 , 10 , 10 , 13 , 14 , 16 , 19 , 20 , 21 , 22 , 23$

حال با توجه به اینکه تعداد داده‌ها $N = 16$ است، داریم:

محل چارک اول: $\frac{N}{4}+\frac{1}{2}=\frac{16}{4}+\frac{1}{2}=4+\frac{1}{2} \to Q_{(1)} = x_{(4)}+\frac{1}{2}\left(x_{(5)}-x_{(4)}\right)=2+\frac{1}{2}(2-2)=2$

محل چارک سوم: $\frac{3N}{4}+\frac{1}{2}=\frac{3\times16}{4}+\frac{1}{2}=12+\frac{1}{2}\to Q_3 = x_{(12)}+\frac{1}{2}\left(x_{(13)}-x_{(12)}\right)=19+\frac{1}{2}(20-19)=19.5$

حال باید واریانس داده‌های بین $x_{(4)} = 2$ و 19.5 را محاسبه کنیم، یعنی:

$x_i : 2,4,10,10,13,14,16,19$

$$\sigma^2=\frac{\sum(x_i-\mu)^2}{N}=\frac{(2-11)^2+(4-11)^2+(10-11)^2+(10-11)^2+(13-11)^2+(14-11)^2+(16-11)^2+(19-11)^2}{8}=29.25$$

$$\mu=\frac{\sum x_i}{N}=\frac{2+4+10+10+13+14+16+19}{8}=\frac{88}{8}=11$$

۹۳.گزینه ۲ درست است.

ابتدا نمودار هیستوگرام (با فراوانی نسبی) را به صورت جدول فراوانی زیر درمی‌آوریم:

طول عمر	$0-100$	$100-200$	$200-300$	$300-400$	$400-500$	$500-600$	$600-700$	
(فراوانی نسبی) f_i	0.45	0.25	0.2	0.04	0.03	0.02	0.01	$\Sigma f_i = 1$

صدک 20ام طول عمر برابر با عددی است که 20% قطعات کمتر و یا مساوی آن عمر می‌کنند و ما به دنبال آن عدد هستیم.

محل صدک 20ام: دسته‌ای که فراوانی تجمعی نسبی آن بزرگ‌تر یا مساوی 0.20 باشد (دسته اول):

$$P_{20} = \text{حد پایین دسته} + \frac{0.2-f_{c_{i-1}}}{f_i}\times \text{طول دسته} = 0+\frac{0.2-0}{0.45}\times100=\frac{20}{0.45}=\frac{2000}{45}=44.4$$

خودآزمایی

تعاریف

۱. **کدام مقیاس‌ها دارای صفر قراردادی است؟**

۱) نسبی ۲) فاصله‌ای ۳) اسمی ۴) رتبه‌ای

۲. **یکی از موارد بررسی در استنباط آماری، کدام است؟**

۱) برآورد پارامترها ۲) محاسبه معیارهای عددی

۳) جمع‌آوری اطلاعات ۴) ارائه جداول و نمودارها

۳. **«هر خصوصیت عددی از توزیع جامعه» چه نامیده می‌شود؟**

۱) آماره ۲) پارامتر ۳) آزمایش ۴) فرض آماری

۴. **در چه مرحله‌ای از یک تحقیق علمی معلوم می‌شود که حدس یا نظریه موجود، با داده‌ها در تناقض است یا نه؟**

۱) فرضیه تحقیق ۲) بیان یافته‌ها

۳) تعیین زمینه و موضوع تحقیق ۴) تجزیه و تحلیل داده‌ها

فراوانی

۵. **اگر $80-89$ و $90-99$ دو طبقه از یک جدول طبقه‌بندی باشند، اندازه طول طبقه کدام است؟**

۱) مساوی با عرض طبقه ۲) یک واحد کمتر از عرض طبقه

۳) یک واحد بیشتر از عرض طبقه ۴) مساوی با فاصله طبقات

۶. **اگر در جدول توزیع فراوانی‌ها، فراوانی مطلق طبقه سوم 12 و فراوانی نسبی همان طبقه 0.48 باشد، فراوانی تجمعی طبقه آخر کدام است؟**

۱) 25 ۲) 50 ۳) 96 ۴) 100

معیارهای تمرکز

مد (نما)

۷. مد این جدول کدام است؟

x_i	-2	-1	0	1	2	3
F_i	10	20	10	15	15	30

۱) 0
۲) 3
۳) 2.5
۴) 2

۸. مد جدول زیر کدام است؟

C – L	10 – 20	20 – 30	30 – 40	40 – 50
f_i	0.10	0.20	0.30	0.40

۱) 40
۲) 39
۳) 42
۴) 45

میانه

۹. میانه این جدول کدام است؟

C – L	2 – 4	4 – 6	6 – 8	8 – 10
F_i	5	10	5	10

۱) 4
۲) 5
۳) 6
۴) 7

۱۰. میانه جدول زیر کدام است؟

x_i	-2	-1	0	1	2	3
F_i	10	20	10	15	15	30

۱) $\frac{1}{2}$
۲) 0
۳) 1
۴) 1.5

۱۱. میانه توزیع آماری 40 مشاهده 32.5 است. اگر $I = 5$ و فراوانی طبقه میانه‌دار 10 و مجموع فراوانی‌های ماقبل طبقه میانه‌دار 14 باشد، حدود کرانه طبقه میانه‌دار کدام است؟

۱) $34.5 - 29.5$ ۲) $39 - 29$ ۳) $40 - 30$ ۴) $39 - 35$

۱۲. در مجموعه اعداد 40 ,80 ,60 ,50 ,60 عدد 60 کدام پارامتر است؟

۱) میانه ۲) میانه و نما ۳) میانگین ۴) نما و میانگین

چندک

۱۳. چارک اول این داده‌ها کدام است؟

C – L	110 – 119	120 – 129	130 – 139
F_i	10	20	70

۱) 129
۲) 130
۳) 127
۴) 132.70

۱۴. در داده‌های زیر چارک اول کدام است؟

X_i:	10	5	7	11	14	15
F_i:	3	2	1	7	4	2

۱) 5 ۲) 7 ۳) 9 ۴) 10

میانگین هارمونیک

۱۵. با بررسی پرسشنامه‌های طرح هزینه و درآمد خانوار شهری در چهار سال متوالی، معلوم شد قیمت نفت سفید یک خانوار به ترتیب 1.6, 1.8, 2.1 و 2.5 ریال در لیتر است. اگر خانواری برای هر سال 20 هزار ریال هزینه در نظر بگیرد، متوسط مصرف سوخت سالانه این خانوار برحسب ریال در لیتر چقدر است؟

۱) 1.75 ۲) 2.5 ۳) 2.25 ۴) 1.94

۱۶. میانگین هارمونیک داده‌های آماری جدول زیر کدام است؟

x_i	3	4	6	9
F_i	2	4	3	1

۱) 4.39
۲) 4.47
۳) 4.51
۴) 4.90

میانگین هندسی

۱۷. قیمت سهام یک کارخانه از 100 ریال در سال ۱۳۸۰ به 3200 ریال در سال ۱۳۸۵ افزایش یافته است. متوسط نرخ افزایش قیمت سهام در این دوره چقدر بوده است؟

۱) %80 ۲) %120 ۳) %125 ۴) %100

۱۸. میانگین هندسی اعداد 45, 24, 30, 25 کدام است؟

۱) 30 ۲) 26 ۳) 32 ۴) 28

میانگین حسابی

۱۹. اگر میانگین داده‌های $x_1, x_2, ..., x_{10}$ برابر 5 باشد، میانگین $x_1, x_2, ..., x_{10}$ و 16 کدام است؟

۱) 5 ۲) 6 ۳) 6.6 ۴) 10.5

۲۰. در جدول داده‌های زیر مقدار میانگین کدام است؟

x_i	1	3	5	7	9
F_i	2	4	8	5	1

۱) 4.7
۲) 4.9
۳) 5.1
۴) 5.2

۲۱. میانگین داده‌ها با توزیع فراوانی تجمعی زیر به صورت $90 + 6\bar{u}$ محاسبه شده است. $\bar{u}$ کدام است؟

نشان دسته	72	78	84	90	96	102
فراوانی تجمعی	3	9	14	22	29	35

۱) -0.1 ۲) -0.2 ۳) 0.1 ۴) 0.3

خواص میانگین حسابی

۲۲. اگر میانگین $x_1, x_2, ..., x_N$ برابر μ_x و میانگین $y_1, y_2, ..., y_k$ مساوی μ_y باشد و داشته باشیم $\mu_y = a\mu_x$، در آن صورت مقدار $\frac{\sum x_i}{\sum y_i}$ کدام است؟

۱) $\frac{N}{Ka}$ ۲) $N.a$ ۳) $N.\mu_x$ ۴) $N.\mu_y$

۲۳. اگر میانگین $x_1, x_2, ..., x_N$ مساوی μ_x باشد، مقدار $\sum(x_i - \mu_x)$ کدام است؟

۱) $N.\mu_x$ ۲) صفر ۳) N ۴) یک

۲۴. اگر μ_x میانگین $x_1, ..., x_N$ باشد، میانگین $\left(-\frac{x_1}{2}+3\right), \left(-\frac{x_2}{2}+3\right), ..., \left(-\frac{x_N}{2}+3\right)$ کدام است؟

۱) $-\mu_x + 3$ ۲) $\mu_x + 3$ ۳) $-\frac{1}{2}\mu_x + 3$ ۴) $\frac{1}{2}\mu_x$

۲۵. اگر $\mu_x = 10$ و $\mu_y = 22$ و $Z = y - x$ باشد، μ_z کدام است؟

۱) 32 ۲) 16 ۳) 12 ۴) 6

معیارهای پراکندگی

دامنه تغییرات

۲۶. چه عدد دیگری بین $6, 5, 7, 3$ و 9 قرار گیرد تا بدون ایجاد تغییر در پارامترهای میانگین، میانه و دامنه تغییرات داده‌ها، مد مجموعه محسوب شود؟

۱) 5 ۲) 6 ۳) 7 ۴) 8

انحراف چارکی

۲۷. فرض کنید $Q_1 = 100$، $Md = 140$ و $Q_3 = 180$ باشد. نیم‌دامنه کدام است؟

۱) 40 ۲) 30 ۳) 60 ۴) 80

۲۸. در جدول داده‌های آماری زیر، انحراف چارکی کدام است؟

حدود دسته	15−18	18−21	21−24	24−27
فراوانی	12	15	19	14

۱) 2.1 ۲) 2.4 ۳) 2.6 ۴) 2.9

واریانس نمونه

۲۹. انحراف‌های مقادیر مشاهده‌شده از میانگین در 6 مورد از یک نمونه 7 تایی به صورت اعداد $4, 3, -1, -2, -4, -5$ محاسبه شده است. انحراف معیار نمونه چقدر است؟

۱) 2 ۲) 3 ۳) 4 ۴) 5

۳۰. در نمونه‌ای به چه حجم مجموع مجذورات داده‌ها 61، میانگین 3 و انحراف معیار 2 است؟

۱) 4 ۲) 5 ۳) 6 ۴) 7

خواص واریانس

۳۱ . اگر از هر یک از مشاهدات سه واحد کم کنیم، در انحراف معیار مشاهدات چه وضعی پیش می‌آید؟

۱) ثابت می‌ماند.
۲) سه واحد کاسته می‌شود.
۳) 9 واحد کاسته می‌شود.
۴) هیچ‌کدام

۳۲ . اگر واریانس مقادیر $x_1, x_2, \ldots, x_N$ برابر 16 باشد، انحراف معیار $\frac{x_1}{4}, \frac{x_2}{4}, \ldots, \frac{x_N}{4}$ کدام است؟

۱) 4
۲) 2
۳) 16
۴) 1

قضیه چی‌بی‌شف

۳۳ . اگر X یک متغیر تصادفی با متوسط صفر و واریانس یک باشد، آن‌گاه:

۱) $P(|X| \geq 1) \leq \frac{1}{2}$
۲) $P(|X| \geq 1) \leq \frac{1}{2}$
۳) $P(|X| \geq 1) \leq 1$
۴) $P(|X| \geq 1) \geq \frac{1}{4}$

۳۴ . فرض کنید X یک متغیر تصادفی نامنفی با میانگین 3 و واریانس 4 باشد، یک کران پایین برای $P(X < 6)$ عبارت است از:

۱) $\frac{6}{13}$
۲) $\frac{5}{9}$
۳) $\frac{3}{4}$
۴) $\frac{5}{6}$

میانگین و واریانس کل

۳۵ . باتوجه به این اطلاعات میانگین کل کدام است؟

N_i	100	50	150
μ_i	20	50	30

۱) 100
۲) 30
۳) 50
۴) 33.33

۳۶ . در یک شرکت، میانگین حقوق ماهیانه کارکنان مرد 120000، کارکنان زن 70000 و تمام کارکنان 100000 تومان بوده، چند درصد کارکنان، زن هستند؟

۱) 25
۲) 30
۳) 40
۴) 75

معیارهای پراکندگی نسبی

ضریب تغییرات

۳۷ . اطلاعات نمونه‌ای از کیفیت قطعات تولیدشده در دو خط تولید عبارت است از:

خط ۱: 3, 4, 6, 5, 5, 7 و **خط ۲:** 2, 3, 5, 4, 4, 2

کدام خط تولید از دقت بیشتری در تولید کالاهای همگون برخوردار است؟

۱) دقت در خط ۱ کمتر است.
۲) دقت در خط ۱ بیشتر است.
۳) دقت در دو خط برابر است.
۴) نیاز به اطلاعات بیشتری داریم.

۳۸ . اگر $\sum x_i = 60$ ، $\sum x_i^2 = 400$ و $N = 10$ باشد، ضریب پراکندگی کدام است؟

۱) 0.40
۲) 0.33
۳) 0.70
۴) 0.62

۳۹. در چه صورت ضریب تغییرات حقوق کارکنان یک مؤسسه، کاهش می‌یابد؟

۱) افزایش 5000 تومان به حقوق هر شخص
۲) کاهش 5 درصدی از حقوق هر شخص
۳) کاهش 5000 تومان از حقوق هر شخص
۴) افزایش 5 درصدی حقوق هر شخص

چولگی

۴۰. فرض کنید $\mu_x = 37$، $Mo = 43.6$ و واریانس 144 باشد. ضریب چولگی پیرسون کدام است؟

۱) 0.95 ۲) -0.55 ۳) 0.045 ۴) -0.045

۴۱. فرض کنید $Q_1 = 223.55$، $Md = 246$ و $Q_3 = 271.55$ باشد. ضریب چولگی کدام است؟

۱) 0.60 ۲) 0.551 ۳) 0.065 ۴) 0.089

۴۲. در یک توزیع متمایل به چپ، کدام‌یک از این گزینه‌ها درست است؟

۱) $\mu_x < Md < Mo$ ۲) $Md < \mu_x < Mo$ ۳) $Mo < Md < \mu_x$ ۴) $\mu_x < Mo < Md$

رابطه سه معیار تمرکز (در چولگی ضعیف)

۴۳. میانگین و میانه یک جامعه آماری به ترتیب 30 و 50 است و توزیع جامعه از چولگی معقولی برخوردار است. مد کدام است؟

۱) 90 ۲) 40 ۳) 25 ۴) مد ندارد.

کشیدگی

۴۴. اگر ضریب کشیدگی توزیعی -0.71 باشد، کدام عبارت درباره پراکندگی این جامعه صادق است؟

۱) پراکندگی جامعه نسبت به توزیع نرمال بیشتر و تفاوت آن فاحش است.
۲) پراکندگی جامعه نسبت به توزیع نرمال کمتر و تفاوت آن فاحش است.
۳) پراکندگی جامعه نسبت به توزیع نرمال بیشتر و تفاوت آن اندک است.
۴) پراکندگی جامعه نسبت به توزیع نرمال کمتر و تفاوت آن اندک است.

۴۵. اگر $N = 10$، $\sum (x_i - \mu_x)^4 = 7680$ و انحراف معیار جامعه برابر 4 باشد، ضریب کشیدگی کدام است؟

۱) 1 ۲) 3 ۳) 4 ۴) 0

۴۶. اگر گشتاور مرتبه چهارم حول میانگین مساوی 162 و واریانس برابر 9 باشد، ضریب کشیدگی کدام است؟

۱) 54 ۲) 18 ۳) 2.5 ۴) -1

نمودارها

۴۷. کدام‌یک از این نمودارها برای تحلیل اکتشافی مشاهدات استفاده می‌شود؟

۱) بافت نگار ۲) دایره‌ای ۳) پارتو ۴) جعبه‌ای

۴۸. کدام‌یک از این نمودارها برای تحلیل مشاهدات کمّی استفاده می‌شود؟

۱) شاخه و برگ ۲) دایره‌ای ۳) پارتو ۴) ستونی

۴۹. کدام‌یک از این نمودارها برای نمایش مشاهداتی با مقیاس رتبه‌ای مناسب است؟

۱) دایره‌ای ۲) چند ضلعی ۳) بافت نگار ۴) جعبه‌ای

۵۰ . کدام‌یک از این نمودارها برای نمایش مشاهداتی با مقیاس نسبی مناسب است؟

۱) بافت نگار ۲) چند ضلعی ۳) جعبه‌ای ۴) هر سه مورد

۵۱ . در رسم نمودارهای بافت‌نگار، محور x را بر اساس کدام اندازه مدرج می‌کنند؟

۱) فراوانی‌های نسبی ۲) کرانه‌های طبقات ۳) متوسط طبقات ۴) فراوانی‌های تجمعی

۵۲ . در رسم نمودار تجمعی، محور x را بر اساس کدام اندازه مدرج می‌کنند؟

۱) متوسط طبقات ۲) کرانه‌های طبقات ۳) حد پایین طبقات ۴) موارد ۱ و ۲

۵۳ . در کدام‌یک از این نمودارها ارزش مشاهدات هر طبقه یکسان تلقی می‌شود؟

۱) منحنی فراوانی تجمعی ۲) بافت‌نگار

۳) پلی‌گن فراوانی تجمعی ۴) هر سه مورد

پاسخنامه

۱	۲
۲	۱
۳	۲
۴	۲
۵	۴
۶	۱
۷	۲
۸	۳
۹	۳
۱۰	۳
۱۱	۱
۱۲	۲
۱۳	۳
۱۴	۴
۱۵	۴
۱۶	۱
۱۷	۴
۱۸	۱
۱۹	۲
۲۰	۲
۲۱	۲
۲۲	۱
۲۳	۲
۲۴	۳
۲۵	۳
۲۶	۲
۲۷	۴
۲۸	۳
۲۹	۳
۳۰	۲
۳۱	۱
۳۲	۴
۳۳	۳
۳۴	۲
۳۵	۲
۳۶	۳
۳۷	۲
۳۸	۲
۳۹	۱
۴۰	۲
۴۱	۳
۴۲	۳
۴۳	۱
۴۴	۱
۴۵	۴
۴۶	۴
۴۷	۴
۴۸	۱
۴۹	۱
۵۰	۴
۵۱	۲
۵۲	۴
۵۳	۳
۵۴	
۵۵	
۵۶	
۵۷	
۵۸	
۵۹	
۶۰	

برای دریافت پاسخ تشریحی سؤالات به سایت www.Tourani.ir **مراجعه کنید.**

فصل ۲ آنالیز ترکیبی و احتمال

آنالیز ترکیبی

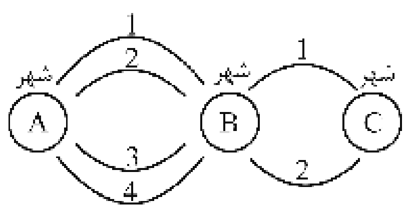
گاهی حالت‌های مختلف در وقوع یک رویداد آن‌قدر زیاد است که شمارش آن‌ها به صورت مستقیم عملاً زمان‌گیر و غیر ممکن است، مانند حالات مختلف پرتاب 1000 سکه و 400 تاس که اگر بخواهیم به طور مستقیم تمام حالات این رویداد را بنویسیم شاید چندین سال زمان صرف شود! در این‌گونه موارد از اصولی برای راحت‌تر شدن شمارش استفاده می‌کنیم که به آن **آنالیز ترکیبی** گویند.

بحث آنالیز ترکیبی شامل مباحث اصل ضرب، جایگشت، ترکیب و ترتیب است که در ادامه با ذکر مثال‌های مختلف به بررسی هر یک از آن‌ها می‌پردازیم.

اصل ضرب (اصل اساسی شمارش)

تعریف: هرگاه عملی به n_1 راه انجام شود و به ازای هر یک از این راه‌ها عمل دیگری به n_2 راه صورت گیرد و برای هر یک از راه‌های این دو عمل، n_3 راه برای عمل دیگری وجود داشته باشد و الی آخر، آن‌گاه $n_1 \times n_2 \times n_3 \times \cdots \times n_K$ راه برای انجام K عمل وجود خواهد داشت.

مثال ۱ شخصی می‌خواهد از شهر A به شهر C مسافرت کند و باید حتماً از شهر B عبور کند؛ یعنی عمل مسافرت این شخص در دو مرحله انجام می‌شود. اگر مسافرت از شهر A به شهر B به 4 طریق و از شهر B به شهر C به 2 طریق ممکن باشد، مسافرت از شهر A به شهر C به چند طریق ممکن است؟



حل: این عمل بنا بر اصل ضرب به $4 \times 2 = 8$ طریق ممکن است.

مثال ۲ یک سکه و یک تاس را با هم پرتاب می‌کنیم. تعداد کل حالات ظاهر شدن سکه و تاس کدام است؟

۱) 36 ۲) 6 ۳) 8 ۴) 12

حل: گزینه ۴ درست است.

سکه می تواند به یکی از 2 طریق {شیر ، خط} ظاهر شود و برای هریک از این دو طریق، تاس می‌تواند به یکی از 6 طریق $\{1,2,3,4,5,6\}$ ظاهر شود؛ بنابراین تاس و سکه می‌توانند به $2\times 6=12$ طریق مختلف ظاهر شوند.

مثال ۳ تهیه گزارش ستونی از اقلام حساب‌های اسناد پرداختنی، ذخیره مالیات بر درآمد، ذخیره مزایای پایان خدمت کارکنان و پیش‌دریافت‌ها، به طوری که بدهی جاری و بدهی بلندمدت از لحاظ سرفصل به تفکیک ارائه شوند، به چند صورت امکان‌پذیر است؟

۱) 4 ۲) 6 ۳) 8 ۴) 12

حل: گزینه ۳ درست است.

دقت کنید که دو سرفصل بدهی جاری و بلندمدت داریم که هر کدام باید به چهار مورد اسناد پرداختی، ذخیره مالیات بر درآمد، ذخیره مزایای پایان خدمت کارکنان و پیش‌دریافت‌ها تفکیک شوند؛ درنتیجه بنابر اصل ضرب، کل حالات برابر است با $2\times 4=8$ حالت.

	۱) بدهی بلندمدت	۲) بدهی جاری
۱) اسناد پرداختی	-	-
۲) ذخیره مالیات بر درآمد	-	-
۳) ذخیره مزایای پایان خدمت کارکنان	-	-
۴) پیش‌دریافت‌ها	-	-

$\longleftarrow 8=2\times 4$

شمارش اعداد

شمارش اعداد یکی از مسایلی است که به کمک اصل شمارش (ضرب) انجام می‌شود، فقط باید دقت کنیم که همواره میان یک عدد n رقمی و یک رمز یا شماره سریال n رقمی تفاوت وجود دارد.

اگر بخواهیم شماره‌ای چندرقمی برای رمز، پلاک ماشین و یا شماره سریال بسازیم، رقم یکان آن (رقم سمت چپ) می‌تواند صفر هم باشد اما اگر بخواهیم عددی چندرقمی بسازیم آن‌گاه دیگر رقم یکان(رقم سمت چپ) نمی‌تواند صفر باشد.

همچنین می‌دانیم **جایگاه اعداد** 0 تا 9 یعنی یکان، دهگان، صدگان و .... را **رقم** می‌نامیم؛ برای مثال می‌گوییم عدد 1 در رقم دهگان یا عدد 6 در رقم هزارگان.

✓ **دقت کنید!**

هنگام شمارش اعداد، به طور پیش‌فرض تکرار ارقام مجاز است، مگر آنکه به صورت صریح عبارت «بدون تکرار» ذکر شود.

مثال ۱ چند عدد چهاررقمی می‌توان ساخت؟

حل:

با تکرار: ابتدا یکی از ارقام 1 تا 9 را برای رقم هزارگان انتخاب می‌کنیم (9 حالت)، حال با توجه به مجاز بودن تکرار، برای هر یک از ارقام باقی‌مانده می‌توان یکی از ارقام 0 تا 9 را انتخاب کرد (10 حالت).

هزارگان صدگان دهگان یکان

$$\boxed{9}\times\boxed{10}\times\boxed{10}\times\boxed{10}=9000$$

بدون تکرار: ابتدا یکی از ارقام 1 تا 9 را برای رقم هزارگان انتخاب می‌کنیم (9 حالت)، حال با توجه به غیر مجاز بودن تکرار، 9 انتخاب باقی می‌ماند (تمام ارقام 0 تا 9 به غیر از عدد استفاده شده در رقم هزارگان) که به ترتیب برای ارقام باقی‌مانده تا یکان، 9 و 8 و 7 حالت خواهیم داشت؛ بنابراین:

$$\underset{\text{هزارگان}}{\boxed{9}} \times \underset{\text{صدگان}}{\boxed{9}} \times \underset{\text{دهگان}}{\boxed{8}} \times \underset{\text{یکان}}{\boxed{7}} = 4536$$

مثال ۲ چند عدد چهاررقمی با ارقام زوج می‌توان ساخت؟

حل: انتخاب‌ها فقط از میان ارقام زوج $8,6,4,2,0$ است. فقط رقم صفر نمی‌تواند در مکان هزارگان قرارگیرد؛ بنابراین:

$$\underset{\text{هزارگان}}{\boxed{4}} \times \underset{\text{صدگان}}{\boxed{5}} \times \underset{\text{دهگان}}{\boxed{5}} \times \underset{\text{یکان}}{\boxed{5}} = 500$$

مثال ۳ چند عدد چهاررقمی زوج می‌توان ساخت؟

حل:

با تکرار: ابتدا یکی از ارقام 0، 2، 4، 6 یا 8 را در رقم یکان برای زوج بودن عدد در نظر می‌گیریم (5 حالت). حال با توجه به مجاز بودن تکرار، 9 انتخاب برای هزارگان و 10 انتخاب برای بقیه ارقام وجود دارد؛ بنابراین:

$$\underset{\text{هزارگان}}{\boxed{9}} \times \underset{\text{صدگان}}{\boxed{10}} \times \underset{\text{دهگان}}{\boxed{10}} \times \underset{\text{یکان}}{\boxed{5}} = 4500$$

بدون تکرار: برای حل این مسئله دو حالت به وجود می‌آید:

الف) اگر رقم یکان برای زوج بودن 0 انتخاب شود (1 حالت)، آن‌گاه برای رقم هزارگان 9 انتخاب (یکی از اعداد 1 تا 9) را خواهیم داشت، در این وضعیت بقیه ارقام به ترتیب 8 و 7 انتخاب خواهد داشت، زیرا تکرار ارقام مجاز نیست؛ بنابراین:

$$\underset{\text{هزارگان}}{\boxed{9}} \times \underset{\text{صدگان}}{\boxed{8}} \times \underset{\text{دهگان}}{\boxed{7}} \times \underset{\text{یکان}}{\boxed{1}} = 504 \rightarrow$$ تعداد اعداد زوج که به صفر ختم می‌شوند.

ب) اگر رقم یکان برای زوج بودن یکی از ارقام 2 ، 4 ، 6 یا 8 انتخاب شود (4 حالت) آن‌گاه برای رقم هزارگان 8 انتخاب داریم (رقم 0 و رقمی که در یکان انتخاب شده است، دیگر مجاز نیست). در این حالت بعد از حذف یک انتخاب از رقم هزارگان برای ارقام باقی‌مانده 8 حالت باقی می‌ماند (تمام اعداد 0 تا 9 به غیر از عدد استفاده شده در رقم هزارگان و یکان) که به ترتیب 8 و 7 انتخاب خواهیم داشت؛ بنابراین:

$$\underset{\text{هزارگان}}{\boxed{8}} \times \underset{\text{صدگان}}{\boxed{8}} \times \underset{\text{دهگان}}{\boxed{7}} \times \underset{\text{یکان}}{\boxed{4}} = 1792 \rightarrow$$ تعداد اعداد زوج که به 2 ، 4 ، 6 یا 8 ختم می‌شوند.

در مجموع تعداد کل اعداد زوج چهاررقمی بدون تکرار برابر است با:

$$504 + 1792 = 2296$$

مثال ۴ چند عدد چهاررقمی با ارقام یکان و هزارگان یکسان می‌توان ساخت؟

حل: ابتدا یکی از ارقام 1 تا 9 را برای رقم هزارگان انتخاب می‌کنیم (9 حالت). حال ازآنجاکه رقم یکان و هزارگان باید یکسان باشد، 1 انتخاب برای رقم یکان وجود دارد (همان انتخابی که برای رقم هزارگان کرده‌ایم). درنهایت هر یک از ارقام باقی‌مانده می‌توانند یکی از ارقام 0 تا 9 را انتخاب کنند (10 حالت).

$$\underset{\text{هزارگان}}{\boxed{9}} \times \underset{\text{صدگان}}{\boxed{10}} \times \underset{\text{دهگان}}{\boxed{10}} \times \underset{\text{یکان}}{\boxed{1}} = 900$$

مثال ۵ چند عدد چهاررقمی می‌توان ساخت که فقط ارقام یکان و هزارگان یکسان دارند؟

حل: دقت کنید با توجه به ذکر کلمه «فقط» هیچ‌کدام از ارقام دهگان و صدگان نه با خودشان و نه با یکان و هزارگان نباید یکسان باشند.

ابتدا یکی از ارقام 1 تا 9 را برای رقم هزارگان انتخاب می‌کنیم (9حالت). حال ازآنجاکه رقم یکان و هزارگان باید یکسان باشد، 1 انتخاب برای رقم یکان وجود دارد (همان انتخابی که برای رقم هزارگان کرده‌ایم). درنهایت برای ارقام باقی‌مانده، 9انتخاب باقی می‌ماند (تمام ارقام 0 تا 9 به غیر از رقم استفاده شده در هزارگان) که با توجه به غیر مجاز بودن تکرار برای آن‌ها به ترتیب 9 و 8 انتخاب خواهیم داشت.

هزارگان صدگان دهگان یکان

$\boxed{9} \times \boxed{9} \times \boxed{8} \times \boxed{1} = 648$

مثال ۶ چند عدد چهاررقمی می‌توان ساخت که فقط سه رقم یکان و دهگان و صدگان یکسان دارند؟

حل: ابتدا یکی از ارقام 1 تا 9 را برای رقم هزارگان انتخاب می‌کنیم (9حالت). حال با توجه به ذکر کلمه «فقط»، برای ارقام باقی‌مانده 9 حالت باقی می‌ماند (تمام ارقام 0 تا 9 به غیر از عدد استفاده شده در رقم هزارگان). حال اگر 9 انتخاب را برای یکی از ارقام باقی‌مانده در نظر بگیریم، برای سایر ارقام 1 انتخاب بیشتر نخواهیم داشت چرا که باید ارقام صدگان و دهگان و یکان یکسان باشند.

هزارگان صدگان دهگان یکان

$\boxed{9} \times \boxed{9} \times \boxed{1} \times \boxed{1} = 81$

مفهوم متمایز و نامتمایز

وقتی گفته می‌شود «چند شیء متمایز» یعنی حتماً اشیا دارای تفاوت‌هایی هستند که این تفاوت‌ها به هنگام چیده شدن آن‌ها در کنار هم مشخص می‌شوند، مانند 6 کارت که از یک تا 6 شماره‌گذاری شده‌اند. اما در اشیای نامتمایز هیچ ویژگی که تفاوت میان اشیا را به هنگام چیده شدن آن‌ها کنارهم، مشخص کند وجود ندارد، مانند 6 کارت که بر روی همه آن‌ها عدد 2 نوشته شده است.

توجه کنید که افراد (اشخاص) همیشه متمایزند. همچنین اگر از نامتمایز بودن اشیا در صورت سؤال صحبتی نشود و هیچ استنتاجی نیز درباره نامتمایز بودن اشیا نتوان کرد، آن‌ها را به طور پیش‌فرض متمایز می‌دانیم.

جایگشت

تعریف: جایگشت‌های n شیء متمایز، حالات قرار گرفتن آن اشیا در کنار هم است.

حالت‌های مختلفی که در جایگشت بررسی می‌شود به صورت زیر است:

جایگشت در یک ردیف

تعداد جایگشت‌های n شیء متمایز در یک ردیف (صف) کنار یکدیگر برابر است با:

$$\boxed{n \times (n-1) \times \cdots \times 2 \times 1 = n!}$$

یادآوری: فاکتوریل عدد n برابر است با حاصل‌ضرب اعداد طبیعی از 1 تا n و آن را به صورت زیر نشان می‌دهند:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

همچنین **فاکتوریل عدد صفر، برابر یک است**؛ یعنی: $0! = 1$

مثال ۱ 5 نفر با 5 صندلی در یک ردیف داریم. به چند طریق افراد می‌توانند بر روی صندلی‌ها بنشینند؟

۱) 2^6 ۲) $6!$ ۳) 6^2 ۴) $5!$

حل: گزینه ۴ درست است.

$$\text{تعداد حالات} = 5 \times 4 \times 3 \times 2 \times 1 = 5!$$

مثال ۲ 5 دانشجوی حسابداری و 3 دانشجوی کامپیوتر در یک صف ایستاده‌اند. تعداد حالاتی که اول و آخر صف دانشجوی حسابداری باشد، چقدر است؟

۱) $7!$ ۲) $20 \times 6!$ ۳) $6!$ ۴) $10 \times 6!$

حل: گزینه ۲ درست است.

ابتدا به 5 حالت یکی از دانشجویان حسابداری را در ابتدا یا انتهای صف قرار می‌دهیم.
سپس از 4 حسابدار باقی‌مانده یکی را به 4 حالت انتخاب می‌کنیم و در طرف دیگر صف قرار می‌دهیم.
درنهایت 3 دانشجوی حسابداری و 3 دانشجوی کامپیوتر داریم که به $6!$ حالت می‌توانند بین ابتدا و انتهای صف جابه‌جا شوند، بنابراین:

$$\text{تعداد حالات} = 5 \times 6! \times 4$$

مثال ۳ 4 مرد و 4 زن را در یک صف در نظر بگیرید. مطلوب است محاسبه:

الف) تعداد کل حالاتی که آن‌ها در صف قرار می‌گیرند.
ب) تعداد حالاتی‌که همه مردها کنار هم باشند.
ج) تعداد حالاتی‌که همه مردها و همه زن‌ها کنار هم باشند.
د) تعداد حالاتی‌که مردها و زن‌ها یکی در میان قرار بگیرند (هیچ دو مرد یا دو زنی کنار هم نباشند).
ه) تعداد حالاتی که این 8 نفر 4 زوج باشند و باید هر زوج پهلوی هم بنشینند.

حل:

با توجه به آنکه کل افراد در صف 8 نفر هستند:

الف) $\text{حالات ممکن} = 8!$

ب)

$$W_1\,W_2\,W_3\,W_4\ \overset{M}{\boxed{M_1\,M_2\,M_3\,M_4}}$$

جایگشت 4 زن و 1 مرد $= 5!$
جایگشت 4 مرد کنار هم $= 4!$
$\rightarrow$ تعداد حالات مساعد $= 5! \times 4!$

ج)

$$\overset{W}{\boxed{W_1\,W_2\,W_3\,W_4}}\ \overset{M}{\boxed{M_1\,M_2\,M_3\,M_4}}$$

جایگشت 1 زن و 1 مرد $= 2!$
جایگشت 4 مرد کنار هم $= 4!$
جایگشت 4 زن کنار هم $= 4!$
$\rightarrow$ تعداد حالات مساعد $= 2! \times 4! \times 4!$

د) ابتدا از یک مرد شروع می‌کنیم و زن‌ها و مردها را یکی در میان قرار می‌دهیم.

$$M_1\ \ W_1\ \ M_2\ \ W_2\ \ M_3\ \ W_3\ \ M_4\ \ W_4$$

جایگشت 4 مرد در وضعیت‌های مشخص‌شده $= 4!$
جایگشت 4 زن در وضعیت‌های مشخص‌شده $= 4!$
$\rightarrow$ تعداد حالات مساعد $= 4! \times 4!$

حال می‌توانیم از یک زن شروع کنیم و مردها و زن‌ها را یکی در میان قرار دهیم:

$$W_1\ \ M_1\ \ W_2\ \ M_2\ \ W_3\ \ M_3\ \ W_4\ \ M_4$$

جایگشت 4 مرد در وضعیت‌های مشخص‌شده $= 4!$
جایگشت 4 زن در وضعیت‌های مشخص‌شده $= 4!$
$\rightarrow$ تعداد حالات مساعد $= 4! \times 4!$

بنابراین تعداد کل حالات مساعد برابر است با $2 \times 4! \times 4!$.

دقت کنید اگر 4 مرد و 3 زن وجود داشت برای حل قسمت (د) حتماً باید از یک مرد شروع می‌کردیم، به عبارت دیگر:

$$M_1\ \ W_1\ \ M_2\ \ W_2\ \ M_3\ \ W_3\ \ M_4$$

جایگشت 4 مرد در وضعیت‌های مشخص‌شده $= 4!$
جایگشت 3 زن در وضعیت‌های مشخص‌شده $= 3!$
$\rightarrow$ تعداد حالات مساعد $= 4! \times 3!$

ه) ابتدا هر زوج را به طور جداگانه کنار هم قرار می‌دهیم (جایگشت هر زوج در کنار هم $2!$ است). حال هر 4 زوج را به $4!$ حالت کنار هم قرار می‌دهیم:

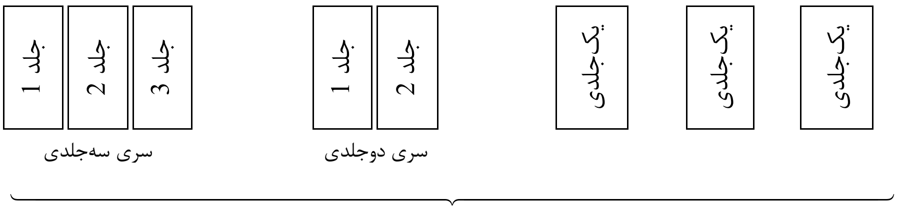
$$(2!)^4 \times 4! = 16 \times 24 = 384$$

مثال ۴ در یک ردیف از قفسه‌ای، 8 کتاب است که سه تای آن‌ها یک‌جلدی و بقیه دوجلدی و سه‌جلدی هستند. به چند طریق می‌توان کتاب‌ها را در این ردیف جا داد به طوری که کتاب‌های سه‌جلدی و دوجلدی به ترتیب شماره جلدها قرار گیرند؟

۱) 120 ۲) 180 ۳) 240 ۴) 360

حل: گزینه ۱ درست است.

8 کتاب داریم که 3 تای آن‌ها یک‌جلدی، 2 تای آن‌ها مربوط به یک سری دوجلدی و 3 تای آن‌ها نیز مربوط به یک سری سه‌جلدی است. می‌خواهیم سری دوجلدی و سه‌جلدی به ترتیب شماره جلدها کنار هم باشند. ابتدا این دو سری را جای می‌دهیم، یعنی سری سه‌جلدی را به ترتیب شماره می‌چینیم و چون می‌خواهیم حتماً به همین شیوه کنار هم باشند این 3 جلد را با هم یک بسته در نظر می‌گیریم. برای سری دوجلدی نیز همین کار را انجام می‌دهیم. حال دو بسته و 3 جلد کتاب یک جلدی داریم که جایگشت آن‌ها در یک ردیف $5! = 120$ است.



$5! =$ جایگشت 5 تا در یک ردیف

توجه به یک نکته ضروری است، در بسته سری سه‌جلدی و دوجلدی، جای سه جلد و دو جلد کتاب ثابت است زیرا به ترتیب شماره جلد قرار گرفته‌اند، بنابراین کتاب‌های درون بسته جایگشت ندارند.

جایگشت دایره‌ای (مدور)

تعریف: جایگشت‌های n شیء متمایز را روی محیط یک منحنی بسته، ترتیب دایره‌ای گویند که تعداد این جایگشت‌ها برابر است با $(n-1)!$.

مثال ۱ 4 نفر به چند طریق می‌توانند دور یک میز بنشینند؟

۱) $4!$ ۲) 4 ۳) 6 ۴) 2

حل: گزینه ۳ درست است.

چون در مثال به «دور میز» اشاره شده است، داریم:

$(4-1)! = 3! = 6 =$ تعداد حالات

مثال ۲ به چند طریق 7 نفر می‌توانند دور هم بنشینند به طوری که یک فرد همواره در جای ثابتی باشد؟

۱) $5!$ ۲) $7!$ ۳) $7 \times 5!$ ۴) $6!$

حل: گزینه ۳ درست است.

می‌خواهیم 7 نفر دور هم بنشینند به طوری که یک نفر در جای ثابتی باشد؛ بنابراین:

ابتدا به 7 **حالت** یکی از افراد را انتخاب می‌کنیم و در یک جای ثابت دلخواه قرار می‌دهیم، سپس 6 نفر باقی‌مانده می‌توانند به $(6-1)!$ **حالت** دور هم بنشینند؛ در این صورت تعداد حالات برابر است با:

$7(6-1)! = 7 \times 5!$

مثال ۳ در یک کنفرانس شامل نمایندگان پنج کشور، همه افراد دور یک میز گرد نشسته‌اند. هر کشور سه نماینده به کنفرانس فرستاده است که رئیس آن‌ها در وسط می‌نشیند. به چند طریق نمایندگان این پنج کشور می‌توانند دور میز بنشینند؟

۱) 320 ۲) 768 ۳) 780 ۴) 1440

حل: گزینه ۲ درست است.

در این سؤال هر کشور سه نماینده به کنفرانس می‌فرستد که از این سه نفر یک نفر رئیس است و بین دو نفر دیگر می‌نشیند، یعنی به صورت زیر:

نماینده رئیس نماینده

می‌توانند جابه‌جا شوند $\Rightarrow$ تعداد حالات قرار گرفتن سه نماینده یک کشور $= 2!$

سه نماینده هر کشور به $2!$ حالت می‌توانند در کنار هم قرار گیرند؛ بنابراین برای 5 کشور داریم:

$$(2!)^5 = 32$$

اما توجه کنید که این تنها جایگشت سه نماینده هر کشور در کنار هم بود و جایگشت این دسته‌های سه نفری (5 کشور) دور یک میز برابر است با:

$$(5-1)! = 4!$$

درواقع هر سه نماینده مربوط به یک کشور را یک گروه در نظر می‌گیریم؛ حال این 5 گروه به $(5-1)!$ حالت می‌توانند دور یک میز گرد بنشینند.

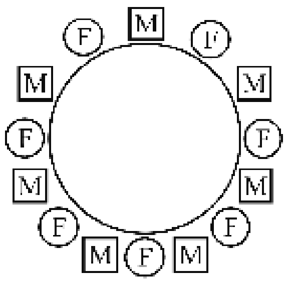
بنابراین کل حالات نشستن نمایندگان برابر است با:

$$(2!)^5 \times (5-1)! = 32 \times 24 = 768$$

نکته: هرگاه بخواهیم n شیء و m شیء را یک در میان روی یک منحنی بسته (دایره) بچینیم، تعداد حالات قرار گرفتن آن‌ها با شرط $m = n$ برابر است با $m!(n-1)!$.

مثال ۴ تعداد حالات نشستن 7 مرد و 7 زن را، به صورت یک در میان دور یک میز، به دست آورید.

حل:



$$\text{تعداد حالات مساعد} = m!(n-1)! = (7!)(6!)$$

توجه کنید درصورتی‌که مردها 7 نفر و زن‌ها 6 نفر باشند، نمی‌توان آن‌ها را یکی در میان دور میز نشاند، اما می‌توان در یک ردیف یک در میان به صورت زیر قرار داد که تعداد حالات آن برابر است با $7! \times 6!$.

MFMFMFMFMFMFM

اگر مردها 7 نفر و زن‌ها 5 نفر باشند، به هیچ صورتی نمی‌توان آن‌ها را یکی در میان قرار داد، نه دور یک میز و نه در یک ردیف.

نکته: اگر بخواهیم دو نوع شیء متمایز (مانند دختر و پسر، مهره‌های سفید و سیاه، کتاب‌های ریاضی و آمار) به طور متناوب (یک در میان) در یک ردیف کنار هم قرار بگیرند، حداکثر اختلاف بین تعداد آن‌ها باید «یک» باشد.

جایگشت با تکرار (افرازهای مرتب)

تعداد جایگشت‌های n شیء که n_1 تای آن‌ها از نوع اول (نامتمایز)، n_2 تای آن‌ها از نوع دوم (نامتمایز)، ... و n_k تای آن‌ها از نوع kام (نامتمایز) باشد، برابر است با:

$$\boxed{\frac{n!}{n_1!n_2!\cdots n_k!} \quad ; \quad n_1 + n_2 + \cdots + n_k = n}$$

مثال ۱ با حروف کلمه «حسابداری» چند کلمه 8 حرفی می‌توان نوشت؟

۱) $8!$ ۲) $4!$ ۳) $\frac{8!}{2!}$ ۴) $\frac{8!}{4!}$

حل: گزینه ۳ درست است.

کلمه «حسابداری» شامل 8 حرف است که حرف «الف»، دوبار در آن تکرار شده است، بنابراین دو حرف از این 8 حرف نامتمایز و از یک نوع‌اند؛ بنابراین:

$$\text{تعداد کلمات} = \frac{8!}{2!} = 20160$$

مثال ۲ با حروف کلمه ENGINEERING چند کلمه می‌توان ساخت؟

۱) 27720 ۲) 72720 ۳) 727200 ۴) 277200

حل: گزینه ۴ درست است.

حروف کلمه ENGINEERING، شامل 11 حرف با 3 تکرار E، 3 تکرار N، 2 تکرار I، 2 تکرار G و یک R است:

$$\{ EEE , NNN , GG , II , R \}$$

بنابراین جایگشت حروف آن برابر است با:

$$\frac{11!}{3!\,3!\,2!\,2!\,1!} = 277200$$

مثال ۳ هرگاه m نسخه از هر n کتاب مختلف وجود داشته باشد، به چند طریق می‌توان آن‌ها را در یک ردیف قرار داد؟

۱) $\frac{(m+n)!}{n \times m!}$ ۲) $\frac{(mn)!}{(m!)^n}$ ۳) $\frac{(m-n)!}{(m!)^n}$ ۴) $\frac{(m-n)!}{n \times m!}$

حل: گزینه ۲ درست است.

توجه کنید که n نوع کتاب مختلف، اما از هر نوع m تا شبیه هم موجود است یعنی در کل $n \times m$ کتاب داریم؛ بنابراین خواهیم داشت:

$$\frac{\text{جایگشت کل کتاب‌ها در یک ردیف}}{\text{از هر نوع کتاب } m \text{ تا شبیه به هم}} = \frac{(n \times m)!}{\underbrace{m!m!\ldots m!}_{n}} = \frac{(nm)!}{(m!)^n}$$

مثال ۴ چند عدد 6 رقمی می‌توان با رقم‌های $4,3,2,2,2,1$ درست کرد؟

۱) 120 ۲) 240 ۳) 720 ۴) 1440

حل: گزینه ۱ درست است.

تعداد اعداد 6 رقمی با این ارقام عبارت است از جایگشت 6 رقم که 3 تای آن‌ها شبیه به هم است؛ بنابراین:

$$\text{تعداد اعداد 6 رقمی با این ارقام} = \frac{6!}{3!} = 120$$

انتخاب

انتخاب اشیا، با توجه به با جایگذاری یا بدون جایگذاری بودن آن و اهمیت داشتن یا نداشتن ترتیب، به حالات زیر تقسیم می‌شود:

- انتخاب
  - بدون جایگذاری
    - ترکیب (ترتیب مهم نیست)
    - تبدیل (ترتیب مهم است)
  - با جایگذاری (ترتیب مهم است)

ترکیب (Combination)

تعریف: ترکیب عبارت است از انتخاب r شیء از n شیء متمایز به طوری که اولاً، ترتیب انتخاب مهم نباشد (بدون ترتیب)، ثانیاً، بدون جایگذاری انجام شود. در این صورت، تعداد انتخاب‌ها برابر است با:

$$C_n^r=\binom{n}{r}=\frac{n!}{r!(n-r)!}$$

✓ **دقت کنید!**

در رابطه بالا ازآنجاکه ترتیب در « r شیء انتخاب شده» و همین‌طور « $n-r$ شیء باقی‌مانده» مهم نیست، جایگشت آن‌ها در مخرج قرار می‌گیرد تا در حالات انتخاب محاسبه نشوند.

مثال ۱ به چند طریق می‌توان از 12 کتاب، 3 کتاب را به عنوان کتاب سال برگزید؟

۱) 36 ۲) 1320 ۳) 72 ۴) 220

حل: گزینه ۴ درست است.

تعداد انتخاب‌های 3 کتاب از 12 تا، وقتی ترتیب کتاب‌ها در انتخاب تأثیری ندارد برابر است:

$$\binom{12}{3}=\frac{12!}{3!(12-3)!}=220$$

مثال ۲ دبیر حسابداری در امتحان آخر ترم 20 سؤال به دانش‌آموزان خود داده است که آن‌ها به طور دلخواه به 18 سؤال پاسخ دهند. دانش‌آموزان به چند طریق می‌توانند سؤالات خود را برگزینند؟

۱) 190 ۲) 380 ۳) 360 ۴) 180

حل: گزینه ۱ درست است.

انتخاب 18 سؤال از 20 سؤال که ترتیب انتخاب در آن اهمیتی ندارد، برابر است با:

$$\binom{20}{18}=\frac{20!}{18!(20-18)!}=\frac{19\times 20}{2}=190$$

مثال ۳ به چند طریق می‌توان از 12 کتاب که 5 تای آن آمار و بقیه ریاضی هستند، یک کتاب آمار و 2 کتاب ریاضی را به عنوان کتاب سال برگزید؟

۱) 220 ۲) 110 ۳) 205 ۴) 105

حل: گزینه ۴ درست است.

تعداد حالات انتخاب 1 کتاب آمار از 5 تا $=\binom{5}{1}$

تعداد حالات انتخاب 2 کتاب ریاضی از 7 تا $=\binom{7}{2}$

بنا بر اصل ضرب، کل حالات انتخاب برابر است با:

$$\binom{5}{1}\binom{7}{2}=\frac{5!}{1!4!}\times\frac{7!}{2!5!}=5\times 21=105$$

مثال ۴ اگر دایره بازاریابی و فروش شرکتی بخواهد یکی از 5 متن تهیه‌شده را با یکی از 4 وسیله تبلیغاتی رادیو، تلویزیون، مجله و روزنامه آگهی کند، این کار به چند طریق امکان‌پذیر است؟

۱) 9 ۲) 5 ۳) 20 ۴) 24

حل: گزینه ۳ درست است.

$\binom{5}{1}=$ تعداد حالات انتخاب یکی از 5 متن

$\binom{4}{1}=$ تعداد حالات انتخاب یکی از 4 وسیله تبلیغاتی

بنا بر اصل ضرب، کل حالات انتخاب برابر است با:

$$\binom{5}{1}\binom{4}{1}=5\times 4=20$$

مثال ۵ دانشجویی موظف است به 10 پرسش از 13 پرسش داده‌شده پاسخ دهد با این شرط که باید حداقل 3 پرسش از 5 پرسش اول را پاسخ دهد. پاسخ به سؤالات به چند طریق ممکن است؟

۱) 264 ۲) 272 ۳) 276 ۴) 286

حل: گزینه ۳ درست است.
در این سؤال دانشجو باید به 10 سؤال از 13 سؤال پاسخ دهد به‌گونه‌ای که حداقل 3 تا از 5 تای اول را پاسخ دهد و بقیه را از 8 سؤال بعدی انتخاب کند.

$$\underbrace{\binom{5}{3}\binom{8}{7}}_{\text{3 تا از 5 تای اول}}+\underbrace{\binom{5}{4}\binom{8}{6}}_{\text{4 تا از 5 تای اول}}+\underbrace{\binom{5}{5}\binom{8}{5}}_{\text{5 تا از 5 تای اول}}=10\times 8+5\times 28+1\times 56=276$$

مثال ۶ 4 نفر از 7 کارگر یک کارگاه مرد هستند. به چند طریق می‌توان 3 نفر از بین کارگران انتخاب کرد به طوری که حداکثر یک نفر زن باشد؟

۱) 21 ۲) 22 ۳) 18 ۴) 19

حل: گزینه ۲ درست است.

$$\text{حداکثر یک زن} \equiv \text{هیچ زن یا یک زن} \Rightarrow \binom{3}{0}\binom{4}{3}+\binom{3}{1}\binom{4}{2}=22$$

مثال ۷ فردی 8 دوست دارد که می‌خواهد 5 نفر آن‌ها را به یک مهمانی دعوت کند. چند انتخاب وجود دارد اگر:

الف) دو نفر از دوستان وی با هم اختلاف داشته باشند و نخواهند با هم شرکت کنند؟

ب) دو نفر از دوستان وی درصورتی‌که با هم دعوت شوند در مهمانی شرکت کنند؟

حل:

الف) **راه حل اول**: وضعیت‌های زیر قابل پیش‌بینی است:

۱- هر دو نفر را دعوت نمی‌کنیم: $\binom{2}{0}$؛ یعنی انتخاب 5 دوست از $8-2=6$ نفر باقی‌مانده: $\binom{6}{5}$

۲- یکی از دو نفر را دعوت می‌کنیم: $\binom{2}{1}$ و 4 نفر دیگر را از $8-2=6$ نفر باقی‌مانده انتخاب می‌کنیم: $\binom{6}{4}$

بنابراین تعداد کل حالات مساعد برابر است با:

$$\binom{2}{0}\binom{6}{5}+\binom{2}{1}\binom{6}{4}=6+2\times 15=36$$

راه حل دوم: از حالت مکمل آن استفاده می‌کنیم، یعنی حالاتی که دو نفر با هم دعوت شوند را از کل حالات انتخاب کم می‌کنیم.

کل حالات انتخاب (دعوت) 5 نفر از 8 نفر برابر است با: $\binom{8}{5}$

حالات انتخاب (دعوت) دو نفر مورد نظر عبارت است از انتخاب دو نفر و انتخاب 3 مهمان دیگر از $8-2=6$ نفر باقی‌مانده؛ یعنی:

$$\binom{2}{2}\binom{6}{3}=20$$

بنابراین تعداد کل حالات مساعد برابر است با:

$$\binom{8}{5}-20=56-20=36$$

ب) وضعیت‌های زیر قابل پیش‌بینی است:

۱- دو نفر خاص را دعوت کرده و سه مهمان دیگر را از بین 6 نفر باقی‌مانده انتخاب کنیم: $\binom{2}{2}\binom{6}{3}$

۲- دو نفر خاص را دعوت نکنیم و هر 5 مهمان را از میان 6 نفر باقی‌مانده انتخاب کنیم: $\binom{2}{0}\binom{6}{5}$

بنابراین تعداد کل حالات مساعد برابر است با:

$$\binom{2}{2}\binom{6}{3}+\binom{2}{0}\binom{6}{5}=20+6=26$$

مثال ۸ با حروف کلمه «حسابداری» چند کلمه 4 حرفی می‌توان نوشت؟

۱) $4!$ ۲) $8!$ ۳) 20160 ۴) 1020

حل: گزینه ۴ درست است.

تعداد حروف متمایز کلمه حسابداری، 7 حرف است اما در مسئله، تعداد کلمات 4 حرفی خواسته شده است، بنابراین ابتدا باید 4 حرف از بین 7 حرف متمایز انتخاب کرد و در تعداد جایگشت‌های این 4 حرف ضرب کرد؛ جواب حاصل تعداد کلماتی است که در آن‌ها حرف «الف» یک بار می‌تواند باشد. حال باید کلماتی را در نظر بگیریم که حرف «الف» دو بار در آن‌ها تکرار شده است، بنابراین به غیر از دو حرف «الف»، 2 حرف دیگر باقی می‌ماند که باید از بین 6 حرف انتخاب و در تعداد جایگشت‌های آن ضرب کرد. جواب حاصل، تعداد کلماتی است که در آن‌ها دو بار حرف «الف» آمده است. در نهایت باید دو جواب به دست آمده را با هم جمع کنیم.

$$\underbrace{\binom{7}{4}\times 4!}_{\text{تعداد کلمات 4 حرفی دارای حداکثر یک حرف «الف»}}+\underbrace{\binom{6}{2}\times \frac{4!}{2!}}_{\text{تعداد کلمات 4 حرفی دارای دو حرف «الف»}}=1020$$

مثال ۹ حروف کلمه حسابداران را در نظر بگیرید. مطلوب است محاسبه:

الف) تعداد کلمات 9 حرفی با حروف کلمه مورد نظر.

ب) تعداد کلمات 4 حرفی با حروف کلمه مورد نظر.

ج) تعداد کلمات 4 حرفی که حرف «س» در آن باشد.

د) تعداد کلمات 4 حرفی که با حرف «س» شروع شود.

حل:

کلمه «حسابداران» دارای 9 حرف است که حرف «الف» در آن 3 بار تکرار شده است.

الف) تعداد جایگشت 9 حرف با 3 حرف نامتمایز برابر است با $\frac{9!}{3!}$.

ب) ابتدا تعداد کلمات 4 حرفی را که از حروف (ح، س، ا، ب، د، ر، ن) تشکیل شده و حرف «الف» در آن حداکثر یک بار تکرار شده است، به دست می‌آوریم.

$$\left.\begin{array}{l}\binom{7}{4}: \text{انتخاب 4 حرف از بین 7 حرف متمایز} \\ 4!: \text{جایگشت 4 حرف متمایز در کلمه}\end{array}\right\} \Rightarrow \text{تعداد حالات} = \binom{7}{4} \times 4!$$

سپس تعداد کلمات 4 حرفی را که حرف «الف» در آن 2 بار تکرار شده است، به دست می‌آوریم:

$$\underbrace{(\text{ا} ، \text{ا})}_{\text{2 حرف}} \quad \text{و} \quad \underbrace{(\text{ح} ، \text{س} ، \text{ب} ، \text{د} ، \text{ر} ، \text{ن})}_{\text{2 حرف}} \Rightarrow \underset{\substack{\text{انتخاب 2 حرف} \\ \text{از 6 حرف متمایز}}}{\binom{6}{2}} \times \underset{\substack{\text{جایگشت 4 حرف} \\ \text{با 2 حرف نامتمایز «الف»}}}{\frac{4!}{2!}}$$

در نهایت تعداد کلمات 4 حرفی را که حرف «الف» در آن 3 بار تکرار شده است، به دست می‌آوریم:

$$\underbrace{(\text{ا} ، \text{ا} ، \text{ا})}_{\text{3 حرف}} \quad \text{و} \quad \underbrace{(\text{ح} ، \text{س} ، \text{ب} ، \text{د} ، \text{ر} ، \text{ن})}_{\text{1 حرف}} \Rightarrow \underset{\substack{\text{انتخاب 1 حرف} \\ \text{از 6 حرف متمایز}}}{\binom{6}{1}} \times \underset{\substack{\text{جایگشت 4 حرف} \\ \text{با 3 حرف نامتمایز «الف»}}}{\frac{4!}{3!}}$$

بنابراین تعداد کلمات 4 حرفی برابر است با:

$$\binom{7}{4} \times 4! + \binom{6}{2} \times \frac{4!}{2!} + \binom{6}{1} \frac{4!}{3!}$$

ج) درصورتی‌که حرف «س» یکی از 4 حرف باشد، باید 3 حرف دیگر را انتخاب کرده، سپس جایگشت‌ها را بررسی کنیم.

ابتدا فرض می‌کنیم حرف «الف» حداکثر یک بار تکرار شده است، بنابراین:

$$\underbrace{(\text{س})}_{\text{1 حرف}} \quad \text{و} \quad \underbrace{(\text{ح} ، \text{ا} ، \text{ب} ، \text{د} ، \text{ر} ، \text{ن})}_{\text{3 حرف}} \Rightarrow \binom{6}{3} \times 4!$$

سپس فرض می‌کنیم حرف «الف»، 2 بار تکرار شده است، بنابراین:

$\underbrace{(\text{س})}_{\text{حرف } 1}$ و $\underbrace{(\text{ا ، ا})}_{\text{حرف } 2}$ و $\underbrace{(\text{ح ، ب ، د ، ر ، ن})}_{\text{حرف } 1}$ $\Rightarrow \binom{5}{1}\times\frac{4!}{2!}$

درنهایت فرض می‌کنیم حرف «الف»، 3 بار تکرار شده است، بنابراین:

$\underbrace{(\text{س})}_{\text{حرف } 1}$ و $\underbrace{(\text{ا ، ا ، ا})}_{\text{حرف } 3}$ $\Rightarrow \frac{4!}{3!}$

بنابراین تعداد کلمات 4 حرفی شامل حرف «س» برابر است با:

$$\binom{6}{3}\times 4!+\binom{5}{1}\times\frac{4!}{2!}+\frac{4!}{3!}$$

د) درصورتی‌که قرار باشد، حرف «س» ابتدای کلمه 4 حرفی باشد، محل آن ثابت است و باید 3 حرف دیگر را انتخاب کرده، جایگشت آن‌ها را بررسی کنیم:

ابتدا تعداد کلماتی را به دست می‌آوریم که حرف «الف» حداکثر یک بار در آن‌ها تکرار شده است، بنابراین:

$\underbrace{(\text{س})}_{\text{حرف (ثابت) } 1}$ و $\underbrace{(\text{ح ، ا ، ب ، د ، ر ، ن})}_{\text{حرف } 3}$ $\Rightarrow \binom{6}{3}\times 3!$

سپس فرض می‌کنیم حرف «الف»، 2 بار تکرار شده است، بنابراین:

$\underbrace{(\text{س})}_{\text{حرف (ثابت) } 1}$ و $\underbrace{(\text{ا ، ا})}_{\text{حرف } 2}$ و $\underbrace{(\text{ح ، ب ، د ، ر ، ن})}_{\text{حرف } 1}$ $\Rightarrow \binom{5}{1}\times\frac{3!}{2!}$

درنهایت فرض می‌کنیم حرف «الف»، 3 بار تکرار شده است، بنابراین:

$\underbrace{(\text{س})}_{\text{حرف (ثابت) } 1}$ و $\underbrace{(\text{ا ، ا ، ا})}_{\text{حرف } 3}$ $\Rightarrow \frac{3!}{3!}=1$

بنابراین تعداد کلمات 4 حرفی که با حرف «س» شروع شده است، برابر است با:

$$\binom{6}{3}\times 3!+\binom{5}{1}\frac{3!}{2!}+\frac{3!}{3!}$$

✓ **دقت کنید!**

همواره به طور پیش‌فرض انتخاب، بدون جایگذاری در نظر گرفته می‌شود، مگر آنکه به طور صریح قید شود انتخاب با جایگذاری است.

روابط مهم در ترکیب

1) $\boxed{C_n^r = C_n^{n-r} \quad \text{یا} \quad \binom{n}{r}=\binom{n}{n-r}}$

برای مثال برای $r = 0,1,2$ داریم:

$$r=0 \to C_n^0 = C_n^n = 1 \quad \text{یا} \quad \binom{n}{0}=\binom{n}{n}=1$$

$$r=1 \to C_n^1 = C_n^{n-1} = n \quad \text{یا} \quad \binom{n}{1}=\binom{n}{n-1}=n$$

$$r=2 \to C_n^2 = C_n^{n-2} = \frac{n(n-1)}{2} \quad \text{یا} \quad \binom{n}{2}=\binom{n}{n-2}=\frac{n(n-1)}{2}$$

2) $$\boxed{C_{n+1}^r = C_n^{r-1} + C_n^r \quad \text{یا} \quad \binom{n+1}{r}=\binom{n}{r-1}+\binom{n}{r}}$$

3) $$\boxed{(a+b)^n = C_n^0 a^n b^0 + C_n^1 a^{n-1} b + \cdots + C_n^n a^0 b^n}$$

4) $$\boxed{C_n^0 + C_n^1 + C_n^2 + \cdots + C_n^n = 2^n \quad \text{یا} \quad \binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^n}$$

به عبارت دیگر، مجموع تمام ترکیبات (انتخاب‌ها) از یک مجموعه n عضوی، 2^n است.

$$\boxed{n \text{ زوج}: \binom{n}{0}+\binom{n}{2}+\cdots+\binom{n}{n}=\binom{n}{1}+\binom{n}{3}+\cdots+\binom{n}{n-1}}$$

برای مثال برای $n = 4$ داریم:

$$\binom{4}{0}+\binom{4}{2}+\binom{4}{4}=\binom{4}{1}+\binom{4}{3}$$

$$\boxed{n \text{ فرد}: \binom{n}{0}+\binom{n}{2}+\cdots+\binom{n}{n-1}=\binom{n}{1}+\binom{n}{3}+\cdots+\binom{n}{n}}$$

برای مثال برای $n = 3$ داریم:

$$\binom{3}{0}+\binom{3}{2}=\binom{3}{1}+\binom{3}{3}$$

5) $$\boxed{\sum_{i:\text{زوج}}\binom{n}{i}=\sum_{i:\text{فرد}}\binom{n}{i}=2^{n-1}}$$

مثال ۱ از بین 6 نفر به چند طریق می‌توان یک تیم حداقل 2 نفره انتخاب کرد؟

۱) 247 ۲) 216 ۳) 164 ۴) 57

حل: گزینه ۴ درست است.

با توجه به رابطه (۴)، تعداد کل حالات انتخاب r شیء $(r = 0,1,\ldots,n)$ از n شیء برابر 2^n است. بنابراین در این سؤال که حداقل 2 انتخاب خواسته شده است می‌توان از مکمل آن استفاده کرد:

راه حل اول:

$$\text{حداقل 2 انتخاب} = \text{کل حالات} - (\text{یک انتخاب} + \text{هیچ انتخاب}) = 2^6 - \binom{6}{0} - \binom{6}{1} = 64-1-6=57$$

راه حل دوم:

$$\text{حداقل 2 انتخاب} = \binom{6}{2}+\binom{6}{3}+\binom{6}{4}+\binom{6}{5}+\binom{6}{6}=57$$

مثال ۲ اگر n زوج باشد، آن‌گاه مقدار $C_n^0 - C_n^1 + C_n^2 - \cdots + C_n^n$ چقدر است؟

۱) 0 ۲) 1 ۳) 2^n ۴) 2^{n-1}

حل: گزینه ۱ درست است.

راه حل اول: با توجه به رابطه (۴) برای n زوج داریم:

$$C_n^0+C_n^2+\cdots+C_n^n=C_n^1+C_n^3+\cdots+C_n^{n-1} \quad \rightarrow \quad C_n^0-C_n^1+C_n^2-\cdots+C_n^n=0$$

راه حل دوم: با یک عددگذاری ساده نیز می‌توان به نتیجه رسید. مثلاً اگر $n=2$ را در نظر بگیریم، داریم:

$$C_2^0-C_2^1+C_2^2=1-2+1=0$$

توزیع n شیء متمایز در k سلول

تعداد تقسیمات n شیء در k سلول به طوری که n_1 تای آن‌ها در سلول اول، n_2 تای آن‌ها در سلول دوم، ... و n_k تای آن‌ها در سلول kام قرار گیرد، برابر است با:

$$\boxed{\binom{n}{n_1\ n_2 \cdots\ n_k}=\frac{n!}{n_1!n_2!\ldots n_k!}}$$

مثال ۱ به چند طریق می‌توان 9 اسباب‌بازی را بین 4 بچه تقسیم کرد به شرط آنکه به کوچک‌ترین بچه 3 اسباب‌بازی و به هر کدام از بچه‌های دیگر 2 اسباب‌بازی برسد؟

۱) 27 ۲) 108 ۳) 5674 ۴) 7560

حل: گزینه ۴ درست است.

$$\frac{9!}{3!2!2!2!}=7560$$

مثال ۲ به چند طریق می‌توان 9 نفر کارمند را در یک اتاق 4 نفره، یک اتاق 3 نفره و یک اتاق 2 نفره نشاند؟ (مدیریت ـ ۸۲)

۱) 1400 ۲) 1260 ۳) 72 ۴) 24

حل: گزینه ۲ درست است.

$$\binom{9}{4\ 3\ 2}=\frac{9!}{4!\,3!\,2!}=1260$$

مثال ۳ اگر 8 نفر شام را با هم صرف کنند، به چند طریق می‌توانند 3 خوراک مرغ، 4 خوراک ماهی و یک خوراک میگو سفارش دهند؟

۱) 280 ۲) 80 ۳) $8!$ ۴) 360

حل: گزینه ۱ درست است.

$$\frac{8!}{3!\,4!\,1!}=280$$

تبدیل (Permutation)

تعریف: هرگاه در **ترکیب** (انتخاب r شیء از n شیء)، **ترتیب r** انتخاب اهمیت داشته باشد، آن‌گاه حالات مختلف این ترتیب‌ها با نماد P_n^r نشان داده و به صورت زیر محاسبه می‌شود:

$$P_n^r = \frac{n!}{(n-r)!} \quad ; \quad n \geq r$$

✓ **دقت کنید!**

دو ترتیب وقتی متمایز هستند که یا مجموعه اشیای به کار رفته در آن‌ها متفاوت باشد و یا در صورت یکسان بودن، ترتیب آن‌ها متفاوت باشد.

تفاوت اصلی ترکیب و تبدیل

به طور کلی، در مسایلی که تغییر در ترتیب انتخاب اشیا نتواند حالت جدیدی ایجاد کند، از ترکیب و در غیر این صورت از ترتیب استفاده می‌شود؛ بنابراین:

در مسایلی مانند انتخاب چند نماینده از بین چند نفر، انتخاب چند شیء از میان اشیای مختلف و مخلوط کردن رنگ‌های مختلف برای ساختن رنگ‌های جدید، از **ترکیب** استفاده می‌کنیم.

اما در مسایلی مانند انتخاب سه نفر برای پست‌های مدیر، حسابدار، معاون، یا انتخاب سه جایزه برای شاگرد اول، دوم و سوم، از **تبدیل** استفاده می‌کنیم.

مثال ۱ مطلوب است محاسبه تعداد حالات انتخاب 3 دانشجو از بین 10 دانشجو به طوری که:

الف) به عنوان شاگردان ممتاز معرفی شوند. (ترتیب اهمیت ندارد)

ب) به عنوان شاگرد اول، دوم، سوم معرفی شوند. (ترتیب اهمیت دارد)

حل:

الف) $$C_{10}^3 = \binom{10}{3} = \frac{10!}{3!7!} = 120$$

ب) $$P_{10}^3 = \frac{10!}{(10-3)!} = 720$$

مثال ۲ از فهرست نام 10 عضو یک باشگاه، برای انتخاب رییس، نایب رییس، خزانه‌دار و منشی، 4 نام استخراج می‌شود. به چند راه مختلف می‌توان این کار را انجام داد؟

۱) 210 ۲) 5040 ۳) 8400 ۴) 2450

حل: گزینه ۲ درست است.

$$P_{10}^4 = \frac{10!}{(10-4)!} = 5040$$

مثال ۳ شرکتی با امکان 20 نوع گزینش متفاوت برای استخدام، به ترتیب یک کارمند اداری و یک فروشنده انتخاب کرده است. داوطلبان این دو شغل چند نفر بوده‌اند؟

۱) 40 ۲) 20 ۳) 10 ۴) 5

حل: گزینه ۴ درست است.

این شرکت از میان n داوطلب این دو شغل، **به ترتیب** یک کارمند و یک فروشنده را با 20 بار گزینش متفاوت انتخاب کرده است. دقت کنید که چون دو شغل کارمند و فروشنده متفاوت است مهم است که کدام‌یک کارمند باشد و کدام‌یک فروشنده؛ بنابراین ترتیب انتخاب اهمیت دارد (اولی کارمند و دومی فروشنده یا اولی فروشنده و دومی کارمند).

$$P_n^2 = \frac{n!}{(n-2)!} \rightarrow 20 = \frac{n!}{(n-2)!} \rightarrow \frac{n(n-1)(n-2)!}{(n-2)!} = 20 \rightarrow n(n-1) = 20 \rightarrow n = 5$$

مثال ۴ اگر 9 نفر در یک مسابقه شرکت کنند، به چند طریق ممکن است جوایز اول، دوم و سوم را دریافت کنند؟

(حسابداری ـ ۸۱ و مدیریت ـ ۷۹)

۱) 84 ۲) 504 ۳) 635 ۴) 3024

حل: گزینه ۲ درست است.

ترتیب جوایز مهم است، بنابراین:

$$P_9^3 = \frac{9!}{(9-3)!} = \frac{9!}{6!} = 9 \times 8 \times 7 = 504$$

انتخاب با جایگذاری

در انتخاب r شیء از n شیء متمایز به روش با جایگذاری، در هر انتخاب ابتدا انتخاب قبلی را برمی‌گردانیم، سپس انتخاب جدید انجام می‌شود؛ درنتیجه ترتیب انتخاب مهم است و داریم:

با جایگذاری n شیء از r تعداد حالات انتخاب $= \underbrace{n \times n \times \cdots \times n \times n}_{r \text{ بار}} = n^r$

نکته: در انتخاب با جایگذاری، در هر مرحله از r انتخاب، همواره n شیء وجود دارد.

مثال ۱ از بین سه شیء a، b و c، دو شی ء را یک بار با جایگذاری و بار دیگر بدون جایگذاری انتخاب کنید.

حل:

با جایگذاری		بدون جایگذاری		
		تبدیل		ترکیب
$n^r = 3^2 = 9$ حالت		$P_n^r = \frac{3!}{1!} = 6$ حالت		$C_n^r = \binom{n}{r} = \binom{3}{2} = 3$ حالت
انتخاب اول	انتخاب دوم	انتخاب اول	انتخاب دوم	
↓	↓	↓	↓	
a	a	a	b	
b	b	b	a	
c	c	a	c	a , b
a	b	c	a	a , c
b	a	b	c	b , c
a	c	c	b	
c	a			
b	c			
c	b			

مثال ۲ تعداد نمونه‌های سه‌تایی با جایگذاری و بدون جایگذاری از جامعه‌ای که دارای 5 عنصر است به ترتیب کدام است؟

۱) 5 , 125 ۲) 10 , 125 ۳) 10 , 243 ۴) 20 , 243

حل: گزینه ۲ درست است.

انتخاب با جایگذاری : $n^r = 5^3 = 125$

انتخاب بدون جایگذاری : $C_n^r = C_5^3 = \binom{5}{3} = \frac{5!}{3!2!} = 10$

مسئله انطباق (جورها)

تعریف: اگر سه حرف a، b و c مفروض باشند، به ترتیب قرارگرفتن آن‌ها به طوری که هیچ‌یک در جایگاه فعلیشان قرار نگیرند، ترتیب ناسازگار گفته می‌شود.

۱- تعداد ترتیب‌های ناسازگار n شیء متمایز برابر است با:

$$\boxed{n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}}$$

۲- تعداد ترتیب‌های ناسازگار r شیء از n شیء متمایز برابر است با:

$$\boxed{\frac{n!}{(n-r)!}\sum_{k=0}^{r}\frac{(-1)^k}{k!}}$$

✓ **دقت کنید!**

r ناسازگاری برابر $n-r$ سازگاری است.

مثال ۱ 3 نفر که پالتوهای خود را در محلی آویزان کرده‌اند، هر یک به تصادف یک پالتو برمی‌دارند. مطلوب است تعداد حالاتی که هیچ‌یک از آن‌ها پالتوی خود را برنداشته باشد.

۱) 8 ۲) 6 ۳) 4 ۴) 2

حل: گزینه ۴ درست است.

بنا بر حالت (۱) داریم:

$$n!\sum_{k=0}^{n}\frac{(-1)^k}{k!} = 3!\sum_{k=0}^{3}\frac{(-1)^k}{k!} = 6\left(1-1+\frac{1}{2}-\frac{1}{6}\right) = 6\times\frac{2}{6} = 2$$

مثال ۲ برای 10 نفر، 10 نامه می‌فرستیم. تعداد حالاتی که فقط 7 نفر نامه خودشان را دریافت کنند، کدام است؟

۱) $\frac{10!}{7!}$ ۲) 7 ۳) 240 ۴) 70

حل: گزینه ۳ درست است.

فقط 7 نفر نامه خودشان را دریافت کنند یعنی 3 نفر نامه خودشان را دریافت نکنند، درنتیجه بنا بر حالت (۲) داریم:

$$\frac{n!}{(n-r)!}\sum_{k=0}^{r}\frac{(-1)^k}{k!} = \frac{10!}{(10-3)!}\sum_{k=0}^{3}\frac{(-1)^k}{k!} = 720\left(1-1+\frac{1}{2}-\frac{1}{6}\right) = 240$$

احتمال (Probability)

برای بیان مفهوم احتمال، ابتدا باید عناوین زیر مطرح شوند:

آزمایش (Experiment)

تعریف: در نظریه احتمال، هر عملی که نتیجه آن را نتوان از پیش تعیین کرد آزمایش نامیده می‌شود. به عبارت دیگر آزمایش فرآیندی است با نتایج محدود و غیر قابل پیش‌بینی.

فضای نمونه (Sample Space)

تعریف: مجموعه تمام پیامدهای ممکن یک آزمایش را فضای نمونه آن آزمایش گویند.

برای مثال، فضای نمونه پرتاب دو سکه به صورت زیر است:

S = { (ش و ش)، (ش و خ)، (خ و ش)، (خ و خ)}

فضای نمونه محدود و نامحدود

اگر تعداد اعضای فضای نمونه‌ای نامتناهی باشد، آن فضای نمونه نامحدود است. برای مثال شرکتی را در نظر بگیرید که تولیدکننده لامپ است. مأمور کنترل کیفیت این شرکت می‌خواهد آن‌قدر لامپ‌ها را آزمایش کند تا به اولین لامپ معیوب برسد. فضای نمونه این آزمایش به صورت زیر است:

$S=\{1,2,3,\ldots\}$

اگر فضای نمونه تعداد محدودی عضو داشته باشد، آن فضای نمونه محدود است، مانند فضای نمونه پرتاب یک تاس که به صورت زیر است:

$S=\{1,2,3,4,5,6\}$

فضای نمونه گسسته و پیوسته

فضای نمونه گسسته

فضای نمونه‌ای را که شامل تعداد متناهی یا نامتناهی ولی شمارش پذیر عضو است، فضای نمونه گسسته گویند؛ مانند فضای نمونه پرتاب یک سکه یا یک تاس که تعداد عناصر فضای نمونه آن‌ها متناهی است و فضای نمونه تعداد لامپ‌های انتخاب‌شده تا مشاهده اولین لامپ معیوب، که تعداد عناصر آن نامتناهی ولی شمارش‌پذیر است.

فضای نمونه پیوسته

اگر فضای نمونه شامل مجموعه تمام اعداد بین دو حد مشخص (یک فاصله) باشد، اصطلاحاً آن را فضای نمونه پیوسته گویند؛ مانند مدت زمانی که کارگری برای کار روی قطعه‌ای صرف می‌کند.

$S=\{t: 4\leq t\leq 6\}$

پیشامد (Event)

تعریف: در نظریه احتمال، پیشامد زیرمجموعه‌ای از فضای نمونه S است.

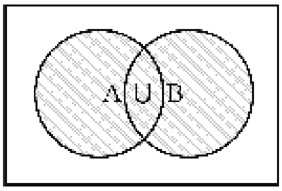
برای مثال، در پرتاب یک تاس با فضای نمونه $S=\{1,2,3,4,5,6\}$:

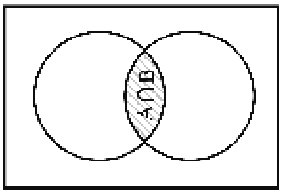
$A=\{2,4,6\}$: پیشامد زوج آمدن

$B=\{\}$: پیشامد 7 آمدن

اجتماع و اشتراک دو پیشامد

مفهوم اجتماع و اشتراک با توجه به دو نمودار زیر کاملاً آشکار است:





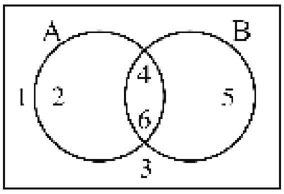
✓ دقت کنید!

اجتماع با نماد $\bigcup$ به مفهوم **«یا»** و اشتراک با نماد $\bigcap$ به مفهوم **«و»** است.

برای مثال اگر در پرتاب یک تاس پیشامد های A و B به صورت زیر باشند، داریم:

A : زوج آمدن

B : بیشتر از 3 آمدن



$A = \{2,4,6\}$

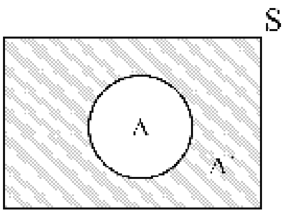
$B = \{4,5,6\}$

$A \bigcap B = \{4,6\}$

$A \bigcup B = \{2,4,5,6\}$

مکمل (متمم) پیشامد

اگر A پیشامد وقوع یک حادثه باشد، آن‌گاه مکمل A که با یکی از علایم «A' , A^C , $\bar{A}$» نشان داده می‌شود، عدم وقوع حادثه را نشان می‌دهد، که:

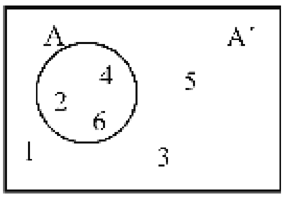


$A \bigcap A' = \{\varnothing\}$

برای مثال اگر در پرتاب یک تاس، پیشامد های A و A' به صورت زیر باشند، داریم:

A : زوج آمدن

A' : فرد آمدن



$A = \{2,4,6\}$

$A' = \{1,3,5\}$

$A \bigcap A' = \{\varnothing\}$

$A \bigcup A' = \{S\}$

احتمال

تعریف: اندازه شانس وقوع پیشامد A را که با $P(A)$ نشان داده می‌شود، احتمال پیشامد A گویند. به عبارت دیگر شانس وقوع پیشامد خاصی را گویند که $0 \le P(A) \le 1$.

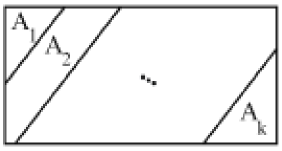
انواع پیشامد از نظر احتمال وقوع

پیشامد	احتمال وقوع	مثال
۱) پیشامد غیرممکن	$P(A)=0$	پیشامد ظاهر شدن 7 در پرتاب تاس
۲) پیشامد تصادفی	$0<P(A)<1$	پیشامد ظاهر شدن شیر در پرتاب سکه
۳) پیشامد حتمی (یقینی)	$P(A)=1$	پیشامد وقوع فضای نمونه (S)

گروه کامل حوادث (افراز فضای نمونه‌ای)

تعریف: هرگاه نتیجه وقوع حادثه A یکی از پیشامدهای **جدا از هم** (مجزای) $A_1, A_2, \dots, A_n$ باشد، آن‌گاه $A_1, A_2, \dots, A_n$ را گروه کامل حوادث گویند.

به عبارت دیگر:



$$\begin{cases} S = A_1 \cup A_2 \cup \dots \cup A_k \\ A_i \cap A_j = \phi \qquad i \neq j \end{cases} \longrightarrow \sum_{i=1}^{n} P(A_i) = P(A_1 \cup A_2 \cup \dots \cup A_k) = 1$$

نکته: **پیشامدهای مکمل، گروه کامل حوادث** هستند و داریم:

$$P(A \cup A') = 1$$

نکته: اگر احتمال وقوع حادثه B، k برابر احتمال وقوع حادثه A باشد و این دو حادثه، کل را بپوشانند $(P(A)+P(B)=1)$، داریم:

$$P(A) = \frac{1}{k+1} \quad , \quad P(B) = \frac{k}{k+1}$$

مثال هرگاه فضای نمونه شامل چهار پیشامد باشد و $P(e_1)=2P(e_2)$ و $P(e_3)=P(e_4)=\frac{1}{4}$ باشد، $P(e_1)$ و $P(e_2)$ را پیدا کنید.

۱) $\frac{1}{2}, \frac{1}{2}$ ۲) $\frac{1}{2}, \frac{1}{4}$ ۳) $\frac{2}{3}, \frac{1}{3}$ ۴) $\frac{1}{3}, \frac{1}{6}$

حل: گزینه ۴ درست است.

$$P(e_1) + P(e_2) + P(e_3) + P(e_4) = 1$$

$$2P(e_2) + P(e_2) + \frac{1}{4} + \frac{1}{4} = 1 \rightarrow 3P(e_2) = \frac{1}{2} \rightarrow P(e_2) = \frac{1}{6}, P(e_1) = \frac{1}{3}$$

پیشامدهای هم‌تراز (هم‌شانس)

هرگاه نتیجه وقوع حادثه A یکی از پیشامدهای $A_1, A_2, ..., A_n$ باشد، آن‌گاه به طور پیش‌فرض، این پیشامدها هم‌تراز (هم‌شانس) در نظر گرفته می‌شوند و احتمال وقوع آن‌ها برابر با $\frac{1}{n}$ است؛ به عبارت دیگر:

$$P(A_i) = \frac{1}{n} \quad ; \quad i = 1, 2, ..., n$$

برای مثال در پرتاب یک تاس، احتمال وقوع هر یک از اعداد روی تاس با هم برابر و مساوی $\frac{1}{6}$ است.

$$P(A_i) = \frac{1}{6} \quad ; \quad i = 1, 2, \cdots, 6$$

یا در پرتاب یک سکه سالم، احتمال وقوع شیر آمدن یا خط آمدن با هم برابر و مساوی $\frac{1}{2}$ است.

$$P(A_i) = \frac{1}{2} \quad ; \ i = \text{خط و شیر}$$

نکته: پیشامدهای ممکن (گروه کامل حوادث) برای هر حادثه به طور پیش‌فرض هم‌تراز در نظرگرفته می‌شوند؛ بنابراین درصورتی‌که تعداد پیشامدهای ممکن برای یک حادثه برابر k باشد، آن‌گاه احتمال وقوع هر یک برابر $\frac{1}{k}$ است.

انواع بیان احتمال

۱- احتمال کلاسیک

تعریف: احتمال وقوع پیشامد خاصی مانند A عبارت است از تعداد عضوهای پیشامد A (تعداد حالات مساعد) به تعداد عضوهای فضای نمونه (تعداد حالات ممکن).

به عبارت دیگر:

$$\boxed{P(A) = \frac{n(A)}{n(S)} = \frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات ممکن}}}$$

مثال ۱ اعداد $1, 2, 3, 4, 5, 6$ بر روی 6 مهره یکسان نوشته شده‌اند. اگر دو مهره را با هم بیرون بیاوریم، با کدام احتمال مجموع اعداد این دو مهره مضرب 3 خواهد بود؟

۱) $\frac{1}{3}$ ۲) $\frac{1}{4}$ ۳) $\frac{2}{5}$ ۴) $\frac{4}{15}$

حل: گزینه ۱ درست است.

$$\text{حالات ممکن : انتخاب 2 مهره از 6 مهره} \rightarrow n(S) = \binom{6}{2} = 15$$

$$\text{حالات مساعد} : A = \{(1,2), (1,5), (2,4), (3,6), (4,5)\} \rightarrow n(A) = 5$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{5}{15} = \frac{1}{3}$$

مثال ۲ دو تاس را به طور متوالی پرتاب می‌کنیم. احتمال اینکه عدد تاس دوم حداقل 2 واحد از عدد تاس اول بیشتر باشد کدام است؟

۱) $\frac{2}{9}$ ۲) $\frac{5}{18}$ ۳) $\frac{7}{18}$ ۴) $\frac{11}{26}$

حل: گزینه ۲ درست است.

کل حالات پرتاب دو تاس : حالات ممکن $\rightarrow \quad n(S)=36$

عدد تاس اول − عدد تاس دوم ≥ 2 : حالات مساعد

تاس اول \ تاس دوم	1	2	3	4	5	6
1	–	–	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
2	–	–	–	$(2,4)$	$(2,5)$	$(2,6)$
3	–	–	–	–	$(3,5)$	$(3,6)$
4	–	–	–	–	–	$(4,6)$
5	–	–	–	–	–	–
6	–	–	–	–	–	–

$\longrightarrow n(A)=10$

$$P(A)=\frac{n(A)}{n(S)}=\frac{10}{36}=\frac{5}{18}$$

مثال ۳ 5 دانشجوی برق و 3 دانشجوی کامپیوتر در یک صف ایستاده‌اند. احتمال اینکه اول و آخر صف دانشجوی برق باشد، چقدر است؟

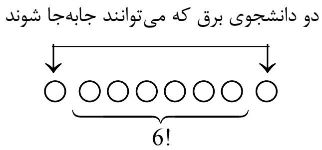
۱) $\frac{3}{14}$ ۲) $\frac{5}{14}$ ۳) $\frac{9}{28}$ ۴) $\frac{9}{14}$

حل: گزینه ۲ درست است.

حالات ممکن: تعداد کل حالات قرار گرفتن این 8 نفر در یک ردیف برابر $8!$ است.

$$n(S)=8!$$

حالات مساعد: می‌خواهیم دو نفر اول و آخر صف دانشجوی برق باشند. بنابراین ابتدا 2 نفر را از 5 دانشجوی برق انتخاب می‌کنیم $\binom{5}{2}$ و در دو انتهای صف قرار می‌دهیم. باید دقت کنید که این دو نفر می‌توانند با هم جابه‌جا شوند یعنی به $2!$ حالت قرار گیرند. حال 6 دانشجوی باقیمانده (3 برق، 3 کامپیوتر) را به $6!$ حالت ردیفی در کنار هم بین این دو نفر قرار می‌دهیم:



$$\begin{cases} n(A)=2!\binom{5}{2}\times 6! \\ P(A)=\dfrac{n(A)}{n(S)}=\dfrac{2!\binom{5}{2}6!}{8!}=\dfrac{\frac{5!}{2!\,3!}\times 2!}{7\times 8}=\dfrac{20}{56}=\dfrac{5}{14} \end{cases}$$

مثال ۴ پنج رقم $1,1,1,2,2$ را به تصادف در کنار هم قرار می‌دهیم. احتمال اینکه عدد 5 رقمی حاصل زوج باشد کدام است؟

۱) 0.3 ۲) 0.4 ۳) 0.5 ۴) 0.6

حل: گزینه ۲ درست است.

حالات ممکن : 1,1,1,2,2 تمام اعداد 5 رقمی با ارقام $\rightarrow \quad n(S)=\dfrac{5!}{3!\,2!}$

حالات مساعد : 1,1,1,2,2 تمام اعداد 5 رقمی زوج با ارقام $\rightarrow \quad n(A)=\dfrac{4!}{3!}$

$$P(A)=\frac{n(A)}{n(S)}=\frac{\frac{4!}{3!}}{\frac{5!}{3!\,2!}}=\frac{4}{10}=0.4$$

دقت کنید، برای اعداد 5 رقمی زوج با ارقام 1,1,1,2,2، کافی است یک رقم 2 را سمت راست ثابت نگه داریم و بقیه ارقام جایگشت با تکرار داشته باشند.

$$\underbrace{\boxed{1\,,1\,,1\,,2}}_{\frac{4!}{3!}}\boxed{2}$$

مثال ۵ پنج دانشجو را در اتاق‌های 2 و 3 نفره به تصادف جای می‌دهیم. احتمال اینکه دو نفر مورد نظر از آنان در یک اتاق جای نداشته باشند کدام است؟

۱) 0.3　　۲) 0.4　　۳) 0.5　　۴) 0.6

حل: گزینه ۴ درست است.

حالات ممکن: تعداد حالات ممکن تقسیم 5 نفر بین دو اتاق 2 نفره و 3 نفره برابر است با:

$$n(S)=\binom{5}{2\ \ 3}=\frac{5!}{2!\,3!}=10$$

حالات مساعد: ابتدا آن دو نفر خاص را به $2!$ حالت در دو اتاق جای می‌دهیم. حال 3 نفر باقی‌مانده‌اند که طبق $\binom{3}{2\ \ 1}=\frac{3!}{2!\,1!}=3$ در دو اتاق جای خواهند گرفت؛ بنابراین داریم:

$$P(A)=\frac{n(A)}{n(S)}=\frac{2\times 3}{10}=0.6$$

مثال ۶ یک قفل رمزدار از سه رقم 0 تا 9 تشکیل شده است (ارقام نمی‌توانند تکرار شوند). رمزی را به تصادف امتحان می‌کنیم؛ احتمال آنکه قفل باز شود چقدر است؟

حل:

$$\begin{cases} P(A)=\dfrac{\text{تعداد مساعد}}{\text{تعداد کل حالات}}=\dfrac{n(A)}{n(S)}=\dfrac{1}{720}\ ; \quad \text{احتمال کلاسیک} \\ n(S)=\text{تعداد کل حالات}=10\times 9\times 8=720 \\ n(A)=\text{تعداد مساعد}=1 \quad \rightarrow \quad \text{فقط یکی از رمزها قفل را باز می‌کند.} \end{cases}$$

مثال ۷ فرض کنید 6 کارت وجود دارد که شماره‌های 1 تا 6 روی آن‌ها نوشته شده است. با قرار دادن این کارت‌ها به ترتیب‌های مختلف، شماره‌های 6 رقمی ساخته می‌شود. شماره‌ای به کمک آن‌ها ساخته شده است، مطلوب است محاسبه احتمال آنکه شماره ساخته‌شده:

الف) زوج باشد.

ب) بزرگ‌تر از 300 هزار باشد.

ج) رقم یکان آن مضرب 3 باشد.

حل:

الف) تعداد کل حالات اعداد شش رقمی با ارقام 1 تا 6 بدون محدودیت $6!=720$ است.

$$n(A)=\boxed{5}\times\boxed{4}\times\boxed{3}\times\boxed{2}\times\boxed{1}\times\overset{2,4,6}{\boxed{3}}=360 \;\rightarrow\; P(A)=\frac{n(A)}{n(S)}=\frac{360}{720}=\frac{1}{2}$$

توجه کنید که چون شماره‌ها بر روی کارت نوشته شده‌اند وقتی یک کارت را برای یکی از ارقام انتخاب کنیم، برای ارقام دیگر نمی‌توان آن را استفاده کرد در واقع همانند احتمال بدون تکرار ارقام است.

ب)

$$n(B)=\overset{3,4,5,6}{\boxed{4}}\times\boxed{5}\times\boxed{4}\times\boxed{3}\times\boxed{2}\times\boxed{1}=480 \;\rightarrow\; P(B)=\frac{480}{720}=\frac{2}{3}$$

ج)

$$n(D)=\boxed{5}\times\boxed{4}\times\boxed{3}\times\boxed{2}\times\boxed{1}\times\overset{3,6}{\boxed{2}}=240 \;\rightarrow\; P(D)=\frac{240}{720}=\frac{1}{3}$$

مثال ۸ در کمدی 10 جفت کفش نگهداری می‌شود. اگر 8 لنگه کفش به تصادف انتخاب شود، احتمال پیشامدهای زیر را به دست آورید:

الف) هیچ جفت کفش انتخاب نشود.

ب) درست یک جفت کفش انتخاب شود.

حل:

حالات ممکن: برای هر دو قسمت «الف» و «ب» عبارت است از انتخاب 8 لنگه کفش از بین 20 لنگه کفش (10جفت)، یعنی:

$$n(S)=\binom{20}{8}$$

الف) می‌خواهیم این 8 لنگه کفش هیچ‌کدام جفت نباشند، بنابراین ابتدا باید 8 جفت از 10 جفت را انتخاب کنیم، سپس یک لنگه از هر یک برداریم:

$$n(A)=\binom{10}{8}\underbrace{\binom{2}{1}\binom{2}{1}\cdots\binom{2}{1}}_{\text{8 تا}}=\binom{10}{8}2^8 \xrightarrow{n(S)=\binom{20}{8}} P(A)=\frac{\binom{10}{8}\times 2^8}{\binom{20}{8}}$$

ب) می‌خواهیم در این 8 لنگه کفش دقیقاً یک جفت کفش موجود باشد. یعنی یک جفت از 10جفت انتخاب می‌کنیم و برای انتخاب 6 لنگه متفاوت ابتدا 6 جفت از 9 جفت باقی‌مانده را انتخاب می‌کنیم، سپس از هرکدام یک لنگه برمی‌داریم:

$$n(A)=\underbrace{\binom{10}{1}\binom{2}{2}}_{\text{یک جفت}}\times\overbrace{\binom{9}{6}\underbrace{\binom{2}{1}\binom{2}{1}\cdots\binom{2}{1}}_{\text{6 تا}}}^{\text{6 لنگه}} \xrightarrow{n(S)=\binom{20}{8}} P(A)=\frac{10\times\binom{9}{6}\times 2^6}{\binom{20}{8}}$$

۲- احتمال هندسی

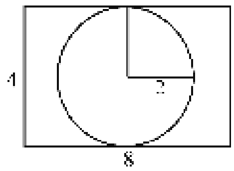
احتمال هندسی به صورت زیر محاسبه می‌شود:

$$P(A)=\frac{\text{طول، سطح یا حجم مساعد}}{\text{طول، سطح یا حجم کل}}$$

مثال تیراندازی به تصادف تیری به سمت صفحه‌ای شکل زیر پرتاب می‌کند. احتمال آنکه تیر با دایره برخورد کند کدام است؟

۱) $\frac{\pi}{32}$ ۲) $\frac{\pi}{8}$

۳) $\frac{\pi}{4}$ ۴) $\frac{\pi}{2}$



حل: گزینه ۲ درست است.

توجه کنید در مسایلی که با شکل مواجه هستیم (فضای نمونه پیوسته) باید به‌جای تعداد فضای نمونه (حالات کل) و تعداد حالات مطلوب، مساحت کل و مساحت مطلوب را به دست آوریم. مثلاً در این سؤال داریم:

$$P(\text{تیر به دایره بخورد})=\frac{\text{مساحت دایره}}{\text{مساحت مستطیل}}=\frac{\pi r^2}{\text{عرض}\times\text{طول}}=\frac{\pi\times 2^2}{8\times 4}=\frac{\pi}{8}$$

۳- احتمال آماری

در آزمایش‌هایی که پیشامدهای مقدماتی هم‌شانس نیستند، تعریف احتمال به صورت زیر است:

$$P(A)=A\ \text{فراوانی نسبی پیشامد}=\frac{\text{تعداد دفعاتی که } A \text{ در } N \text{ تکرار آزمایش روی می‌دهد}}{N}$$

در صورتی می‌توان از فراوانی نسبی به عنوان مبنای احتمال استفاده کرد که تعداد تکرارهای آزمایش (N) به سمت بی‌نهایت میل کند، بنابراین:

$$P(A)=\lim_{N\to\infty}(\text{فراوانی نسبی پیشامد } A \text{ در } N \text{ تکرار})=\lim_{N\to\infty} f_i$$

مثال در 100 بار پرتاب یک سکه، 65 بار شیر ظاهر شده است. احتمال شیر آمدن سکه با توجه به یافته‌های آزمایش چقدر است؟

حل:

$$P(A)=\text{فراوانی نسبی شیرآمدن}=\frac{\text{تعداد دفعاتی که از 100 بار آزمایش شیر آمده}}{N}=\frac{65}{100}=0.65$$

قانون اعداد بزرگ (به صورت برنولی)

احتمال اینکه قدرمطلق تفاضل فراوانی نسبی یک حادثه از احتمالش، کوچک‌تر از هر عدد مثبتی مانند ε باشد، با افزایش n برابر با 1 است، یعنی:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1$$

مثال کدام‌یک از عبارات زیر بیان قانون اعداد بزرگ به صورت برنولی است؟

۱) $\lim\limits_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=0$
۲) $\lim\limits_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=1$
۳) $\lim\limits_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1$
۴) $\lim\limits_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\ge\varepsilon\right)=0$

حل: گزینه ۳ درست است.

پیشامدهای مستقل و وابسته (Independent & Dependent Events)

تعریف ۱: دو پیشامد را **مستقل** گویند هرگاه وقوع یا عدم وقوع یک پیشامد، تأثیری در وقوع یا عدم وقوع پیشامد دیگر نداشته باشد و داریم:

مستقل $A , B \iff P(A\cap B)=P(A)\times P(B)$

تعریف ۲: دو پیشامد را **وابسته** گویند هرگاه مستقل نباشند؛ به عبارت دیگر وقوع یا عدم وقوع یک پیشامد بتواند در وقوع یا عدم وقوع پیشامد دیگر اثر بگذارد، و داریم:

وابسته $A , B \iff P(A\cap B)\ne P(A)\times P(B)$

نکته: اگر (A , B) **مستقل** باشند، بدیهی است تمام زوج پیشامدهای (A , B') و (A' , B) و (A' , B') نیز **مستقل** هستند و برعکس.

مثال ۱ اگر $P(A)=0.3$ و $P(B)=0.2$ و $P(A\cap B)=0.06$، رویدادهای (حوادث) A و B چگونه‌اند؟

۱) مکمل ۲) مستقل ۳) ناسازگار ۴) وابسته

حل: گزینه ۲ درست است.

$P(A\cap B)=P(A).P(B) \to 0.06=0.3\times 0.2 \to$ A و B مستقل‌اند

دقت کنید، درصورتی‌که مقدار احتمال $P(A\cap B)$ هر عددی غیر از 0.06 باشد، آن‌گاه پیشامدهای A و B وابسته خواهند بود.

مثال ۲ اگر $P(A)=0.3$ و $P(B)=0.2$ و A و B مستقل باشند، $P(\bar{A}\cap\bar{B})$ برابر است با:

۱) 0.44 ۲) 0.50 ۳) 0.56 ۴) 0.667

حل: گزینه ۳ درست است.

مستقل$(A , B)\to(\bar{A} , \bar{B})$ مستقل

$$P(\bar{A}\cap\bar{B})=P(\bar{A})\cdot P(\bar{B})=(1-P(A))\cdot(1-P(B))=(1-0.3)(1-0.2)=(0.7)(0.8)=0.56$$

مثال ۳ اگر $P(A)=0.3$ و $P(B)=0.2$، A و B مستقل باشند، $P(A\cap B')$ برابر است با:

۱) 0.24 ۲) 0.50 ۳) 0.56 ۴) 0.667

حل: گزینه ۱ درست است.

مستقل $(A , B) \to$ مستقل (A , B')

$$P(A\cap B')=P(A)\times P(B')=0.3\times 0.8=0.24$$

$$P(A)=0.3$$

$$P(B)=0.2 \to P(B')=1-0.2=0.8$$

سه پیشامد مستقل

سه پیشامد A، B و C مفروض هستند:

$$(1)\begin{cases}P(A\cap B)=P(A).P(B)\\P(A\cap C)=P(A).P(C)\\P(B\cap C)=P(B).P(C)\end{cases}$$

$$(2)\ P(A\cap B\cap C)=P(A).P(B).P(C)$$

✓ دقت کنید!

هرگاه روابط (1) و (2) همزمان برقرار باشند، سه پیشامد A، B و C مستقل هستند.

هرگاه فقط رابطه (1) برقرار باشد، سه پیشامد A، B و C دوبه‌دو مستقل هستند.

مثال ۴ شرط استقلال سه واقعه A و B و C تعریف شده در یک فضای نمونه از یکدیگر چیست؟ (اقتصاد ـ ۷۹)

۱) $P(A\cap B\cap C)=P(A).P(B).P(C)$

۲) $\begin{matrix}P(A\cap B)=P(A).P(B)\\P(A\cap C)=P(A).P(C)\\P(B\cap C)=P(B).P(C)\end{matrix}$

۳) $\begin{matrix}P(A\cap B\cap C)=P(A).P(B).P(C)\\P(A\cap B)=P(A).P(B)\\P(A\cap C)=P(A).P(C)\\P(B\cap C)=P(B).P(C)\end{matrix}$

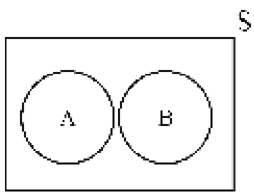
۴) یکی از دو شرط ۱ یا ۲ جاری باشد.

حل: گزینه ۳ درست است.

پیشامدهای ناسازگار (Exclusive Events)

تعریف: هرگاه وقوع همزمان دو پیشامد **غیر ممکن** باشد، پیشامدها را ناسازگار گویند.

درنتیجه:



$$\text{ناسازگار } A\,,\,B\ \Leftrightarrow\begin{cases}A\cap B=\phi\ \rightarrow\ P(A\cap B)=0\\P(A'\cap B)=P(B)\\P(B'\cap A)=P(A)\end{cases}$$

نکته: پیشامدهای مکمل، ناسازگارند و داریم:

$$A\cap A'=\{\varnothing\}\longrightarrow P(A\cap A')=0$$

هرگاه دو پیشامد A و B ناسازگار باشند، آن‌گاه بدیهی است هر یک، زیرمجموعه مکمل دیگری است؛ به عبارت دیگر:

$$\text{ناسازگار } A\,,\,B\ \Rightarrow\begin{cases}A\subset B'\rightarrow\ P(A\cap B')=P(A)\\B\subset A'\rightarrow\ P(A'\cap B)=P(B)\end{cases}$$

رابطه پیشامدهای ناسازگار و وابسته

هرگاه دو پیشامد تصادفی ناسازگار باشند، بدیهی است که وابسته‌اند زیرا اگر دو پیشامد نتوانند با هم اتفاق بیفتند حتماً به هم وابسته هستند.

ناسازگار $A , B : \ P(A \cap B) = 0 \neq \underbrace{P(A)}_{>0} \times \underbrace{P(B)}_{>0} \ \rightarrow$ وابسته‌اند A , B

یادآوری: احتمال هر پیشامد تصادفی، حتماً عددی بین صفر و یک است.

$0 < P(A) < 1 \ , 0 < P(B) < 1$

مثال $P(A) = 0.1$ و $P(B) = 0.2$ و $P(A \cap B) = 0$ باشد، A و B نسبت به هم چگونه‌اند؟

۱) مستقل‌اند ۲) ناسازگارند ۳) وابسته‌اند ۴) ۲ و ۳

حل: گزینه ۴ درست است.

$P(A \cap B) = 0 \ \rightarrow$ ناسازگارند B , A $\rightarrow$ وابسته نیز هستند. B , A

رابطه پیشامدهای ناسازگار و مستقل

اگر یکی از دو پیشامد مستقل، غیرممکن باشد آن دو پیشامد، ناسازگار و درنتیجه وابسته می‌شوند.

برای مثال، در پرتاب یک سکه و ریختن یک تاس به صورت همزمان، دو پیشامد «A : شیر آمدن سکه» و «$B : 7$ آمدن تاس» مستقل‌اند، درنتیجه داریم:

$$P(A \cap B) = P(A)P(B)$$

اما ازآنجاکه پیشامد 7 آمدن تاس غیرممکن است $(P(B) = 0)$، داریم:

$$P(A \cap B) = P(A)P(B) = P(A) \times 0 = 0$$

که در این صورت دو پیشامد A و B ناسازگار و در نتیجه، وابسته‌اند.

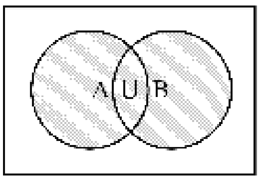
احتمال اجتماع دو پیشامد

احتمال اجتماع دو پیشامد A و B به صورت $P(A \cup B)$ می‌تواند یکی از مفاهیم زیر را داشته باشد:

الف) حداقل یکی از پیشامدهای A و B اتفاق بیفتد.

ب) وقوع پیشامد A یا B

ج) حادثه متأثر از پیشامد A و B اتفاق بیفتد.



روش محاسبه

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \begin{cases} \xrightarrow{\text{A و B مستقل}} P(A \cup B) = P(A) + P(B) - P(A).P(B) \\ \xrightarrow{\text{A و B ناسازگار}} P(A \cup B) = P(A) + P(B) \end{cases}$$

نکته: ازآنجاکه **پیشامدهای مکمل** هم **ناسازگار** $(P(A \cap A') = 0)$ و هم **گروه کامل حوادث** $(P(A \cup A') = 1)$ هستند، بنابراین:

$$P(A) + P(A') = 1 \quad \rightarrow \quad \begin{cases} P(A) = 1 - P(A') \\ P(A') = 1 - P(A) \end{cases}$$

مثال ۱ اگر A و B دو پیشامد مستقل و $P(A)=0.4$ و $P(A\cap B)=0.2$ باشد، احتمال اجتماع این دو پیشامد چقدر است؟

۱) 0.6 ۲) 0.8 ۳) 0.7 ۴) 0.9

حل: گزینه ۳ درست است.

$$\begin{cases} P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.4+0.5-0.2=0.7 \\ \text{مستقلاند } B,A \rightarrow P(A\cap B)=P(A)P(B)\rightarrow 0.2=0.4P(B)\rightarrow P(B)=0.5 \end{cases}$$

مثال ۲ اگر $P(A\cup B)=0.7$ و $P(A)=0.3$ باشد، در صورتی که A و B دو پیشامد ناسازگار باشند، احتمال وقوع B کدام است؟

۱) 0.4 ۲) 0.7 ۳) 0.21 ۴) 0.3

حل: گزینه ۱ درست است.

$$\text{ناسازگارند } B, A \rightarrow P(A\cup B)=P(A)+P(B) \quad\rightarrow\quad 0.7=0.3+P(B) \quad\rightarrow\quad P(B)=0.4$$

قوانین دمورگان

$$1)\ P\left(\bigcup_{i=1}^{n} A'_i\right)=P\left(\bigcap_{i=1}^{n} A_i\right)'$$

$$2)\ P\left(\bigcap_{i=1}^{n} A'_i\right)=P\left(\bigcup_{i=1}^{n} A_i\right)'$$

بنابراین روابط زیر همواره برقرار هستند:

$$\begin{cases} P(A'\cap B')=P(A\cup B)'=1-P(A\cup B) \\ P(A'\cup B')=P(A\cap B)'=1-P(A\cap B) \end{cases}$$

مثال ۳ عددی را به تصادف از فضای نمونه $\{1, 2, 3, \dots, 9\}$ انتخاب می‌کنیم. احتمال آنکه عدد انتخاب‌شده زوج یا مضرب 3 باشد، کدام است؟

۱) $\frac{1}{3}$ ۲) $\frac{2}{3}$ ۳) $\frac{2}{9}$ ۴) $\frac{5}{9}$

حل: گزینه ۲ درست است.

حرف اضافه «**یا**» بین دو پیشامد به معنی اجتماع آن‌هاست.

یک عدد از اعداد $1, 2, \dots, 9$ انتخاب کرده‌ایم، حال می‌خواهیم عدد انتخاب‌شده، زوج یا مضرب 3 باشد:

$$\begin{cases} P(\text{عدد زوج یا مضرب } 3)=P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{4}{9}+\frac{3}{9}-\frac{1}{9}=\frac{6}{9}=\frac{2}{3} \\ A: \text{زوج}=\{2, 4, 6, 8\} \quad\rightarrow\quad P(A)=\frac{4}{9} \\ B: \text{مضرب } 3=\{3, 6, 9\} \quad\rightarrow\quad P(B)=\frac{3}{9} \\ A\cap B: \text{زوج و مضرب } 3=\{6\} \quad\rightarrow\quad P(A\cap B)=\frac{1}{9} \end{cases}$$

مثال ۴ احتمال اینکه حسن یک مسئله ریاضی را حل کند، 0.4 و احتمال اینکه حسین آن را حل کند، 0.5 است. احتمال اینکه مسئله حل شود برابر است با:

۱) 0.2 ۲) 0.7 ۳) 0.8 ۴) 0.9

حل: گزینه ۲ درست است.

$$P(\text{مسئله حل شود}) = P(\text{حداقل یکی از دو نفر مسئله را حل کنند}) = P(\text{حسن یا حسین مسئله را حل کنند})$$

$$P(A\cup B) = P(A) + P(B) - P(A\cap B) = 0.4 + 0.5 - 0.4 \times 0.5 = 0.7$$

ازآنجاکه حل مسئله توسط دو نفر، مستقل از یکدیگر است، پس:

$$P(A\cap B) = P(A).P(B)$$

مثال ۵ اگر A و B دو پیشامد ناسازگار باشند و احتمال آن دو به ترتیب a و b باشد، مقدار $P\left(A^c \cap B^c\right)$ کدام است؟ (A^c متمم A است.)

۱) $b-a$ ۲) $1-ab$ ۳) $1-a-b$ ۴) $1+a+b$

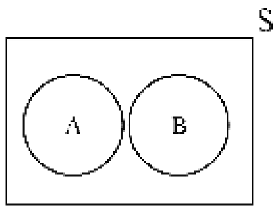
حل: گزینه ۳ درست است.

$$\begin{cases} P(A'\cap B') = P(A\cup B)' = 1 - P(A\cup B) = 1 - \left(P(A) + P(B)\right) = 1 - a - b \\ P(A\cup B) = P(A) + P(B) - P(A\cap B) = P(A) + P(B) \\ \text{ناسازگار } B\,,\,A \rightarrow P(A\cap B) = 0 \\ P(A) = a\,,\, P(B) = b \end{cases}$$

مثال ۶ اگر A و B دو پیشامد ناسازگار و $P(A) = 0.4$ و $P(B') = 0.5$ باشد، کدام‌یک از موارد زیر درست نیست؟

۱) $P(A\cup B) = 0.9$ ۲) $P(A'\cup B) = 0.6$ ۳) $P(A') = 0.6$ ۴) $P(A'\cap B') = 0$

حل: گزینه ۴ درست است.



$$\text{ناسازگارند } A,B \leftrightarrow \begin{cases} P(A\cap B) = 0 \\ P(A'\cap B) = P(B) \\ P(B'\cap A) = P(A) \end{cases}$$

$$\text{(گزینه ۱)} \begin{cases} P(A\cup B) = P(A) + P(B) - \overset{0}{\cancel{P(A\cap B)}} = 0.4 + 0.5 = 0.9 \checkmark \\ P(B') = 0.5 \rightarrow P(B) = 1 - P(B') = 1 - 0.5 = 0.5 \end{cases}$$

$$\text{(گزینه ۲)}\, P(A'\cup B) = P(A') + P(B) - \underset{P(B)}{\cancel{P(A'\cap B)}} = P(A') = 0.6 \checkmark$$

$$\text{(گزینه ۳)}\, P(A') = 1 - P(A) = 1 - 0.4 = 0.6 \checkmark$$

$$\text{(گزینه ۴)}\, P(A'\cap B') = P(A\cup B)' = 1 - P(A\cup B) = 1 - 0.9 = 0.1 \times$$

قاعده کلی برای اجتماع پیشامدها

اگر A_1، A_2، ...، A_n پیشامدهای دلخواه باشند، آن‌گاه:

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum P(A_i)-\sum P(A_i \cap A_j)+\sum P(A_i \cap A_j \cap A_k)-\cdots(-1)^{n+1} P(A_1 \cap A_2 \ldots \cap A_n)$$

برای مثال، برای سه پیشامد A، B و C داریم:

$$P(A \cup B \cup C)=P(A)+P(B)+P(C)-P(A \cap B)-P(A \cap C)-P(B \cap C)+P(A \cap B \cap C)$$

اجتماع و ناسازگاری

هرگاه پیشامدهای $A_1, A_2, \ldots, A_k$ دو به دو ناسازگار باشند، آن‌گاه:

$$P(A_1 \cup A_2 \ldots \cup A_k)=P(A_1)+P(A_2)+\ldots+P(A_k) \quad \rightarrow \quad P\left(\bigcup_{i=1}^{k} A_i\right)=\sum_{i=1}^{k} P(A_i)$$

اجتماع و استقلال

هرگاه پیشامدهای $A_1, A_2, \ldots, A_k$ از هم مستقل باشند، آن‌گاه:

$$P(A_1 \cup A_2 \ldots \cup A_k)=1-P(A_1 \cup A_2 \ldots \cup A_k)'=1-P(A'_1 \cap A'_2 \ldots \cap A'_k)=1-P(A'_1)P(A'_2)\ldots P(A'_k)$$

مثال ۱ احتمال به صدا درآمدن هر یک از سه آژیر خطر مستقلی که در یک فروشگاه نصب شده‌اند، به هنگام آتش‌سوزی برابر 0.95 است. احتمال آنکه به هنگام بروز آتش‌سوزی حداقل یکی از سه آژیر خطر به صدا در آید، چقدر است؟ (اقتصاد ـ ۸۲)

۱) 0.15 ۲) 0.95^3 ۳) $1-(0.05)^3$ ۴) $1-(0.95)^3$

حل: گزینه ۳ درست است.

درصورتی‌که احتمال به صدا درآمدن هر یک از سه آژیر $P(A_i)=0.95$ باشد، احتمال به صدا درنیامدن هرکدام $P(A'_i)=0.05$ است، بنابراین:

$$P(\text{حداقل یک آژیر به صدا درآید})=1-P(\text{هیچ‌کدام از سه آژیر به صدا درنیاید})$$

$$P(A_1 \cup A_2 \cup A_3)=1-P(A'_1 \cap A'_2 \cap A'_3)=1-P(A'_1)P(A'_2)P(A'_3)=1-(0.05)^3$$

مثال ۲ احتمال زنده بودن یک زن و شوهر در 20 سال آینده به ترتیب $\frac{3}{5}$ و $\frac{1}{2}$ است. احتمال اینکه در این مدت دست‌کم یکی از آن‌ها زنده بماند، چقدر است؟ (مدیریت ـ ۸۰)

۱) 0.8 ۲) 0.2 ۳) 0.6 ۴) 0.3

حل: گزینه ۱ درست است.

اولاً: پیشامدهای زنده‌ماندن زن و شوهر مستقل از یکدیگر هستند.

ثانیاً: احتمال دست‌کم یا حداقل، همان احتمال اجتماع دو پیشامد است.

$$\begin{cases} \text{احتمال زنده ماندن زن} = P(A) = \dfrac{3}{5} & \rightarrow \quad \text{احتمال زنده نماندن زن} = P(A') = \dfrac{2}{5} \\ \text{احتمال زنده ماندن مرد} = P(B) = \dfrac{1}{2} & \rightarrow \quad \text{احتمال زنده نماندن مرد} = P(B') = \dfrac{1}{2} \end{cases}$$

$$P(A\cup B) = P(\text{حداقل یکی از دو نفرزنده بمانند}) = 1 - P(\text{هر دو بمیرند})$$

$$P(A\cup B) = 1 - P(A\cup B)' = 1 - P(A'\cap B') = 1 - P(A')\times P(B') = 1 - \frac{2}{5}\times\frac{1}{2} = 0.8$$

مثال ۳ سه نفر A و B و C به ترتیب با احتمال 0.4 و 0.7 و 0.5 یک مسئله را حل می‌کنند. مطلوب است محاسبه احتمال آنکه:

الف) فقط یکی مسئله را حل کند.

ب) مسئله حل شود.

حل:

الف) احتمال فقط یکی یعنی یک نفر مسئله را حل کند و دو نفر دیگر حل نکنند.

$$P(\text{B و C حل نکنند، A حل کند}) + P(\text{A و C حل نکنند، B حل کند}) + P(\text{A و B حل نکنند، C حل کند}) =$$

$$P(A\cap B'\cap C') \quad + \quad P(A'\cap B\cap C') \quad + \quad P(A'\cap B'\cap C) =$$

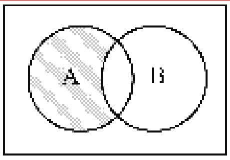
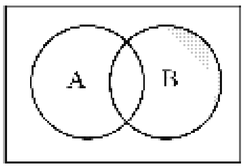
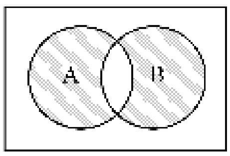
$$(0.4\times 0.3\times 0.5) \quad + \quad (0.6\times 0.7\times 0.5) \quad + \quad (0.6\times 0.3\times 0.5) = 0.36$$

ب) حل شدن مسئله به این معناست که حداقل یکی مسئله را حل کند، بنابراین از مکمل آن استفاده می‌کنیم.

$$P(A\cup B\cup C) = 1 - P(A'\cap B'\cap C') = 1 - P(A')P(B')P(C') = 1 - 0.6\times 0.3\times 0.5 = 0.91$$

تفاضل دو پیشامد

تفاضل دو پیشامد همواره وقوع فقط یکی از پیشامدها را مطرح می‌کند، به شکل‌های زیر توجه کنید:

 (A -B)	احتمال وقوع فقط A (A باشد، B نباشد): $P(A-B) = P(A) - P(A\cap B) = P(A\cap B')$
 (B -A)	احتمال وقوع فقط B (B باشد، A نباشد): $P(B-A) = P(B) - P(A\cap B) = P(B\cap A')$
 (A Δ B)=(A-B)+(B-A)	احتمال وقوع فقط یکی از دو پیشامد A و B (تفاضل متقارن): $P(A\Delta B) = P(A-B) + P(B-A) = P(A\cap B') + P(B\cap A')$ $= P(A\cup B) - P(A\cap B)$

نکته: اگر دو پیشامد مستقل باشند، آن‌گاه:

$$P(A-B)=P(A\cap B')=P(A)\times P(B')$$

$$P(B-A)=P(B\cap A')=P(B)\times P(A')$$

$$P(A\Delta B)=P(A\cap B')+P(B\cap A')=P(A)\times P(B')+P(B)\times P(A')$$

نکته: اگر دو پیشامد ناسازگار باشند، آن‌گاه ازآنجاکه $A\subset B'$, $B\subset A'$ است، داریم:

$$P(A-B)=P(A\cap B')=P(A)$$

$$P(B-A)=P(B\cap A')=P(B)$$

$$P(A\Delta B)=P(A\cap B')+P(B\cap A')=P(A)+P(B)$$

مثال ۱ اگر $P(A)=0.59$ و $P(B)=0.3$ و $P(A\cap B)=0.21$، آن‌گاه مقدار $P(A\cap \overline{B})$ کدام است؟

۱) 0.56 ۲) 0.38 ۳) 0.28 ۴) 0.18

حل: گزینه ۲ درست است.

با توجه به آنکه $P(A\cap B)\neq P(A)\times P(B)$ است دو پیشامد مستقل نیستند، بنابراین از تعریف تفاضل دو پیشامد استفاده می‌کنیم.

$$\begin{cases} P(A\cap B')=P(A-B)=P(A)-P(A\cap B)=0.59-0.21=0.38 \\ P(A)=0.59\ ,\ P(A\cap B)=0.21 \end{cases}$$

مثال ۲ احتمال کار کردن یک نفر در شرکتی برای مدت بیشتر از 10 سال برابر $\frac{1}{6}$ است. اگر دو فرد A و B کار خود را همزمان در این شرکت شروع کنند احتمال اینکه فقط یک نفر از آن‌ها بیشتر از 10 سال در شرکت بماند کدام است؟

۱) $\frac{11}{36}$ ۲) $\frac{5}{36}$ ۳) $\frac{25}{36}$ ۴) $\frac{5}{18}$

حل: گزینه ۴ درست است.

احتمال کار کردن هر فرد در شرکت، مستقل از دیگری است؛ یعنی A و B مستقل‌اند.

$$P(\text{فقط یک نفر})=P(\text{فقط }A)+P(\text{فقط }B)=P(A-B)+P(B-A)=P(A\cap B')+P(B\cap A')$$

$$=P(A)P(B')+P(B)P(A')=\frac{1}{6}\times\frac{5}{6}+\frac{1}{6}\times\frac{5}{6}=\frac{10}{36}=\frac{5}{18}$$

نتیجه کلی:

پیشامد	احتمال
وقوع هر دو پیشامد A و B	$P(A\cap B)$
وقوع حداقل یکی از دو پیشامد A یا B	$P(A\cup B)$
عدم وقوع حداقل یکی از دو پیشامد A یا B	$P(A'\cup B')$
وقوع هیچ‌کدام از دوپیشامد(نه A و نه B)	$P(A'\cap B')$
وقوع فقط پیشامد A (A و نه B)	$P(A-B)=P(A\cap B')$
وقوع فقط پیشامد B (B و نه A)	$P(B-A)=P(A'\cap B)$
وقوع فقط یکی از دو پیشامد (فقط B یا فقط A)	$P(B-A)+P(A-B)$

مثال ۱ در دانشکده‌ای، 50% دانشجویان فوتبال، 40% بسکتبال و 30% هم فوتبال و هم بسکتبال بازی می‌کنند. احتمال آنکه دانشجویی در این دانشکده ورزش نکند، کدام است؟

۱) 0 ۲) 0.1 ۳) 0.4 ۴) 0.6

حل: گزینه ۳ درست است.

$P(\text{دانشجویی ورزش نکند}) = P(\text{نه فوتبال و نه بستکبال}) = P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B) = ?$

$\text{فوتبال} : A \to P(A) = 0.5$

$\text{بسکتبال} : B \to P(B) = 0.4$

$\text{هم فوتبال و هم بسکتبال} : A \cap B \to P(A \cap B) = 0.3$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.4 - 0.3 = 0.6$

$P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B) = 1 - 0.6 = 0.4$

مثال ۲ در یک کارخانه قرار است یک نفر استخدام شود. 40% افراد مراجعه‌کننده حداقل مدرک کارشناسی دارند، 60% از آن‌ها سابقه کار دارند و 15% از آن‌ها حداقل مدرک کارشناسی و سابقه کار نیز دارند. احتمال اینکه فردی که استخدام می‌شود سابقه کار نداشته باشد و حداقل مدرک کارشناسی داشته باشد، چقدر است؟

۱) %15 ۲) %20 ۳) %25 ۴) %60

حل: گزینه ۳ درست است.

$A : \text{حداقل مدرک کارشناسی} \rightarrow P(A) = 0.4$

$B : \text{سابقه کار} \rightarrow P(B) = 0.6$

$A \cap B : \text{حداقل کارشناسی و سابقه کار} \rightarrow P(A \cap B) = 0.15$

$P(\text{سابقه نداشته و مدرک داشته باشد}) = P(A \cap B') = P(A - B) = P(A) - P(A \cap B) = 0.4 - 0.15 = 0.25 = \%25$

مثال ۳ فرض کنید احتمال آنکه زن و شوهری در بیست سال آینده زنده بمانند به ترتیب 0.7 و 0.4 باشد. مطلوب است محاسبه احتمال آنکه:

الف) هردو زنده بمانند.

ب) هیچ‌کدام زنده نمانند.

ج) حداقل یکی زنده بماند (در بیست سال آینده شخصی زنده بماند).

د) فقط یکی زنده بماند.

ه) فقط زن زنده بماند.

حل:

الف) هر دو زنده بمانند یعنی زن زنده بماند و شوهر هم زنده بماند.

$A : \text{پیشامد زنده ماندن شوهر} \rightarrow P(A) = 0.4$

$B : \text{پیشامد زنده ماندن زن} \rightarrow P(B) = 0.7$

$P(A \cap B) = P(A) \times P(B) = 0.4 \times 0.7 = 0.28$

دقت کنید که پیشامد زنده ماندن زن و شوهر مستقل از یکدیگر است.

ب) هیچ‌کدام زنده نمانند یعنی نه زن و نه شوهر زنده نمانند.

$\text{مستقل } A , B \to \text{مستقل } A', B' \to P(A' \cap B') = P(A') \times P(B') = 0.6 \times 0.3 = 0.18$

ج) حداقل به معنی اجتماع دو پیشامد است.

راه حل اول:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.4+0.7-(0.4)(0.7)=0.82$$

راه حل دوم:

$$P(A\cup B)=1-P(A'\cap B')=1-P(\text{هیچ‌کدام زنده نمانند})=1-0.18=0.82$$

د)

راه حل اول:

$$P(A'\cap B)+P(A\cap B')=P(A').P(B)+P(A).P(B')=(0.6)(0.7)+(0.4)(0.3)=0.54$$

راه حل دوم:

$$P(A\Delta B)=P(A\cup B)-P(A\cap B)=0.82-0.28=0.54$$

یادآوری: (B,A) مستقل‌اند $\leftarrow$ (B,A') مستقل‌اند و (B',A) مستقل‌اند.

ه) $P(A'\cap B)=P(A').P(B)=(0.6)(0.7)=0.42$

مثال ۴ دانشگاهی در یکی از استان‌ها واقع شده است. $\frac{1}{3}$ دانشجویان آن دانشگاه خارج از خوابگاه دانشجویی زندگی می‌کنند. $\frac{5}{9}$ دانشجویان اهل آن استان‌اند و $\frac{3}{4}$ دانشجویان اهل آن استان نیستند و در خوابگاه دانشجویی زندگی می‌کنند. احتمال اینکه دانشجویی که به صورت تصادفی از این دانشگاه انتخاب می‌شود، اهل آن استان نباشد یا در خوابگاه زندگی کند، چقدر است؟

۱) $\frac{13}{36}$ ۲) $\frac{8}{27}$ ۳) $\frac{3}{4}$ ۴) $\frac{5}{36}$

حل: گزینه ۱ درست است.

$$A:\text{دانشجویان غیر خوابگاهی}\rightarrow P(A)=\frac{1}{3}\rightarrow P(A')=1-P(A)=1-\frac{1}{3}=\frac{2}{3}$$

$$B:\text{دانشجویان اهل استان}\rightarrow P(B)=\frac{5}{9}\rightarrow P(B')=1-P(B)=1-\frac{5}{9}=\frac{4}{9}$$

$$A'\cap B':\text{اهل استان نیستند و در خوابگاه زندگی می‌کنند}\rightarrow P(A'\cap B')=\frac{3}{4}$$

$$P(\text{اهل استان نباشد یا در خوابگاه زندگی کند})=P(A'\cup B')=P(A')+P(B')-P(A'\cap B')=\frac{2}{3}+\frac{4}{9}-\frac{3}{4}=\frac{13}{36}$$

دقت کنید، طبق قانون دمورگان داریم:

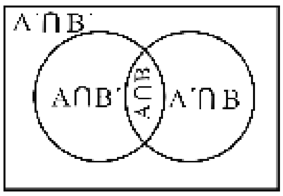
$$P(A'\cup B')=P(A\cap B)'=1-P(A\cap B)$$

اما چون در این سؤال از $P(A\cap B)$ هیچ صحبتی به میان نیامده، نمی‌توان از این رابطه استفاده کرد.

وضعیت‌های مختلف دو پیشامد

درصورتی‌که A و B دو پیشامد دلخواه باشند روابط زیر برای آن‌ها برقرار است:

	B	**B′**	
A	$P(A\cap B)$	$P(A\cap B')$	$\mathbf{P(A)}$
A′	$P(A'\cap B)$	$P(A'\cap B')$	$\mathbf{P(A')}$
	$\mathbf{P(B)}$	$\mathbf{P(B')}$	1



درنتیجه همواره داریم:

$$\begin{cases} P(A) = P(A \cap B) + P(A \cap B') \\ P(B) = P(B \cap A) + P(B \cap A') \\ P(A') = P(A' \cap B) + P(A' \cap B') \\ P(B') = P(B' \cap A) + P(B' \cap A') \end{cases}$$

مثال ۱ جدول زیر مفروض است؛ $P(A)$ کدام است؟

	B	B'
A	0.2	?
A'	0.3	0.1

۱) 0.6 ۲) 0.3

۳) 0.4 ۴) 0.2

حل: گزینه ۱ درست است.

با توجه به آنکه مجموع احتمال‌ها باید برابر 1 باشد، داریم:

	B	B'	
A	0.2 +	$\boxed{0.4}$	0.6
A'	0.3	0.1	

$$0.2+0.3+?+0.1=1 \quad \to \quad ?=0.4$$

$$P(A)=P(A \cap B)+P(A \cap B')=0.2+0.4=0.6$$

نکته: اگر پیشامدهای A و B به صورتی باشند که یکی زیرمجموعه دیگری باشد، آن‌گاه:

$$A \subset B \to P(A)<P(B)$$
$$B \subset A \to P(B)<P(A)$$

مثال ۲ در آزمایشی حادثه A باعث وقوع حادثه B می‌گردد. کدام‌یک از عبارات زیر درباره احتمال‌های این حوادث صحیح است؟ (مدیریت ـ ۷۱)

۱) $P(A)>P(B)$ ۲) $P(A) \le P(B)$ ۳) $P(A) \ne P(B)$ ۴) $P(A) \ge P(B)$

حل: گزینه ۲ درست است.

چون حادثه A باعث وقوع حادثه B شده است پس:

$$A \subset B \quad \to \quad P(A) \le P(B)$$

کران های $P(A \cap B)$ و $P(A \cup B)$

برای آنکه مشخص کنیم هر یک از احتمال‌های $P(A \cap B)$ و $P(A \cup B)$ در چه محدوده‌ای قرار دارند، می‌توانیم از قوانین زیر استفاده کنیم:

قانون۱:

$$P(A)+P(B)-1 \le P(A \cap B) \le \underbrace{P(A),P(B)}_{\min(P(A),P(B))}$$

دقت کنید که:

$$P(A\cup B)\le 1 \;\rightarrow\; P(A)+P(B)-P(A\cap B)\le 1 \;\rightarrow\; P(A\cap B)\ge P(A)+P(B)-1$$

قانون۲:

$$\underbrace{P(A),P(B)}_{\max(P(A),P(B))}\le P(A\cup B)\le P(A)+P(B)$$

مثال اگر $P(A)=0.4$ و $P(B)=0.7$ باشد، حداقل مقدار $P(A\cap B)$ کدام است؟

۱) 0 ۲) 0.35 ۳) 0.1 ۴) 0.3

حل: گزینه ۳ درست است.

با توجه به قانون اول داریم:

$$P(A\cap B)\ge P(A)+P(B)-1 \quad\rightarrow\quad P(A\cap B)\ge 0.4+0.7-1=0.1$$

مسایل مهم احتمال

مهم‌ترین مسایل مربوط به احتمال عبارت‌اند از:

پرتاب تاس

پرتاب سکه

پرتاب تاس و سکه

مسئله مهره‌ها

مدارهای سری و موازی

مسئله کلاسیک روز تولد

پرتاب تاس

در پرتاب m تاس، تعداد عناصر فضای نمونه 6^m است.

مثال ۱ در پرتاب دوتاس، احتمال ظاهر شدن مجموع 6، چقدر است؟

حل:

حالات مساعد: $\{(1,5),(5,1),(4,2),(2,4),(3,3)\}$

حالات ممکن: 6^2

$P(A)=\frac{5}{36}$

توجه کنید که زوج $(3,3)$ باید یک بار نوشته شود.

مثال ۲ در پرتاب دو تاس، احتمال ظاهر شدن مجموع 6 و تاس اول کمتر از 3، چقدر است؟

حل: توجه کنید که «و» به معنی اشتراک دو مجموعه است.

حالات مساعد: $\{(1,5),(2,4)\}$

حالات ممکن: 6^2

$P(A)=\frac{2}{36}$

مثال ۳ در پرتاب دو تاس، احتمال آنکه مجموع کمتر از 5 باشد و یکی از تاس‌ها کمتر از 2 باشد، چقدر است؟

حل:

حالات مساعد: $\{(1,1),(1,2),(1,3),(2,1),(3,1)\}$

حالات ممکن: 6^2

$P(A)=\frac{5}{36}$

مثال ۴ در پرتاب سه تاس، احتمال آنکه نتایج متفاوت باشند، چقدر است؟

حل: $\frac{6}{6}\times\frac{5}{6}\times\frac{4}{6}=\frac{20}{36}$

مثال ۵ در پرتاب سه تاس، احتمال آنکه تاس اول و سوم یکسان و با تاس دوم متفاوت باشند، چقدر است؟

حل: $\frac{6}{6}\times\frac{5}{6}\times\frac{1}{6}=\frac{5}{36}$

نکته: در پرتاب چند تاس، اگر بخواهیم در مراحل مختلف نتایج یکسان داشته باشیم، کافی است مرحله اول را محاسبه کرده و به‌جای مراحل یکسان دیگر، 1 بگذاریم.

مثال ۶ در پرتاب چهار تاس، احتمال آنکه تاس اول عدد 4 باشد، چقدر است؟

حل:

$$\frac{1}{6}\times\frac{6}{6}\times\frac{6}{6}\times\frac{6}{6}=\frac{1}{6}$$

مثال ۷ در پرتاب چهار تاس، احتمال آنکه تاس اول عدد 4 باشد و تاس دوم و چهارم برابر باشند، چقدر است؟

حل:

$$\frac{1}{6}\times\frac{6}{6}\times\frac{6}{6}\times\frac{1}{6}=\frac{1}{36}$$

مثال ۸ در پرتاب چهار تاس، احتمال آنکه تاس اول و سوم یکسان و تاس دوم و چهارم یکسان و عدد دو تاس اول و سوم و دو تاس دوم و چهارم با هم متفاوت باشند، چقدر است؟

حل:

$$\frac{6}{6}\times\frac{5}{6}\times\frac{1}{6}\times\frac{1}{6}=\frac{5}{216}$$

مثال ۹ در پرتاب پنج تاس، چقدر احتمال دارد:

الف) همه شماره‌ها فرد باشند.

ب) شماره‌های اول و سوم و پنجم یکسان باشند.

ج) فقط شماره‌های اول و سوم و پنجم یکسان باشند.

د) فقط سه شماره یکسان باشند.

ه) همه شماره‌ها متفاوت باشند.

حل:

الف)

$$\frac{3}{6}\times\frac{3}{6}\times\frac{3}{6}\times\frac{3}{6}\times\frac{3}{6}=\frac{1}{2^5}=\frac{1}{32}$$

ب)

$$\underset{\substack{\downarrow\\ \text{اول}}}{\frac{6}{6}}\times\frac{6}{6}\times\underset{\substack{\downarrow\\ \text{سوم}}}{\frac{1}{6}}\times\frac{6}{6}\times\underset{\substack{\downarrow\\ \text{پنجم}}}{\frac{1}{6}}=\frac{1}{36}$$

ج)

$$\underset{\substack{\downarrow\\ \text{اول}}}{\frac{6}{6}}\times\frac{5}{6}\times\underset{\substack{\downarrow\\ \text{سوم}}}{\frac{1}{6}}\times\frac{4}{6}\times\underset{\substack{\downarrow\\ \text{پنجم}}}{\frac{1}{6}}=\frac{20}{6^4}$$

د) چون در سؤال مشخص نشده است کدام شماره‌ها یکسان باشند، ابتدا با استفاده از ترکیب $\binom{5}{3}$،تعداد حالات مختلف سه تاس از پنج تاس را که می‌خواهیم یکسان باشند، در نظر می‌گیریم.

$$\binom{5}{3}\times\frac{6}{6}\times\frac{5}{6}\times\frac{1}{6}\times\frac{4}{6}\times\frac{1}{6}=\frac{200}{6^4}$$

ه)

$$\frac{6}{6}\times\frac{5}{6}\times\frac{4}{6}\times\frac{3}{6}\times\frac{2}{6}=\frac{120}{6^4}$$

پرتاب سکه

در پرتاب n سکه، فضای نمونه 2^n است.

مثال ۱ در پرتاب 2 سکه، احتمال ظاهر شدن نتایج یکسان، چقدر است؟

حل:

{(خ و خ) ، (ش و ش)} : حالات مساعد

2^2 : حالات ممکن

$$P(\text{نتایج یکسان})=\frac{2}{4}=\frac{1}{2}$$

مثال ۲ در پرتاب 3 سکه، احتمال ظاهر شدن حداقل یک خط، چقدر است؟

حل:

{(ش و ش و ش)} : مکمل حالات مساعد

2^3 : حالات ممکن

$$P(\text{حداقل یک خط})=1-P(\text{همه شیر})=1-\frac{1}{8}=\frac{7}{8}$$

مثال ۳ در پرتاب 4 سکه، احتمال آنکه سکه اول خط ظاهر شود، چقدر است؟

حل:

$$P(\text{سکه اول خط})=\frac{1}{2}\times\frac{2}{2}\times\frac{2}{2}\times\frac{2}{2}=\frac{1}{2}$$

مثال ۴ در پرتاب 2 سکه، احتمال آنکه نتایج متفاوت باشد، چقدر است؟

حل:

{(ش و خ) ، (خ و ش)} : حالات مساعد

2^2 : حالات ممکن

$$P(\text{نتایج متفاوت})=\frac{2}{4}=\frac{1}{2}$$

پرتاب تاس و سکه

توجه کنید که تاس و سکه مستقل از هم بررسی می‌شوند.

مثال ۱ در پرتاب یک تاس و یک سکه، احتمال آنکه تاس، 5 و سکه، خط ظاهر شود، چقدر است؟

حل:

$$P(\text{تاس 5 و سکه خط})=P(\text{تاس 5})\times P(\text{سکه خط})=\frac{1}{6}\times\frac{1}{2}=\frac{1}{12}$$

مثال ۲ در پرتاب 3 تاس و یک سکه، احتمال آنکه سکه، خط بیاید و تاس اول و سوم یکسان ظاهر شوند، چقدر است؟

حل:

$$\frac{1}{2}\times\underbrace{\frac{6}{6}}_{\text{اول}}\times\frac{6}{6}\times\underbrace{\frac{1}{6}}_{\text{سوم}}=\frac{1}{12}$$

مسئله مهره‌ها

مثال ۱ کیسه‌ای دارای 10 مهره از شماره 1 تا 10 است. مهره‌ای از این کیسه انتخاب می‌کنیم.مطلوب است احتمال آنکه:

الف) مهره 4 باشد.

ب) مهره یک عدد بین 1 تا 10باشد.

ج) مهره زوج باشد.

د) مهره زوج و کمتر از 6 باشد.

حل:

الف)

$\{4\}$:حالات مساعد

$\{1, 2, \dots, 10\}$:حالات کل

$$P(A)=\frac{1}{10}$$

ب)

$\{2,3,4,5,6,7,8,9\}$: حالات مساعد

$\{1, 2, \dots, 10\}$:حالات کل

$$P(A)=\frac{8}{10}$$

ج)

$\{2, 4, 6, 8, 10\}$:حالات مساعد

$\{1, 2, \dots, 10\}$:حالات کل

$$P(A)=\frac{5}{10}$$

د)

$\{2, 4\}$:حالات مساعد

$\{1, 2, \dots, 10\}$:حالات کل

$$P(A)=\frac{2}{10}$$

مثال ۲ کیسه‌ای حاوی 4 مهره قرمز و 5 مهره آبی است. مهره‌ای از آن خارج می‌کنیم، احتمال آنکه قرمز باشد، چقدر است؟

حل:

$$P(\text{قرمز})=\frac{\text{تعداد مهره‌های قرمز}}{\text{تعداد کل مهره‌ها}}=\frac{4}{9}$$

مثال ۳ کیسه‌ای حاوی 4 مهره قرمز و 5 مهره آبی است. 2 مهره آبی از آن خارج می‌کنیم، سپس مهره دیگری خارج می‌کنیم؛ احتمال آنکه مهره قرمز باشد، چقدر است؟

حل:

آبی	قرمز
5	4

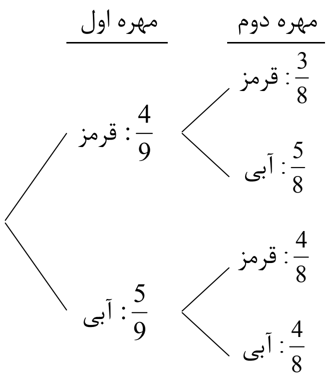
2 مهره آبی خارج می‌کنیم ⟶

آبی	قرمز
3	4

$\rightarrow P(\text{مهره قرمز}) = \frac{4}{7}$

مثال ۴ کیسه‌ای حاوی 4 مهره قرمز و 5 مهره آبی است. یک مهره از آن خارج می‌کنیم، سپس مهره دیگری از آن خارج می‌کنیم؛ احتمال آنکه مهره دوم خارج‌شده قرمز باشد، چقدر است؟

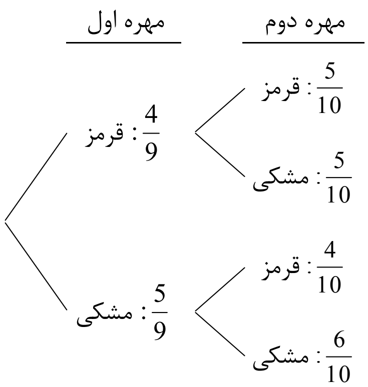
حل:



$$\Rightarrow P(\text{قرمز}) = \frac{4}{9}\times\frac{3}{8}+\frac{5}{9}\times\frac{4}{8}=\frac{32}{72}=\frac{4}{9}$$

مثال ۵ کیسه‌ای حاوی 4 مهره قرمز و 5 مهره مشکی است. یک مهره از آن خارج می‌کنیم و به همراه یک مهره همرنگ آن دوباره داخل ظرف می‌گذاریم، سپس مهره‌ای خارج می‌کنیم. احتمال آنکه این مهره مشکی باشد چقدر است؟

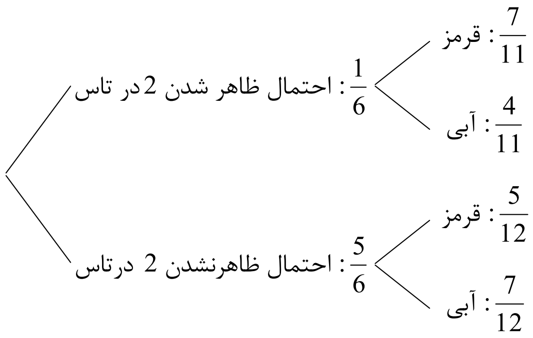
حل:



$$\Rightarrow P(\text{مشکی}) = \frac{4}{9}\times\frac{5}{10}+\frac{5}{9}\times\frac{6}{10}=\frac{50}{90}=\frac{5}{9}$$

مثال ۶ کیسه‌ای شامل 5 مهره قرمز و 4 مهره آبی است. تاسی را پرتاب می‌کنیم؛ اگر 2 آمد، 2 مهره قرمز و در غیر این صورت 3 مهره آبی به کیسه اضافه می‌کنیم، سپس مهره‌ای از ظرف خارج می‌کنیم. احتمال آنکه این مهره قرمز باشد چقدر است؟

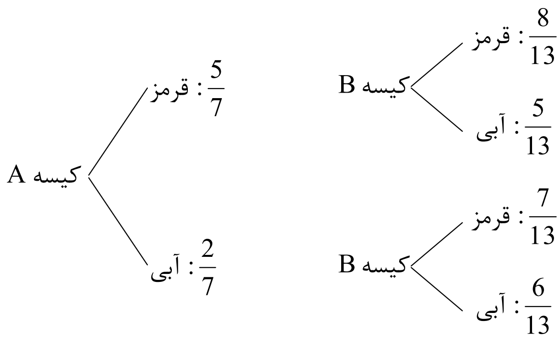
حل:



$$\Rightarrow P(\text{قرمز}) = \frac{1}{6}\times\frac{7}{11}+\frac{5}{6}\times\frac{5}{12}=\frac{359}{792}$$

مثال ۷ کیسه A شامل 5 مهره قرمز و 2 مهره آبی و کیسه B شامل 7 مهره قرمز و 5 مهره آبی است. مهره‌ای از کیسه A خارج کرده‌ایم و به کیسه B ریخته‌ایم، سپس مهره‌ای از کیسه B خارج می‌کنیم. احتمال آنکه این مهره قرمز باشد چقدر است؟

حل:



$$\Rightarrow P(\text{قرمز}) = \frac{5}{7}\times\frac{8}{13}+\frac{2}{7}\times\frac{7}{13}=\frac{54}{91}$$

انتخاب با جایگذاری و بدون جایگذاری

مثال ۱ در جعبه‌ای 3 مهره قرمز، 2 مهره سبز و 5 مهره سفید وجود دارد. اگر 3 مهره به تصادف انتخاب کنیم، مطلوب است احتمال آنکه: (انتخاب پیش‌فرض بدون جایگذاری است.)

الف) هر سه قرمز باشند.

ب) هر سه همرنگ باشد.

ج) مهره‌ها همرنگ نباشند.

د) حداقل یک مهره قرمز باشد.

حل:

الف)

$$\frac{\binom{3}{3}\binom{2}{0}\binom{5}{0}}{\binom{10}{3}}=\frac{1}{120}$$

ب)

$$\frac{\binom{3}{3}+\binom{5}{3}}{\binom{10}{3}}=\frac{11}{120}$$

ج)

$$\frac{\binom{3}{1}\binom{2}{1}\binom{5}{1}}{\binom{10}{3}}=\frac{30}{120}=\frac{1}{4}$$

د)

$$P(\text{حداقل یک قرمز})=1-P(\text{قرمز نداشته باشیم})=1-\frac{\binom{5+2}{3}}{\binom{10}{3}}=1-\frac{\binom{7}{3}}{\binom{10}{3}}$$

مثال ۲ در کلاسی 10 پسر و 5 دختر شرکت دارند. 3 دانشجو به طور تصادفی یکی پس از دیگری انتخاب می‌شوند. مطلوب است احتمال آنکه:

الف) دو دانشجوی اول پسر و سومی دختر باشد.

ب) اولی و سومی دختر باشد.

ج) اولی و سومی همجنس و دومی از جنس مخالف باشد.

حل:

الف)

$$\underbrace{\frac{10}{15}}_{\text{اولی پسر}}\times\underbrace{\frac{9}{14}}_{\text{دومی پسر}}\times\underbrace{\frac{5}{13}}_{\text{سومی دختر}}$$

ب)

$$\left(\underbrace{\frac{5}{15}}_{\text{اولی دختر}}\times\underbrace{\frac{10}{14}}_{\text{دومی پسر}}\times\underbrace{\frac{4}{13}}_{\text{سومی دختر}}\right)+\left(\underbrace{\frac{5}{15}}_{\text{اولی دختر}}\times\underbrace{\frac{4}{14}}_{\text{دومی دختر}}\times\underbrace{\frac{3}{13}}_{\text{سومی دختر}}\right)$$

ج)

$$\left(\underbrace{\frac{5}{15}}_{\text{اولی دختر}}\times\underbrace{\frac{10}{14}}_{\text{دومی پسر}}\times\underbrace{\frac{4}{13}}_{\text{سومی دختر}}\right)+\left(\underbrace{\frac{10}{15}}_{\text{اولی پسر}}\times\underbrace{\frac{5}{14}}_{\text{دومی دختر}}\times\underbrace{\frac{9}{13}}_{\text{سومی پسر}}\right)$$

مثال ۳ در جعبه‌ای 3 مهره قرمز، 2 مهره سبز و 5 مهره سفید وجود دارد. اگر 3 مهره به تصادف با جایگذاری انتخاب کنیم، مطلوب است احتمال آنکه: (دقت کنید که در مسایل با جایگذاری، ترتیب مهم است و مسئله را مرحله به مرحله حل می‌کنیم.)

الف) هر سه قرمز باشند. ب) هر سه همرنگ باشند.

ج) مهره‌ها همرنگ نباشند. د) دو مهره قرمز باشند.

ه) حداقل یک مهره قرمز باشد.

حل:

الف)

$$\underset{\text{قرمز}}{\frac{3}{10}} \times \underset{\text{قرمز}}{\frac{3}{10}} \times \underset{\text{قرمز}}{\frac{3}{10}} = \frac{27}{1000}$$

ب)

$$\underbrace{\frac{3}{10}\times\frac{3}{10}\times\frac{3}{10}}_{\text{هر سه قرمز}} + \underbrace{\frac{2}{10}\times\frac{2}{10}\times\frac{2}{10}}_{\text{هر سه سبز}} + \underbrace{\frac{5}{10}\times\frac{5}{10}\times\frac{5}{10}}_{\text{هر سه سفید}} = \frac{160}{1000} = 0.16$$

در انتخاب با جایگذاری می‌توان از هر رنگ هر تعداد مهره خارج کرد.

ج) انتخاب‌های مجاز (سفید، سبز، قرمز)، (سبز، سفید، قرمز)، ... (6 حالت)

$$\frac{3}{10}\times\frac{2}{10}\times\frac{5}{10} + \frac{3}{10}\times\frac{5}{10}\times\frac{2}{10} + ... = 3!\left(\frac{3}{10}\times\frac{2}{10}\times\frac{5}{10}\right) = \frac{180}{1000} = 0.18$$

د) انتخاب‌های مجاز (سفید یا سبز، قرمز، قرمز)، (قرمز، سفید یا سبز، قرمز)، ...(3 حالت)

$$\frac{3}{10}\times\frac{3}{10}\times\frac{7}{10} + ... = \frac{3!}{2!}\left(\frac{3}{10}\times\frac{3}{10}\times\frac{7}{10}\right) = \frac{189}{1000}$$

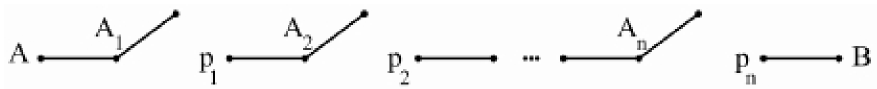
ه)

$$P(\text{حداقل یک قرمز}) = 1 - P(\text{قرمز نداشته باشیم}) = 1 - \frac{7}{10}\times\frac{7}{10}\times\frac{7}{10} = 1 - \left(\frac{7}{10}\right)^3$$

مدارهای سری و موازی

مدار سری

در مدارهای سری احتمال برقراری ارتباط بین دو نقطه برابر با احتمال برقراری همه اتصالات بین آن دو نقطه است، به عبارت دیگر:



فرض کنید p_i احتمال برقراری ارتباط در اتصال A_i باشد، آن‌گاه به دلیل مستقل بودن اتصالات داریم:

$$\begin{cases} P(\text{برقراری ارتباط بین A و B}) = P(A_1 \cap A_2 \cap \ldots \cap A_n) = p_1 \times p_2 \times \ldots \times p_n \\ P(\text{عدم برقراری ارتباط بین A و B}) = 1 - P(A_1 \cap A_2 \cap \ldots \cap A_n) = 1 - (p_1 \times p_2 \times \ldots \times p_n) \end{cases}$$

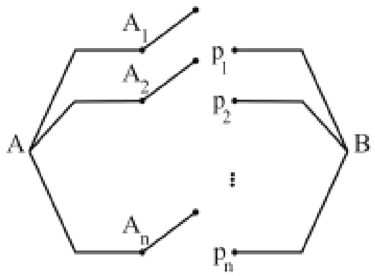
نکته: احتمال عدم برقراری ارتباط در مدارهای سری برابر با قطع حداقل یکی از اتصالات است، به عبارت دیگر:

$$P(\text{عدم برقراری ارتباط بین A و B}) = P(A_1' \cup A_2' \ldots A_n') = P(A_1 \cap A_2 \ldots A_n)' = 1 - P(A_1 \cap A_2 \ldots A_n)$$

نتیجه: در مدارهای سری بهتر است ابتدا احتمال برقراری ارتباط را محاسبه کنیم، سپس در صورت نیاز، احتمال به‌دست‌آمده را از یک کم کنیم تا بتوانیم احتمال عدم برقراری ارتباط را حساب کنیم.

مدار موازی

در مدارهای موازی احتمال عدم برقراری ارتباط بین دو نقطه برابر با احتمال قطع تمام مدارهای بین آن دو نقطه است، به عبارت دیگر:



فرض کنید p_i و $1-p_i$ به ترتیب احتمال برقراری و قطع ارتباط در اتصال A_i باشد، آن‌گاه به دلیل مستقل بودن اتصالات داریم:

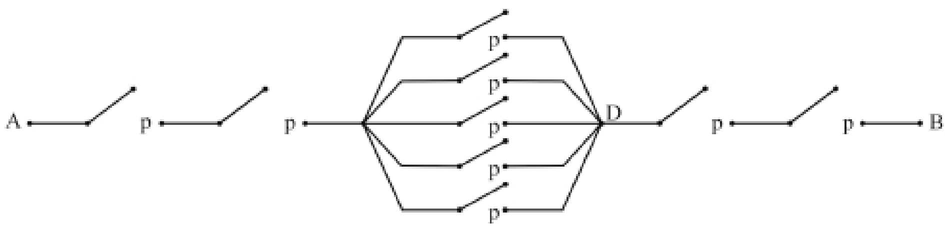
$$\begin{cases} P(\text{عدم برقراری ارتباط بین A و B}) = P(A'_1 \cap A'_2 \cap \ldots \cap A'_n) = (1-p_1)(1-p_2)\ldots(1-p_n) \\ P(\text{برقراری ارتباط بین A و B}) = 1 - P(A'_1 \cap A'_2 \cap \ldots \cap A'_n) = 1-(1-p_1)(1-p_2)\ldots(1-p_n) \end{cases}$$

نکته: احتمال برقراری ارتباط در مدارهای موازی برابر با احتمال برقراری حداقل یکی از اتصالات است، به عبارت دیگر:

$$P(\text{برقراری ارتباط بین A و B}) = P(A_1 \cup A_2 \ldots A_n) = 1 - P(A_1 \cup A_2 \ldots A_n)' = 1 - P(A'_1 \cap A'_2 \ldots A'_n)$$

نتیجه: در مدارهای موازی بهتر است ابتدا احتمال عدم برقراری ارتباط را محاسبه کنیم، سپس در صورت نیاز احتمال به‌دست‌آمده را از یک کم کنیم تا بتوانیم احتمال برقراری ارتباط را حساب کنیم.

مثال ۱ میان دو نقطه A و B که در شکل مشخص شده‌اند، ارتباط از طریق شبکه‌ای که احتمال وصل بودن هر قطعه بر روی آن نوشته شده است برقرار است. احتمال برقراری ارتباط میان A و B کدام است؟



حل:

برای محاسبه احتمال برقراری ارتباط میان A و B به احتمالات زیر نیاز داریم:

۱- احتمال برقراری ارتباط بین A و C (سری) برابر است با $p \times p = p^2$.

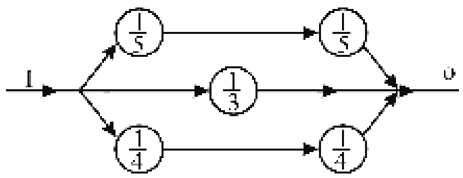
۲- احتمال برقراری ارتباط بین C و D (موازی) برابر است با $\left(1-(1-p)^5\right)$.

۳- احتمال برقراری ارتباط بین B و D (سری) برابر است با $p \times p = p^2$.

بنابراین:

$$P(\text{برقراری ارتباط بین A و B}) = p^2 \times \left(1-(1-p)^5\right) \times p^2$$

مثال ۲ در مدار الکتریکی مقابل، احتمال‌های از کار افتادن اتصالات مختلفی که از هم مستقل‌اند، نشان داده شده است. احتمال اینکه جریان برق برقرار باشد، چقدر است؟



حل:

توجه کنید که در مدار، احتمال کار نکردن اجزا داده شده است.

در این مدار موازی، به ترتیب:

احتمال کار کردن جزء بالا برابر است با $\frac{4}{5}\times\frac{4}{5}=\frac{16}{25}$.

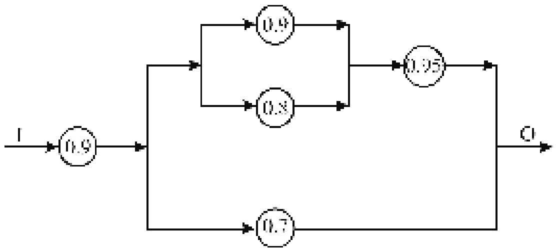
احتمال کار کردن جزء وسط برابر است با $\frac{2}{3}$.

احتمال کار کردن جزء پایین برابر است با $\frac{3}{4}\times\frac{3}{4}=\frac{9}{16}$.

بنابراین:

$$P(\text{برقراری ارتباط بین } I \text{ و } O)=1-\left(1-\frac{16}{25}\right)\times\left(1-\frac{2}{3}\right)\times\left(1-\frac{9}{16}\right)$$

مثال ۳ در مدار مقابل، اگر اعداد نشان‌دهنده احتمال کار کردن هر جزء باشد و کارکرد هر جزء مستقل از دیگری باشد، احتمال برقراری جریان در مدار چقدر است؟



حل:

احتمال برقراری جریان در کل مدار برابر است با:

$$P(\text{برقراری ارتباط بین } I \text{ و } O)=0.9\times P(\text{برقراری جریان در مدار موازی})$$

در مدار موازی به ترتیب:

احتمال کار کردن اجزای بالا برابر است با $\left(1-(1-0.9)(1-0.8)\right)\times 0.95=0.931$.

احتمال کار کردن جزء پایین برابر است با 0.7.

بنابراین:

$$P(\text{احتمال برقراری جریان در مدار موازی})=1-(1-0.931)\times(1-0.7)\simeq 0.98$$

و درنتیجه:

$$P(\text{برقراری ارتباط بین } I \text{ و } O)=0.9\times P(\text{برقراری جریان در مدار موازی})=0.9\times 0.98=0.882$$

مسئله کلاسیک روز تولد

با توجه به حالات ممکن برای تولد n نفر در یک سال (365 روز):

حالات ممکن : $n(S)=365\times 365\times\cdots\times 365=365^n$

مثال ۱ اگر n نفر در یک اتاق حضور داشته باشند، مطلوب است احتمال آنکه:

الف) هیچ دو نفری در یک روز متولد نشده باشند.

ب) حداقل دو نفر در یک روز متولد شده باشند.

ج) فقط دو نفر خاص در یک روز متولد شده باشند.

حل:

الف) برای آنکه هیچ دو نفری در یک روز متولد نشوند، داریم:

A : هیچ دو نفری در یک روز متولد نشده باشند

نفر اول : $365-0=365$ حالت

نفر دوم : $365-1=364$ حالت

$\vdots$

نفر nام : $365-(n-1)=365-n+1$ حالت

$$P(A)=\frac{n(A)}{n(S)}=\frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات ممکن}}=\frac{365\times 364\times 363\times\cdots\times(365-n+1)}{365^n}$$

ب)

$$P(\text{حداقل دونفر در یک روز به دنیا بیایند})=1-P(\text{هیچ دو نفری در یک روز به دنیا نیایند})$$
$$=1-\frac{365\times 364\times 363\times\cdots\times(365-n+1)}{365^n}$$

ج) برای آنکه فقط دو نفر خاص در یک روز به دنیا بیایند، درصورتی‌که نفر اول 365 حالت داشته باشد، نفر دوم 1 حالت دارد (چون باید با نفر اول در یک روز به دنیا بیاید):

A : فقط دو نفر خاص در یک روز متولد شده باشند

$$P(A)=\frac{n(A)}{n(S)}=\frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات ممکن}}=\frac{365\times 1\times 364\times\cdots\times(365-n+2)}{365^n}$$

مثال ۲ چند نفر باید در یک اتاق حضور داشته باشند تا احتمال اینکه حداقل دو نفر از آن‌ها تولدشان را در یک ماه جشن بگیرند بیش از $\frac{1}{2}$ باشد؟ (فرض کنید تولد در ماه‌های مختلف، هم‌شانس است.)

حل:

بهتر است ابتدا احتمال مکمل را به دست آوریم و سپس از یک کم کنیم؛ یعنی احتمال اینکه هیچ دو نفری در اتاق، ماه تولدشان یکسان نباشد.

A: هیچ دو نفری ماه تولدشان یکسان نباشد.

A′: حداقل دو نفر دارای ماه تولد یکسان باشند.

$$P(A)=\frac{12\times 11\times 10\times\ldots(12-n+1)}{(12)^n}$$

حال می‌خواهیم مکمل احتمال بالا بیش از $\frac{1}{2}$ باشد یعنی:

$$P(A')=1-\frac{12\times 11\times\ldots\times(12-n+1)}{(12)^n}>\frac{1}{2}\ \rightarrow\ \frac{12\times 11\times\ldots\times(12-n+1)}{(12)^n}<\frac{1}{2}\ \rightarrow\ n\geq 5$$

احتمال شرطی (Conditional Probability)

هرگاه بخواهیم احتمال وقوع پیشامد A را محاسبه کنیم درصورتی‌که بدانیم پیشامد دیگری مانند B قبل از آن اتفاق افتاده است، احتمال، به صورت شرطی مطرح می‌شود.

مثال کیسه ای حاوی 5 مهره قرمز، 3 مهره سفید و 2 مهره سیاه است. 1 مهره از آن کیسه خارج می‌کنیم:

الف) احتمال آنکه سفید باشد چقدر است؟ (غیر شرطی)

ب) اگر سیاه نباشد، احتمال آنکه سفید باشد چقدر است؟ (شرطی)

حل:

الف) $\frac{3}{2+3+5}=\frac{3}{10}$

ب) $\frac{3}{3+5}=\frac{3}{8}$

✓ **دقت کنید!**

لزوماً نباید کلمه «اگر» در صورت سؤال وجود داشته باشد، تا بدانیم احتمال، شرطی است بلکه همین‌قدر که بفهمیم یک اطلاع اضافه در سؤال وجود دارد برای شرطی بودن مسئله کافی است.

برای مثال اگر قسمت (ب) از سؤال بالا به صورت زیر مطرح می‌شد باز هم احتمال شرطی بود:

1 مهره از آن کیسه خارج می‌کنیم، **می‌بینیم سیاه نیست**. احتمال آنکه سفید باشد چقدر است؟

تعریف: احتمال وقوع پیشامد A را، به شرط آنکه بدانیم B قبلا رخ داده است، به صورت $P(A|B)$ نشان می‌دهیم و آن از رابطه زیر محاسبه می‌کنیم:

$$\text{احتمال A به شرط B}=P(A|B)=\frac{P(A\cap B)}{P(B)}\quad ;\quad P(B)\neq 0$$

همچنین می‌توان احتمال وقوع پیشامد B را به شرط وقوع پیشامد A، به صورت زیر بیان کرد:

$$\text{احتمال B به شرط A}=P(B|A)=\frac{P(A\cap B)}{P(A)}\quad ;\quad P(A)\neq 0$$

نکته: در مسئله مهره‌ها اگر از رنگ مهره‌های قبلی خارج‌شده، هیچ اطلاعی ندارید مسئله را از راه شرطی کردن بر روی رنگ مهره قبلی حل نکنید بلکه احتمال خواسته‌شده را بدون در نظر گرفتن مهره‌های خارج‌شده حل کنید؛ به عبارت دیگر، فرض کنید هیچ مهره‌ای از ظرف خارج نشده و این انتخاب اول است.

مثال ۱ یک تاس را پرتاب می‌کنیم، می‌دانیم عدد بزرگ‌تر از 4 رخ داده است، احتمال آن که عدد 6 رخ دهد کدام است؟

۱) $\frac{1}{2}$ ۲) $\frac{1}{6}$ ۳) $\frac{1}{12}$ ۴) $\frac{1}{3}$

حل: گزینه ۱ درست است.

$$\text{فضای نمونه}: S=\{1,2,3,4,5,6\}$$

$$\begin{cases}\text{پیشامد A}: A=\{6\}\ \to\ P(A)=\frac{1}{6}\\ \text{پیشامد B}: B=\{5,6\}\ \to\ P(B)=\frac{2}{6}\end{cases}\ \Rightarrow\ A\cap B=\{6\}\to P(A\cap B)=\frac{1}{6}$$

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{6}}{\frac{2}{6}}=\frac{1}{2}$$

نکته: راه حل تستی مسایل احتمال شرطی این است که فضای نمونه را به فضای شرط‌شده محدود کرده، سپس حالات خواسته‌شده را از آن انتخاب و احتمالش را محاسبه کنیم.

برای مثال، در این سؤال داریم:

$$\frac{\text{حالات مساعد}}{\text{فضای شرط شده}}=\frac{\text{عدد 6 ظاهر شود}}{\text{عدد بزرگ‌تر از 4 ظاهر شده است}}=\frac{1}{2}$$

مثال ۲ یک جفت تاس را یک بار می‌ریزیم؛ می‌بینیم دو عددی که آمده‌اند، یکسان نیستند. احتمال آنکه مجموع 7 باشد چقدر است؟

۱) $\frac{2}{36}$ ۲) $\frac{1}{5}$ ۳) $\frac{1}{6}$ ۴) $\frac{2}{5}$

حل: گزینه ۲ درست است.

حالات ممکن:

$$n(S)=6^2=36$$

حالات مساعد:

A : مجموع دو عدد 7 باشد $=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

B : دو عدد یکسان نباشند $=36-\{(1,1),\dots,(6,6)\}=36-6=30$

$A\cap B$: مجموع دو عدد 7 باشد و دو عدد یکسان نباشند $=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

$$P(\text{مجموع دو عدد 7 باشد}\mid\text{دو عدد یکسان نباشند})=P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{6}{36}}{\frac{30}{36}}=\frac{6}{30}=\frac{1}{5}$$

راه حل تستی:

$$\frac{\text{حالات مساعد}}{\text{فضای شرط شده}}=\frac{\text{مجموع دو عدد 7 باشد}}{\text{دو عدد ظاهرشده یکسان نباشند}}=\frac{6}{30}=\frac{1}{5}$$

مثال ۳ احتمال اینکه احمد در مرحله اول کنکور دانشگاه‌ها قبول شود، 0.7 است و احتمال اینکه در مرحله اول قبول شده و در مرحله دوم نیز قبول شود 0.8 است. احتمال اینکه احمد در هر دو مرحله قبول شود، چقدر است؟

۱) 0.94 ۲) 0.8 ۳) 0.14 ۴) 0.56

حل: گزینه ۴ درست است.

A : قبول شدن در مرحله اول $\rightarrow\ P(A)=0.7$

$B|A$: در مرحله دوم قبول شود به شرطی که بدانیم در مرحله اول قبول شده است $\rightarrow\ P(B|A)=0.8$

$A\cap B$: در هر دو مرحله قبول شود $\rightarrow\ P(A\cap B)=P(B|A)P(A)=0.8\times0.7=0.56$

مثال ۴ از جعبه‌ای حاوی 3 خودکار سبز، 4 خودکار قرمز و 5 خودکار مشکی هم‌اندازه، یک خودکار به تصادف برداشته شده است که مشکی نیست؛ احتمال قرمز بودن آن چقدر است؟

۱) $\frac{4}{7}$ ۲) $\frac{4}{12}$ ۳) $\frac{4}{8}$ ۴) $\frac{7}{12}$

حل: گزینه ۱ درست است.

$$P(\text{قرمز} \mid \text{مشکی نبودن}) = \frac{P(\text{قرمز} \cap \text{قرمز یا سبز})}{P(\text{قرمز یا سبز})} = \frac{P(\text{قرمز})}{P(\text{قرمز یا سبز})} = \frac{\frac{4}{12}}{\frac{7}{12}} = \frac{4}{7}$$

مثال ۵ فرض کنید احتمال آمدن برف، امروز 0.2 و فردا 0.22 باشد. اگر احتمال برف آمدن فردا به شرط آنکه امروز برف بیاید 0.7 باشد. احتمال برف نیامدن فردا به شرط آنکه امروز برف نیاید، چقدر است؟

۱) 0.3 ۲) 0.72 ۳) 0.78 ۴) 0.9

حل: گزینه ۴ درست است.

$A:$ امروز برف بیاید. $;\ P(A) = 0.2$

$B:$ فردا برف بیاید. $;\ P(B) = 0.22$

$P(B \mid A) = 0.7$

$P(B' \mid A') = ?$

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \to 0.7 = \frac{P(B \cap A)}{0.2} \to P(B \cap A) = 0.14$$

$$P(B' \mid A') = \frac{P(B' \cap A')}{P(A')} = \frac{1 - P(A \cup B)}{0.8} = \frac{1 - [0.2 + 0.22 - 0.14]}{0.8} = 0.9$$

مثال ۶ در یک مدرسه با 100 دانش‌آموز دختر و پسر که در کلاس‌های اول و دوم حضور دارند، از 45 دانش‌آموز کلاس اول 30 نفر پسر و در کلاس دوم، 25 نفر دختر هستند. مطلوب است احتمال آنکه:

الف) یک دانش‌آموز کلاس دوم باشد.

ب) اگر یک دانش‌آموز کلاس دوم باشد، پسر باشد.

ج) اگر یک دانش‌آموز دختر باشد، کلاس اول باشد.

حل: اطلاعات مسئله در جدول زیر خلاصه شده است:

	اول	دوم	
پسر	30	30	60
دختر	15	25	40
	45	55	100

الف) $P(\text{دوم}) = \frac{55}{100}$

ب) $P(\text{پسر} \mid \text{دوم}) = \frac{P(\text{دوم و پسر})}{P(\text{دوم})} = \frac{\frac{30}{100}}{\frac{55}{100}} = \frac{30}{55}$

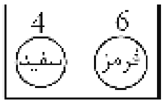
ج) $P(\text{اول} \mid \text{دختر}) = \frac{P(\text{دختر و اول})}{P(\text{دختر})} = \frac{\frac{15}{100}}{\frac{40}{100}} = \frac{15}{40}$

مثال ۷ در کیسه‌ای چهار مهره سفید و شش مهره قرمز وجود دارد. مهره‌ای از کیسه بیرون می‌آوریم و بدون نگاه کردن به رنگش، آن را کنار می‌گذاریم؛ مهره دومی بیرون می‌آوریم. احتمال آنکه این مهره سفید باشد، چقدر است؟

۱) $\frac{3}{9}$ ۲) $\frac{4}{9}$ ۳) $\frac{3}{10}$ ۴) $\frac{4}{10}$

حل: گزینه ۴ درست است.

چون از رنگ مهره اول بی‌اطلاع هستیم، آن را نادیده می‌گیریم و فرض می‌کنیم از کیسه مهره‌ای خارج نشده است. حال احتمال سفید بودن مهره‌ای که خارج می کنیم (مهره دوم) $\frac{4}{10}$ است.



مثال ۸ اگر برای دو پیشامد A و B داشته باشیم $P(A)=\frac{1}{2}$، $P(A|B)=\frac{1}{3}$ و $P(B|A')=\frac{1}{4}$، مقدار $P(B)$ کدام است؟ (اقتصاد ـ ۷۹)

۱) $\frac{3}{16}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{2}$ ۴) $\frac{3}{4}$

حل: گزینه ۱ درست است.

$$\begin{cases} (I)\quad P(A|B)=\dfrac{P(A\cap B)}{P(B)} \;\rightarrow\; P(A\cap B)=\dfrac{1}{3}\times P(B) \\ (II)\quad P(B|A')=\dfrac{P(A'\cap B)}{P(A')} \;\rightarrow\; P(A'\cap B)=\dfrac{1}{4}\times\dfrac{1}{2}=\dfrac{1}{8} \\ P(A')=1-P(A)=1-\dfrac{1}{2}=\dfrac{1}{2} \end{cases}$$

$$P(B)=P(A\cap B)+P(A'\cap B)\xrightarrow{I\ ,\ II}P(B)=\frac{1}{3}P(B)+\frac{1}{8} \;\rightarrow\; P(B)=\frac{3}{16}$$

مثال ۹ اگر $P(A)=0.4$ و $P(B)=0.6$ و $P(B|A)=0.1$ باشد، آن‌گاه $P(A|B)$ کدام است؟ (مدیریت ـ ۷۶)

۱) 0.0153 ۲) 0.04 ۳) 0.05 ۴) 0.067

حل: گزینه ۴ درست است.

$$\begin{cases} P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{0.04}{0.6}=0.067 \\ P(B|A)=\dfrac{P(A\cap B)}{P(A)} \;\rightarrow\; P(A\cap B)=0.1\times 0.4=0.04 \end{cases}$$

مثال ۱۰ اگر $P(A')=\%70$، $P(B)=\%50$ و $P(A|B)=\%60$ باشد، $P(A\cup B)$ چقدر است؟

۱) 0.8 ۲) 0.5 ۳) 0.3 ۴) 0.7

حل: گزینه ۲ درست است.

$$\begin{cases} P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.3+0.5-0.3=0.5 \\ P(A)=1-P(A')=1-0.7=0.3 \\ P(A|B)=\dfrac{P(A\cap B)}{P(B)} \;\rightarrow\; 0.6=\dfrac{P(A\cap B)}{0.5} \;\rightarrow\; P(A\cap B)=0.3 \end{cases}$$

مثال ۱۱ اگر $P(A)=\frac{1}{5}$ و $P(B)=\frac{1}{2}$ و $P(A|B)=\frac{1}{4}$ باشد، $P(A'\cup B')$ کدام است؟ (مدیریت ـ ۷۳)

۱) $\frac{19}{20}$ ۲) $\frac{7}{8}$ ۳) $\frac{21}{20}$ ۴) $\frac{3}{8}$

حل: گزینه ۲ درست است.

$$\begin{cases} P(A'\cup B')=P(A\cap B)'=1-P(A\cap B)=1-\frac{1}{8}=\frac{7}{8} \\ P(A|B)=\frac{P(A\cap B)}{P(B)} \rightarrow P(A\cap B)=P(A|B).P(B)=\frac{1}{4}\times\frac{1}{2}=\frac{1}{8} \end{cases}$$

مکمل احتمال شرطی

مکمل احتمال وقوع پیشامد A به شرط وقوع پیشامد B:

$$P(A'|B)=\frac{P(A'\cap B)}{P(B)}=\frac{P(B-A)}{P(B)}=\frac{P(B)-P(A\cap B)}{P(B)}=1-\frac{P(A\cap B)}{P(B)}=1-P(A|B)$$

مکمل احتمال وقوع پیشامد B به شرط وقوع پیشامد A:

$$P(B'|A)=\frac{P(B'\cap A)}{P(A)}=\frac{P(A-B)}{P(A)}=\frac{P(A)-P(A\cap B)}{P(A)}=1-\frac{P(A\cap B)}{P(A)}=1-P(B|A)$$

بنابراین:

$$\boxed{P(A'|B)=1-P(A|B)}$$
$$\boxed{P(B'|A)=1-P(B|A)}$$

✓ **دقت کنید!**

$$\begin{cases} P(A|B')\neq 1-P(A|B) \\ P(B|A')\neq 1-P(B|A) \end{cases}$$

مثال ۱۲ برای دو پیشامد A و B داریم $P(A'|B')=0.2$ و $P(A|B)=0.1$ و $P(B)=0.5$. مقدار $P(A)$ کدام است؟

۱) 0.25 ۲) 0.5 ۳) 0.4 ۴) 0.45

حل: گزینه ۴ درست است.

$$\begin{cases} (I)\ \ P(A|B)=\frac{P(A\cap B)}{P(B)}=0.1 \rightarrow P(A\cap B)=0.05 \\ (II)\ P(A'|B')=0.2 \rightarrow P(A|B')=\frac{P(A\cap B')}{P(B')}=1-P(A'|B')=0.8 \rightarrow P(A\cap B')=0.4 \\ P(B)=0.5 \rightarrow P(B')=0.5 \end{cases}$$

$$(I),(II) \rightarrow P(A)=P(A\cap B)+P(A\cap B')=0.45$$

مثال ۱۳ اگر برای دو پیشامد A و B داشته باشیم $P(A)=\frac{1}{2}$ و $P(A|B)=\frac{1}{3}$ و $P(B|A')=\frac{1}{4}$، مقدار $P(B)$ کدام است؟

۱) $\frac{3}{16}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{2}$ ۴) $\frac{3}{4}$

حل: گزینه ۱ درست است.

با توجه به آنکه $P(A')=1-P(A)=1-\frac{1}{2}=\frac{1}{2}$، داریم:

$$\begin{cases}(I)\ P(B|A')=\dfrac{P(A'\cap B)}{P(A')}=\dfrac{1}{4} \ \rightarrow\ P(A'\cap B)=\dfrac{1}{4}\times\dfrac{1}{2}=\dfrac{1}{8}\\ (II)\ \ P(A'|B)=1-P(A|B)=\dfrac{2}{3}\end{cases}$$

$$(I),(II)\ \rightarrow\ \overbrace{P(A'|B)}^{\frac{2}{3}}=\dfrac{\overbrace{P(A'\cap B)}^{\frac{1}{8}}}{P(B)}\ \rightarrow\ P(B)=\dfrac{3}{16}$$

احتمال شرطی و پیشامدهای وابسته، ناسازگار و مستقل

بدیهی است احتمال‌های شرطی $P(A|B)$ و $P(B|A)$ زمانی مفهوم دارند که دو پیشامد A و B به هم وابسته باشند، زیرا در صورت مستقل بودن آن‌ها، وقوع هر یک ربطی به وقوع یا عدم وقوع دیگری ندارد و درنتیجه احتمال شرطی مفهومی نخواهد داشت.

$$\text{مستقل}\ \ A,B\xrightarrow{P(A\cap B)=P(A)\times P(B)}\begin{cases}P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{P(A)\times P(B)}{P(B)}=P(A)\\ P(B|A)=\dfrac{P(A\cap B)}{P(A)}=\dfrac{P(A)\times P(B)}{P(A)}=P(B)\end{cases}$$

بنابراین:

$$\text{مستقل}\ \ A,B\Leftrightarrow\begin{cases}P(A|B)=P(A)\\ P(B|A)=P(B)\end{cases}$$

$$\text{وابسته}\ \ A\ ,\ B\longrightarrow\begin{cases}P(A|B)=\dfrac{P(A\cap B)}{P(B)}\neq P(A)\\ P(B|A)=\dfrac{P(A\cap B)}{P(A)}\neq P(B)\end{cases}$$

$$\text{ناسازگار}\ \ A\ ,\ B\xrightarrow{P(A\cap B)=0}\begin{cases}P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{0}{P(B)}=0\\ P(B|A)=\dfrac{P(A\cap B)}{P(A)}=\dfrac{0}{P(A)}=0\end{cases}$$

مثال ۱۴ اگر $P(A)=0.30$، $P(B)=0.50$ و $P(A|B)=0.30$ باشد، می‌توان گفت A و B:

۱) مستقل‌اند ۲) ناسازگارند ۳) وابسته‌اند ۴) شرطی‌اند

حل: گزینه ۱ درست است.

$P(A|B)=P(A)\ \rightarrow$ A , B مستقل‌اند

مثال ۱۵ اگر $P(A)=0.3$، $P(B)=0.7$ و $P(A|B)=0$ باشد، می‌توان گفت A و B:

۱) مستقل‌اند ۲) ناسازگارند ۳) وابسته‌اند ۴) شرطی‌اند

حل: گزینه ۲ درست است.

$P(A|B)=0\ \rightarrow\ P(A\cap B)=0\ \rightarrow$ A , B ناسازگاراند

مثال ۱۶ اگر $P(A)=0.3$ ، $P(B)=0.7$ و $P(A\mid B)=0.1$ باشد، می‌توان گفت A و B:

۱) مستقل‌اند ۲) ناسازگارند ۳) وابسته‌اند ۴) مکمل‌اند

حل: گزینه ۳ درست است.

چون $P(A|B)\neq P(A)$ پس A و B مستقل نیستند و همچنین $P(A|B)\neq 0$ پس A و B ناسازگار نیز نیستند؛ بنابراین A و B دو حادثه وابسته‌اند.

مثال ۱۷ اگر $P(A)=\frac{1}{3}$ ، $P(B)=\frac{1}{5}$ و $P(A\mid B)=\frac{1}{3}$ باشد، $P(A\cup B)$ چقدر است؟

۱) $\frac{8}{15}$ ۲) $\frac{6}{15}$ ۳) $\frac{7}{15}$ ۴) $\frac{9}{15}$

حل: گزینه ۳ درست است.

راه حل اول:

$$\begin{cases} P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{1}{3}+\frac{1}{5}-\frac{1}{3}\times\frac{1}{5}=\frac{7}{15} \\ P(A\mid B)=P(A)=\frac{1}{3} \ \rightarrow \ \text{مستقل‌اند } B,A \ \rightarrow \ P(A\cap B)=P(A)P(B) \end{cases}$$

راه حل دوم:

$$\begin{cases} P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{1}{3}+\frac{1}{5}-\frac{1}{15}=\frac{7}{15} \\ P(A\mid B)=\frac{P(A\cap B)}{P(B)} \ \rightarrow \ \frac{1}{3}=\frac{P(A\cap B)}{\frac{1}{5}} \ \rightarrow \ P(A\cap B)=\frac{1}{15} \end{cases}$$

مثال ۱۸ اگر A و B دو پیشامد مفروض باشند و $P(A)=0.2$, $P(B)=0.3$ و $P(A'|B')=0.8$ باشد، آن‌گاه احتمال $P(B\mid A')$ کدام است؟

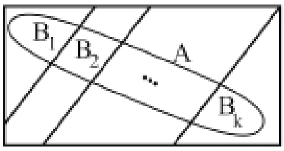
۱) 0.3 ۲) 0.2 ۳) 0.45 ۴) 0.28

حل: گزینه ۱ درست است.

$$P(A)=0.2 \ , P(A')=0.8 \ , \ P(B)=0.3$$
$$P(A'|B')=0.8=P(A') \ \rightarrow \ \text{مستقل‌اند } A,B$$
$$P(B|A')=P(B)=0.3$$

احتمال متوسط

هرگاه پیشامد A درنتیجه وقوع هر یک از پیشامدهای B_1، B_2، ....B_k بتواند اتفاق بیفتد، آن‌گاه وقوع پیشامد A به طور متوسط به صورت زیر محاسبه می‌شود:

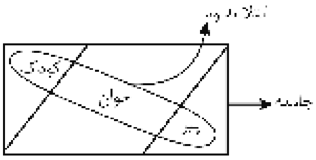


$$P(A)=P(A\cap B_1)+P(A\cap B_2)+\ldots+P(A\cap B_k)=\sum_{i=1}^{k}P(A\cap B_i)$$

که با توجه به قانون احتمال شرطی داریم:

$$P(A)=P(B_1)P(A|B_1)+\cdots+P(B_k)P(A|B_k)=\sum_{i=1}^{k}P(B_i)P(A|B_i)$$

برای مثال، جامعه‌ای شامل **کودک، جوان و پیر** را درنظر بگیرید که در آن بیماری وبا شایع شده باشد. اگر بخواهیم **احتمال مبتلا بودن به بیماری وبا** را برای یک شخص محاسبه کنیم، باید در نظر بگیریم که این شخص می‌تواند **کودک یا جوان یا پیر** باشد؛ بنابراین داریم:



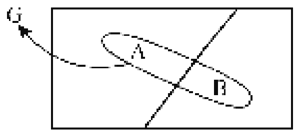
$$P(\text{مبتلا بودن به وبا})=P(\text{مبتلا بودن به وبا و کودک})+P(\text{مبتلا بودن به وبا و جوان})+P(\text{مبتلا بودن به وبا و پیر})$$

$$P(\text{مبتلا بودن به وبا})=P(\text{کودک})P(\text{وبا | کودک})+P(\text{جوان})P(\text{وبا | جوان})+P(\text{پیر})P(\text{وبا | پیر})$$

مثال ۱ اگر $P(A)=P(B)=0.4$ و $P(G|A)=P(G|B)=0.1$ مقدار $P(G)$ چقدر است؟ (حسابداری ـ ۷۷)

۱) 0.08 ۲) 0.21 ۳) 0.51 ۴) 0.9

حل: گزینه ۱ درست است.

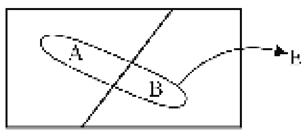


$$\begin{cases} P(G)=\overbrace{P(A)\times P(G|A)}^{P(G\cap A)}+\overbrace{P(B)\times P(G|B)}^{P(G\cap B)} \\ P(G)=0.4\times 0.1+0.4\times 0.1=0.04+0.04=0.08 \end{cases}$$

مثال ۲ اگر $P(A)=0.3$ و $P(B)=0.4$ و $P(E|A)=0.1$ و $P(E'|B)=0.8$ مطلوب است $P(E)$؟ (مدیریت ـ ۸۰)

۱) 0.11 ۲) 0.18 ۳) 0.3 ۴) 0.35

حل: گزینه ۱ درست است.



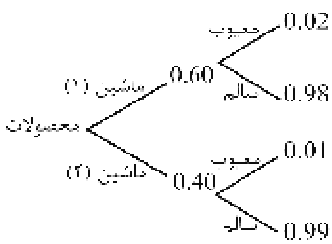
$$\begin{cases} P(E)=\overbrace{P(A)\times P(E|A)}^{P(A\cap E)}+\overbrace{P(B)\times P(E|B)}^{P(B\cap E)} \\ P(E)=0.3\times0.1+0.4\times0.2=0.03+0.08=0.11 \\ P(E|B)=1-P(E'|B)=1-0.8=0.2 \end{cases}$$

مثال ۳ 60 درصد محصولات کارخانه‌ای به وسیله ماشین شماره یک و 40 درصد به وسیله ماشین شماره دو تولید می‌شود. 0.02 محصولات ماشین شماره یک و 0.01 محصولات ماشین شماره دو معیوب‌اند. اگر یک کالا از محصولات کارخانه انتخاب شود، احتمال سالم بودن آن چقدر است؟

۱) 0.016 ۲) 0.996 ۳) 0.012 ۴) 0.984

حل: گزینه ۴ درست است.

E : سالم بودن
A : ماشین شماره یک
B : ماشین شماره دو



$$P(E)=P(E|A)P(A)+P(E|B)P(B)=0.98\times0.6+0.99\times0.4=0.984$$

مثال ۴ مأمور کنترل کیفی کارخانه‌ای از بین دو انبار به طور تصادفی یک انبار و سپس کالایی را انتخاب و بازرسی می‌کند. انبار شماره یک دارای 30 واحد کالاست که 3 واحد آن‌ها معیوب است و انبار شماره دو دارای 100 واحد کالاست که 10 واحد آن معیوب است. احتمال معیوب بودن کالای انتخابی چقدر است؟

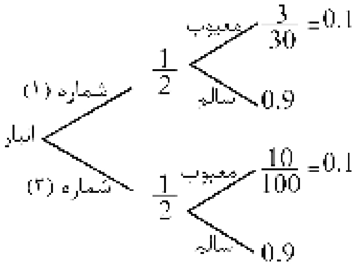
۱) 0.01   ۲) 0.20   ۳) 0.10   ۴) 0.0025

حل: گزینه ۳ درست است.

E : معیوب بودن

A : انبار شماره (۱)

B : انبار شماره (۲)



$$P(E)=P(E\,|\,A)P(A)+P(E\,|\,B)P(B)=0.1\times\frac{1}{2}+0.1\times\frac{1}{2}=0.1$$

مثال ۵ جعبه‌ای حاوی 3 سکه است. یکی از آن‌ها سالم، یکی دیگر دوشیره و دیگری نیز ناسالم است، به طوری که احتمال شیر آمدن آن $\frac{1}{3}$ است. سکه‌ای را به صورت تصادفی انتخاب و پرتاب می‌کنیم. احتمال آنکه شیر ظاهر شود چقدر است؟

۱) $\frac{1}{6}$ ۲) $\frac{5}{6}$ ۳) $\frac{5}{18}$ ۴) $\frac{11}{18}$

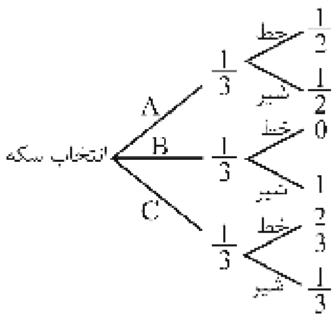
حل: گزینه ۴ درست است.

E : ظاهر شدن شیر

A : سکه سالم

B : سکه دوشیره

C : سکه ناسالم



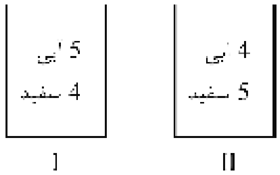
$$P(E)=P(E\,|\,A)P(A)+P(E\,|\,B)P(B)+P(E\,|\,C)P(C)=\frac{1}{2}\times\frac{1}{3}+1\times\frac{1}{3}+\frac{1}{3}\times\frac{1}{3}=\frac{1}{3}\left(\frac{1}{2}+1+\frac{1}{3}\right)=\frac{11}{18}$$

مثال ۶ فرض کنید 5 مهره آبی و 4 مهره سفید در کیسه اول و 4 مهره آبی و 5 مهره سفید درکیسه دوم باشد. اگر مهره‌ای به تصادف از کیسه اول به دوم انتقال یابد و سپس از کیسه دوم یک مهره انتخاب شود، احتمال آبی بودن آن چیست؟ (اقتصاد ـ ۸۱)

۱) $\frac{4}{9}$ ۲) $\frac{5}{10}$ ۳) $\frac{41}{90}$ ۴) $\frac{41}{81}$

حل: گزینه ۳ درست است.

E : آبی بودن مهره خروجی از کیسه دوم
A : سفید بودن مهره خروجی از کیسه اول
B : آبی بودن مهره خروجی از کیسه اول

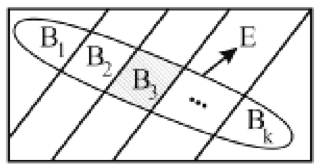


$$P(E)=P(A)P(E|A)+P(B)P(E|B)=\left(\frac{4}{9}\times\frac{4}{10}\right)+\left(\frac{5}{9}\times\frac{5}{10}\right)=\frac{41}{90}$$

قضیه بیز (Bayes' Rule)

اگر $B_1, B_2, \ldots, B_k$ گروه کامل حوادث و E حادثه‌ای باشد که روی این حوادث اتفاق بیفتد آن‌گاه احتمال شرطی وقوع B_i به شرط آنکه بدانیم حادثه E اتفاق افتاده است، تحت عنوان قضیه بیز مطرح می‌شود.
در این شرایط احتمال وقوع B_i را قبل از آنکه اطلاعی درباره حادثه E داشته باشیم احتمال پیشین می‌گوییم و احتمال وقوع B_i را به شرط دانستن وقوع حادثه E احتمال پسین می‌نامیم.
اگر فضای نمونه‌ای توسط پیشامدهای $B_1, B_2, \ldots, B_k$ افراز شود، یعنی:

$$\begin{cases} B_i \cap B_j = \varnothing & i \neq j \\ \bigcup_{i=1}^{k} B_i = S \\ B_i \neq \varnothing \end{cases}$$



و E یک پیشامد در این فضا باشد، داریم:

$$\begin{cases} P(B_i|E)=\dfrac{P(B_i)P(E|B_i)}{P(E)} \quad ; \quad k=1,2,\ldots,n \\ P(E)=\sum_{i=1}^{k} P(B_i)P(E|B_i) \end{cases}$$

توجه کنید که $P(E)$، احتمال متوسط است.

مثال ۱ اطلاعات زیر داده شده است:

$P(B_1)=0.2\ ,\ P(A|B_1)=0.01$

$P(B_2)=0.3\ ,\ P(A|B_2)=0.02$

$P(B_3)=0.5\ ,\ P(A|B_3)=0.05$

مطلوب است احتمال $P(B_2|A)$. (مدیریت ـ ۷۵)

۱) $\frac{2}{33}$ ۲) $\frac{1}{16}$ ۳) $\frac{2}{11}$ ۴) $\frac{3}{4}$

حل: گزینه ۳ درست است.

بنا بر قضیه بیز داریم:

$$P(B_2|A)=\frac{P(B_2).P(A|B_2)}{P(B_1).P(A|B_1)+P(B_2).P(A|B_2)+P(B_3).P(A|B_3)}$$

$$P(B_2|A)=\frac{(0.3)(0.02)}{(0.2)(0.01)+(0.3)(0.02)+(0.5)(0.05)}=\frac{0.006}{0.033}=\frac{2}{11}$$

مثال ۲ کارخانه‌ای سه خط تولید دارد که به ترتیب 60%، 25% و 15% تولیدات را انجام می‌دهند. می‌دانیم 3% تولیدات خط اول، 5% تولیدات خط دوم و 7% تولیدات خط سوم معیوب هستند. کالای معیوبی از این کارخانه در دست است؛ احتمال آنکه متعلق به خط سوم باشد، چقدر است؟

۱) 0.07 ۲) 0.25 ۳) 0.01 ۴) 0.041

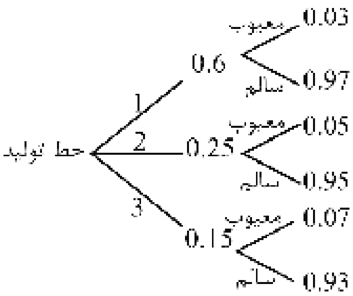
حل: گزینه ۲ درست است.

E : معیوب بودن

A_1 : خط اول

A_2 : خط دوم

A_3 : خط سوم



$$\begin{cases}P(A_3|E)=\dfrac{P(A_3)P(E|A_3)}{P(E)}=\dfrac{0.15\times0.07}{0.041}=0.25\\ P(E)=P(A_1)P(E|A_1)+P(A_2)P(E|A_2)+P(A_3)P(E|A_3)=0.6\times0.03+0.25\times0.05+0.15\times0.07=0.041\end{cases}$$

مثال ۳ شخصی را که با احتمال 0.6 مبتلا به نوعی سرطان است، با یک تست مورد آزمایش قرار دادیم. این تست، 10% افراد سالم را سرطانی و 2% افراد سرطانی را سالم نشان می دهد. چقدر احتمال دارد شخصی که تست او جواب مثبت داشته است، واقعاً مبتلا باشد؟

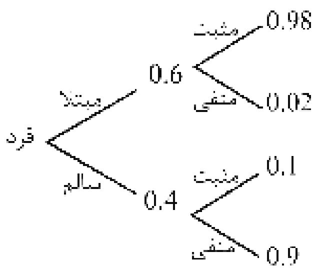
۱) 0.628 ۲) 0.588 ۳) 0.94 ۴) 0.6

حل: گزینه ۳ درست است.

E : جواب مثبت

A : واقعاً مبتلا

A' : واقعاً سالم



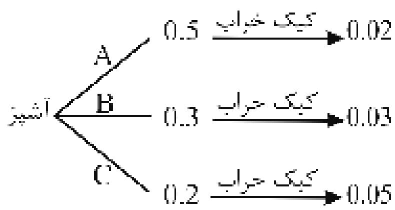
$$\begin{cases}P(A|E)=\dfrac{P(A)P(E|A)}{P(E)}=\dfrac{0.6\times0.98}{0.628}=0.94\\ P(E)=P(A)P(E|A)+P(A')P(E|A')=0.6\times0.98+0.4\times0.1=0.628\end{cases}$$

مثال ۴ سه آشپز A، B و C هر کدام کیک خاصی را تهیه می‌کنند؛ با احتمال‌های 0.02، 0.03 و 0.05 کیک آن‌ها هنگام پخت خراب می‌شود. اگر در رستورانی که آن‌ها کار می‌کنند، آشپز A، 50 درصد، آشپز B ، 30 درصد و آشپز C، 20 درصد از کیک‌ها را بپزند، چه نسبتی از کیک‌های خراب توسط آشپز A تهیه می‌شود؟

۱) $\frac{1}{100}$ ۲) $\frac{2}{100}$ ۳) $\frac{10}{29}$ ۴) $\frac{1}{3}$

حل: گزینه ۳ درست است.

E : کیک خراب



$$P(A|E)=\frac{P(E|A)P(A)}{P(E|A)P(A)+P(E|B)P(B)+P(E|C)P(C)}$$

$$=\frac{0.02\times0.5}{0.02\times0.5+0.03\times0.3+0.05\times0.2}=\frac{10}{10+9+10}=\frac{10}{29}$$

مثال ۵ در جعبه‌ای 3 سکه وجود دارد که یکی از آن‌ها هر دو طرف شیر، دیگری یک سکه سالم و سومی سکه‌ای اریب است که هنگام پرتاب با احتمال 0.75 شیر ظاهر می‌شود. وقتی که یکی از سکه‌ها را به تصادف انتخاب کرده و پرتاب می‌کنیم شیر ظاهر می‌شود. احتمال اینکه سکه دو طرف شیر انتخاب شده باشد، چقدر است؟

۱) $\frac{1}{3}$ ۲) $\frac{4}{9}$ ۳) $\frac{1}{5}$ ۴) $\frac{4}{7}$

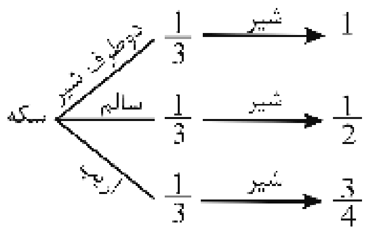
حل: گزینه ۲ درست است.

A : سکه دو طرف شیر

B : سکه سالم

C : سکه اریب

H : شیر آمدن



$$P(A|H)=\frac{P(H|A)P(A)}{P(H|A)P(A)+P(H|B)P(B)+P(H|C)P(C)}=\frac{1\times\frac{1}{3}}{1\times\frac{1}{3}+\frac{1}{2}\times\frac{1}{3}+\frac{3}{4}\times\frac{1}{3}}=\frac{4}{9}$$

مباحث اضافی احتمال

این مباحث در هیچ‌یک از منابع آمار انسانی نیامده است و مربوط به رشته‌های فنی است. در صورت تمایل به مطالعه بیشتر می‌توانید از آن‌ها استفاده کنید. درعین‌حال مبحث «چند جمله‌ای» در منابع ریاضی بررسی شده است و سؤالات مربوط به آن در آزمون ریاضی مطرح می‌شود.

بسط چندجمله‌ای

الف) $(x_1+\cdots+x_k)^n=\sum\binom{n}{n_1\cdots n_k}x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}$

ب) ضریب هر جمله مانند $x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}$ در بسط $(x_1+x_2\cdots+x_k)^n$ برابر است با:

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

ج) ضریب هر جمله مانند $x_1^{n_1}x_2^{n_2}\ldots x_k^{n_k}$ در بسط $(a_1x_1+\cdots+a_kx_k)^n$ برابر است با:

$$a_1^{n_1}\times\cdots\times a_k^{n_k}\times\frac{n!}{n_1!n_2!\ldots n_k!}$$

د) برای محاسبه مجموع ضرایب بسط یک چندجمله‌ای می‌توان به‌جای مجهولات، عدد 1 قرار داد.

مثال ۱ ضریب $x_1^3x_2x_3^2$ در بسط $(x_1+x_2+x_3)^6$ کدام است؟

۱) 6 ۲) 30 ۳) 60 ۴) 20

حل: گزینه ۳ درست است.

با توجه به قسمت (ب) از موارد ذکر شده، داریم:

$$n_1=3\ ,\ n_2=1\ ,\ n_3=2\ \rightarrow\begin{cases}n=n_1+n_2+n_3=6\\ \dfrac{n!}{n_1!n_2!n_3!}=\dfrac{6!}{3!1!2!}=60\end{cases}$$

مثال ۲ در بسط $(2x_1+x_2+x_3)^6$ ضریب $x_1^3x_2x_3^2$ کدام است؟

۱) 60 ۲) 180 ۳) 240 ۴) 480

حل: گزینه ۴ درست است.

با توجه به قسمت (ج) از موارد ذکر شده، داریم:

$$n_1=3\ ,\ n_2=1\ ,\ n_3=2\ \rightarrow\begin{cases}n=n_1+n_2+n_3=3+1+2=6\\ a_1^{n_1}\times\cdots\times a_k^{n_k}\times\dfrac{n!}{n_1!n_2!\ldots n_k!}=2^3\dfrac{6!}{3!1!2!}=480\end{cases}$$

مثال ۳ مجموع ضرایب بسط $(5x_1-2x_2+x_3)^3$ کدام است؟

۱) 64 ۲) 160 ۳) 280 ۴) 118

حل: گزینه ۱ درست است.

با توجه به قسمت (د) از موارد ذکر شده، داریم:

$$\text{مجموع ضرایب}=(5\times1-2\times1+1)^3=4^3=64$$

توزیع n شیء نامتمایز (مشابه) در k سلول

الف) تعداد تقسیمات n شیء مشابه در k سلول بدون هیچ محدودیتی برابر است با:

$$\binom{n+k-1}{k-1}=\binom{n+k-1}{n}=\frac{(n+k-1)!}{(k-1)!n!}$$

ب) تعداد تقسیمات $n_1, n_2, \ldots$ شیء مشابه در k سلول بدون هیچ محدودیتی برابر است با:

$$\binom{n_1+k-1}{k-1}\times\binom{n_2+k-1}{k-1}\times\cdots$$

ج) تعداد تقسیمات n شیء مشابه در k سلول به طوری که در هر سلول حداقل r شیء قرار گیرد برابر است با:

$$\binom{n-k(r-1)-1}{k-1}$$

د) تعداد تقسیمات n شیء مشابه در k سلول به طوری که در هر سلول حداقل $r=1$ شیء قرار گیرد برابر است با:

$$\binom{n-1}{k-1}$$

مثال ۱ به چند طریق می‌توان 35 سیب مشابه را بین 4 جعبه تقسیم کرد؟

۱) 8436 ۲) 52360 ۳) 26180 ۴) 82251

حل: گزینه ۱ درست است.

$$\binom{n+k-1}{k-1}=\binom{35+4-1}{4-1}=\binom{38}{3}=\frac{38!}{3!\,35!}=8436$$

مثال ۲ به چند طریق می‌توان 8 خودکار سبز، 11 خودکار آبی و 5 خودکار قرمز را بین 3 نفر تقسیم کرد؟

۱) 36855 ۲) 73710 ۳) 1320 ۴) 5670

حل: گزینه ۲ درست است.

$$\binom{8+3-1}{3-1}\binom{11+3-1}{3-1}\binom{5+3-1}{3-1}=\binom{10}{2}\binom{13}{2}\binom{7}{2}=73710$$

دقت کنید، اگر بخواهیم هر نفر حداقل یک رنگ از هر خودکار داشته باشد، بنا بر حالت (د) داریم:

$$\binom{n-1}{k-1}\rightarrow\binom{8-1}{3-1}\binom{11-1}{3-1}\binom{5-1}{3-1}=\binom{7}{2}\binom{10}{2}\binom{4}{2}=21\times45\times6=5670$$

مثال ۳ می‌خواهیم 10 معلم (مشابه) را بین 3 مدرسه به‌گونه‌ای تقسیم کنیم که حداقل 2 معلم در هر مدرسه باشد. تعداد حالات این تقسیم کدام است؟

۱) 66 ۲) 36 ۳) 440 ۴) 15

حل: گزینه ۴ درست است.

بنا بر مورد (ج) داریم:

$$n=10\ ,\ k=3\ ,\ r=2$$

بنابراین تعداد حالات برابر است با:

$$\binom{10-3(2-1)-1}{3-1}=\binom{6}{2}=15$$

مثال ۴ یک آسانسور از طبقه همکف با 8 مسافر (بدون مسئول آسانسور) حرکت کرده و تا طبقه ششم همه را پیاده می‌کند.
الف) اگر مسافران از نظر مسئول آسانسور یکسان باشند، او به چند طریق مختلف می‌تواند شاهد پیاده شدن مسافران باشد؟
ب) اگر 5 نفر از مسافران مرد و 3 نفر زن باشند و مسئول آسانسور بتواند مرد و زن را تشخیص دهد آن‌گاه به چند طریق می‌تواند شاهد پیاده شدن مسافران باشد؟

حل:
الف) تعداد تقسیمات $n = 8$ مسافر مشابه در $k = 6$ طبقه برابر است با:

$$\binom{8+6-1}{6-1} = \binom{13}{5}$$

ب)
تعداد تقسیمات $n = 5$ مرد مشابه در $k = 6$ طبقه برابر است با:

$$\binom{5+6-1}{5} = \binom{10}{5}$$

تعداد تقسیمات $n = 3$ زن مشابه در $k = 6$ طبقه برابر است با:

$$\binom{3+6-1}{5} = \binom{8}{5}$$

بنابراین تعداد کل حالات برابر است با:

$$\binom{10}{5} \times \binom{8}{5} = 252 \times 56 = 14112$$

دقت کنید، در قسمت (الف) هر 8 مسافر از نظر مسئول آسانسور مشابه‌اند و در قسمت (ب) تنها نقطه تمایز این 8 نفر، مرد و زن بودن آن‌هاست. درواقع از نظر مسئول آسانسور همه مردها مشابه و همه زن‌ها مشابه هستند. بنابراین مردها و زن‌ها را به طور جداگانه بررسی می‌کنیم.

نکته:
الف) تعداد جملات بسط $(x_1 + x_2 \cdots + x_k)^n$ و یا به عبارت دیگر تعداد جواب‌های صحیح و نامنفی $(x_i \geq 0)$ معادله $x_1 + x_2 + \ldots + x_k = n$ برابر است با:

$$\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$$

ب) تعداد جواب‌های صحیح و مثبت $(x_i > 0)$ معادله $x_1 + x_2 + \ldots + x_k = n$ برابر است با:

$$\binom{n-1}{k-1}$$

مثال در معادله $x_1 + x_2 + x_3 + x_4 = 7$:
الف)تعداد جواب‌های صحیح نامنفی را به دست آورید.
ب) تعداد جواب‌های صحیح بزرگ‌تر از صفر را به دست آورید.

حل:

الف) $\binom{7+4-1}{4-1} = \binom{10}{3} = 120$ = تعداد جواب‌های صحیح نامنفی

ب) $\binom{7-1}{4-1} = \binom{6}{3} = 20$ = تعداد جواب‌های صحیح بزرگ‌تر از صفر

نکته:

الف) می‌خواهیم با n شیء ترکیب‌هایی حداکثر r تایی و حداقل 1 تایی که در آن‌ها تکرار مجاز است بسازیم؛ تعداد حالات ممکن برابر است با:

$$n + n^2 + n^3 + \ldots + n^r = \frac{n^r - 1}{n - 1} \times n$$

ب) تعداد تقسیمات n شیء **متمایز** در k سلول برابر است با k^n.

ج) تعداد تقسیمات n شیء **نامتمایز** در k سلول به طوری که در هر سلول حداکثر یک شیء قرار گیرد برابر است با $\binom{n}{k} = C_n^k$.

مثال با ارقام 4 ,3 ,2 ,1 ,6 چند عدد حداکثر 10رقمی می‌توان ساخت؟

حل:

$$\begin{cases} 5^1 + 5^2 + \ldots + 5^{10} = \dfrac{5^{10} - 1}{5 - 1} \times 5 \\ r = 10 \;, \; n = 5 \end{cases}$$

تقسیم n شیء متمایز در k سلول (بدون محدودیت)

الف) تعداد تقسیمات n شیء متمایز در k سلول (ترتیب توزیع n شیء مهم باشد) برابر است با:

$$P_{n+k-1}^{k-1} = \frac{(n+k-1)!}{(k-1)!}$$

ب) تعداد تقسیمات n شیء متمایز در k سلول به طوری که در هر سلول حداقل 1 شیء قرار گیرد برابر است با:

$$P_{n-1}^{k-1} = \frac{(\overset{n-k}{\cancel{n}} + k - 1)!}{(k-1)!} = \frac{(n-1)!}{(k-1)!}$$

در این حالت، k شیء از n شیء را برمی‌داریم و $n-k$ شیء باقی‌مانده را که ترتیبشان مهم است، توزیع می‌کنیم.

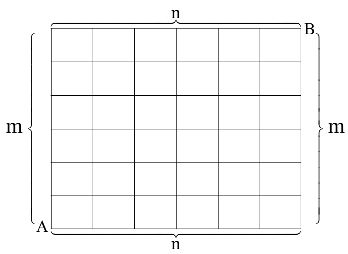
ج) تعداد تقسیمات n شیء متمایز در k سلول به طوری که در هر سلول حداکثر یک شیء قرار گیرد، برابر است با:

$$P_k^n = \frac{n!}{(n-k)!}$$

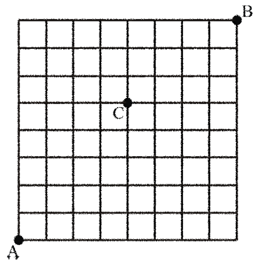
تعداد مسیرهای بین دو نقطه

درصورتی‌که شخصی بخواهد با حرکت‌های قائم و افقی (بالا و جلو) از نقطه A به نقطه B برود تعداد راه‌های ممکن برابر است با:

$$\frac{(m+n)!}{m!\; n!} \qquad ; \qquad \begin{cases} n: \text{تعداد واحدهای افقی} \\ m: \text{تعداد واحدهای قائم} \end{cases}$$



مثال در شکل زیر، تعداد مسیرها از نقطه A به نقطه B به طوری که از نقطه C بگذرد کدام است؟ (حرکت‌ها فقط به سمت بالا یا به سمت راست امکان‌پذیر است.)



۱) $\binom{8}{4}^2$ ۲) $\binom{8}{4}\binom{7}{4}$

۳) $\binom{9}{4}\binom{7}{4}$ ۴) $\binom{9}{5}\binom{8}{4}$

حل: گزینه ۳ درست است.

برای رفتن از A به C باید 9 حرکت انجام شود که شامل 4 حرکت به راست (مشابه) و 5 حرکت به بالا (مشابه) است، پس تعداد کل حالات برابر است با $\frac{9!}{4!5!}$. همچنین برای رفتن از C به B نیز باید 7 حرکت انجام شود که شامل 4 حرکت به راست (مشابه) و 3 حرکت به بالا (مشابه) است که کل حالات برابر است با $\frac{7!}{4!3!}$. درنتیجه کل حالات رفتن از A به B به شرط عبور از C برابر است با:

$$\frac{9!}{4!5!}\times\frac{7!}{4!\,3!}=\binom{9}{4}\times\binom{7}{4}$$

نکته: تعداد مربع‌های موجود در یک جدول $n\times n$ برابر است با:

$$\sum_{i=1}^{n} i^2=1^2+2^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6}$$

مسئله بازی‌ها

اشخاص A و B را در نظر بگیرید که عمل مشخصی را تا رسیدن به اولین موفقیت یکی پس از دیگری تکرار می‌کنند. در این وضعیت احتمال موفقیت برای هر کدام از آن‌ها بستگی به این موضوع دارد که کدام‌یک بازی را شروع کرده است؛ چراکه در ادامه خواهیم دید احتمال موفقیت بازیکنی که بازی را شروع کند، بیشتر است.

مسئله محدود (انتخاب بدون جایگذاری)

مثال ظرفی شامل 2 توپ سفید و 3 توپ سیاه است. بازیکن‌های A و B به طور متوالی و هر دفعه یک توپ به تصادف خارج می‌کنند تا زمانی‌که یک توپ سفید انتخاب شود. اگر A بازی را شروع کند و انتخاب بدون جایگذاری باشد:

الف) احتمال برنده شدن A چقدر است؟

ب) احتمال برنده شدن B چقدر است؟

حل:

الف) ازآنجاکه A بازی را شروع می‌کند، وضعیت‌های زیر برای محاسبه احتمال برنده شدن $P(S_A)$ قابل پیش‌بینی است:

۱) A در همان بار اول مهره سفید را با احتمال $\frac{2}{5}$ خارج کند و موفق شود:

$$P(A)=\frac{2}{5}$$

۲) A در بار اول مهره سیاه را با احتمال $\frac{3}{5}$ خارج کند سپس B نیز مهره سیاه دیگری را با احتمال $\frac{2}{4}$ خارج کند تا A در مرحله بعدی مهره سفید را با احتمال $\frac{2}{3}$ انتخاب کرده و موفق شود:

$$P(A'B'A)=\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}=\frac{1}{5}$$

دقت کنید وضعیت دیگری برای برنده شدن A وجود ندارد، چراکه اگر A در دومین انتخاب خود (سومین انتخاب در کل) مهره سیاه خارج کند، ازآنجاکه انتخاب بدون جایگذاری است و 3 مهره سیاه بیشتر نداریم، B در دومین انتخاب خود (چهارمین انتخاب درکل) مهره سفید را برداشته و موفق می‌شود:

$$\left(\begin{array}{cccc} A' & B' & A' & B \end{array}\right)$$

سیاه سیاه سیاه سفید

بنابراین درنهایت احتمال برنده شدن A با توجه به موارد (۱) و (۲) برابر است با:

$$P(S_A)=P(A)+P(A'B'A)=\frac{2}{5}+\frac{1}{5}=\frac{3}{5}$$

ب) با توجه به آنکه A بازی را شروع می‌کند، وضعیت‌های زیر برای محاسبه احتمال برنده شدن B قابل پیش‌بینی است:

۱) A در مرتبه اول سیاه خارج کرده و سپس B مهره سفید را خارج می‌کند:

$$P(A'B)=\frac{3}{5}\times\frac{2}{4}=\frac{3}{10}$$

۲) A و B در مرحله اول انتخاب خود سیاه انتخاب کرده، سپس A دوباره سیاه انتخاب می‌کند تا B سفید را خارج کند:

$$P(A'B'A'B)=\frac{3}{5}\times\frac{2}{4}\times\frac{1}{3}\times 1=\frac{1}{10}$$

بنابراین درنهایت احتمال برنده شدن B با توجه به موارد (۱) و (۲) برابر است با:

$$P(S_B)=P(A'B)+P(A'B'A'B)=\frac{3}{10}+\frac{1}{10}=\frac{4}{10}=\frac{2}{5}$$

مسئله نامحدود (انتخاب با جایگذاری)

مثال ظرفی شامل 2 توپ سفید و 3 توپ سیاه است. بازیکن‌های A و B به طور متوالی و هر دفعه یک توپ به‌تصادف خارج می‌کنند تا زمانی‌که یک توپ سفید انتخاب شود. اگر A بازی را شروع کند و انتخاب با جایگذاری باشد:

الف) احتمال برنده شدن A چقدر است؟

ب) احتمال برنده شدن B کدام است؟

حل:

ازآنجاکه انتخاب با جایگذاری است، تعداد مهره‌ها در ظرف ثابت است و درنتیجه احتمال انتخاب سفید و سیاه در هر بار انتخاب ثابت و برابر است با:

$$P(\text{سفید})=\frac{2}{5} \quad \text{و} \quad P(\text{سیاه})=\frac{3}{5}$$

الف) راه حل اول: ازآنجاکه A بازی را شروع می‌کند، وضعیت‌های زیر برای محاسبه احتمال برنده شدن A به صورت زیر قابل پیش‌بینی است:

$$P(S_A)=\overset{A}{\frac{2}{5}}+\overset{A'}{\frac{3}{5}}\times\overset{B'}{\frac{3}{5}}\times\overset{A}{\frac{2}{5}}+\ldots=\frac{\frac{2}{5}}{1-\frac{9}{25}}=\frac{\frac{2}{5}}{\frac{16}{25}}=\frac{5}{8}$$

راه حل دوم: برای برنده شدن A دو حالت وجود دارد؛ یا در مرحله اول A می‌برد یا هیچ‌کدام از A و B در مرحله اول نمی‌برند و A در مراحل بعدی می‌برد:

$$P(S_A)=\overset{A}{\frac{2}{5}}+\overset{A'}{\frac{3}{5}}\times\overset{B'}{\frac{3}{5}}\times P(S_A) \rightarrow \frac{16}{25}P(S_A)=\frac{2}{5} \rightarrow P(S_A)=\frac{5}{8}$$

ب) راه حل اول: ازآنجاکه A بازی را شروع می‌کند، وضعیت‌های زیرا برای محاسبه احتمال برنده شدن B وجود دارد:

$$P(S_B)=\overset{A'}{\frac{3}{5}}\times\overset{B}{\frac{2}{5}}+\overset{A'}{\frac{3}{5}}\times\overset{B'}{\frac{3}{5}}\times\overset{A'}{\frac{3}{5}}\times\overset{B}{\frac{2}{5}}+\dots=\frac{\frac{6}{25}}{1-\frac{9}{25}}=\frac{\frac{6}{25}}{\frac{16}{25}}=\frac{3}{8}$$

راه حل دوم: برای آنکه B برنده شود دو حالت وجود دارد؛ یا در مرحله اول A می‌بازد و B می‌برد یا در مرحله اول هر دو می‌بازند و در مراحل بعدی B می‌برد:

$$P(S_B)=\overset{A'}{\frac{3}{5}}\times\overset{B}{\frac{2}{5}}+\overset{A'}{\frac{3}{5}}\times\overset{B'}{\frac{3}{5}}P(S_B) \rightarrow P(S_B)=\frac{3}{8}$$

راه حل سوم: احتمال برنده شدن B برابر با $1-P(S_A)$ است بنابراین با توجه به قسمت (الف) داریم:

$$P(S_B)=1-P(S_A)=1-\frac{5}{8}=\frac{3}{8}$$

تست‌های طبقه‌بندی شده

آنالیز ترکیبی

۱. شش نفر کارشناس مدیریت را به چند طریق می‌توان به 3 شهر اعزام کرد به طوری که تعداد افراد اعزامی به دو شهر برابر نباشند؟ (حسابداری و مدیریت ــ ۸۸)

۱) 60 ۲) 180 ۳) 240 ۴) 360

۲. تعداد حالات ممکن از تقسیم 10 نفر به سه گروه 5 ,3 و 2 نفری برابر است با: (اقتصاد ــ ۸۶)

۱) 1440 ۲) 2520 ۳) 2630 ۴) 5040

۳. در یک مسابقه دوچرخه‌سواری 43 دوچرخه‌سوار قرار است در یک جاده کمربندی دور شهری مسابقه دهند. در چند مورد یا حالت دوچرخه‌سواران می‌توانند مقام اول، دوم و سوم را کسب نمایند؟ (اقتصاد ــ ۸۷)

۱) 129 ۲) 1763 ۳) 12341 ۴) 74046

۴. از بین حروف کلمه MANAGEMENT به چند طریق می‌توان سه حرف انتخاب کرد؟ (برنامه‌ریزی شهری ــ ۸۷)

۱) 30 ۲) 32 ۳) 36 ۴) 40

۵. با حروف کلمه APPLICATION به چند طریق می‌توان یک رمز عبوری سه‌حرفی ساخت؟ (برنامه‌ریزی شهری ــ ۸۸)

۱) 360 ۲) 378 ۳) 399 ۴) 420

۶. به چند طریق می‌توان فقط به 10 پرسش از 12 پرسش داده‌شده پاسخ داد به شرط آنکه حداقل 4 پرسش اول از 5 پرسش اجباری باشد؟ (GIS ــ ۸۶)

۱) 56 ۲) 65 ۳) 120 ۴) 140

۷. از حروف کلمه OPERATOR به چند طریق می‌توان 4 حرف کنار گذاشت؟ (GIS ــ ۸۷)

۱) 32 ۲) 36 ۳) 70 ۴) 72

۸. 5 تیم دونفره بدمینتون در یک مسابقه شرکت می‌کنند. فرض کنید این افراد قرار است دور یک میز طوری قرار بگیرند که نفرات هر تیم در کنار هم قرار داشته باشند. این افراد به چند طریق می‌توانند دور این میز قرار بگیرند؟ (محیط زیست ـ ۸۷)

۱) $4!\times 2^4$ ۲) $4!\times 2^5$ ۳) $5!\times 2^4$ ۴) $5!\times 2^5$

روابط احتمالاتی

۹. در یک آزمایش مهارت احتمال موفقیت دو نفر به ترتیب $\frac{3}{5}$ و $\frac{1}{2}$ است. با کدام احتمال لااقل یکی از آن دو موفق می‌شوند؟ (حسابداری و مدیریت ـ ۸۷)

۱) $\frac{2}{3}$ ۲) $\frac{3}{4}$ ۳) $\frac{4}{5}$ ۴) $\frac{5}{6}$

۱۰. اگر $A\cup B$ برابر فضای نمونه، $P(A)=0.7$ و $P(B)=0.6$ باشد، مقدار $P(B-A')$ چقدر است؟ (اقتصاد ـ ۸۸)

۱) 0.18 ۲) 0.3 ۳) 0.42 ۴) 0.7

۱۱. در یک کلاس 40 نفری 15 نفر علاقه‌مند به ادامه تحصیل می‌باشند. 20 نفر به ورزش فوتبال علاقه‌مند می‌باشند. 5 نفر علاقه‌مند به ادامه تحصیل و علاقه‌مند به ورزش فوتبال می‌باشند. یک نفر از افراد این کلاس را به تصادف انتخاب می‌کنیم. احتمال اینکه این فرد نه علاقه‌مند به ورزش فوتبال و نه علاقه‌مند به ادامه تحصیل باشد چقدر است؟ (محیط زیست ـ ۸۶)

۱) %12.5 ۲) %25 ۳) %37.5 ۴) %75

۱۲. در یک کارخانه قرار است یک نفر استخدام شود. 40% افراد مراجعه‌کننده حداقل مدرک کارشناسی دارند، 60% از افراد مراجعه‌کننده سابقه شغل قبلی دارند، 15% از افراد مراجعه‌کننده حداقل مدرک کارشناسی داشته و سابقه شغل قبلی نیز دارند. احتمال اینکه فردی که استخدام می‌شود قبلاً سابقه شغل نداشته باشد و حداقل مدرک کارشناسی داشته باشد چقدر است؟ (محیط زیست ـ ۸۷)

۱) %15 ۲) %20 ۳) %25 ۴) %60

۱۳. در یک کارخانه 100 نفر شاغل هستند که 56 نفر سابقه شغلی بیش از 5 سال در این کارخانه دارند، همچنین 44 نفر قبلاً در جایی دیگر نیز کار کرده‌اند، همچنین 15 نفر سابقه شغلی بیش از 5 سال در این کارخانه دارند و قبلاً نیز در جایی دیگر کار کرده‌اند. فردی را به تصادف از میان 100 نفر انتخاب می‌کنیم. احتمال اینکه سابقه شغلی بیش از 5 سال نداشته باشد و قبلاً نیز در جایی دیگر کار کرده باشد، کدام است؟ (محیط زیست ـ ۸۸)

۱) $\frac{56}{100}$ ۲) $\frac{44}{100}$ ۳) $\frac{29}{100}$ ۴) $\frac{15}{100}$

احتمال

۱۴. اعداد $1, 2, 3, 4, 5, 6$ بر روی 6 مهره یکسان نوشته شده‌اند، اگر دو مهره را با هم بیرون آوریم با کدام احتمال مجموع اعداد این دو مهره مضرب 3 خواهد بود؟ (حسابداری و مدیریت ـ ۸۶)

۱) $\frac{1}{3}$ ۲) $\frac{1}{4}$ ۳) $\frac{2}{5}$ ۴) $\frac{4}{15}$

۱۵. ارقام $1, 2, 2, 3, 3$ و 1 به‌تصادف کنار هم قرار می‌گیرند. با کدام احتمال بین هر دو رقم یکسان دو رقم متمایز قرار می‌گیرند؟ (حسابداری و مدیریت ـ ۸۷)

۱) $\frac{1}{10}$ ۲) $\frac{1}{12}$ ۳) $\frac{1}{15}$ ۴) $\frac{1}{18}$

۱۶. شش نفر دانشجو که دو نفر آنان از گروه مدیریت، دو نفر از گروه حسابداری و دو نفر دیگر از گروه آمار می‌باشند دور یک میز گرد می‌نشینند. با کدام احتمال افراد هم‌گروه کاملاً مقابل هم قرار می‌گیرند؟
(حسابداری و مدیریت ـ ۸۸)

۱) $\frac{1}{15}$ ۲) $\frac{1}{12}$ ۳) $\frac{1}{8}$ ۴) $\frac{1}{6}$

۱۷. در یک رمز عبور شش‌رقمی بدون صفر با کدام احتمال دقیقاً سه رقم مضرب 3 و یک رقم مضرب 4 می‌باشد؟
(حسابداری و مدیریت ـ ۸۸)

۱) $\frac{320}{81\times 81}$ ۲) $\frac{640}{81\times 81}$ ۳) $\frac{80}{27\times 27}$ ۴) $\frac{160}{27\times 27}$

۱۸. در پرتاب دو سکه سالم به هوا، احتمال حصول حداقل یک شیر (Head) چند است؟ (اقتصاد ـ ۸۷)

۱) $\frac{3}{8}$ ۲) $\frac{3}{4}$ ۳) $\frac{1}{4}$ ۴) $\frac{1}{2}$

۱۹. ارقام $1, 2, 2, 3$ و 1 را تصادفی کنار هم قرار می‌دهیم با کدام احتمال سه رقم متوالی به ترتیب صعودی، در عدد 5 رقمی حاصل دیده می‌شوند؟ (GIS ـ ۸۶)

۱) $\frac{4}{15}$ ۲) $\frac{2}{5}$ ۳) $\frac{1}{5}$ ۴) $\frac{1}{3}$

۲۰. شش نفر کارمند را به طور تصادفی در اتاق‌های 1 نفره، 2 نفره و 3 نفره جای می‌دهیم. با کدام احتمال دو فرد مورد نظر در یک اتاق جای می‌گیرند؟ (GIS ـ ۸۷)

۱) $\frac{1}{6}$ ۲) $\frac{1}{12}$ ۳) $\frac{2}{15}$ ۴) $\frac{4}{15}$

۲۱. هر یک از ارقام $1, 2, 3, 5$ را بر روی 4 گوی یکسان نوشته و در ظرفی ریخته‌ایم. به تصادف یک گوی خارج کرده و با ثبت شماره آن دوباره به ظرف بر می‌گردانیم. با تکرار این عمل در سه بار متوالی به ترتیب ارقام عدد سه‌رقمی حاصل می‌شود. با کدام احتمال در این عدد سه‌رقمی لااقل یک بار عدد 2 وجود دارد؟ (GIS ـ ۸۷)

۱) $\frac{37}{64}$ ۲) $\frac{27}{64}$ ۳) $\frac{17}{32}$ ۴) $\frac{13}{32}$

۲۲. با حروف کلمه ABADAN یک کلمه رمز عبور 4 حرفی می‌سازیم. با کدام احتمال هر سه حرف A به کار رفته است؟ (GIS ـ ۸۸)

۱) $\frac{1}{3}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{6}$ ۴) $\frac{1}{9}$

احتمال باجایگذاری و بدون جایگذاری

۲۳. از بین 5 دانشجوی دختر و 3 دانشجوی پسر، سه دانشجو را برای شرکت در یک سمینار به تصادف انتخاب می‌کنیم. احتمال اینکه حداقل یک دانشجوی دختر انتخاب شود، برابر است با: (اقتصاد ـ ۸۶)

۱) $\frac{3}{56}$ ۲) $\frac{5}{56}$ ۳) $\frac{54}{56}$ ۴) $\frac{55}{56}$

۲۴.ظرفی شامل 4 مهره سفید و n مهره سیاه است $(n>1)$. دو مهره پی‌درپی بدون جایگذاری انتخاب می‌کنیم. n چقدر باشد تا احتمال اینکه مهره اول سفید و مهره دوم سیاه باشد، برابر $\frac{1}{5}$ شود؟ (اقتصاد ـ ۸۷)

۱) 1 ۲) 4 ۳) 5 ۴) 12

۲۵.از 10 پست در یک اداره، می‌خواهند 3 پست را به‌علت کمی مراجعه‌کننده حذف کنند. احتمال اینکه پست به‌خصوصی حذف نشود، چقدر است؟ (اقتصاد ـ ۸۸)

۱) $\frac{7}{9}$ ۲) $\frac{7}{10}$ ۳) $\frac{3}{9}$ ۴) $\frac{3}{10}$

۲۶.در یک جنگل کوچک 50 درخت وجود دارد به طوری که 12 درخت سرو، 8 درخت صنوبر، 4 درخت کاج و بقیه درخت اقاقیا. 4 درخت را به تصادف بریده‌ایم؛ احتمال اینکه دقیقاً 2 درخت سرو و 2 درخت کاج باشد چقدر است؟ (محیط زیست ـ ۸۶)

۱) $\frac{4}{50}$ ۲) $\frac{2}{12}\times\frac{2}{4}$ ۳) $\frac{12}{50}\times\frac{4}{50}$ ۴) $\frac{396}{\binom{50}{4}}$

۲۷.در یک سالن ورزشی 40 نفر ورزشکار وجود دارند که 10 نفر در حال انجام ورزش والیبال، 25 نفر در حال انجام ورزش فوتبال و بقیه در حال انجام ورزش بسکتبال هستند. می‌خواهیم یک کمیته 10 نفری انتخاب کنیم به طوری که 3 نفر از تیم والیبال، 5 نفر از تیم فوتبال و بقیه از تیم بسکتبال باشند. احتمال این رویداد کدام است؟ (محیط زیست ـ ۸۷)

۱) $\frac{\binom{10}{3}\binom{25}{5}\binom{5}{2}}{\binom{40}{10}}$ ۲) $\frac{\binom{10}{3}\binom{25}{5}\binom{5}{2}}{40^{10}}$ ۳) $\frac{10!25!5!}{40^{10}}$ ۴) $\frac{10!25!5!}{\binom{40}{10}}$

۲۸.فرض کنید از یک کلاسی که شامل 20 دانشجو می‌باشد (12 پسر و 8 دختر) یک کمیته 5 نفری که شامل 3 دانشجوی دختر و 2 دانشجوی پسر می‌باشد، انتخاب نماییم. احتمال چنین پیشامدی کدام است؟ (محیط زیست ـ ۸۸)

۱) $\frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}}$ ۲) $\frac{12!\,8!}{2!\,20!\,3!}$ ۳) $\binom{20}{5}\left(\frac{12}{20}\right)^2\left(\frac{8}{20}\right)^3$ ۴) $\binom{20}{5}\left(\frac{3}{8}\right)^3\left(\frac{2}{12}\right)^2$

۲۹.در کیسه اول 4 مهره قرمز، 2 مهره سفید و 4 مهره آبی وجود دارد و در کیسه دوم 3 مهره قرمز، 6 مهره سفید و 1 مهره آبی وجود دارد. از یکی از کیسه‌ها به تصادف مهره‌ای بیرون می‌آوریم، احتمال اینکه مهره سفید باشد، کدام است؟ (محیط زیست ـ ۸۸)

۱) $\frac{2}{10}$ ۲) $\frac{4}{10}$ ۳) $\frac{6}{10}$ ۴) $\frac{8}{10}$

احتمال متوسط

۳۰. در ظرف اول 1 مهره سفید و 4 مهره سیاه و در ظرف دوم 3 مهره سفید و 2 مهره سیاه وجود دارد. به‌تصادف یک مهره از ظرف اول برداشته بدون رؤیت در ظرف دوم قرار می‌دهیم. سپس از ظرف دوم دو مهره با هم خارج می‌کنیم. با کدام احتمال هر دو مهره خارج‌شده سفید است؟ (حسابداری و مدیریت ـ ۸۸)

۱) 0.12 ۲) 0.18 ۳) 024 ۴) 0.36

۳۱. **احتمال وجود سفره زیرزمینی نفتی در مناطق مختلف یک استان 0.4 است. احتمال برخورد چاه حفرشده به نفت حتی در حالت وجود سفره نفتی تنها 0.3 است. اگر یک چاه در این استان به‌تصادف حفر شود، احتمال عدم برخورد آن به نفت چقدر است؟** **(اقتصاد ـ ۸۸)**

۱) 0.12 ۲) 0.28 ۳) 0.30 ۴) 0.88

۳۲. **در کیسه A، 4 مهره سفید و 6 مهره آبی و در کیسه B، 7 مهره سفید و 3 مهره آبی وجود دارد. به‌تصادف از یکی از دو کیسه، مهره‌ای بیرون آورده شده و کنار گذاشته می‌شود (بدون نگاه کردن به رنگ آن). مهره دومی را بیرون می‌آوریم. احتمال اینکه این مهره سفید باشد، چقدر است؟** **(محیط زیست ـ ۸۷)**

۱) $\frac{4}{20}$ ۲) $\frac{7}{20}$ ۳) $\frac{11}{20}$ ۴) $\frac{5}{39}$

قضیه بیز

۳۳. **تعداد دانشجویان کلاس A دو برابر دانشجویان کلاس B است و نسبت دختران در این دو کلاس به ترتیب 0.4 و 0.6 است. اگر دختری به‌تصادف از این دو کلاس انتخاب شود، احتمال اینکه متعلق به کلاس A باشد چقدر است؟** **(اقتصاد ـ ۸۶)**

۱) 0.4 ۲) 0.6 ۳) 0.57 ۴) 0.86

۳۴. **اگر $P(A_1)=0.4$ و $P(A_2)=0.6$، $P(B|A_1)=0.2$ و $P(B|A_2)=0.05$ باشد، احتمال $P(A_1|B)$ عبارت است از:** **(اقتصاد ـ ۸۷)**

۱) 0.72 ۲) 0.27 ۳) 0.11 ۴) 0.03

۳۵. **در ظرفی 4 مهره سفید و 5 مهره سیاه قرار دارد. یک مهره از ظرف خارج کرده و مهره‌ای با رنگ دیگر به داخل ظرف می‌اندازیم؛ بار دوم مهره دیگری از ظرف خارج می‌کنیم. اگر هر دو مهره خارج‌شده همرنگ باشند با کدام احتمال هر دو سفید هستند؟** **(برنامه‌ریزی شهری ـ ۸۸)**

۱) $\frac{1}{4}$ ۲) $\frac{1}{3}$ ۳) $\frac{3}{8}$ ۴) $\frac{5}{16}$

۳۶. **سه ماشین A, B و C به ترتیب 35, 50 و 15 درصد محصولات کارخانه‌ای را تولید می‌کنند. محصولات آن‌ها به ترتیب 1, 2 و 3 درصد معیوب هستند. از میان محصولات این کارخانه یک محصول به‌تصادف انتخاب می‌کنیم. اگر این محصول معیوب باشد با کدام احتمال با ماشین C تولید شده است؟** **(GIS ـ ۸۷)**

۱) $\frac{1}{3}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{5}$ ۴) $\frac{1}{6}$

۳۷. **دستگاه فشارسنج در %90 اوقات روزهای بارانی را درست پیش‌بینی می‌کند. همچنین در %70 موارد روزهای آفتابی را درست پیش‌بینی می‌کند. می‌دانیم در یک شهر %40 روزها هوا بارانی است. فرض کنید دستگاه فشارسنج روز شنبه را بارانی پیش‌بینی کند، احتمال اینکه واقعاً باران ببارد چقدر خواهد بود؟** **(محیط زیست ـ۸۶)**

۱) %40 ۲) %90 ۳) $\frac{1}{3}$ ۴) $\frac{2}{3}$

پاسخ‌های تشریحی

آنالیز ترکیبی

۱. گزینه ۴ درست است.

یادآوری: تعداد تقسیمات n شیء در k سلول به طوری که n_1 تای آن‌ها در سلول اول، n_2 تای آن‌ها در سلول دوم، … و n_k تای آن‌ها در سلول kام قرار گیرد، برابر است با:

$$\binom{n}{n_1\ \ n_2\ \ \dots\ \ n_k}=\frac{n!}{n_1!\,n_2!\dots n_k!}$$

تنها حالت تقسیم 6 نفر به 3 شهر به طوری که تعداد افراد دو شهر یکسان نباشند $(1, 2, 3)$ است که البته با جایگشت $3!$ در بین سه شهر 6 حالت امکان‌پذیر است.

بنابراین می‌خواهیم 6 نفر را در 3 شهر تقسیم کنیم به طوری که در یکی از شهرها 1 و در دیگری 2 و در شهر دیگر 3 نفر قرار گیرد:

$$3!\times\binom{6}{1\ \ 2\ \ 3}=6\times\frac{6!}{1!\ 2!\ 3!}=6\times 60=360$$

۲. گزینه ۲ درست است.

یادآوری: تعداد کل حالات تقسیم n شیء متمایز در k گروه به طوری که گروه اول شامل n_1 عضو متمایز، … و گروه kام شامل n_k عضو متمایز باشد برابر است با:

$$\binom{n}{n_1 n_2 \dots n_k}=\frac{n!}{n_1!\,n_2!\dots n_k!}$$

حال تعداد حالات تقسیم 10 نفر به سه گروه $3, 5$ و 2 نفری برابر است با:

$$\binom{10}{5\ 3\ 2}=\frac{10!}{5!3!2!}=2520$$

۳. گزینه ۴ درست است.

کل دوچرخه‌سواران 43 نفر هستند که از بین این 43 نفر، 3 نفر مقام‌های اول و دوم و سوم را کسب می‌کنند. دقت کنید که ترتیب مقام‌ها برای ما مهم است، بنابراین تعداد حالات انتخاب این سه نفر ترتیب 3 از 43 خواهد بود.

$$P_n^r = \frac{n!}{(n-r)!} \rightarrow P_{43}^3 = \frac{43!}{(43-3)!} = \frac{43!}{40!} = 41 \times 42 \times 43 = 74046$$

به عبارت دیگر چون ترتیب مقام‌ها مهم است داریم:

$$\underset{\text{رتبه اول}}{41} \times \underset{\text{رتبه دوم}}{42} \times \underset{\text{رتبه سوم}}{43} = 74046$$

۴. گزینه ۴ درست است.

حروف کلمه MANAGEMENT شامل 6 حرف متمایز است که حروف E, N, A, M هر یک 2 بار تکرار شده است.

انتخاب سه حرف از 6 حرف متمایز: $\binom{6}{3} = \frac{6!}{3!3!} = 20$

حال چون هیچ محدودیتی در انتخاب این 3 حرف نیست پس این سه حرف می‌توانند مشابه نیز باشند.

اگر هر کدام از چهار حرف E, N, A, M دو بار تکرار شوند می‌توانند با یک حرف از 5 حرف متمایز دیگر 3 حرف بسازند که دو تای آن تکراری است، مثلاً: $(M,M,A),(M,M,N),(M,M,E)(M,M,G)(M,M,T)$

همین تکرارها برای E, N, A نیز وجود دارد، بنابراین:

انتخاب 3 حرف با 2 حرف تکراری و یک حرف غیر تکراری $= 4 \times 5 = 20$

بنابراین تعداد کل سه‌حرفی‌های انتخابی برابر است با $20+20=40$.

۵. گزینه ۳ درست است.

کلمه APPLICATION شامل 8 حرف متمایز (A,P,L,I,C,T,O,N) است که حروف P, A و I هر یک 2 بار تکرار شده‌اند.

می‌خواهیم رمز عبور سه‌حرفی تشکیل دهیم، بدون هیچ محدودیتی در مشابه بودن حروف رمز؛

تعداد حالات انتخاب 3 حرف متمایز از 8 حرف برابر است با:

$$\binom{8}{3} \times \underbrace{3!}_{\text{جایگشت 3 حرف}} = \frac{8!}{3!5!} \times 3! = 6 \times 7 \times 8 = 336$$

تعداد حالات انتخاب 2 حرف تکراری A یا P یا I با یک حرف متمایز دیگراز بین 7 حرف متمایز باقی‌مانده برابر است با:

$$\underbrace{3}_{\text{A یا P یا I}} \times \binom{7}{1} \times \underbrace{\frac{3!}{2!1!}}_{\text{جایگشت 3 حرف با 2 حرف مشابه}} = 3 \times 7 \times 3 = 63$$

بنابراین تعداد کل انتخاب‌ها برابر است با $336 + 63 = 399$.

۶. گزینه ۱ درست است.

یادآوری: انتخاب k شیء از n شیء بدون جایگذاری و بدون توجه به ترتیب انتخاب آن‌ها ترکیب k از n است:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

درکل باید به 10 پرسش از 12 تا پاسخ دهیم اما باید این 12 پرسش را به 5 تای اول و 7 تای بقیه تفکیک کنیم. سپس حداقل 4 تا از 5 تای اول را پاسخ دهیم و بقیه پاسخ‌ها باید از 7 تای باقی‌مانده باشد:

$$\binom{5}{4}\binom{7}{6} + \binom{5}{5}\binom{7}{5} = 5 \times 7 + 1 \times 21 = 56$$

۷. گزینه ۲ درست است.

کلمه OPERATOR شامل 8 حرف است که حروف O و R هر یک 2 بار تکرار شده‌اند.

انتخاب 4 حرف از 6 حرف متمایز (OPERAT) $\binom{6}{4}=15$

از 4 حرف انتخابی، 2 حرف O و 2 حرف دیگر از (PERAT) $\binom{2}{2}\binom{5}{2}=10$

از 4 حرف انتخابی، 2 حرف R، 2 حرف دیگر از (OPEAT) $\binom{2}{2}\binom{5}{2}=10$

4 حرف انتخابی شامل 2 حرف O و 2 حرف R باشد. $\binom{2}{2}\binom{2}{2}=1$

مجموع انتخاب‌ها: $15+10+10+1=36$

۸. گزینه ۲ درست است.

یادآوری: جایگشت n نفر در یک ردیف برابر با $n!$ و جایگشت n نفر دور یک میز برابر است با $(n-1)!$.

در این سؤال ابتدا افراد هر تیم را در کنار هم قرار می‌دهیم $(2!)^5$. حال 5 تیم به صورت $(5-1)!=4!$ می‌توانند دور میز قرار گیرند.

بنابراین کل حالات برابر است با: $4!(2!)^5=4!2^5$

روابط احتمالاتی

۹. گزینه ۳ درست است.

با توجه به اینکه موفقیت هر نفر مستقل از دیگری است، داریم:

راه حل اول: $P(\text{لااقل یکی موفق})=1-P(\text{هیچ موفق})=1-\left(1-\frac{1}{2}\right)\times\left(1-\frac{3}{5}\right)=1-\frac{1}{2}\times\frac{2}{5}=\frac{4}{5}$

راه حل دوم: $P(\text{لااقل یکی موفق})=P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{3}{5}+\frac{1}{2}-\frac{3}{5}\times\frac{1}{2}=\frac{8}{10}=\frac{4}{5}$

۱۰. گزینه ۲ درست است.

یادآوری: $P(A-B)=P(A\cap B')$

بنابراین: $P(B-A')=P\left(B\cap(A')'\right)=P(B\cap A)\rightarrow$ کافی است $P(A\cap B)$ را پیدا کنیم.

حال با توجه به اینکه $A\cup B$ برابر فضای نمونه است یعنی $P(A\cup B)=1$ داریم:

$P(A)+P(B)-P(A\cap B)=1\ \rightarrow\ 0.7+0.6-P(A\cap B)=1\ \rightarrow\ P(A\cap B)=0.3$

۱۱. گزینه ۲ درست است.

$A:$ علاقه‌مند به ادامه تحصیل $\rightarrow\ P(A)=\frac{15}{40}$

$B:$ علاقه‌مند به فوتبال $\rightarrow\ P(B)=\frac{20}{40}$

$A\cap B:$ علاقه‌مند به ادامه تحصیل و فوتبال $\rightarrow\ P(A\cap B)=\frac{5}{40}$

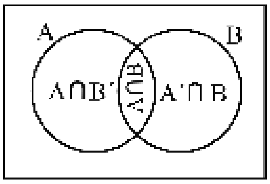
$P($علاقه‌مند نه به تحصیل و نه به فوتبال$) = P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$

$= 1 - P(A) - P(B) + P(A \cap B) = 1 - \frac{15}{40} - \frac{20}{40} + \frac{5}{40} = \frac{10}{40} = 0.25 = \%25$

یادآوری: قانون دمورگان

$$P\left(\bigcap_{i=1}^{n} A_i'\right) = P\left(\bigcup_{i=1}^{n} A_i\right)'$$

۱۲.گزینه ۳ درست است.



$A:$ حداقل مدرک کارشناسی $\rightarrow P(A) = 0.4$

$B:$ سابقه شغلی $\rightarrow P(B) = 0.6$

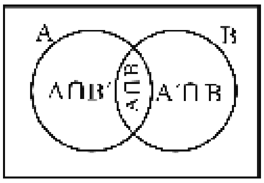
$A \cap B:$ حداقل کارشناسی و سابقه $\rightarrow P(A \cap B) = 0.15$

$P($سابقه نداشته و مدرک داشته$) = P(A \cap B') = P(A - B) = P(A) - P(A \cap B) = 0.4 - 0.15 = 0.25 = \%25$

۱۳. گزینه ۳ درست است.

$B:$ قبلاً در جایی دیگر کار کرده‌اند $\rightarrow P(B) = \frac{44}{100}$

$A \cap B:$ سابقه شغلی بیش از 5 سال در این کارخانه و قبلاً در جایی دیگر کار کرده‌اند $\rightarrow P(A \cap B) = \frac{15}{100}$



$P($سابقه شغلی بیش از 5 سال در این کارخانه نداشته و قبلاً در جایی دیگر کار کرده است$) = P(B \cap A') = P(B - A)$

$= P(B) - P(A \cap B) = \frac{44}{100} - \frac{15}{100} = \frac{29}{100}$

احتمال

۱۴. گزینه ۱ درست است.

$A:$ مجموع دو عدد مضرب 3 باشد.

$A = \{(1,2),(1,5),(2,4),(3,6),(4,5)\}$

$$P(A) = \frac{n(A)}{n(S)} = \frac{5}{\binom{6}{2}} = \frac{5}{15} = \frac{1}{3}$$

۱۵. گزینه ۳ درست است.

یادآوری: جایگشت n شیء که n_1 تای آن‌ها شبیه هم، n_2 تای آن‌ها شبیه هم، … و n_k تای آن‌ها شبیه هم باشد برابر است با:

$$\binom{n}{n_1\ n_2\ \dots\ n_k} = \frac{n!}{n_1!n_2!\dots n_k!}$$

بنابراین جایگشت کل ارقام $1,1,2,2,3,3$ برابر است با:
$$n(s)=\frac{6!}{2!2!2!}=\frac{720}{8}=90$$
برای تعداد حالات مساعد بهتر است حالات را بررسی کنیم. می‌خواهیم بین هر دو رقم یکسان، دو رقم متمایز باشد یعنی:

$$\overbrace{\quad\quad}^{\text{بین دو } 2}$$
$$\boxed{1\ \ \underbrace{2\ \ \ 3}_{\text{بین دو } 1}}\ \boxed{\underbrace{1\ \ \ 2}_{\text{بین دو } 3}\ \ 3}$$

دقت کنید که سه رقم اول در این حالت با سه رقم دوم دقیقاً برابر و در جای یکسان هستند. در تمامی حالات مساعد این اتفاق می‌افتد پس تعداد حالات مساعد برابر جایگشت 3 رقم اول است. به حالات مساعد نگاه کنید:

1	2	3	1	2	3
1	3	2	1	3	2
2	3	1	2	3	1
2	1	3	2	1	3
3	1	2	3	1	2
3	2	1	3	2	1

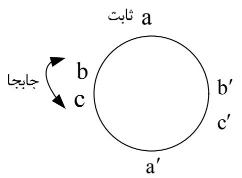
$$\begin{cases} P(A)=\dfrac{n(A)}{n(s)}=\dfrac{6}{90}=\dfrac{1}{15} \\ n(A)=3!=6 \end{cases}$$

۱۶. گزینه ۱ درست است.

یادآوری: جایگشت n نفر دور یک میز (دایره‌ای) برابر $(n-1)!$ است، بنابراین کل این 6 نفر به $(6-1)!$ حالت می‌توانند دور میز قرار گیرند:

کل حالات $=(6-1)!=5!$

حالات مساعد: برای قرار گرفتن افراد هر گروه مقابل یکدیگر ابتدا یک نفر را در جای ثابتی می‌نشانیم، سپس نفر دیگر گروهش را مقابلش قرار می‌دهیم:



حال برای دو گروه باقی‌مانده 2 حالت بیشتر وجود ندارد که مقابل هم باشند و افراد هر گروه نیز به !2 می‌توانند با هم جابجا شوند؛ بنابراین تعداد حالات مساعد برابر است با $2\times(2!)^2$.

$$\text{احتمال مورد نظر}=\frac{\text{حالات مساعد}}{\text{کل حالات}}=\frac{2\times(2!)^2}{5!}=\frac{8}{120}=\frac{1}{15}$$

۱۷. گزینه ۲ درست است.

می‌خواهیم با ارقام $1, 2, \dots, 9$ یک رمز 6 رقمی بسازیم.

کل حالات $=\boxed{9}\times\boxed{9}\times\boxed{9}\times\boxed{9}\times\boxed{9}\times\boxed{9}=9^6$

حال می‌خواهیم در این رمز فقط 3 رقم مضرب 3 ، یک رقم مضرب 4 و 2 رقم دیگر از بین مابقی ارقام باشد، بنابراین:

ارقام مضرب 3 : $3 , 6 , 9$

ارقام مضرب 4 : $4 , 8$

مابقی ارقام : $1 , 2 , 5 , 7$

ابتدا 3 مکان از 6 مکان را انتخاب می‌کنیم $\binom{6}{3}$ و ارقام مضرب 3 را در آن قرار می‌دهیم، سپس 1 مکان از 3 مکان باقی‌مانده انتخاب می‌کنیم $\binom{3}{1}$ و ارقام مضرب 4 را در آن قرار می‌دهیم و در آخر 2 مکان باقی می‌ماند که ارقام باقی‌مانده را در آن جایگذاری می‌کنیم:

$$\underbrace{\boxed{3}\boxed{3}\boxed{3}}_{\text{مضرب 3}}\quad \underbrace{\boxed{2}}_{\text{مضرب 4}}\quad \underbrace{\boxed{4}\boxed{4}}_{\text{مابقی}}$$

$$\text{حالات مساعد}: \underbrace{\binom{6}{3}\times 3^3}_{\text{3 مکان مضرب 3}} \times \underbrace{\binom{3}{1}\times 2}_{\text{1 مکان مضرب 4}} \times \underbrace{\binom{2}{2}\times 4^2}_{\text{2 مکان مابقی}}$$

$$\text{احتمال مورد نظر} = \frac{\text{حالات مساعد}}{\text{کل حالات}} = \frac{\binom{6}{3}3^3\times\binom{3}{1}2\times 4^2}{9^6} = \frac{20\times \cancel{3^4}\times 2\times 4^2}{\cancel{9^6}\,9^4} = \frac{640}{81\times 81}$$

۱۸. گزینه ۲ درست است.

یادآوری: فضای نمونه پرتاب n سکه برابر 2^n است:

$$S:\{(T,T),\underbrace{(T,H),(H,T),(H,H)}_{\text{حداقل یک شیر}}\}\to n(S)=2^2=4$$

$$P(\text{حداقل یک شیر})=1-P(\text{هیچ شیر})=1-P(\text{هردو خط})=1-\frac{1}{4}=\frac{3}{4}$$

۱۹. گزینه ۳ درست است.

یادآوری: جایگشت n شیء با n_1 شیء مشابه، n_2 شیء مشابه، ... و n_k شیء مشابه برابر است با: $\frac{n!}{n_1!n_2!\ldots n_k!}$

$1,1,2,2,3$ تعداد اعداد 5 رقمی حاصل از ارقام : $n(s)=\frac{5!}{2!2!1!}=30$

حال برای به دست آوردن تعداد حالات مساعد باید سه رقم متوالی صعودی 123 را در یک دسته کنار هم قرار دهیم؛ حال جایگشت این دسته را با دو رقم باقی‌مانده و 2 محاسبه کنیم: $123\,,1\,,2\to n(A)=3!=6$

$$P(A)=\frac{n(A)}{n(S)}=\frac{6}{30}=\frac{1}{5}$$

۲۰. گزینه ۴ درست است.

یادآوری: تعداد حالت تقسیم n شیء متمایز در k سلول به طوری که در سلول اول n_1 شیء، در سلول دوم n_2 شیء و ... در سلول kام n_k شیء قرار گیرد برابر است با: $\frac{n!}{n_1!n_2!\ldots n_k!}$

کل حالات: تقسیم 6 نفر در اتاق‌های 2, 1 و 3 نفره: $n(s)=\frac{6!}{3!2!1!}=60$

حالات مساعد: یک بار 2 نفر را در اتاق دونفره قرار می‌دهیم و 4 نفر باقی‌مانده را در اتاق سه‌نفره و یک‌نفره تقسیم می‌کنیم $\left(\frac{4!}{3!\,1!}\right)$. بار دیگر 2 نفر را در اتاق سه‌نفره قرار می‌دهیم و 4 نفر باقی‌مانده را در اتاق یک‌نفره و دونفره و 1 نفره باقی‌مانده از اتاق سه‌نفره جای می‌دهیم $\left(\frac{4!}{1!\,2!\,1!}\right)$.

$$n(A)=\frac{4!}{3!\,1!}+\frac{4!}{2!\,1!\,1!}=4+12=16$$

$$p(A)=\frac{n(A)}{n(s)}=\frac{16}{60}=\frac{4}{15}$$

دقت کنید که چون دو نفر مورد نظر مشخص است دیگر به انتخاب 2 نفر از 6 نفر $\binom{6}{2}$ نیازی نیست.

۲۱.گزینه ۱ درست است.

کل حالات: تعداد کل اعداد سه‌رقمی با ارقام $1, 2, 3, 5$ برابر است با: $\boxed{4}\times\boxed{4}\times\boxed{4}=4^3$

حالات مساعد: برای تعداد حالات مساعد یعنی عدد سه‌رقمی با حداقل یک رقم 2 بهتر است تعداد حالاتی را که در عدد سه رقمی، رقم 2 وجود ندارد، به دست آورده و سپس از کل حالات کم کنیم.

عدد سه‌رقمی با ارقام $1, 3, 5$: $\boxed{3}\times\boxed{3}\times\boxed{3}=3^3 \rightarrow$ عدد سه‌رقمی با حداقل یک رقم 2 $=4^3-3^3=37$

$$P(\text{عدد سه رقمی با حداقل یک }2)=\frac{\text{حالات مساعد}}{\text{کل حالات}}=\frac{37}{64}$$

۲۲.گزینه ۳ درست است.

حروف کلمه ABADAN شامل 6 حرف است با 4 حرف متمایز (A ,B ,D ,N) و 3 تکرار A

کل حالات: تمام حالاتی که می‌توان یک رمز 4حرفی ساخت عبارت‌اند از:

انتخاب 4 حرف از بین 4 حرف متمایز: $\binom{4}{4}\times\underbrace{4!}_{\text{جایگشت 4 حرف متمایز}}=24$

انتخاب 2 حرف A و 2 حرف متمایز از بین سه حرف B, D, N: $\binom{3}{2}\times\underbrace{\frac{4!}{2!\,1!\,1!}}_{\text{جایگشت 4 حرف با 2 حرف تکراری}}=3\times12=36$

انتخاب 3 حرف A و یک حرف متمایز از بین سه حرف B, D, N: $\binom{3}{1}\times\underbrace{\frac{4!}{3!\,1!}}_{\text{جایگشت 4 حرف با 3 حرف تکراری}}=3\times4=12$ **(حالات مساعد)**

$$\text{احتمال مورد نظر}=\frac{\text{حالات مساعد}}{\text{کل حالات}}=\frac{12}{24+36+12}=\frac{12}{72}=\frac{1}{6}$$

احتمال باجایگذاری و بدون جایگذاری

۲۳.گزینه ۴ درست است.

$$P(\text{حداقل یک دختر}) = 1 - P(\text{هیچ دختر}) = 1 - \frac{\binom{5}{0}\binom{3}{3}}{\binom{8}{3}} = 1 - \frac{1}{56} = \frac{55}{56}$$

۲۴.گزینه ۴ درست است.

جعبه شامل 4 مهره سفید و n مهره سیاه است بنابراین کل مهره‌های جعبه برابر $n+4$ است. دقت کنید که انتخاب مهره‌ها بدون جایگذاری است؛ بنابراین در هر بار انتخاب، یک مهره از کل مهره‌های جعبه کم خواهد شد. همچنین چون ترتیب رنگ مهره‌ها مهم است بنابراین تک تک حالات را بررسی می‌کنیم:

$$P(\text{اولی سفید، دومی سیاه}) = \frac{\text{تعداد سفیدها}}{\text{کل مهره‌ها}} \times \frac{\text{تعداد سیاه‌ها}}{\text{کل مهره‌ها} - 1} = \frac{4}{4+n} \times \frac{n}{4+n-1} = \frac{1}{5}$$

$$\rightarrow 20n = (4+n)(3+n) \rightarrow n^2 - 13n + 12 = 0 \rightarrow (n-12)(n-1) = 0 \begin{cases} n = 12 \checkmark \\ n = 1 \end{cases}$$

$n=12$ قابل قبول است زیرا $n>1$ است (اگر حل معادله برایتان سخت بود می‌توانید گزینه‌ها را در معادله جایگزین کنید).

۲۵.گزینه ۲ درست است.

کل حالات: انتخاب 3 پست از 10 پست: $\binom{10}{3}$

حالات مساعد: پستی را که می‌خواهیم حذف نشود کنار می‌گذاریم و 3 پست را از بین 9 تای باقی‌مانده انتخاب می‌کنیم $\binom{9}{3}$.

$$P(A) = \frac{\text{حالات مساعد}}{\text{کل حالات}} = \frac{\binom{9}{3}}{\binom{10}{3}} = \frac{\frac{9!}{3!\,6!}}{\frac{10!}{3!7!}} = \frac{7}{10}$$

۲۶.گزینه ۴ درست است.

انتخاب k شیء از n شیء بدون جایگذاری و بدون مهم بودن ترتیب انتخاب آن‌ها ترکیب $\binom{n}{k}$ خواهد بود:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

کل حالات : انتخاب 4 تا از 50 $= \binom{50}{4}$

حالات مساعد : انتخاب 2 درخت سرو و 2 درخت کاج $= \binom{12}{2}\binom{4}{2} = \frac{12!}{2!\,10!} \times \frac{4!}{2!\,2!} = 66 \times 6 = 396$

$$\text{احتمال مورد نظر} = \frac{\text{حالات مساعد}}{\text{کل حالات}} = \frac{396}{\binom{50}{4}}$$

۲۷. گزینه ۱ درست است.

انتخاب k نفر از n نفر بدون جایگذاری و بدون مهم بودن ترتیب انتخاب آن‌ها، ترکیب $\binom{n}{k}$ خواهد بود:

کل حالات : انتخاب 10 نفر از 40 نفر $=\binom{40}{10}$

حالت مساعد: انتخاب 3 والیبالیست، 5 فوتبالیست و 2 بسکتبالیست

$40-10-25=$ بقیه از 40 ورزشکار

$$\binom{10}{3}\binom{25}{5}\binom{5}{2}$$

$10-3-5=$ بقیه از 10 نفر

$$\text{احتمال مورد نظر}=\frac{\text{حالات مساعد}}{\text{کل حالات}}=\frac{\binom{10}{3}\binom{25}{5}\binom{5}{2}}{\binom{40}{10}}$$

۲۸. گزینه ۱ درست است.

انتخاب k نفر از n نفر بدون جایگذاری و بدون مهم بودن ترتیب انتخاب آن‌ها، ترکیب $\binom{n}{k}$ خواهد بود:

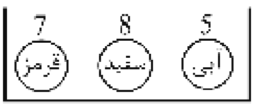
کل حالات : انتخاب 5 نفر از 20 نفر $=\binom{20}{5}$

حالات مساعد : انتخاب 3 دختر از 8 دختر و 2 پسر از 12 پسر $=\binom{12}{2}\binom{8}{3}$

$$\text{احتمال مورد نظر}=\frac{\text{حالات مساعد}}{\text{کل حالات}}=\frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}}$$

۲۹. گزینه ۲ درست است.

توجه: با توجه به اینکه انتخاب کیسه‌ها تصادفی است دو کیسه روی هم ریخته و یک کیسه با 7 مهره قرمز، 8 مهره سفید و 5 مهره آبی (جمعاً 20 مهره) در نظر می‌گیریم:

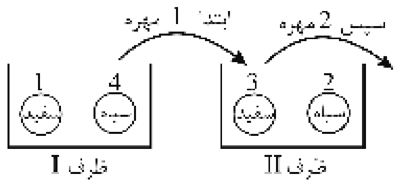


$$P(\text{سفید})=\frac{\text{تعداد مهره‌های سفید}}{\text{تعداد کل مهره‌ها}}=\frac{8}{20}=\frac{4}{10}$$

احتمال متوسط

۳۰. گزینه ۳ درست است.

با شرط بر روی سفید یا سیاه بودن مهره خارج‌شده از ظرف اول (I) و با استفاده از فرمول احتمال متوسط داریم:



$$P(\text{هر دو مهره سفید}) = P(\text{اولی سفید})P(\text{اولی سفید} \mid \text{2مهره سفید}) + P(\text{اولی سیاه})P(\text{اولی سیاه} \mid \text{2مهره سفید})$$

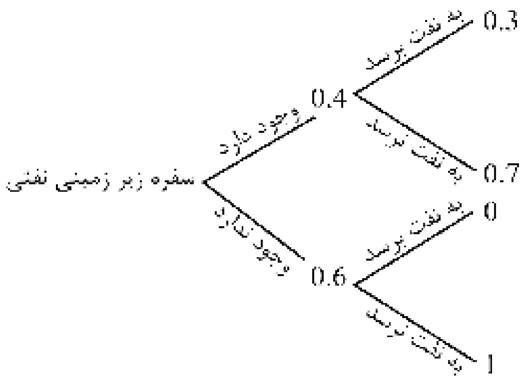
$$= \frac{1}{5} \times \frac{\binom{4}{2}}{\binom{6}{2}} + \frac{4}{5} \times \frac{\binom{3}{2}}{\binom{6}{2}} = \frac{1}{5} \times \frac{6}{15} + \frac{4}{5} \times \frac{3}{15} = \frac{6+12}{5\times 15} = \frac{18}{5\times 15} = \frac{6}{25} = 0.24$$

دقت کنید که ظرف (II) با اضافه شدن مهره خارج‌شده از ظرف (I)، 6 مهره خواهد داشت.

۳۱. گزینه ۴ درست است.

E : چاه به نفت نرسد

A : وجود سفره زیر زمینی



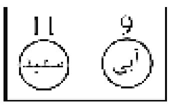
با توجه به فرمول احتمال متوسط با شرط بر روی وجود سفره زیر زمینی داریم:

$$P(E) = P(E \mid A)P(A) + P(E \mid A')P(A') = 0.7 \times 0.4 + 1 \times 0.6 = 0.28 + 0.6 = 0.88$$

۳۲. گزینه ۳ درست است.

راه حل تستی: با توجه به اینکه از رنگ مهره اول اطلاعی نداریم بنابراین مهره دوم مستقل از مهره اول انتخاب شده و احتمال آن همانند خارج شدنش در دفعه اول محاسبه می‌شود.

دقت کنید که چون انتخاب هر جعبه با شانس مساوی است می‌توانیم دو جعبه را روی هم ریخته یک جعبه با 11 مهره سفید و 9 مهره آبی (جمعاً 20 مهره) در نظر بگیریم:



$$P(\text{سفید}) = \frac{\text{تعداد مهره‌های سفید}}{\text{تعداد کل مهره‌ها}} = \frac{11}{20}$$

روش شرطی: با توجه به اینکه احتمال سفید بودن مهره دوم را می‌خواهیم باید بر روی رنگ مهره اول شرط بگذاریم؛ از فرمول احتمال متوسط داریم:

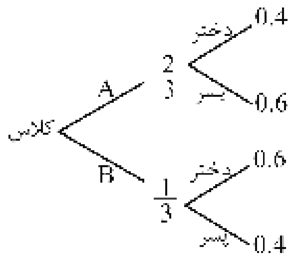
$$P(\text{دومی سفید}) = P(\text{اولی سفید})P(\text{دومی سفید} \mid \text{اولی سفید}) + P(\text{اولی آبی})P(\text{دومی سفید} \mid \text{اولی آبی})$$

$$= \frac{11}{20}\times\frac{10}{19}+\frac{9}{20}\times\frac{11}{19}=\frac{209}{380}=\frac{11}{20}$$

قضیه بیز

۳۳. گزینه ۳ درست است.

تعداد دانشجویان کلاس A دو برابر کلاس B است $\leftarrow$ $P(B\ \text{انتخاب کلاس}) = \frac{1}{3}\ ,\ P(A\ \text{انتخاب کلاس}) = \frac{2}{3}$



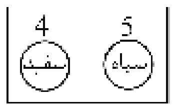
$$P(A\ \text{کلاس} \mid \text{دختر}) = \frac{P(\text{دختر} \mid A\ \text{کلاس})P(A\ \text{کلاس})}{P(\text{دختر} \mid A\ \text{کلاس})P(A\ \text{کلاس}) + P(\text{دختر} \mid B\ \text{کلاس})P(B\ \text{کلاس})} = \frac{0.4\times\frac{2}{3}}{0.4\times\frac{2}{3}+0.6\times\frac{1}{3}}=\frac{8}{14}=0.57$$

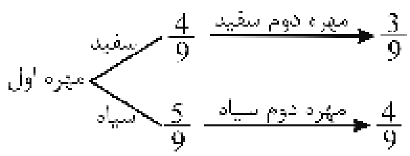
۳۴. گزینه ۱ درست است.

با توجه به قضیه بیز داریم:

$$\begin{cases} P(A_1 \mid B) = \dfrac{P(B\mid A_1)P(A_1)}{P(B\mid A_1)P(A_1)+P(B\mid A_2)P(A_2)} = \dfrac{0.2\times 0.4}{0.2\times0.4+0.05\times0.6}=\dfrac{8}{11}=0.72 \\ P(B\mid A_1)=0.2\ ,\ P(A_1)=0.4\ ,\ P(B\mid A_2)=0.05\ ,\ P(A_2)=0.6 \end{cases}$$

۳۵. گزینه ۳ درست است.





با توجه به قضیه بیز داریم:

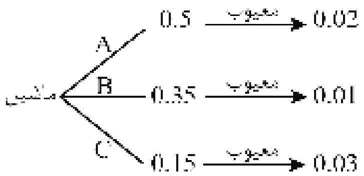
$$P(\text{هر دو سفید} \mid \text{هر دو همرنگ}) = \frac{P(\text{هر دو همرنگ} \cap \text{هر دو سفید})}{P(\text{هر دو همرنگ})} = \frac{P(\text{هر دو سفید})}{P(\text{هر دو سفید})+P(\text{هر دو سیاه})}$$

$$= \frac{P(\text{اولی سفید})\times P(\text{دومی سفید} \mid \text{اولی سفید})}{P(\text{اولی سفید})\times P(\text{دومی سفید} \mid \text{اولی سفید}) + P(\text{اولی سیاه})\times P(\text{دومی سیاه} \mid \text{اولی سیاه})} = \frac{\frac{4}{9}\times\frac{3}{9}}{\frac{4}{9}\times\frac{3}{9}+\frac{5}{9}\times\frac{4}{9}}=\frac{12}{12+20}=\frac{12}{32}=\frac{3}{8}$$

دقت کنید که وقتی مهره اول با احتمال $\frac{4}{9}$ سفید خارج می‌شود یک مهره سیاه به‌جای آن مهره سفید خارج‌شده به ظرف برمی‌گردد یعنی الان در ظرف 3 سفید و 6 سیاه موجود است که احتمال سفید خارج شدنش در این مرحله (بار دوم) $\frac{3}{9}$ است.

همچنین وقتی مهره اول با احتمال $\frac{5}{9}$ سیاه خارج می‌شود یک مهره سفید به‌جای آن مهره سیاه خارج‌شده به ظرف برمی‌گردد یعنی الان در ظرف 4 سیاه و 5 سفید موجود است که احتمال سیاه خارج شدنش در این مرحله $\frac{4}{9}$ است.

۳۶.گزینه ۲ درست است.

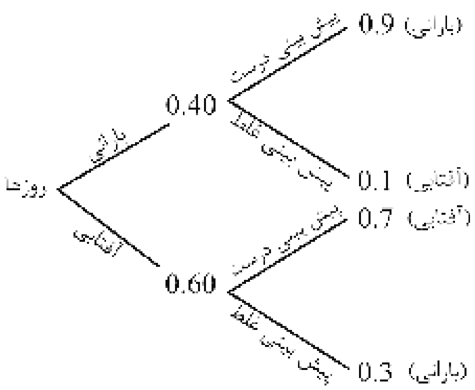


E : معیوب بودن

با توجه به قضیه بیز داریم:

$$P(C|E)=\frac{P(E|C)P(C)}{P(E|A)P(A)+P(E|B)P(B)+P(E|C)P(C)}$$
$$=\frac{0.03\times0.15}{0.02\times0.5+0.01\times0.35+0.03\times0.15}=\frac{45}{180}=\frac{1}{4}$$

۳۷. گزینه ۴ درست است.



A : پیش‌بینی روز بارانی
E : واقعاً باران ببارد

با توجه به قضیه بیز داریم:

$$P(\text{واقعاً بارانی} \mid \text{پیش‌بینی بارانی})=P(E|A)=\frac{P(A\cap E)}{P(A)}=\frac{P(A|E)P(E)}{P(A|E)P(E)+P(A|E')P(E')}$$
$$=\frac{0.9\times0.4}{0.9\times0.4+0.3\times0.6}=\frac{36}{36+18}=\frac{36}{54}=\frac{2}{3}$$

خودآزمایی

آنالیز ترکیبی

۱. به چند طریق می‌توان با اعداد صفر تا 9، شماره تلفن 6 رقمی ساخت؟

۱) 10^6 ۲) 9^6 ۳) 9×10^5 ۴) 6×10^5

۲. به چند طریق می‌توان از 12 کتاب که 5 تای آن آمار و بقیه ریاضی هستند، یک کتاب آمار و 2 کتاب ریاضی را به عنوان کتاب سال برگزید؟

۱) 220 ۲) 205 ۳) 110 ۴) 105

۳. دانشجویی موظف است از 5 سؤال اول به 3 سؤال و از 15 سؤال بعد به 12 سؤال جواب دهد. به چند طریق می‌تواند به سؤالات جواب دهد؟

۱) 5054 ۲) 5540 ۳) 4550 ۴) 5450

۴. برای چراغانی کردن سر در یک شرکت، 2 لامپ قرمز، 3 لامپ زرد و 4 لامپ آبی موجود است. با قرار دادن این لامپ‌ها در یک ردیف به چند شکل می‌توان چراغانی کرد؟

۱) 1080 ۲) 1260 ۳) 1170 ۴) 1350

۵. با حروف کلمه EHSAN چند کلمه 3 حرفی می‌توان ساخت به‌گونه‌ای که شامل حرف A باشند؟

۱) 24 ۲) 30 ۳) 12 ۴) 36

۶. به چند طریق می‌توان یک گروه حداقل 2 نفره از بین 7 نفر انتخاب کرد؟

۱) 119 ۲) 120 ۳) 101 ۴) 146

۷. به چند طریق می‌توان از بین اعضای 12 نفره تیمی، 3 نفر را برای مقام‌های اول تا سوم انتخاب کرد؟

۱) 110 ۲) 220 ۳) 1100 ۴) 1320

۸. چند عدد 5 رقمی یا 6 رقمی می‌توان با رقم‌های $4, 3, 2, 2, 2, 1$ درست کرد؟

۱) 120 ۲) 240 ۳) 720 ۴) 1440

روابط احتمالاتی

۹. اگر $P(A) = 0.59$ و $P(B) = 0.3$ و $P(A \cap B) = 0.21$، آن‌گاه مقدار $P(A \cap \bar{B})$ کدام است؟ ($\bar{B}$ متمم B است.)

۱) 0.56 ۲) 0.38 ۳) 0.28 ۴) 0.18

۱۰. اگر A و B دو پیشامد ناسازگار باشند و احتمال آن دو به ترتیب a و b باشد مقدار $P(A^c \cap B^c)$ کدام است؟ (A^c متمم A است.)

۱) $b - a$ ۲) $1 - ab$ ۳) $1 - a - b$ ۴) $1 + a + b$

۱۱. در یک دانشکده، %50 دانشجویان فوتبال، %40 بسکتبال، %30 هم فوتبال و هم بسکتبال بازی می‌کنند. احتمال آنکه دانشجویی در این دانشکده ورزش نکند کدام است؟

۱) 0 ۲) 0.1 ۳) 0.4 ۴) 0.6

۱۲. سه سکه را که شانس شیر آمدن آن‌ها $\frac{1}{2}$، $\frac{1}{3}$ و $\frac{1}{4}$ است با هم می‌اندازیم. احتمال اینکه دست‌کم یک شیر دیده شود برابر است با:

۱) $\frac{3}{4}$ ۲) $\frac{3}{8}$ ۳) $\frac{1}{12}$ ۴) $\frac{5}{12}$

۱۳. احتمال کار کردن یک نفر در یک شرکت معین برای مدت بیشتر از 10 سال برابر $\frac{1}{6}$ است. اگر دو فرد A و B کار خود را همزمان در این شرکت شروع کنند، احتمال اینکه فقط یک نفر از آن‌ها بیشتر از 10 سال در شرکت بماند، کدام است؟

۱) $\frac{11}{36}$ ۲) $\frac{5}{36}$ ۳) $\frac{25}{36}$ ۴) $\frac{5}{18}$

احتمال

۱۴. احتمال آنکه مقداری که به‌صورت تصادفی از جامعه‌ای انتخاب می‌شود بیشتر از میانگین آن جامعه باشد چقدر است؟

۱) 0.25 ۲) 0.5 ۳) 1 ۴) 0.67

۱۵. سیستمی دارای دو جزء است که احتمال کار نکردن هر یک از آن‌ها 0.20 است. اگر اجزا به صورت موازی قرار گرفته باشند، احتمال کار کردن سیستم چقدر است؟

۱) 0.96 ۲) 0.04 ۳) 0.40 ۴) 0.64

۱۶. فرض کنید احتمال داشتن فرزند پسر و دختر مساوی باشد. احتمال اینکه خانواده‌ای که 3 فرزند دارد، حداقل یک فرزند پسر داشته باشد چقدر است؟

۱) $\frac{5}{8}$ ۲) $\frac{6}{8}$ ۳) $\frac{7}{8}$ ۴) 1

۱۷. دو تاس را با هم پرتاب می‌کنیم. احتمال اینکه عدد بالا قرارگرفته از تاس اول کوچک‌تر از عدد بالا قرارگرفته از تاس دوم باشد چقدر است؟

۱) 0.42 ۲) 0.55 ۳) 0.20 ۴) 0.69

۱۸. یک عدد 3 رقمی را به‌تصادف انتخاب می‌کنیم. کدام‌یک از موارد زیر، احتمال آن است که این عدد حداقل یک رقم 1، داشته باشد؟

۱) 0.72 ۲) 0.029 ۳) 0.28 ۴) 0.69

احتمال با جایگذاری و بدون جایگذاری

۱۹. از فارغ‌التحصیلان کارشناسی ارشد دانشکده‌ای در یک دوره، 6 نفر شاغل هستند و 4 نفر سر کار نرفته‌اند. در صورت تماس تصادفی با 7 نفر آنان، چند درصد احتمال دارد بتوان از 2 نفرشان دعوت به کار کرد؟

۱) 30 ۲) 50 ۳) 40 ۴) 20

۲۰. از 10 واحد کالای برگشتی از فروش هر روزه در یک شرکت، تعداد 7 واحد به‌تصادف انتخاب و بازرسی می‌شوند. هرگاه در یک روز، تعداد 6 واحد از کالاهای برگشتی واقعاً معیوب باشد، احتمال وجود 3 واحد کالای معیوب در نمونه انتخابی چقدر است؟

۱) $\frac{5}{6}$ ۲) $\frac{6}{7}$ ۳) $\frac{1}{6}$ ۴) $\frac{3}{7}$

۲۱. از 12 عدد کالای همگن داخل جعبه‌ای، 4 عدد معیوب است؛ یک مشتری 3 عدد آن‌ها را به‌طور تصادفی یکی بعد از دیگری برداشته و بررسی می‌کند. احتمال آنکه 2 عدد اول سالم و سومی معیوب باشد چقدر است؟

۱) $\frac{28}{165}$ ۲) $\frac{4}{27}$ ۳) $\frac{28}{55}$ ۴) $\frac{4}{16}$

۲۲. از حروف کلمه ORIGIN به طور تصادفی دو حرف حذف می‌کنیم. با کدام احتمال حداقل یکی از حروف حذف‌شده، I است؟

۱) $\frac{2}{5}$ ۲) $\frac{3}{5}$ ۳) $\frac{4}{10}$ ۴) $\frac{5}{11}$

۲۳. در جعبه‌ای، 4 توپ سفید و 8 توپ سیاه قرار دارد. دو توپ به‌تصادف و بدون جایگذاری از آن خارج می‌کنیم. کدام‌یک از موارد زیر، احتمال آن را نشان می‌دهد که توپ دوم سفید باشد؟

۱) $\frac{1}{1}$ ۲) $\frac{1}{12}$ ۳) $\frac{1}{3}$ ۴) $\frac{2}{11}$

۲۴. اگر درون کیسه‌ای 20 عدد مهره سفید و 30 عدد مهره سیاه باشد و از درون آن 2 عدد مهره با جایگزینی برداریم، احتمال اینکه هر دو مهره سفید باشد، برابر است با:

۱) $\frac{2}{25}$ ۲) $\frac{3}{25}$ ۳) $\frac{4}{25}$ ۴) $\frac{1}{5}$

۲۵. در یک میهمانی، شش زوج ازدواج کرده، شامل 6 مرد و همسران آن‌ها شرکت دارند. اگر به‌طور تصادفی دو نفر از بین آن‌ها انتخاب کنیم، احتمال آنکه این دو نفر زن و شوهر باشند، کدام است؟

۱) $\frac{1}{12}$ ۲) $\frac{1}{11}$ ۳) $\frac{2}{11}$ ۴) $\frac{1}{6}$

احتمال شرطی

۲۶. ظرفی حاوی 5 مهره قرمز و 7 مهره سفید است. مهره‌ای را به طور تصادفی از ظرف خارج می‌کنیم؛ اگر قرمز بود به همراه آن یک مهره قرمز دیگر و اگر سفید بود به همراه آن 2 مهره سفید دیگر به داخل ظرف می‌اندازیم و سپس مهره دوم را برحسب تصادف خارج می‌کنیم. اگر بدانیم مهره خارج‌شده در بار اول سفید بوده است، احتمال آنکه مهره دوم قرمز باشد چقدر است؟

۱) $\frac{9}{14}$ ۲) $\frac{5}{12}$ ۳) $\frac{9}{12}$ ۴) $\frac{5}{14}$

۲۷. اگر $P(A)=\frac{1}{3}\,,\,P(B)=\frac{1}{5}\,,\,P(A|B)=\frac{1}{3}$ باشد، $P(A\cup B)$ چقدر است؟

۱) $\frac{6}{15}$ ۲) $\frac{7}{15}$ ۳) $\frac{8}{15}$ ۴) $\frac{9}{15}$

۲۸. 120 دانشجو طبق جدول زیر توزیع شده‌اند. یک دانشجو به‌تصادف انتخاب می‌کنیم که مرد است؛ احتمال اینکه رشته ریاضی باشد، چقدر است؟

۱) $\frac{1}{12}$ ۲) $\frac{1}{3}$ ۳) $\frac{2}{3}$ ۴) $\frac{1}{6}$

جنس / رشته	مرد	زن	جمع
فیزیک	20	30	50
ریاضی	10	60	70
جمع	30	90	120

۲۹. از جعبه‌ای شامل 3 خودکار سبز، 4 خودکار قرمز و 5 خودکار مشکی هم‌اندازه، یک خودکار به‌تصادف برداشته و مشاهده شده که مشکی نیست؛ احتمال قرمز بودن آن چقدر است؟

۱) $\frac{4}{7}$ ۲) $\frac{4}{12}$ ۳) $\frac{4}{8}$ ۴) $\frac{7}{12}$

۳۰. کیسه‌ای شامل دو گلوله سفید و دو گلوله سیاه است. از این کیسه دو گلوله به طور تصادفی خارج می‌کنیم. احتمال سفید بودن هر دو گلوله به شرطی که اقلاً یکی از آن‌ها سفید باشد برابر است با:

۱) $\frac{1}{2}$ ۲) $\frac{1}{3}$ ۳) $\frac{1}{4}$ ۴) $\frac{1}{5}$

۳۱. در یک بار پرتاب تاسی که احتمال نتیجه زوج داشتن آن سه برابر احتمال فرد داشتن آن است، اگر بدانیم نتیجه‌ای بزرگ‌تر از سه به دست آمده، با چه احتمالی نتیجه پرتاب تاس مربع کامل است؟

۱) $\frac{2}{5}$ ۲) $\frac{1}{9}$ ۳) $\frac{4}{12}$ ۴) $\frac{3}{7}$

احتمال متوسط

۳۲. ظرفی حاوی 5 مهره قرمز و 7 مهره سفید است. مهره‌ای را به طور تصادفی از ظرف خارج می‌کنیم؛ اگر قرمز بود به همراه آن یک مهره قرمز دیگر و اگر سفید بود به همراه آن 2 مهره سفید دیگر به داخل ظرف می‌اندازیم و سپس مهره دوم را برحسب تصادف خارج می‌کنیم. احتمال آنکه مهره دوم قرمز باشد چقدر است؟

۱) $\frac{125}{312}$ ۲) $\frac{5}{24}$ ۳) $\frac{37}{161}$ ۴) $\frac{7}{32}$

۳۳. در سؤال قبل، احتمال همرنگ نبودن 2 مهره چقدر است؟

۱) $\frac{45}{104}$ ۲) $\frac{59}{104}$ ۳) $\frac{125}{312}$ ۴) $\frac{187}{312}$

قضیه بیز

۳۴.احتمال اینکه یک حادثه اتومبیل ناشی از نقص ترمز باشد 0.04 و احتمال اینکه آن را به درستی ناشی از نقص ترمز بدانند 0.82 و احتمال اینکه آن را به غلط به نقص ترمز نسبت دهند 0.03 است. احتمال آنکه یک حادثه اتومبیل را که به نقص ترمز نسبت داده‌اند واقعاً ناشی از نقص ترمز باشد، چقدر است؟

۱) 0.0328 ۲) 0.5325 ۳) 0.0288 ۴) 0.3525

۳۵. ظرف A، 5 توپ سفید و 7 توپ سیاه و ظرف B، 3 توپ سفید و 12 توپ سیاه دارد. سکه‌ای را پرتاب کرده اگر شیر ظاهر شود یک توپ از ظرف A و اگر خط ظاهر شود یک توپ از ظرف B انتخاب می‌کنیم. فرض کنید که توپ انتخاب‌شده سفید باشد، احتمال اینکه سکه خط آمده باشد را به دست آورید.

۱) $\frac{3}{8}$ ۲) $\frac{1}{10}$ ۳) $\frac{25}{37}$ ۴) $\frac{12}{37}$

پاسخنامه

۱	۱	۲	●	۴
۲	۱	۲	۳	●
۳	۱	۲	●	۴
۴	۱	●	۳	۴
۵	۱	۲	۳	●
۶	۱	●	۳	۴
۷	۱	۲	۳	●
۸	۱	●	۳	۴
۹	۱	●	۳	۴
۱۰	۱	۲	●	۴
۱۱	۱	۲	●	۴
۱۲	●	۲	۳	۴
۱۳	۱	۲	۳	●
۱۴	۱	●	۳	۴
۱۵	●	۲	۳	۴
۱۶	۱	۲	●	۴
۱۷	●	۲	۳	۴
۱۸	۱	۲	●	۴
۱۹	●	۲	۳	۴
۲۰	۱	۲	●	۴

۲۱	●	۲	۳	۴
۲۲	۱	●	۳	۴
۲۳	۱	۲	●	۴
۲۴	۱	۲	●	۴
۲۵	۱	●	۳	۴
۲۶	۱	۲	۳	●
۲۷	۱	●	۳	۴
۲۸	۱	●	۳	۴
۲۹	●	۲	۳	۴
۳۰	۱	۲	۳	●
۳۱	۱	۲	۳	●
۳۲	●	۲	۳	۴
۳۳	●	۲	۳	۴
۳۴	۱	●	۳	۴
۳۵	۱	۲	۳	●
۳۶	۱	۲	۳	۴
۳۷	۱	۲	۳	۴
۳۸	۱	۲	۳	۴
۳۹	۱	۲	۳	۴
۴۰	۱	۲	۳	۴

برای دریافت پاسخ تشریحی سؤالات به سایت www.Tourani.ir مراجعه کنید.

فصل ۳ متغیرهای تصادفی

مقدمه

کمیتی را در نظر بگیرید که نتیجه انجام یک آزمایش است و در اثر انجام آن آزمایش در زمان‌های مختلف، می‌تواند مقادیر متفاوتی را اختیار کند، مقادیر این کمیت قبل از انجام آزمایش مشخص نیست و به صورت تصادفی در نتیجه انجام آزمایش به دست می‌آید.

مثال کمیت « X : تعداد شیرهای ظاهرشده» در آزمایش پرتاب دو سکه چگونه است؟

حل:

ابتدا فضای نمونه (کل حالات ممکن) را برای پرتاب دو سکه با توجه به وضعیت‌های مختلف سکه اول و سکه دوم به دست می‌آوریم:

$S = \{$ (خ و خ) ، (ش و خ) ، (خ و ش) ، (ش و ش) $\}$

$n(S) =$ تعداد کل حالات ممکن $= 4$

سپس مقادیر ممکن برای متغیر تصادفی(X : تعداد شیرهای ظاهرشده) به همراه احتمال وقوع آن‌ها را به کمک فضای نمونه به دست می‌آوریم:

احتمال مقادیر متغیر تصادفی $P(x)=P(X=x)=\frac{n(X)}{n(S)}$	تعداد حالات مساعد $n(X)$	حالات مساعد	تعداد شیرهای ظاهر شده : X
$P(X=0)=\frac{1}{4}$	1	(خ و خ)	0
$P(X=1)=\frac{2}{4}$	2	(خ و ش)، (ش و خ)	1
$P(X=2)=\frac{1}{4}$	1	(ش و ش)	2
$\sum P(x)=1$			

همان‌طورکه مشاهده می‌شود، متغیر تصادفی X مقادیر خود را به صورت تصادفی می‌گیرد. این مقادیر بر اساس فضای نمونه تعیین شده و با احتمال $P_X(x)$ اختیار می‌شوند.

در واقع ارتباط بین فضای نمونه و مجموعه اعداد $(2,1,0)$ به وسیله متغیر تصادفی(X : تعداد شیرهای ظاهرشده) به صورت زیر برقرار است:

فضای نمونه		مجموعه اعداد
(خ و خ)	$\xrightarrow{\text{تعداد شیر ظاهرشده}}$	0
(ش و خ) (خ و ش)	$\xrightarrow{\text{تعداد شیر ظاهرشده}}$	1
(ش و ش)	$\xrightarrow{\text{تعداد شیر ظاهرشده}}$	2

متغیر تصادفی

تعریف ۱: متغیر تصادفی تابعی از فضای نمونه به مجموعه اعداد حقیقی است.

تعریف ۲: متغیر تصادفی تابعی است که دامنه آن فضای نمونه و حوزه (برد) آن مجموعه اعداد حقیقی است.

تعریف ۳: متغیر تصادفی کمیتی است که مقادیر خود را با احتمال دریافت می‌کند.

قاعده کلی: هر متغیر تصادفی را معمولاً با حرف بزرگ (X) و هر یک از مقادیری را که می‌تواند اختیار کند، با حرف کوچک (x) نمایش می‌دهند.

انواع متغیر تصادفی

متغیرهای تصادفی بر اساس مقادیری که می‌توانند اختیار کنند به دو دسته تقسیم می شوند:

۱) متغیر تصادفی گسسته

۲) متغیر تصادفی پیوسته

متغیر تصادفی گسسته

تعریف: متغیر تصادفی X را گسسته گویند، هرگاه تعداد مقادیری که می‌تواند اختیار کند **متناهی یا نامتناهی شمارش‌پذیر** باشد.

برای مثال:

$x = 1,2,3,...$; **تعداد** آزمایش لازم برای رسیدن به اولین موفقیت : X **نامتناهی شمارش‌پذیر**

$x = 0,1,2,...,n$; **تعداد** موفقیت در n آزمایش : X **متناهی شمارش‌پذیر**

متغیر تصادفی پیوسته

تعریف: متغیر تصادفی X را پیوسته گویند، هرگاه مقادیری که می‌تواند اختیار کند از یک مجموعه نامتناهی (زمان، مکان، ...) باشد؛ به عبارت دیگر مقادیر متغیر تصادفی پیوسته X در یک فاصله یا بازه قرار دارد.

برای مثال:

X : زمان لازم برای ورود مشتری بعدی ; $0 < x < \infty$

تابع احتمال (Probability Function)

تابعی که به کمک آن می‌توان احتمال وقوع وضعیت‌های مختلف یک متغیر تصادفی را به دست آورد، تابع احتمال نامیده می‌شود.

تعریف: تابع احتمال $f(x)$، تابعی است که:

الف) دامنه آن (x)، مقادیر ممکن از یک متغیر تصادفی دلخواه است.

ب) حوزه یا برد آن $(f(x))$، احتمالات مربوط به مقادیر متغیر تصادفی است.

✓ دقت کنید!

۱- همواره مقادیر تابع احتمال $f(x)$ باید مثبت باشد $(f(x)\geq 0)$.

۲- می‌دانیم مجموع احتمالات یک متغیر تصادفی همواره برابر 1 است، درنتیجه ازآنجاکه حوزه $f(x)$ همان احتمالات ممکن برای یک متغیر تصادفی است، باید مجموع مقادیر $f(x)$ به ازای مقادیر ممکن برای متغیر X در دو حالت گسسته و پیوسته، برابر 1 شود؛ به عبارت دیگر:

$$\text{گسسته } X: \begin{cases} f(x)\geq 0 \\ \sum\limits_{\forall x} f(x)=1 \end{cases} \qquad \text{پیوسته } X: \begin{cases} f(x)\geq 0 \\ \int f(x)\,dx=1 \end{cases}$$

تابع احتمال گسسته (Discrete Probability Function)

تعریف: درصورتی‌که X متغیر تصادفی گسسته باشد، آن‌گاه $f(x)$ با داشتن شرایط زیر، تابع احتمال متغیر X خواهد بود:

الف) $\forall x: 0\leq f(x)=P(x)\leq 1$

$f(x)$ به ازای هر مقدار از x برابر با $P(x)$ (احتمال x) است و مقداری بین 0 و 1 دارد.

ب) $\sum f(x)=\sum P(x)=1$

مجموع مقادیر $f(x)$ برابر با مجموع مقادیر احتمال $P(x)$ و برابر 1 است.

x	x_1	$\dots$	x_n	
$P(X=x)=f(x)$	$0\leq f(x_1)\leq 1$	$\dots$	$0\leq f(x_n)\leq 1$	$\sum f(x)=\sum P(x)=1$

محاسبه احتمال

برای محاسبه مقدار احتمال در توابع گسسته روی محدوده A، کافی است به صورت زیر عمل کنیم:

$$f(A)=\sum_A f(x) \quad \text{یا} \quad P(A)=\sum_A P(x)$$

به عبارت دیگر حاصل $f(A)$ یا $P(A)$ برابر با مجموع مقادیر $f(x)$ یا $P(x)$ به ازای هر محدوده A است.

✓ دقت کنید!

تابع احتمال متغیر تصادفی گسسته همان جدول داده‌های طبقه‌بندی شده در فصل آمار توصیفی است و مقادیر $f(x)$ نیز همان فراوانی نسبی داده‌هاست.

یادآوری: قوانین زیر در محاسبه مجموع متغیرها (Σ) برقرار است؛ در صورتی‌که a و b اعداد ثابت مثبت یا منفی باشند داریم:

1) $\sum_{i=1}^{n} a = na$

2) $\sum_{i=a}^{b} i = \frac{(b-a+1)(a+b)}{2} \longrightarrow \sum_{i=1}^{n} i = \frac{n(n+1)}{2} \ , \ \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}$

4) $\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$

5) $\sum_{i=1}^{n} (ax_i \pm b) = a\sum_{i=1}^{n} x_i \pm bn$

مثال ۱ کدام‌یک از توابع زیر یک تابع احتمال است؟

الف)

x	1	2	3	5	
f(x)	0.2	0.3	−0.4	0.9	$\sum f(x)=1$

ب)

x	1	2	3	5	
f(x)	0.1	1.2	0.9	0.4	$\sum f(x)=2.6$

ج) $f(x)=\frac{x+1}{5} \quad ; \quad x=2,3,4$

د) $f(x)=\frac{x+2}{5} \quad ; \quad x=0,1$

حل:

الف) $f(x)$ تابع احتمال نیست، زیرا تابع احتمال همواره مثبت است $(f(x)\geq 0)$، درحالی‌که $P(X=3)=-0.4$ مقداری منفی است.

ب) $f(x)$ تابع احتمال نیست، زیرا:

اولاً، تابع $f(x)$ باید بین 0 و 1 باشد $(0\leq f(x)\leq 1)$ درحالی‌که $P(X=2)=1.2>1$ است؛

ثانیاً، مجموع مقادیر $f(x)$ باید برابر 1 شود $(\sum f(x)=1)$ درحالی که $\sum f(x)=2.6>1$ است.

ج)

x	2	3	4	
f(x)	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$	$\sum f(x)=2.4$

$f(x)$ تابع احتمال نیست، زیرا مجموع مقادیر $f(x)$ باید برابر 1 باشد $(\sum f(x)=1)$، درحالی‌که $\sum f(x)=2.4>1$ است.

د)

x	0	1	
f(x)	$\frac{2}{5}$	$\frac{3}{5}$	$\sum f(x)=1$

$f(x)$ یک تابع احتمال است، زیرا:

اولاً، به ازای تمام x ها، $0\leq f(x)\leq 1$؛

ثانیاً، $\sum f(x)=1$ است.

مثال ۲ تابع احتمال زیر را در نظر بگیرید:

x	0	2	4	5
$P(x)$	0.16	α	0.24	$3\alpha-1$

مطلوب است محاسبه:

الف) مقدار α به صورتی که این جدول یک تابع احتمال باشد.

ب) $P(X\le 4)$

ج) $P(X=3)$

د) $P(X\le 3)$

ه) $P\left(X>\sqrt{5}\right)$

حل:

الف) بنا بر شرط (الف) و (ب) از تابع احتمال داریم:

$$\text{(ب) شرط}: \sum P(x)=1 \;\rightarrow\; 0.16+\alpha+0.24+3\alpha-1=1 \;\rightarrow\; 4\alpha=1.6 \;\rightarrow\; \alpha=0.4$$

$$\text{(الف) شرط}: 0\le P(x)\le 1 \;\rightarrow\; 0\le 3\alpha-1\le 1 \;\rightarrow\; \frac{1}{3}\le\alpha\le\frac{2}{3} \xrightarrow{\alpha=0.4} \text{در شرط صدق می‌کند}$$

درنتیجه تابع احتمال به صورت زیر خواهد شد:

x	0	2	4	5
$P(x)$	0.16	0.4	0.24	0.2

ب)

$$P(X\le 4)=\sum_{x\le 4}P(x)=P(X=0)+P(X=2)+P(X=4)=0.16+0.4+0.24=0.8$$

ج) نقطه $X=3$ وجود ندارد، بنابراین احتمال آن برابر صفر است $\left(P(X=3)=0\right)$.

د)

$$P(X\le 3)=\sum_{x\le 3}P(x)=P(X=0)+P(X=2)+P(X=3)=0.16+0.4+0=0.56$$

ه)

$$P\left(X>\sqrt{5}\right)=P(X>2.24)=P(X=4)+P(X=5)=0.24+0.2=0.44$$

مثال ۳ تابع احتمال متغیر تصادفی X به صورت $f(x)=cx\;;\;x=1,2,\ldots,9$ را در نظر بگیرید.

الف) به ازای کدام مقدار c تابع $f(x)$ یک تابع احتمال است؟

ب) مقدار $P(X>2)$ چقدر است؟

حل:

الف) بنا بر شرط (ب) از تابع احتمال گسسته داریم:

$$\sum_{x=1}^{9}f(x)=1\rightarrow\sum_{x=1}^{9}cx=1\rightarrow c[1+2+\ldots+9]=1\rightarrow c\left[\frac{9(9+1)}{2}\right]=1\rightarrow 45c=1\rightarrow c=\frac{1}{45}$$

یادآوری:

$$\sum_{i=1}^{n}i=1+2+\ldots+n=\frac{n(n+1)}{2}$$

درنتیجه تابع احتمال به صورت زیر خواهد شد:

$$f(x)=\frac{x}{45}\quad;\quad x=1,2,\ldots,9$$

ب)

$$\begin{cases} P(X>2)=1-P(X\le 2)=1-P(X=1)-P(X=2)=1-\frac{1}{45}-\frac{2}{45}=\frac{14}{15} \\ f(x)=\frac{x}{45}\ ;\ x=1,2,...,9 \end{cases}$$

مثال ۴ به ازای کدام مقدار از c، $f(x)=\frac{1}{50}\left[c(x-1)+\frac{1}{2}\right]\ ;\ x=1,2,...,10$ تابع احتمال متغیر تصادفی X است؟

حل:

با توجه به شرط (ب) باید $\sum f(x)=1$ برقرار باشد. بنابراین:

$$\sum_{x=1}^{10} f(x)=1 \;\rightarrow\; \sum_{x=1}^{10}\frac{1}{50}\left[c(x-1)+\frac{1}{2}\right]=1 \;\rightarrow\; \frac{1}{50}\left[c\sum_{x=1}^{10}(x-1)+\sum_{x=1}^{10}\frac{1}{2}\right]=1$$

$$\rightarrow\; \frac{1}{50}\left[c(0+1+...+9)+10\times\frac{1}{2}\right]=1 \;\rightarrow\; \frac{1}{50}[45c+5]=1 \;\rightarrow\; c=1$$

درنتیجه تابع احتمال به صورت زیر خواهد شد:

$$f(x)=\frac{1}{50}\left(x-\frac{1}{2}\right)\quad;\quad x=1,2,...,10$$

مثال ۵ به ازای کدام مقدار a تابع $P(X=x)=\frac{1}{100}[2(10-x)+a]\ ;\ x=1,2,...,10$ یک تابع احتمال است؟

۱) -1 ۲) 0 ۳) 1 ۴) 2

حل: گزینه ۳ درست است.

با توجه به گسسته بودن تابع احتمال، رابطه $\sum_{x=1}^{10}P(x)=1$ باید برقرار باشد؛ بنابراین:

$$\sum_{x=1}^{10}\frac{1}{100}[2(10-x)+a]=1 \;\rightarrow\; \frac{1}{100}\left[2\sum_{x=1}^{10}(10-x)+\sum_{x=1}^{10}a\right]=1$$

$$\frac{1}{100}[2(9+8+...+1+0)+10a]=1 \;\rightarrow\; \frac{1}{100}\left[2\times\frac{9(9+1)}{2}+10a\right]=1 \;\rightarrow\; a=1$$

مد (نما) در تابع احتمال گسسته

برای یک متغیر تصادفی گسسته، مد (نما) کمیتی است که به ازای آن $f(x)$ (تابع احتمال) بیشترین مقدار را دارد.

یادآوری: اگر $f(x)$ به ازای تمام مقادیر x یکسان باشد، مد وجود ندارد.

مثال توابع احتمال گسسته زیر مفروضند؛ مقدار مد (نما) چقدر است؟

الف)

x	-1	0	1	2
$P(x)$	0.1	0.4	0.3	0.2

ب)

x	-1	0	1	2
$P(x)$	0.1	0.4	0.1	0.4

ج)

x	-1	0	1	2
$P(x)$	0.25	0.25	0.25	0.25

حل:

الف) $Mo = 0$

ب) $Mo = 0,2$

ج) Mo نداریم.

✓ **دقت کنید!**

مد در تابع احتمال گسسته، همان مد در جدول داده‌های طبقه‌بندی شده با فراوانی نسبی است.

چندک‌ها در تابع احتمال گسسته

برای یک متغیر تصادفی گسسته، چندک‌ها (چارک، دهک، صدک) و میانه به صورت زیر محاسبه می‌شود:

الف) ابتدا f_c (تجمعی) را تشکیل می‌دهیم.

ب) اولین x که در آن $f_c \geq \frac{1}{2}$ (میانه) یا $f_c \geq \frac{a}{4}$ (چارک) یا $f_c \geq \frac{a}{10}$ (دهک) یا $f_c \geq \frac{a}{100}$ (صدک) باشد، چندک مورد نظر است.

مثال تابع احتمال گسسته زیر مفروض است:

x	-1	0	1	2	
$f(x)$	0.1	0.3	0.15	0.45	$\sum f(x) = 1$
f_c	0.1	0.4	0.55	1	

میانه، چارک اول، دهک اول و صدک هشتادم را محاسبه کنید.

حل:

میانه: اولین x که $f_c \geq \frac{1}{2} \quad \leftarrow \quad Md = 1$

چارک اول: اولین x که $f_c \geq \frac{1}{4} = 0.25 \quad \leftarrow \quad Q_1 = 0$

دهک اول: اولین x که $f_c \geq \frac{1}{10} \quad \leftarrow \quad D_1 = -1$

صدک هشتادم: اولین x که $f_c \geq \frac{80}{100} \quad \leftarrow \quad P_{80} = 2$

تابع احتمال $Y = g(X)$

متغیر تصادفی گسسته X با تابع احتمال $f(x) = P(x)$ را در نظر بگیرید، اگر $Y = g(X)$ تابعی یک‌به‌یک از متغیر X باشد، تابع $f(y)$ به صورت زیر محاسبه می‌شود:

برای مجموعه مقادیر x $(x_1, x_2, ..., x_n)$، مقادیر y $(y_1, y_2, ..., y_n)$ را از رابطه $y_i = g(x_i)$، به دست می‌آوریم، آن‌گاه:

الف) اگر تمام مقادیر y_i با هم متفاوت باشند $(y_1 \neq y_2 \neq ... \neq y_n)$ تابع $f(y)$ همان $f(x)$ است؛ به عبارت دیگر:

$$\begin{array}{c|cccc} x & x_1 & x_2 & \cdots & x_n \\ \hline f(x) & f(x_1) & f(x_2) & \cdots & f(x_n) \end{array} \xrightarrow{y_i = g(x_i)} \begin{array}{c|cccc} y & y_1 & \neq y_2 & \cdots & \neq y_n \\ \hline f(y) & f(x_1) & f(x_2) & \cdots & f(x_n) \end{array}$$

ب) اگر به ازای حداقل دو مقدار متفاوت از x $(x_i \neq x_j)$، مقادیر متناظر از y یکسان شود $(y^* = y_i = y_j)$، $f(y)$ برای $y^* = y_i = y_j$ فقط یک بار در نظر گرفته می‌شود و برابر با $f(y^*) = f(x_i) + f(x_j)$ است؛ به عبارت دیگر:

$$x_i \neq x_j \xrightarrow{Y = g(X)} y^* = y_i = y_j \rightarrow f(y^*) = f(x_i) + f(x_j)$$

مثال ۶ تابع احتمال زیر را در نظر بگیرید. تابع احتمال $Y = X^2$ کدام است؟

x	0	1	2	3
$P(x)=f(x)$	0.2	0.3	0.1	0.4

۱)

y	0	1	4	9
$f(y)$	0.2	0.3	0.1	0.4

۲)

y	0	1	4	9
$f(y)$	0.2	0.5	0	0.1

۳)

y	0	1	2	3
$f(y)$	0.4	0.2	0.1	0.3

۴) هیچ‌کدام

حل: گزینه ۱ درست است.

ابتدا مقادیر y را از رابطه $y_i = x_i^2$ به دست می‌آوریم:

x	0	1	2	3
$y=x^2$	0	1	4	9

ازآنجاکه تمام مقادیر y_i با هم متفاوت هستند، داریم:

x	0	1	2	3
$f(x)$	0.2	0.3	0.1	0.4

$\xrightarrow{f(y)=f(x)}$

y	0	1	4	9
$f(y)$	0.2	0.3	0.1	0.4

مثال ۷ تابع احتمال زیر را در نظر بگیرید. تابع احتمال $Y = X^2$ کدام است؟

x	-2	-1	1	2
$P(x)=f(x)$	0.2	0.3	0.1	0.4

۱)

y	1	4
$f(y)$	0.4	0.6

۲)

y	1	1	4	4
$f(y)$	0.1	0.3	0.2	0.4

۳)

y	1	4
$f(x)$	0.1	0.9

۴)

y	1	2	4
$f(y)$	0.1	0.3	0.6

حل: گزینه ۱ درست است.

ابتدا مقادیر y را از رابطه $y_i = x_i^2$ به دست می‌آوریم:

x	-2	-1	1	2
$y=x^2$	4	1	1	4

ازآنجاکه به ازای مقادیر متفاوت از x، مقادیر متناظر از y یکسان شده است، داریم:

x	-2	-1	1	2
$f(x)$	0.2	0.3	0.1	0.4

$\longrightarrow$

y	1	1	4	4
$f(y)$	$0.3 \;+\; 0.1 \;= 0.4$		$0.2 \;+\; 0.4 \;= 0.6$	

تابع چگالی احتمال پیوسته (Continuous Probability Density Function)

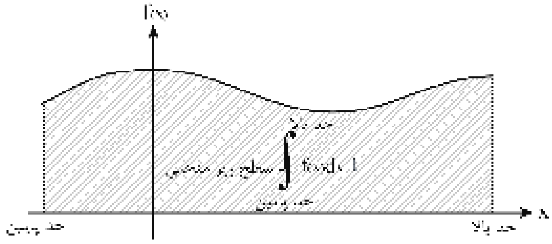
تعریف: $f(x)$ برای متغیر تصادفی پیوسته X در بازه α تا β، تابع چگالی احتمال خواهد بود، اگر:

الف) $f(x) \geq 0$

به ازای تمام مقادیر x، مقدار تابع $f(x)$ همواره بزرگ‌تر یا مساوی صفر است.

ب) $\int_{\alpha=-\infty}^{\beta=+\infty} f(x)dx = 1$

مساحت کل زیرمنحنی چگالی برابر 1 است.



محاسبه احتمال

اولاً: احتمال در بازه a تا b برابر است با سطح زیرمنحنی چگالی در بازه a تا b.

$$P(a < X < b) = \int_a^b f(x)dx$$

ثانیاً: به ازای هر نقطه پیوسته $x = a$، احتمال آنکه متغیر تصادفی دقیقاً مقدار a را اختیار کند، برابر صفر است.

$$P(X = a) = \int_a^a f(x)dx = 0$$

حال از دو رابطه بالا می‌توانیم تساوی زیر را نتیجه بگیریم:

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

مثال تابع چگالی احتمال $f(x)$ که توزیع احتمال یک متغیر تصادفی پیوسته را توصیف می‌کند، کدام ویژگی را ندارد؟
(اقتصاد ـ ۸۰)

۱) به ازای هر نقطه مانند a، $P(X = a) \neq 0$

۲) $f(x) \geq 0$ تابع چگالی مثبت یا صفر است.

۳) مساحت کل زیر منحنی چگالی برابر یک است.

۴) سطح زیر منحنی چگالی بین a و b برابر است با $P(a < X < b)$

حل: گزینه ۱ درست است.

انواع مسایل تابع چگالی احتمال پیوسته

برای حل انواع مسائل مربوط به تابع چگالی پیوسته با توجه به قوانین زیر عمل می‌کنیم:

قانون ۱: هرگاه $f(x)$ (تابع چگالی) داده شود و ضریب ثابتی مانند (k ،c یا ...) را بخواهند، کافی است قرار دهیم $\int_{\text{حد پایین}}^{\text{حد بالا}} f(x)dx=1$ تا مقدار ثابت محاسبه شود.

مثال ۸ به ازای چه مقدار از k تابع زیر می‌تواند تابع چگالی احتمال باشد؟

$$f(x)=\begin{cases} kx^2 & ; \quad 0<x<4 \\ 0 & ; \quad \text{در غیر این‌صورت} \end{cases}$$

۱) $\frac{1}{4}$ ۲) $\frac{3}{16}$ ۳) $\frac{3}{64}$ ۴) $\frac{1}{16}$

حل: گزینه ۳ درست است.

$$\int_0^4 kx^2dx=1 \rightarrow \left[\frac{kx^3}{3}\right]_0^4=1 \rightarrow \frac{64k}{3}=1 \rightarrow k=\frac{3}{64}$$

مثال ۹ تابع چگالی احتمال زیر را در نظر بگیرید. مقدار k کدام است؟

$$f(x)=\begin{cases} \frac{k}{\sqrt{x}} & ; \quad 0<x<9 \\ 0 & ; \quad \text{در غیر این صورت} \end{cases}$$

۱) $\frac{1}{6}$ ۲) $\frac{1}{8}$ ۳) $\frac{1}{12}$ ۴) $\frac{1}{4}$

حل: گزینه ۱ درست است.

$$\int_0^9 \frac{k}{\sqrt{x}}dx=1 \rightarrow \int_0^9 kx^{-\frac{1}{2}}dx=1 \rightarrow \left[2kx^{\frac{1}{2}}\right]_0^9=\left[2k\sqrt{x}\right]_0^9=2k\sqrt{9}=6k=1 \rightarrow k=\frac{1}{6}$$

مثال ۱۰ تابع چگالی احتمال متغیر تصادفی X به صورت زیر داده شده است:

$$f(x)=\begin{cases} x & ; \quad 0\le x\le 1 \\ k-x & ; \quad 1\le x\le 2 \\ 0 & ; \quad \text{برای سایر مقادیر } x \end{cases}$$

مقدار k چقدر است؟ (اقتصاد ـ ۷۶)

۱) $\frac{1}{2}$ ۲) 1 ۳) 2 ۴) 3

حل: گزینه ۳ درست است.

$$\int_{\text{حد پایین}}^{\text{حد بالا}} f(x)dx=1 \rightarrow \int_0^2 f(x)dx=1 \rightarrow \int_0^1 x\,dx+\int_1^2 (k-x)dx=1 \rightarrow \left[\frac{1}{2}x^2\right]_0^1+\left[kx-\frac{1}{2}x^2\right]_1^2=1$$

$$\rightarrow \left(\frac{1}{2}-0\right)+\left(2k-\frac{4}{2}-k+\frac{1}{2}\right)=1 \rightarrow k=2$$

قانون۲: هرگاه $f(x)$ (تابع چگالی) داده شود و احتمال در بازه‌ای مانند (c تا d) خواسته شود، کافی است $\int_c^d f(x)dx$ را محاسبه کنیم؛ به عبارت دیگر «احتمال در هر بازه برابر با انتگرال (سطح زیر منحنی) در بازه مورد نظر است».

مثال ۱ تابع چگالی احتمال زیر را در نظر بگیرید:

$$f(x)=\begin{cases}2x & ; \quad 0<x<1\\ 0 & ; \quad \text{در غیر این‌صورت}\end{cases}$$

مطلوب است محاسبه:

الف) $P\left(X<\frac{1}{3}\right)$ ب) $P\left(X>\frac{1}{4}\right)$ ج) $P\left(\frac{1}{3}<X<\frac{1}{2}\right)$

حل:

الف) $P\left(X<\frac{1}{3}\right)=\int_0^{\frac{1}{3}}2x\,dx=\left[x^2\right]_0^{\frac{1}{3}}=\frac{1}{9}$

ب) $P\left(X>\frac{1}{4}\right)=\int_{\frac{1}{4}}^{1}2x\,dx=\left[x^2\right]_{\frac{1}{4}}^{1}=1-\frac{1}{16}=\frac{15}{16}$

ج) $P\left(\frac{1}{3}<X<\frac{1}{2}\right)=\int_{\frac{1}{3}}^{\frac{1}{2}}2x\,dx=\left[x^2\right]_{\frac{1}{3}}^{\frac{1}{2}}=\frac{1}{4}-\frac{1}{9}=\frac{5}{36}$

مثال ۲ تابع چگالی متغیر تصادفی X به صورت زیر است:

$$f(x)=\begin{cases}6x(1-x) & ; \quad 0<x<1\\ 0 & ; \quad \text{در غیر این صورت}\end{cases}$$

مطلوب است محاسبه:

الف) $P\left(X>\frac{1}{3}\right)$ ب) $P\left(\frac{1}{3}<X<\frac{3}{4}\right)$ ج) $P\left(\frac{1}{2}\le X\le\frac{3}{2}\right)$

حل:

الف)

$$P\left(X>\frac{1}{3}\right)=\int_{\frac{1}{3}}^{1}6x(1-x)dx=6\int_{\frac{1}{3}}^{1}\left(x-x^2\right)dx=6\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_{\frac{1}{3}}^{1}=\left[3x^2-2x^3\right]_{\frac{1}{3}}^{1}=(3-2)-\left(\frac{3}{9}-\frac{2}{27}\right)=1-\frac{7}{27}=\frac{20}{27}$$

ب)

$$P\left(\frac{1}{3}<X<\frac{3}{4}\right)=\int_{\frac{1}{3}}^{\frac{3}{4}}f(x)dx=\int_{\frac{1}{3}}^{\frac{3}{4}}6x(1-x)dx=\left[3x^2-2x^3\right]_{\frac{1}{3}}^{\frac{3}{4}}=\left(\frac{27}{16}-\frac{54}{64}\right)-\left(\frac{3}{9}-\frac{2}{27}\right)=\frac{54}{64}-\frac{7}{27}=\frac{505}{864}$$

ج)

$$P\left(\frac{1}{2}\le X\le\frac{3}{2}\right)=P\left(\frac{1}{2}\le X\le 1\right)=\int_{\frac{1}{2}}^{1}f(x)dx=\int_{\frac{1}{2}}^{1}6x\,(1-x)dx=\left[3x^2-2x^3\right]_{\frac{1}{2}}^{1}=(3-2)-\left(\frac{3}{4}-\frac{2}{8}\right)=1-\frac{4}{8}=\frac{1}{2}$$

مثال ۳ تابع چگالی احتمال متغیر تصادفی X به صورت زیر است. مطلوب است محاسبه $P(X<2)$.

$$f(x)=\begin{cases}\dfrac{k}{\sqrt{x}} & ; \quad 0<x<4\\ 0 & ; \quad \text{در غیر این‌صورت}\end{cases}$$

حل:
ابتدا با استفاده از قانون ۱، ضریب ثابت k را محاسبه می‌کنیم:

$$\int_{-\infty}^{+\infty} f(x)dx=1 \;\rightarrow\; \int_0^4 \frac{k}{\sqrt{x}}dx=1 \;\rightarrow\; \int_0^4 k(x)^{\frac{-1}{2}}dx=1 \;\rightarrow\; \left[2k(x)^{\frac{1}{2}}\right]_0^4=1$$

$$\rightarrow \left[2k\sqrt{x}\right]_0^4=1 \;\rightarrow\; 2k\sqrt{4}=1 \;\rightarrow\; 4k=1 \;\rightarrow\; k=\frac{1}{4}$$

درنتیجه تابع احتمال به صورت زیر خواهد شد:

$$f(x)=\begin{cases}\dfrac{1}{4\sqrt{x}} & ; \quad 0<x<4\\ 0 & ; \quad \text{در غیر این صورت}\end{cases}$$

حال بنا بر قانون (۲) داریم:

$$P(X<2)=P(0<X<2)=\int_0^2 \frac{1}{4\sqrt{x}}dx=\left[\frac{1}{2}\sqrt{x}\right]_0^2=\frac{\sqrt{2}}{2}$$

مثال ۴ تابع چگالی متغیر تصادفی X به صورت زیر است:

$$f(x)=\begin{cases}-kx & ; \quad -2<x<0\\ kx & ; \quad 0\le x<4\\ 0 & ; \quad \text{در غیر این صورت}\end{cases}$$

مطلوب است محاسبه:

الف) مقدار k   ب) $P(-1<X<2.5)$

حل:
الف)

$$\int_{-2}^0 f(x)dx+\int_0^4 f(x)dx=1 \;\rightarrow\; \int_{-2}^0(-kx)dx+\int_0^4(kx)dx=1$$

$$\rightarrow \left[\frac{-kx^2}{2}\right]_{-2}^0+\left[\frac{kx^2}{2}\right]_0^4=1 \;\rightarrow\; \frac{4k}{2}+\frac{16k}{2}=1 \;\rightarrow\; \frac{20k}{2}=1 \;\rightarrow\; k=\frac{1}{10}$$

درنتیجه تابع احتمال به صورت زیر خواهد شد:

$$f(x)=\begin{cases}\dfrac{-x}{10} & ; \quad -2<x<0\\ \dfrac{x}{10} & ; \quad 0\le x<4\\ 0 & ; \quad \text{در غیر این‌صورت}\end{cases}$$

ب) چون بازه $-1<X<2.5$ در ضابطه اول و دوم قرار دارد (قسمتی از آن در محدوده تابع اول و قسمت دیگر در محدوده تابع دوم است)، باید محدوده آن را با توجه به تابع اول و دوم جدا کنیم:

$$P(-1<X<2.5)=\int_{-1}^{0}\frac{-x}{10}dx+\int_{0}^{2.5}\frac{x}{10}dx=\left[-\frac{1}{20}x^2\right]_{-1}^{0}+\left[\frac{1}{20}x^2\right]_{0}^{2.5}=\frac{1}{20}+\frac{6.25}{20}=\frac{7.25}{20}=0.3625$$

قانون۳: هرگاه $f(x)$ (تابع چگالی) داده شود و احتمال آنکه متغیر تصادفی X دقیقاً مقدار مشخصی را اختیار کند خواسته شود، آن‌گاه $P(X=a)=0$.

مثال ۱ کمیت تصادفی X در جامعه‌ای بر طبق قانون نمایی توزیع شده است:

$$f(x)=\frac{1}{20}e^{-\frac{x}{20}} \quad 0<x<\infty$$

احتمال اینکه کمیت تصادفی X، مقداری مساوی با 125 اختیار کند، چقدر است؟ (مدیریت ـ ۷۱)

۱) $\frac{1}{20}$ ۲) 1 ۳) 0 ۴) 0.1

حل: گزینه ۳ درست است.

بنا بر قانون ۳، مقدار احتمال در یک نقطه در تابع چگالی احتمال پیوسته صفر است.

$$P(X=125)=0$$

مثال ۲ اگر متغیر تصادفی X در فاصله $(a=2,b=10)$ بر طبق قانون مستطیلی (یکنواخت) با تابع چگالی $f(x)=\frac{1}{8}\,,\,2<x<10$ توزیع شده باشد، احتمال اینکه X در آزمایش، مقداری مساوی با $\frac{a+b}{2}$ اختیار کند؟

(اقتصاد ـ ۷۰، ۷۲)

۱) $\frac{1}{2}$ ۲) 0 ۳) $\frac{1}{80}$ ۴) $\frac{1}{8}$

حل: گزینه ۲ درست است.

بنا بر قانون ۳، با توجه به پیوسته بودن تابع چگالی، مقدار احتمال در یک نقطه از تابع چگالی احتمال پیوسته برابر صفر است.

$$P\left(X=\frac{a+b}{2}\right)=0$$

مثال ۳ X به عنوان یک متغیر تصادفی معرف طول عمر لامپی است که بین صفر تا 160 ساعت کار می‌کند. احتمال اینکه این لامپ دقیقاً 80 ساعت کار کند برابر است با: (اقتصاد ـ ۷۵)

۱) صفر

۲) 0.5

۳) این احتمال را تا زمانی که تابع چگالی X مشخص نباشد، نمی‌توان محاسبه کرد.

۴) این احتمال را تا زمانی که میانگین و واریانس X مشخص نباشد، نمی‌توان محاسبه کرد.

حل: گزینه ۱ درست است.

متغیر تصادفی X (طول عمر لامپ) پیوسته است پس بنا بر قانون ۳، مقدار احتمال در یک نقطه در تابع چگالی احتمال پیوسته برابر صفر است:

$$P(X=80)=0$$

مد (نما) در تابع چگالی احتمال پیوسته

برای یک متغیر تصادفی پیوسته X، مد (نما) مقداری است که به ازای آن، تابع مقدار ماکزیمم را اختیار می‌کند.

قانون۴: هرگاه $f(x)$ (تابع چگالی) داده شود و مقدار مد (نما) خواسته شود، یکی از روش‌های معمول، آن است که قرار دهیم $f'(x)=0$ و مقدار x را محاسبه کنیم.

مثال ۱ تابع چگالی $f(x)=-\frac{x^2}{3}+2x\ ;\ 0\le x\le 3$ تعریف شده است، نما (مد) آن برابر است با:

۱) 0.6 ۲) 6 ۳) 3 ۴) 4

حل: گزینه ۳ درست است.

$$f'(x)=0 \quad\to\quad -\frac{2x}{3}+2=0 \quad\to\quad x=3$$

مثال ۲ تابع چگالی $0\le x\le 2\ ;\ f(x)=\frac{3}{4}\left(2x-x^2\right)$ تعریف شده است. نما (مد) آن برابر است با:

۱) 0.6 ۲) 1.5 ۳) 1 ۴) 2

حل: گزینه ۳ درست است.

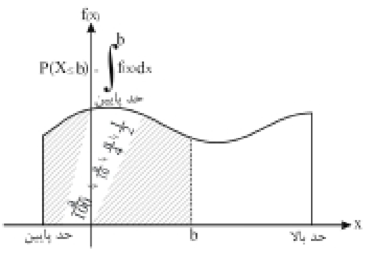
$$f'(x)=0 \quad\to\quad \frac{3}{4}(2-2x)=0 \quad\to\quad x=1$$

✓ **دقت کنید!**

گاهی در $f'(x)=0$، مقدار x به‌دست‌آمده نقطه‌ای است که $f(x)$ در آن مینیمم می‌شود؛ به عبارت دیگر برای محاسبه مد باید تمام مقادیر بحرانی را در تابع $f(x)$ امتحان کرد و مقداری از x را که به ازای آن $f(x)$ ماکزیمم می‌شود، به عنوان مد (نما) انتخاب کرد.

چندک‌ها در تابع چگالی احتمال پیوسته

برای محاسبه چندک‌ها در تابع چگالی متغیر تصادفی پیوسته به صورت زیر عمل می‌کنیم:



$$P(X\le b)=\int_{\text{حد پایین}}^{b} f(x)\,dx=\begin{cases}\frac{1}{2} & & b=Md\ \ (\text{میانه})\\[4pt] \frac{a}{4} & ;\ a=1,2,3 & b=Q_a\ \ (\text{چارک } a\text{ام})\\[4pt] \frac{a}{10} & ;\ a=1,2,\ldots,9 & b=D_a\ \ (\text{دهک } a\text{ام})\\[4pt] \frac{a}{100} & ;\ a=1,2,\ldots,99 & b=P_a\ \ (\text{صدک } a\text{ام})\end{cases}$$

قانون ۵: هرگاه $f(x)$ (تابع چگالی) داده شود و چندک (چارک، دهک، صدک، میانه) خواسته شود، کافی است قرار دهیم $\int_{\text{حد پایین}}^{b} f(x)dx = (\frac{1}{2}$ یا $\frac{a}{4}$ یا $\frac{a}{10}$ یا $\frac{a}{100})$ تا b محاسبه شود؛ در این حالت b همان چندک مورد نظر است.

مثال ۱ در تابع چگالی زیر صدک 80 چقدر است؟ (حسابداری ـ ۸۲)

$$f(x) = \frac{1}{8}x \;\; ; \;\; 0 < x < 4$$

۱) 2.38 ۲) 3.20 ۳) 3.58 ۴) 12.82

حل: گزینه ۳ درست است.

$$\int_{\text{حد پایین}}^{x} f(x)dx = \frac{80}{100}$$

$$\int_0^x \frac{1}{8}x\,dx = \frac{80}{100} \;\rightarrow\; \left[\frac{1}{16}x^2\right]_0^x = 0.8 \;\rightarrow\; \frac{1}{16}x^2 = 0.8 \;\rightarrow\; x^2 = 12.8 \;\rightarrow \begin{cases} x = 3.58 \\ x = -3.58 \end{cases}$$

با توجه به اینکه $0 < x < 4$ است، $x = -3.58$ غیرقابل قبول است.

مثال ۲ تابع چگالی احتمال برای کمیت تصادفی X به صورت $f(x) = \frac{1}{2}x \;;\; 0 < x < 2$ تعریف شده است. مقدار میانه را حساب کنید. (مدیریت ـ ۷۲)

۱) $-\sqrt{2}$ ۲) 1 ۳) $\sqrt{2}$ ۴) $\pm\sqrt{2}$

حل: گزینه ۳ درست است.

$$\int_{\text{حد پایین}}^{me} f(x)dx = \frac{1}{2} \;\rightarrow\; \int_0^{me} \frac{1}{2}x\,dx = \left[\frac{x^2}{4}\right]_0^{me} = \frac{1}{2} \;\rightarrow\; \frac{x^2}{4} = \frac{1}{2} \;\rightarrow\; x^2 = 2 \;\rightarrow \begin{cases} x = \sqrt{2} \\ x = -\sqrt{2} \end{cases}$$

با توجه به منحصر به فرد بودن شاخص میانه، فقط یک مقدار برای آن قابل قبول است و ازآنجاکه $0 < x < 2$، $x = -\sqrt{2}$ غیرقابل قبول است.

مثال ۳ یک تابع توزیع احتمال دارای چگالی $f(x) = 1$ است. اگر حد پایین توزیع 3.4 باشد، میانه توزیع چقدر است؟ (حسابداری ـ ۷۷)

۱) 3.7 ۲) 3.9 ۳) 4 ۴) 6.8

حل: گزینه ۲ درست است.

$$\int_{\text{حد پایین}}^{me} f(x)dx = \frac{1}{2} \;\rightarrow\; \int_{3.4}^{me} 1\times dx = \frac{1}{2} \;\rightarrow\; [x]_{3.4}^{me} = \frac{1}{2} \;\rightarrow\; x - 3.4 = \frac{1}{2} \;\rightarrow\; me = 3.9$$

مثال ۴ اگر X دارای توزیعی با تابع چگالی $x > 0 \;;\; f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$ باشد، آن‌گاه میانه X کدام است؟

۱) $\frac{\ln 2}{\theta}$ ۲) $\theta \ln 2$ ۳) $\frac{1}{\theta \ln 2}$ ۴) $\frac{\theta}{\ln 2}$

حل: گزینه ۲ درست است.

$$\int_{\text{حد پایین}}^{md} f(x)dx=\frac{1}{2}\ \rightarrow\ \int_{0}^{md}\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\frac{1}{2}\ \rightarrow\ \left[-e^{-\frac{x}{\theta}}\right]_{0}^{md}=\frac{1}{2}$$

$$1-e^{-\frac{md}{\theta}}=\frac{1}{2}\ \rightarrow\ e^{-\frac{md}{\theta}}=\frac{1}{2}\ \longrightarrow\ -\frac{md}{\theta}=\ln\frac{1}{2}\ \rightarrow\ \frac{md}{\theta}=\ln 2\ \rightarrow\ md=\theta\ln 2$$

نکته: b میانه $\rightleftarrows\ P(X\le b)=P(X>b)$

زیرا: $P(X\le b)+P(X>b)=1\ \xrightarrow{P(X\le b)=P(X>b)}\ P(X\le b)=P(X>b)=\frac{1}{2}$

تابع چگالی $Y=g(X)$

متغیر تصادفی پیوسته x را در بازه $a<x<b$ با تابع چگالی $f(x)$ در نظر بگیرید، در صورتی‌که $y=g(x)$ تابعی (یک به یک) از متغیر X باشد، آن‌گاه برای محاسبه تابع چگالی $f(y)$ به ترتیب زیر عمل می‌کنیم:

الف) با استفاده از $y=g(x)$، x را برحسب y به‌دست می‌آوریم: $x=h(y)$

ب) با استفاده از رابطه زیر تابع $f_y(y)$ محاسبه می‌شود:

$$\boxed{f_y(y)=\left|h'(y)\right|\cdot f_x(h(y))}$$

در حقیقت $f_y(y)$ از حاصلضرب «قدرمطلق مشتق $h(y)$» در «$f_x(h(y))$» به‌دست می‌آید.

ج) بازه مربوط به y را به ترتیب با قرار دادن $x=a$ و $x=b$ در $y=g(x)$ به‌دست می‌آوریم.

مثال ۱ تابع چگالی متغیر تصادفی پیوسته X را به‌صورت زیر در نظر بگیرید، تابع چگالی $Y=-\frac{X}{3}$ کدام است؟

$$f(x)=\begin{cases}2x & ;\ 0<x<1\\ 0 & ;\ \text{در غیر این‌صورت}\end{cases}$$

۱) $f_Y(y)=\begin{cases}3y & ;\ 0<y<1\\ 0 & ;\ \text{در غیر این صورت}\end{cases}$

۲) $f_Y(y)=\begin{cases}-18y & ;\ -\frac{1}{3}<y<0\\ 0 & ;\ \text{در غیر این صورت}\end{cases}$

۳) $f_Y(y)=\begin{cases}\frac{y}{3} & ;\ 0<y<1\\ 0 & ;\ \text{در غیر این صورت}\end{cases}$

۴) $f_Y(y)=\begin{cases}-y & ;\ 0<y<\frac{1}{2}\\ 0 & ;\ \text{در غیر این صورت}\end{cases}$

حل: گزینه ۲ درست است.

الف) $y=-\frac{x}{3} \rightarrow x=-3y=h(y)$

ب) $\begin{cases} f_y(y)=|h'(y)|\cdot f_x(h(y))=|-3|\cdot 2(-3y)=-18y \\ h'(y)=-3\ ,\ f_x(h(y))=2(-3y) \end{cases}$

ج) $0<x<1 \xrightarrow{y=-\frac{x}{3}} \begin{cases} y=-\frac{0}{3}=0 \\ y=-\frac{1}{3} \end{cases} \rightarrow -\frac{1}{3}<y<0$

در نتیجه:

$$f_Y(y)=\begin{cases} -18y & ; \quad -\frac{1}{3}<y<0 \\ 0 & ; \quad \text{در غیر این‌صورت} \end{cases}$$

مثال ۲ تابع چگالی احتمال برای کمیت تصادفی X به صورت زیر بیان شده است:

$$f(x)=\frac{x^2}{63} \qquad 3<x<6$$

تابع چگالی کمیت Y که بر طبق رابطه $Y=-X$ از X تبعیت دارد، کدام است؟ (اقتصاد ـ ۷۵)

۱) $-6<y<-3\ ,\ f(y)=\frac{y^2}{63}$

۲) $-6<y<-3\ ,f(y)=\frac{y^2-6}{63}$

۳) $-6<y<-3\ ,\ f(y)=\frac{y^2-3}{63}$

۴) $-6<y<-3\ ,\ f(y)=\frac{-y^2}{63}$

حل: گزینه ۱ درست است.

$$y=h(x)=-x \rightarrow x=-y=h(y)$$

$$\begin{cases} h'(y)=-1 \\ f(h(y))=\frac{(-y)^2}{63}=\frac{y^2}{63} \end{cases} \rightarrow \begin{cases} f(y)=|h'(y)|f(h(y))=|-1|\frac{y^2}{63}=\frac{y^2}{63} \\ 3<x<6 \xrightarrow{y=-x} -6<y<-3 \end{cases}$$

امید ریاضی (Expected Value)

مسئله (درک مطلب): در یک شهر 70 درصد روزها بارانی و بقیه روزها آفتابی است. یک چترفروش در روزهای بارانی 2000 تومان سود و در روزهای آفتابی 4000 تومان ضرر می‌کند. «مقدار مورد انتظار» یا «امید ریاضی» یا «میانگین (متوسط)» سود این شخص در هر روز چقدر است؟

امید ریاضی در اصل برای بازی‌های شانسی (تصادفی) به وجود آمده و به مفهوم «مقدار موردانتظار» یا «مقدار متوسط» سود یا ضرر در بازی‌هاست که می‌تواند مقادیر مثبت یا منفی یا صفر را اختیار کند.

درصورتی‌که متغیر تصادفی X مقادیر برد یا باخت را مشخص کند و $f(x)=P(x)$ احتمال متناظر هر مقدار از x را نشان دهد، آن‌گاه می‌توان «مقدار مورد انتظار» متغیر تصادفی X (میزان برد و باخت) را به صورت زیر محاسبه کرد:

مقدار متوسط یا **مقدار مورد انتظار** برابر است با:

مجموع حاصل‌ضرب‌های هر مقدار x (برد یا باخت) در احتمال متناظر آن $(P(x))$ $= \sum x\cdot P(x)$

تعریف: «امید ریاضی» یا «مقدار مورد انتظار» متغیر تصادفی X همان «میانگین» است و در محاسبه آن از رابطه میانگین حسابی وزنی $\mu=\sum f_i x_i$ استفاده می‌شود که در آن احتمالات $P(x)$ مانند فراوانی نسبی (f_i) نقش وزن هر متغیر تصادفی X (برد یا باخت) را بازی می‌کنند.

با تعمیم تعریف ارائه‌شده به حالت پیوسته و استفاده از $\int$ (انتگرال) به‌جای $\sum$ (مجموع) به تعاریف کلی زیر برای دو حالت گسسته و پیوسته می‌رسیم:

اگر X متغیر تصادفی گسسته و $f(x)=P(x)$ تابع احتمال آن به ازای هر x باشد، آن‌گاه مقدار مورد انتظار (امید ریاضی) X به صورت زیر محاسبه می‌شود:

$$E(X)=\sum_{\forall x} x\cdot f(x)=\sum_{\forall x} x\cdot P(x) \qquad (\text{گسسته } X)$$

اگر X متغیر تصادفی پیوسته و $f(x)$ تابع چگالی آن باشد، آن‌گاه مقدار مورد انتظار (امید ریاضی) X به صورت زیر محاسبه می‌شود:

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx \qquad (\text{پیوسته } X)$$

$$-\infty < E(X) < +\infty$$

حل مسئله (درک مطلب): برای مشخص کردن متوسط سود در هر روز به صورت زیر عمل می‌کنیم:

«مقدار مورد انتظار سود» یا «میانگین سود» $= \underbrace{2000\times 0.7}_{x_1\times P(x_1)} + \underbrace{(-4000)\times 0.3}_{x_2\times P(x_2)} = +200$ تومان

به عبارت دیگر «سود مورد انتظار به طور متوسط در هر روز 200 تومان است»، به آن مفهوم است که چون متوسط سود مقداری مثبت است، چترفروش می‌تواند به کار خود ادامه دهد.

مثال ۱ اگر امید ریاضی سود فروش تولیدات یک کارخانه در روز 8 میلیون تومان باشد، کدام گزینه درست است؟

۱) سود روزهای آینده حداقل 8 میلیون تومان است.

۲) انتظار می‌رود شرکت به طور متوسط در هر روز 8 میلیون تومان سود داشته باشد.

۳) سود فردای شرکت دقیقاً 8 میلیون تومان است.

۴) هیچ‌کدام

حل: گزینه ۲ درست است.

مثال ۲ با توجه به تابع احتمال زیر مقدار امید ریاضی (میانگین) کدام است؟

x	-4	2	3	5
$P(x)=f(x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$

۱) $\frac{2}{3}$ ۲) $\frac{13}{6}$

۳) $\frac{1}{6}$ ۴) $\frac{1}{2}$

حل: گزینه ۲ درست است.

$$E(X)=\sum x\cdot f(x)=(-4)\times\frac{1}{6}+2\times\frac{1}{3}+3\times\frac{1}{6}+5\times\frac{1}{3}=\frac{13}{6}$$

مثال ۳ با توجه به تابع احتمال زیر، مقدار امید ریاضی (متوسط) کدام است؟

$$f(x)=\frac{x+1}{16}\ ;\ x=0,2,4,6$$

۱) 4.25 ۲) 5 ۳) 4.0 ۴) 6

حل: گزینه ۱ درست است.

x	0	2	4	6
$P(x)=f(x)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$

$$E(X)=\sum xf(x)=0\times\frac{1}{16}+2\times\frac{3}{16}+4\times\frac{5}{16}+6\times\frac{7}{16}=\frac{68}{16}=\frac{17}{4}=4.25$$

مثال ۴ تابع چگالی متغیر تصادفی پیوسته را به صورت مقابل در نظر بگیرید:

$$f(x)=\frac{x}{8}\quad ;\quad 0<x<4$$

مقدار میانگین جامعه کدام است؟

۱) $\frac{8}{3}$ ۲) $\frac{1}{2}$ ۳) $\frac{4}{5}$ ۴) $\frac{5}{2}$

حل: گزینه ۱ درست است.

$$E(X)=\int_0^4 xf(x)dx=\int_0^4 x\cdot\frac{x}{8}dx=\left[\frac{x^3}{24}\right]_0^4=\frac{64}{24}=\frac{8}{3}$$

مثال ۵ تابع چگالی متغیر تصادفی X را به صورت زیر در نظر بگیرید. $E(X)$ (امید ریاضی) جامعه کدام است؟

$$f(x)=\begin{cases} x & ;\quad 0<x<1 \\ 2-x & ;\quad 1<x<2 \\ 0 & ;\quad \text{سایر نقاط} \end{cases}$$

۱) 2 ۲) 1

۳) 4 ۴) 3

حل: گزینه ۲ درست است.

$$E(X)=\int xf(x)dx=\int_0^1 x\cdot xdx+\int_1^2 x(2-x)dx=\left[\frac{x^3}{3}\right]_0^1+\left[x^2-\frac{x^3}{3}\right]_1^2=\frac{1}{3}+\frac{2}{3}=1$$

مثال ۶ تابع چگالی متغیر تصادفی X را به صورت زیر در نظر بگیرید. $E(X)$ (امید ریاضی) جامعه کدام است؟

$$f(x)=\begin{cases}\frac{x}{2} & 0<x<1\\ \frac{1}{4} & x=1\\ \frac{1}{6} & x=2\\ \frac{1}{3} & 2<x<3\end{cases}$$

۱) 1

۲) $\frac{19}{12}$

۳) $\frac{1}{6}$

۴) $\frac{7}{12}$

حل: گزینه ۲ درست است.

$$E(X)=\int xf(x)dx+\sum xf(x)=\int_0^1 x\times\frac{x}{2}dx+1\times\frac{1}{4}+2\times\frac{1}{6}+\int_2^3 x\times\frac{1}{3}dx=\frac{19}{12}$$

خواص امید ریاضی

با توجه به یکسان بودن مفهوم امید ریاضی با میانگین، تمام خواص میانگین حسابی (μ) درباره امید ریاضی $E(X)$ نیز به شرح زیر برقرار است:

1) $E(a)=a$	a و b مقادیر ثابتی هستند (مثبت یا منفی).
2) $E(X\pm a)=E(X)\pm a$	
3) $E(aX)=aE(X)$	
4) $E(aX+b)=aE(X)+b$	
5) $E(E(E...E(E(X))))=E(X)$	ازآنجاکه امید، مقدار ثابتی است، $E(E(X))$ برابر با $E(X)$ است.
6) $E[X-E(X)]=0$	امید انحرافات از میانگین همیشه صفر است.
7) $E\left[(X-E(X))^2\right]\le E\left[(X-a)^2\right]$	امید مجذور تفاضلات از میانگین همیشه مینیمم است (a عدد ثابتی است).
8) $E(\lvert X-Md\rvert)\le E(\lvert X-a\rvert)$	امید قدرمطلق انحرافات از میانه همیشه مینیمم است(a عدد ثابتی است).

نکته: ازآنجاکه امید ریاضی همان میانگین است، داریم:

$$\boxed{E(X)=\mu}$$

مثال ۱ اگر $Y=3E(X)-2X$ باشد، مقدار $E(Y)$ کدام است؟

۱) $E(X)$ ۲) $5E(X)$ ۳) 0 ۴) قابل محاسبه نیست.

حل: گزینه ۱ درست است.

$$E(Y)=E\left[3E(X)-2X\right]=3E(E(X))-2E(X)=3E(X)-2E(X)=E(X)$$

مثال ۲ اگر a, b, c مقادیر ثابتی باشند، کدام رابطه درست نیست؟

۱) $E(cb+X)=cb+\mu_X$
۲) $E(bX+ca)=b\mu_X+ca$
۳) $E\left(cX^2\pm bX\right)=cE\left(X^2\right)\pm bE(X)$
۴) $E(b(c+X))=bc+E(X)$

حل: گزینه ۴ درست است.

با فرض آنکه همیشه در مبحث خواص امید ریاضی، $E(X)=\mu$ است.

گزینه ۱ : $E(X)=E(cb)+E(X)=cb+\mu_X$

گزینه ۲ : $E(bX+ca)=bE(X)+E(ca)=b\mu_X+ca$

گزینه ۳ : $E\left(cX^2\pm bX\right)=E\left(cX^2\right)\pm E(bX)=cE\left(X^2\right)\pm bE(X)=cE\left(X^2\right)\pm b\mu_X$

گزینه ۴ : $E(b(c+X))=E(bc+bX)=E(bc)+E(bX)=bc+bE(X)=bc+b\mu_X$

مثال ۳ اگر X یک متغیر تصادفی با جدول توزیع احتمال زیر باشد، $E(3X-2E(X))$ چقدر است؟

x	-1	1	2
$P(x)$	$\frac{1}{5}$	$1-a$	$\frac{1}{5}$

۱) $-\frac{1}{5}$
۲) $\frac{1}{5}$
۳) $\frac{2}{5}$
۴) $\frac{4}{5}$

حل: گزینه ۴ درست است.

بنا بر شرط (ب) تابع احتمال گسسته داریم:

$$\sum P(x)=1 \;\rightarrow\; \frac{1}{5}+1-a+\frac{1}{5}=1 \;\rightarrow\; a=\frac{2}{5}$$

بنابراین تابع احتمال به صورت زیر خواهد بود:

x	-1	1	2
$f(x)=P(x)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

$$\begin{cases} E(3X-2E(X))=3E(X)-2E(E(X))=3E(X)-2E(X)=E(X)=\frac{4}{5} \\ E(X)=\sum xP(x)=-1\times\frac{1}{5}+1\times\frac{3}{5}+2\times\frac{1}{5}=\frac{4}{5} \end{cases}$$

امید ریاضی تابعی از X

درصورتی‌که $g(x)$ تابع دلخواهی برحسب X باشد، برای محاسبه $E(g(X))$ به صورت زیر عمل می‌کنیم:

(X گسسته)	$E(g(X))=\sum_{\forall x} g(x)\cdot f(x)$
(X پیوسته)	$E(g(X))=\int_{\text{حد پایین}}^{\text{حد بالا}} g(x)\cdot f(x)\,dx$

✓ **دقت کنید!**

برای محاسبه $E(g(X))$، تابع $f(x)$ تغییری نمی‌کند و به همان شکل اولیه استفاده می‌شود، برای مثال:

$$E\left(X^2\right)=\sum_{\forall x} x^2 f(x) \qquad (\text{گسسته } X)$$

$$E\left(X^2\right)=\int_{\text{حد پایین}}^{\text{حد بالا}} x^2 f(x)\,dx \qquad (\text{پیوسته } X)$$

مثال ۴ تابع احتمال گسسته زیر را در نظر بگیرید:

x	-1	0	1	2
$P(x)=f(x)$	0.4	0.3	0.2	0.1

مطلوب است محاسبه:

الف) $E(X)$ ب) $E\left(X^2\right)$ ج) $E\left(3X^2-5X+2\right)$ د) $E(2X+1)$

حل:

الف) $E(X)=\sum x\cdot f(x)=(-1)\times 0.4+0\times 0.3+1\times 0.2+2\times 0.1=0$

ب) $E\left(X^2\right)=\sum x^2 f(x)=(-1)^2\times 0.4+0^2\times 0.3+1^2\times 0.2+2^2\times 0.1=1$

ج) $E\left(3X^2-5X+2\right)=3\underbrace{E\left(X^2\right)}_{1}-5\underbrace{E(X)}_{0}+2=3+2=5$

د) $E(2X+1)=2E(X)+1=2\times 0+1=1$

مثال ۵ تابع چگالی احتمال $f(x)=2x \;;\; 0\le x\le 1$ مفروض است. مطلوب است محاسبه:

الف) $E(X)$ ب) $E\left(X^2\right)$ ج) $E\left(3X^2-5X+2\right)$ د) $E(2X+1)$

حل:

الف) $E(X)=\int_0^1 xf(x)\,dx=\int_0^1 x\cdot 2x\,dx=\left[\frac{2}{3}x^3\right]_0^1=\frac{2}{3}$

ب) $E\left(X^2\right)=\int_0^1 x^2 f(x)\,dx=\int_0^1 x^2\cdot 2x\,dx=\left[\frac{2x^4}{4}\right]_0^1=\frac{1}{2}$

ج) $E\left(3X^2-5X+2\right)=3\underbrace{E\left(X^2\right)}_{1/2}-5\underbrace{E(X)}_{2/3}+2=\frac{3}{2}-\frac{10}{3}+2=\frac{1}{6}$

د) $E(2X+1)=2\underbrace{E(X)}_{2/3}+1=\frac{4}{3}+1=\frac{7}{3}$

ارزش پولی مورد انتظار (Expected Monetary Value)

معیار ارزش پولی مورد انتظار (EMV) که در واقع همان امیدریاضی است، معیاری برای تصمیم‌گیری در انتخاب بهترین گزینه با شرایط موجود است.

فرض کنید تصمیم‌گیرنده‌ای n گزینه ممکن $(a_1, a_2, ..., a_n)$ داشته و با m حالت طبیعی مواجه باشد. اگر M_{ij} نتیجه ناشی از انتخاب گزینه iام و وقوع حالت j و P_j احتمال وقوع jامین حالت طبیعی باشد (به طوری که $\sum p_j = 1$)، آن‌گاه ارزش پولی مورد انتظار گزینه a_i از رابطه زیر محاسبه می‌شود:

$$EMV(a_i) = p_1 M_{i1} + p_2 M_{i2} + ... + p_m M_{im} = \sum_{j=1}^{m} p_j m_{ij}$$

مثال فرض کنید شرکتی می‌داند که 10 درصد تمام محصولات مشابه قبلی با تقاضای کم، 50 درصد با تقاضای متوسط و 40 درصد با تقاضای زیاد مواجه بوده‌اند $(p_1 + p_2 + p_3 = 1)$. بهترین گزینه در این شرایط با توجه به جدول زیر چیست؟

فرایندهای تولید	سطوح تقاضا		
	کم $p_1 = 0.1$	متوسط $p_2 = 0.5$	زیاد $p_3 = 0.4$
A	70	120	200
B	80	120	180
C	100	125	160

۱) A ۲) B ۳) C ۴) A , B

حل: گزینه ۱ درست است.

$$EMV(A) = 0.1 \times 70 + 0.5 \times 120 + 0.4 \times 200 = 147 \checkmark$$
$$EMV(B) = 0.1 \times 80 + 0.5 \times 120 + 0.4 \times 180 = 140$$
$$EMV(C) = 0.1 \times 100 + 0.5 \times 125 + 0.4 \times 160 = 136.5$$

بهترین گزینه برای انتخاب، فرآیند تولیدی A است، زیرا سود مورد انتظار آن از بقیه گزینه‌ها بیشتر است.

واریانس (Variance)

در فصل اول کتاب، واریانس به صورت $\sigma^2 = \frac{\sum(x-\mu)^2}{N} = \frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2$ تعریف شد. حال با استفاده از مفهوم امید ریاضی می‌توانیم به‌جای $\frac{\sum}{N}$، از E استفاده کنیم و شکل دیگری از تعریف واریانس را ارائه دهیم:

تعریف: متغیر تصادفی X مفروض است، واریانس X از روابط زیر محاسبه می‌شود:

$$(1)\ \sigma^2 = E(X-\mu)^2 = E\left[(X - E(X))^2\right]$$

$$(2)\ \sigma^2 = E(X^2) - [E(X)]^2$$

در منابع مختلف، از علایم متفاوتی برای نمایش واریانس استفاده می‌شود، از جمله:

$$\text{واریانس} = \text{پراش} = \sigma^2 = V(X) = D(X) = Var(X)$$

مثال ۱ تابع احتمال

x	-1	0	1
$P(x)=f(x)$	0.2	0.5	0.3

را در نظر بگیرید. مقدار واریانس X کدام است؟

۱) 0.49 ۲) 0.7 ۳) 0.2 ۴) 0.1

حل: گزینه ۱ درست است.

با توجه به رابطه (2) برای محاسبه واریانس داریم:

$$\begin{cases} \sigma^2 = E(X^2) - [E(X)]^2 = 0.5 - (0.1)^2 = 0.5 - 0.01 = 0.49 \\ E(X) = \sum x f(x) = -1\times 0.2 + 0\times 0.5 + 1\times 0.3 = 0.1 \\ E(X^2) = \sum x^2 f(x) = (-1)^2\times 0.2 + 0^2\times 0.5 + 1^2\times 0.3 = 0.5 \end{cases}$$

مثال ۲ اگر X یک متغیر تصادفی باشد که در آن $P(X=c)=1$ (c مقدار ثابت)، $E(X)$ و $Var(X)$ کدام است؟

۱) $0\,,1$ ۲) $0\,,c$ ۳) $c\,,1$ ۴) $c^2\,,c$

حل: گزینه ۲ درست است.

x	c
$f(x)=P(x)$	1

$$\begin{cases} \sigma^2 = E(X^2) - [E(X)]^2 = c^2 - c^2 = 0 \\ E(X) = \sum x f(x) = c\times 1 = c \\ E(X^2) = \sum x^2 f(x) = c^2\times 1 = c^2 \end{cases}$$

مثال ۳ تابع احتمال $f(x)=\frac{x-1}{6}\ ;\ x=2,3,4$ را در نظر بگیرید، مقدار میانگین و واریانس و انحراف معیار X کدام است؟

(راست به چپ)

۱) $\frac{10}{3}\,,\frac{5}{9}\,,\frac{\sqrt{5}}{3}$ ۲) $\frac{10}{3}\,,\frac{1}{9}\,,\frac{1}{3}$ ۳) $\frac{5}{3}\,,\frac{1}{4}\,,\frac{1}{2}$ ۴) هیچ‌کدام

حل: گزینه ۱ درست است.

$$f(x)=\frac{x-1}{6}\quad ;\quad x=2,3,4 \quad \rightarrow$$

x	2	3	4
$f(x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

$$\begin{cases} \sigma^2 = E(X^2) - [E(X)]^2 = \frac{35}{3} - \left(\frac{10}{3}\right)^2 = \frac{35}{3} - \frac{100}{9} = \frac{5}{9} \\ E(X) = \sum x f(x) = 2\times\frac{1}{6} + 3\times\frac{2}{6} + 4\times\frac{3}{6} = \frac{20}{6} = \frac{10}{3} \\ E(X^2) = \sum x^2 f(x) = 2^2\times\frac{1}{6} + 3^2\times\frac{2}{6} + 4^2\times\frac{3}{6} = \frac{70}{6} = \frac{35}{3} \end{cases}$$

در نتیجه:

$E(X)=\mu=\frac{10}{3}$: میانگین ، $\sigma^2=\frac{5}{9}$: واریانس ، $\sigma=\frac{\sqrt{5}}{3}$: انحراف معیار

مثال ۴ تابع چگالی $f(x)=2x\ ;\ 0<x<1$ را در نظر بگیرید. مقدار واریانس X کدام است؟

۱) $\frac{1}{9}$ ۲) $\frac{1}{18}$ ۳) $\frac{1}{3}$ ۴) $\frac{1}{6}$

حل: گزینه ۲ درست است.

$$\begin{cases}\sigma^2=E(X^2)-[E(X)]^2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{2}-\frac{4}{9}=\frac{1}{18}\\ E(X)=\int_0^1 xf(x)dx=\int_0^1 x\cdot 2xdx=\left[\frac{2}{3}x^3\right]_0^1=\frac{2}{3}\\ E(X^2)=\int_0^1 x^2f(x)dx=\int_0^1 x^2\cdot 2xdx=\left[\frac{2}{4}x^4\right]_0^1=\frac{1}{2}\end{cases}$$

مثال ۵ تابع چگالی احتمال X به صورت $f(x)=2c^2x\ ;\ 0\le x\le\frac{1}{c}$ است. c چقدر باشد تا $\sigma_X^2=\frac{1}{2}$ شود؟

۱) $\frac{1}{3}$ ۲) $\frac{1}{6}$ ۳) $\frac{1}{9}$ ۴) $\frac{1}{2}$

حل: گزینه ۱ درست است.

$$E(X)=\int_0^{\frac{1}{c}} xf(x)dx=\int_0^{\frac{1}{c}} x\cdot(2c^2x)dx=\left[\frac{2}{3}c^2x^3\right]_0^{\frac{1}{c}}=\frac{2}{3c}\qquad (I)$$

$$E(X^2)=\int_0^{\frac{1}{c}} x^2f(x)dx=\int_0^{\frac{1}{c}} x^2\cdot(2c^2x)dx=\left[\frac{2}{4}c^2x^4\right]_0^{\frac{1}{c}}=\frac{1}{2c^2}\qquad (II)$$

$$\sigma^2=E(X^2)-[E(X)]^2=\frac{1}{2}\xrightarrow{(I),(II)}\frac{1}{2c^2}-\frac{4}{9c^2}=\frac{1}{2}\rightarrow\frac{1}{c^2}=9\rightarrow\begin{cases}c=\frac{1}{3}\\ c=-\frac{1}{3}\end{cases}$$

با توجه به اینکه $0\le x\le\frac{1}{c}$ است، $c=-\frac{1}{3}$ غیرقابل قبول است.

مثال ۶ درصورتی‌که $x=-1,0,1$ برای $P(x)=\frac{|x|+1}{5}$ تابع احتمال متغیر تصادفی ناپیوسته X باشد. آن‌گاه امید ریاضی و واریانس X به ترتیب از راست به چپ کدام است؟

۱) $0\ ,\ \frac{4}{5}$ ۲) $1\ ,\ \frac{1}{5}$ ۳) $0\ ,\ \frac{2}{5}$ ۴) $\frac{4}{5}\ ,\ \frac{4}{25}$

حل: گزینه ۱ درست است.

$$P(x)=\frac{|x|+1}{5}\xrightarrow{x=-1,0,1}\begin{array}{c|ccc|c} x & -1 & 0 & 1 & \\ \hline P(x) & \frac{2}{5} & \frac{1}{5} & \frac{2}{5} & \sum P(x)=1\end{array}$$

$$\begin{cases}\sigma^2=E(X^2)-E(X)^2=\frac{4}{5}-0^2=\frac{4}{5}\\ E(X)=\sum xP(x)=-1\times\frac{2}{5}+0\times\frac{1}{5}+1\times\frac{2}{5}=0\\ E(X^2)=\sum x^2P(x)=(-1)^2\times\frac{2}{5}+0^2\times\frac{1}{5}+1^2\times\frac{2}{5}=\frac{4}{5}\end{cases}$$

خواص واریانس

اگر ثابت‌های a و b (مثبت یا منفی) را در نظر بگیریم:

1) $\sigma^2(a)=0$
2) $\sigma^2(bX)=b^2\sigma_X^2$
3) $\sigma^2(bX\overset{0}{\cancel{\pm a}})=b^2\sigma_X^2$

خواص انحراف معیار

اگر ثابت‌های a و b (مثبت یا منفی) را در نظر بگیریم:

1) $\sigma(a)=0$
2) $\sigma(bX)=\lvert b\rvert\sigma_X$
3) $\sigma(bX\overset{0}{\cancel{\pm a}})=\lvert b\rvert\sigma_X$

مثال اگر $V(X)=16$ باشد، واریانس و انحراف معیار متغیر $Y=\frac{-X+3}{2}$ کدام است؟

۱) $2\,,\,4$ ۲) $1\,,\,2$ ۳) $0\,,\,6$ ۴) $4\,,\,16$

حل: گزینه ۱ درست است.

$$\begin{cases}\sigma_X^2=16\rightarrow\sigma_X=4\\ \sigma^2(Y)=\sigma^2\left(-\frac{X}{2}\overset{0}{\cancel{+\frac{3}{2}}}\right)=\left(-\frac{1}{2}\right)^2\sigma_X^2=\frac{1}{4}\sigma_X^2=\frac{16}{4}=4\rightarrow\sigma_Y=2\\ \sigma(Y)=\sigma\left(-\frac{X}{2}\overset{0}{\cancel{+\frac{3}{2}}}\right)=\left|-\frac{1}{2}\right|\sigma_X=\frac{4}{2}=2\end{cases}$$

تابع توزیع تجمعی (Cumulative Distribution Function)

در بسیاری از مسایل، دانستن احتمال $P(X \le x)$ یا به عبارت دیگر «احتمال مقادیر کوچک‌تر یا مساوی x» مورد نظر است؛ در این شرایط می‌توانیم از تابع توزیع تجمعی (احتمال تجمعی) که مقادیر احتمال کمتر یا مساوی x را محاسبه می‌کند، استفاده کنیم.

تعریف: تابع توزیع تجمعی (تراکمی) متغیر تصادفی X، $F_X(x)$، عبارت است از «احتمال مقادیر کوچک‌تر یا مساوی x»؛ به عبارت دیگر:

$$F_X(x) = P(X \le x)$$

ازآنجاکه رابطه $P(X \le x) + P(X > x) = 1$ همواره برقرار است، رابطه $P(X > x) = 1 - F_X(x)$ نیز همواره برقرار است.

مشخصات کلی تابع توزیع تجمعی

اگر $F_X(x)$ تابع توزیع تجمعی متغیر تصادفی X (گسسته یا پیوسته) باشد، خصوصیات زیر همواره برای آن صادق است:

۱- ازآنجاکه $F_X(x) = P(X \le x)$ است، $F_X(x)$ تجمع مقادیر احتمال کوچک‌تر یا مساوی x را محاسبه می‌کند و همواره مجموع احتمال‌ها برابر 1 است؛ بنابراین:

$$\begin{cases} F(-\infty) = P(X \le -\infty) = 0 \\ F(+\infty) = P(X \le +\infty) = 1 \end{cases} \Rightarrow \begin{cases} F(-\infty) = 0 \\ F(+\infty) = 1 \end{cases}$$

۲- با توجه به رابطه 1، مقدار $F_X(x)$ همواره بین 0 و 1 است.

$$0 \le F_X(x) \le 1$$

۳- تابع $F_X(x)$ همواره صعودی است.

$$a < b \quad \to \quad F(a) \le F(b)$$

ازآنجاکه $F_X(x) = P(X \le x)$ مجموع مقادیر احتمال کمتر یا مساوی x را محاسبه می‌کند، طبیعی است که هرچه مقدار x بزرگ‌تر شود، مقدار $F_X(x)$ نیز بزرگ‌تر خواهد شد (زیرا تجمع مقادیر احتمال بیشتر شده و مقدار $F_X(x)$ به 1 نزدیک می‌شود)؛ بنابراین تابع $F_X(x)$ غیر نزولی (صعودی) است.

۴- تابع توزیع تجمعی $F_X(x)$ همواره از راست پیوسته است.

$$\lim_{x \to a^+} F_X(x) = F_X(a^+) = F_X(a)$$

مثال کدام‌یک از موارد زیر همواره در یک تابع احتمال نادرست است؟

۱) $F(5) = 1 \, , \, F(0) = 0.5$
۲) $F(5) = 1 \, , \, F(0) = 1$
۳) $F(5) = 0.5 \, , \, F(0) = 1$
۴) $F(0) = 0 \, , \, F(5) = 0.25$

حل: گزینه ۳ درست است.

بنا بر خصوصیت ۳:

گزینه۱) $0 < 5 \;\to\; (F(0) = 0.5) < (F(5) = 1) \;\to$ درست

گزینه۲) $0 < 5 \;\to\; (F(0) = 1) = (F(5) = 1) \;\to$ درست

گزینه۳) $0 < 5 \;\to\; (F(0) = 1) > (F(5) = 0.5) \;\to$ نادرست

گزینه۴) $0 < 5 \;\to\; (F(0) = 0) < (F(5) = 0.25) \;\to$ درست

تابع توزیع تجمعی متغیر تصادفی گسسته

تعریف: اگر متغیر تصادفی گسسته X مفروض و تابع احتمال آن $f(x)=P(x)$ باشد، تابع توزیع تجمعی (تراکمی) آن $\left(F_X(x)\right)$ به صورت زیر محاسبه می‌شود:

$$F_X(x)=P(X\le x)=\sum_{X\le x}P(x)=\sum f(x)$$

به عبارت دیگر، $F_X(x)$ برای نقطه دلخواه x برابر با «$f(x)$ و مجموع $f(x)$های ماقبل آن» است.

✓ **دقت کنید!**

اگر متغیر تصادفی گسسته X دارای N وضعیت $x_1, x_2, \dots, x_N$ باشد، آن‌گاه:

$$F_X(x_i)=P(X\le x_i)=P(x_1)+P(x_2)+\dots+P(x_{i-1})+P(x_i)$$

به عبارت دیگر تابع توزیع تجمعی در متغیر تصادفی گسسته همان فراوانی تجمعی نسبی در فصل آمار توصیفی است.

مثال ۱ تابع احتمال متغیر تصادفی X را به صورت زیر در نظر بگیرید.

x	0	1	2	6
$P(x)=f(x)$	0.4	0.3	0.1	0.2

مطلوب است محاسبه:

الف) $F(X=0)$, $f(X=0)$ ب) $F(X=2)$, $f(X=2)$

ج) $F(X=1.5)$, $f(X=1.5)$ د) $F(X=\sqrt{5})$, $f(X=\sqrt{5})$

هـ) $F(X=10)$, $f(X=10)$ و) $F(X=-1)$, $f(X=-1)$

حل:

الف) $$\begin{cases} f(X=0)=P(X=0)=0.4 \\ F(X=0)=P(X\le 0)=P(X=0)=0.4 \end{cases}$$

ب) $$\begin{cases} f(X=2)=P(X=2)=0.1 \\ F(X=2)=P(X\le 2)=P(X=0)+P(X=1)+P(X=2)=0.8 \end{cases}$$

ج) $$\begin{cases} f(X=1.5)=P(X=1.5)=0 \\ F(X=1.5)=P(X\le 1.5)=P(X=0)+P(X=1)=0.7 \end{cases}$$

د) $$\begin{cases} f(X=\sqrt{5})=P(X=\sqrt{5})=0 \\ F(X=\sqrt{5})=P(X\le\sqrt{5}\simeq 2.2)=P(X=0)+P(X=1)+P(X=2)=0.8 \end{cases}$$

هـ) $$\begin{cases} f(X=10)=P(X=10)=0 \\ F(X=10)=P(X\le 10)=P(X=0)+P(X=1)+P(X=2)+P(X=6)=1 \end{cases}$$

و) $$\begin{cases} f(X=-1)=P(X=-1)=0 \\ F(X=-1)=P(X\le -1)=0 \end{cases}$$

اگر تابع احتمال متغیر تصادفی گسسته X را به صورت زیر در نظر بگیریم:

x_i	x_1	x_2	$\dots$	x_n	
$f(x_i)=P(x_i)$	$f(x_1)$	$f(x_2)$	$\dots$	$f(x_n)$	$\sum f(x)=1$

آن‌گاه تابع توزیع تجمعی (تراکمی) $\left(P(X\le x_i)=F_X(x_i)\right)$ را می‌توان به یکی از دو صورت زیر نمایش داد:

(1)

x_i	x_1	x_2	$\dots$	x_n
$F_X(x_i)$	$F(x_1)=f(x_1)$	$F(x_2)=f(x_1)+f(x_2)$	$\dots$	$F(x_n)=f(x_1)+\dots+f(x_n)=1$

$$(2)\quad P(X\le x_i)=F_X(x_i)=\begin{cases}0 & x<x_1\\ F(x_1)=f(x_1) & x_1\le x<x_2\\ F(x_2)=f(x_1)+f(x_2) & x_2\le x<x_3\\ \vdots & \vdots\\ F(x_{n-1})=f(x_1)+\cdots+f(x_{n-1}) & x_{n-1}\le x<x_n\\ F(x_n)=f(x_1)+\cdots+f(x_n)=1 & x\ge n\end{cases}$$

با توجه به حالات (1) و (2) می‌توان به روابط زیر برای تابع توزیع تجمعی $F_X(x)$ متغیر تصادفی گسسته X رسید:

$$F_X(x)=\begin{cases}F(x_1)=f(x_1)\\ F(x_i)=f(x_1)+f(x_2)+\cdots+f(x_i) & i=2,3,\dots,n\\ F(x_n)=1\end{cases}$$

مثال ۲ تابع احتمال متغیر تصادفی X را به صورت زیر در نظر بگیرید. تابع توزیع تجمعی X $(F_X(x))$ کدام است؟

x	0	1	2	6
$P(x)=f(x)$	0.4	0.3	0.1	0.2

حل: با توجه به دو صورت تعریف‌شده، داریم:

$$F_X(0)=f(X\le 0)=f(X=0)=0.4$$
$$F_X(1)=f(X\le 1)=f(X=0)+f(X=1)=0.7$$
$$F_X(2)=f(X\le 2)=f(X=0)+f(X=1)+f(X=2)=0.8$$
$$F_X(6)=f(X\le 6)=f(X=0)+f(X=1)+f(X=2)+f(X=6)=1$$

(1)

x	0	1	2	6
$P(X\le x)=F_X(x)$	0.4	0.7	0.8	1

$$(2)\quad F_X(x)=\begin{cases}0 & x<0\\ 0.4 & 0\le x<1\\ 0.7 & 1\le x<2\\ 0.8 & 2\le x<6\\ 1 & x\ge 6\end{cases}$$

محاسبه $F_X(x)$ با استفاده از $f(x)$

با توجه به نحوه محاسبه تابع توزیع تجمعی $F_X(x)$ بر اساس تابع احتمال $f(x)$ داریم:

$$\begin{cases} F(x_1)=f(x_1) \\ F(x_i)=\underbrace{f(x_1)+f(x_2)+\cdots+f(x_{i-1})}_{F(x_{i-1})}+f(x_i) \quad \to \quad f(x_i)=F(x_i)-F(x_{i-1}) \\ i=2,\dots,n \end{cases}$$

بنابراین:

$$\boxed{\begin{array}{l} f(x_1)=F(x_1) \\ f(x_i)=F(x_i)-F(x_{i-1}) \\ i=2,\dots,n \end{array}} \quad \to \quad \boxed{\begin{array}{l} f(\text{مشاهده اول})=F(\text{مشاهده اول}) \\ f(\text{مشاهده i ام})=F(\text{مشاهده i ام})-F(\text{مشاهده i-1 ام}) \\ i=2,\dots,n \end{array}}$$

مثال ۳ اگر تابع توزیع تجمعی متغیر گسسته X به صورت زیر باشد، $f(x)$ کدام است؟

X	2	5	9
$P(X\le x)=F(x)$	0.1	0.4	1

۱)

X	2	5	9
$P(x)$	0.1	0.3	0.6

۲)

X	2	5	9
$P(x)$	0.1	0.4	0.5

۳)

X	2	5	9
$P(x)$	0.1	0	0.9

۴)

X	2	5	9
$P(x)$	0.1	0.4	1

حل: گزینه ۱ درست است.

با توجه به روابط:

$$\begin{cases} f(x_1)=F(x_1) \\ f(x_i)=F(x_i)-F(x_{i-1}) \qquad i=2,3,\dots,n \end{cases}$$

داریم:

$$\begin{cases} f(2)=F(2)=0.1 \\ f(5)=F(5)-F(2)=0.4-0.1=0.3 \\ f(9)=F(9)-F(5)=1-0.4=0.6 \end{cases}$$

درنتیجه:

X	2	5	9
$P(x)=f(x)$	0.1	0.3	0.6

مثال ۴ تابع توزیع تجمعی متغیر تصادفی گسسته X به صورت زیر است. $f(x)$ کدام است؟

$$F_X(x)=\begin{cases} 0 & , \quad x<2 \\ 0.3 & , \quad 2\le x<4 \\ 0.45 & , \quad 4\le x<7 \\ 0.8 & , \quad 7\le x<9 \\ 1 & , \quad x\ge 9 \end{cases}$$

۱)

x	2	4	7	9
$f(x)$	0.3	0.45	0.8	0.1

۲)

x	2	4	7	9
$f(x)$	0.3	0.15	0.35	0.2

۳)

x	2	4	7	9
$f(x)$	0.3	0.15	0.45	0.1

۴) هیچ‌کدام

حل: گزینه ۲ درست است.

با توجه به روابط:

$$\begin{cases} f(x_1)=F(x_1) \\ f(x_i)=F(x_i)-F(x_{i-1}) \qquad i=2,3,\dots,n \end{cases}$$

داریم:

$$\begin{cases} f(X_1=2)=F(X_1=2)=0.3 \\ f(X_2=4)=F(X_2=4)-F(X_1=2)=0.45-0.3=0.15 \\ f(X_3=7)=F(X_3=7)-F(X_2=4)=0.8-0.45=0.35 \\ f(X_4=9)=F(X_4=9)-F(X_3=7)=1-0.8=0.2 \end{cases}$$

درنتیجه:

x	2	4	7	9
$P(x)=f(x)$	0.3	0.15	0.35	0.2

مثال ۵ اگر تابع توزیع کمیت تصادفی ناپیوسته X به‌صورت زیر باشد:

$$F_X(x)=\begin{cases} 0 & x<1 \\ \dfrac{2}{10} & 1\le x<4 \\ 1 & x\ge 4 \end{cases}$$

تابع احتمال آن کدام است؟ (اقتصاد ـ ۷۲)

۱)

X	1	4
$P_X(x)$	0.2	0.8

۲)

X	1	4
$P_X(x)$	0.8	0.4

۳)

X	0	1	4
$P_X(x)$	0.2	0.5	0.3

۴)

X	0	1	4
$P_X(x)$	0.3	0.5	0.2

حل: گزینه ۱ درست است.

با توجه به روابط:

$$\begin{cases} f(x_1)=F(x_1) \\ f(x_i)=F(x_i)-F(x_{i-1}) \qquad i=2,3,\dots,n \end{cases}$$

داریم:

$$\begin{cases} f(X=1)=F(X=1)=0.2 \\ f(X=4)=F(X=4)-F(X=1)=1-0.2=0.8 \end{cases}$$

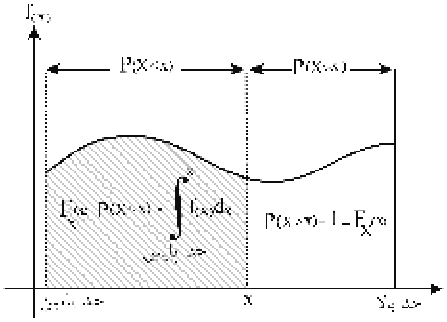
درنتیجه:

X	1	4
$P_X(x)=f(x)$	0.2	0.8

تابع توزیع تجمعی متغیر تصادفی پیوسته

تعریف: هرگاه $f(x)$ تابع چگالی متغیر تصادفی پیوسته X با یک ضابطه باشد، آن‌گاه تابع توزیع تجمعی آن $\left(F_X(x)\right)$ به صورت زیر محاسبه می‌شود:

$$F_X(x) = P(X \le x) = \int_{\text{حد پایین}}^{x} f(x)\,dx$$



مثال ۱ تابع چگالی متغیر تصادفی X به صورت زیر است:

$$f(x) = \begin{cases} 2x & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{سایر نقاط} \end{cases}$$

تابع توزیع تجمعی (تراکمی) X کدام است؟

۱) $F_X(x) = 2x$ ۲) $F_X(x) = x^2$ ۳) $F_X(x) = \frac{x}{2}$ ۴) $F_X(x) = 2x^2$

حل: گزینه ۲ درست است.

$$F_X(x) = P(X \le x) = \int_{\text{حد پایین}}^{x} f(x)\,dx = \int_0^x 2x\,dx = \left[x^2\right]_0^x = x^2$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

مثال ۲ درصورتی‌که تابع چگالی متغیر تصادفی X به صورت زیر باشد، تابع توزیع تجمعی (تراکمی) X کدام است؟

$$f(x) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} & ; \quad x > 0 \\ 0 & ; \quad \text{در غیر این صورت} \end{cases}$$

۱) $e^{-\frac{x}{2}}$ ۲) e^{-x} ۳) $1 - e^{-\frac{x}{2}}$ ۴) $2e^{-\frac{x}{2}}$

حل: گزینه ۳ درست است.

$$F_X(x) = P(X \le x) = \int_{\text{حد پایین}}^{x} f(x)\,dx = \int_0^x \frac{1}{2}e^{-\frac{x}{2}}\,dx = \left[-e^{-\frac{x}{2}}\right]_0^x = -e^{-\frac{x}{2}} - (-1) = 1 - e^{\frac{-x}{2}}$$

$$F_X(x)=\begin{cases}0 & , & x<0\\ 1-e^{-\frac{x}{2}} & , & 0\le x<\infty\\ 1 & , & x\ge\infty\end{cases} \rightarrow F_X(x)=\begin{cases}0 & , & x<0\\ 1-e^{\frac{-x}{2}} & , & x\ge 0\end{cases}$$

مثال ۳ اگر تابع چگالی متغیر تصادفی پیوسته X به صورت زیر باشد، تابع توزیع تجمعی (تراکمی) X کدام است؟

$$f(x)=\begin{cases}\frac{2}{5}x & ; & 2<x<3\\ 0 & ; & \text{در غیر این‌صورت}\end{cases}$$

۱) $\frac{x^2}{5}$ ۲) $\frac{x^2-4}{5}$

۳) x^2-4 ۴) $\frac{x}{5}-4$

حل: گزینه ۲ درست است.

$$F_X(x)=P(X\le x)=\int_{\text{حد پایین}}^{x} f(x)dx=\int_2^x \frac{2}{5}x\,dx=\left[\frac{x^2}{5}\right]_2^x=\frac{x^2}{5}-\frac{4}{5}=\frac{x^2-4}{5}$$

نکته: ازآنجاکه همواره رابطه $F_X(x)=\int_{\text{حد پایین}}^{x} f(x)dx$ در متغیرهای تصادفی پیوسته برقرار است، داریم $f(x)=F'_X(x)$؛ درنتیجه برای حل سؤالاتی مانند مثال‌های ۱ و ۲ یک روش تستی آن است که گزینه‌ها را کنترل کنیم و ببینیم مشتق کدام گزینه $(F'_X(x))$ برابر با $f(x)$ خواهد شد.

$$\boxed{f(x)=F'(x)}$$

در مثال ۱، تنها مشتق تابع توزیع تجمعی گزینه ۲ با $f(x)=2x$ برابر است.

گزینه۲) $F'_X(x)=\left(x^2\right)'=2x=f(x)$ ✓

در مثال ۲، تنها مشتق تابع توزیع تجمعی گزینه ۳ با $f(x)=\frac{1}{2}e^{-\frac{x}{2}}$ برابر است.

گزینه۳) $F'_X(x)=\left(1-e^{-\frac{x}{2}}\right)'=\frac{1}{2}e^{-\frac{x}{2}}=f(x)$ ✓

✓ دقت کنید!

در بعضی مسایل ممکن است مشتق $F_X(x)$ در بیش از یک گزینه برابر $f(x)$ شود؛ در این حالت باید از روشی که در تعریف ارائه شد، استفاده کنیم.

در مثال ۳، مشتق تابع توزیع تجمعی دو گزینه ۱ و ۲ با $f(x)=\frac{2x}{5}$ برابر است، اما با استفاده از روش ارائه شده در تعریف، گزینه ۲ درست است.

گزینه۱) $F'_X(x)=\left(\frac{x^2}{5}\right)'=\frac{2x}{5}=f(x)$ ✓

گزینه۲) $F'_X(x)=\left(\frac{x^2-4}{5}\right)'=\frac{2x}{5}=f(x)$ ✓

محاسبه $F_X(x)$ در توابع چندضابطه‌ای $f(x)$

اگر $f(x)$ چندضابطه‌ای باشد، برای محاسبه تابع توزیع تجمعی $P(X \le x) = F_X(x)$ کافی است در هر ضابطه، مجموع ضابطه‌های قبلی را به تعریف $F_X(x)$ برای ضابطه اضافه کنیم، برای مثال، اگر $f(x)$ دوضابطه‌ای باشد، به صورت زیر عمل می‌کنیم:

$$f(x) = \begin{cases} g(x) & a < x < b \\ h(x) & b < x < c \\ 0 & \end{cases} \rightarrow F_x(x) = \begin{cases} 0 & x < a \\ \int_a^x g(x)dx & a \le x < b \\ \int_a^b g(x)dx + \int_b^x h(x)dx & b \le x < c \\ 1 & x \ge c \end{cases}$$

مثال ۴ اگر تابع چگالی متغیر تصادفی X به صورت زیر باشد، تابع توزیع تجمعی X را به دست آورید.

$$f(x) = \begin{cases} \frac{-x}{10} & ; & -2 < x < 0 \\ \frac{x}{10} & ; & 0 \le x < 4 \\ 0 & ; & \text{در غیر این صورت} \end{cases}$$

حل:

$$\text{ضابطه اول}: F_X(x) = \int_{-2}^{x} \frac{-x}{10} dx = \left[-\frac{x^2}{20}\right]_{-2}^{x} = \frac{1}{5} - \frac{x^2}{20}$$

$$\text{ضابطه دوم}: F_X(x) = \int_{-2}^{0} \frac{-x}{10} dx + \int_{0}^{x} \frac{x}{10} dx = \left[\frac{-x^2}{20}\right]_{-2}^{0} + \left[\frac{x^2}{20}\right]_{0}^{x} = \frac{1}{5} + \frac{x^2}{20}$$

بنابراین تابع توزیع تجمعی X به صورت زیر است:

$$F_X(x) = \begin{cases} 0 & , & x < -2 \\ \frac{1}{5} - \frac{x^2}{20} & , & -2 \le x < 0 \\ \frac{1}{5} + \frac{x^2}{20} & , & 0 \le x < 4 \\ 1 & , & x \ge 4 \end{cases}$$

محاسبه ضریب ثابت با استفاده از $F_X(x)$

اگر $f(x)$ تابع چگالی متغیر تصادفی X باشد، با توجه به آنکه $\int_{\text{حد پایین}}^{\text{حد بالا}} f(x)dx = 1$ است و با توجه به تعریف تابع توزیع تجمعی به صورت زیر:

$$F_X(x) = P(X \le x) = \int_{\text{حد پایین}}^{x} f(x)dx$$

می‌توانیم (حد پایین) F_X و (حد بالا) F_X را به صورت زیر بررسی کنیم:

$$\begin{cases} F_X(\text{حد پایین}) = P(X \le \text{حد پایین}) = \int_{\text{حد پایین}}^{\text{حد بالا}} f(x)dx = 0 \\ F_X(\text{حد بالا}) = P(X \le \text{حد بالا}) = \int_{\text{حد پایین}}^{\text{حد بالا}} f(x)dx = 1 \end{cases}$$

درنتیجه:

$$\begin{cases} F_X(\text{حد پایین}) = 0 \\ F_X(\text{حد بالا}) = 1 \end{cases} \longrightarrow \text{«}F_X(x)\text{، تابع توزیع تجمعی متغیر پیوسته»}$$

مثال ۵ درصورتی‌که تابع توزیع تجمعی زیر مفروض باشد، مقدار k کدام است؟

$$F_X(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{kx^2}{2} & , \quad 0 \le x < 1 \\ 1 & , \quad x \ge 1 \end{cases}$$

۱) 2 ۲) 1

۳) 4 ۴) 5

حل: گزینه ۱ درست است.

$$\begin{cases} F_X(\text{حد پایین}) = 0 \;\rightarrow\; F(0) = \frac{k \times 0^2}{2} = 0 \;\rightarrow\; 0 = 0 \;\; (\text{جواب نمی‌دهد}) \\ F_X(\text{حد بالا}) = 1 \;\rightarrow\; F(1) = \frac{k \times 1^2}{2} = 1 \;\rightarrow\; k = 2 \end{cases}$$

توجه کنید که چون تابع توزیع متغیر تصادفی پیوسته X در تابع‌های تک‌ضابطه‌ای، پیوسته است برای محاسبه (حد بالا) F_X و (حد پایین) F_X از جایگذاری مقدار حد پایین و حد بالا در ضابطه استفاده می‌کنیم.

مثال ۶ درصورتی‌که تابع توزیع تجمعی زیر مفروض باشد، مقدار k کدام است؟

$$F_X(x) = \begin{cases} 1 - \frac{k}{x^2} & ; \quad x > 2 \\ 0 & ; \quad \text{در غیر این صورت} \end{cases}$$

۱) 2 ۲) 1

۳) 4 ۴) 5

حل: گزینه ۳ درست است.

$$\begin{cases} F_X(\text{حد پایین}) = 0 \;\rightarrow\; F(2) = 1 - \frac{k}{2^2} = 0 \;\rightarrow\; k = 4 \\ F_X(\text{حد بالا}) = 1 \;\rightarrow\; F(\infty) = 1 - \frac{k}{\infty^2} = 1 \;\rightarrow\; 1 = 1 \;\; (\text{جواب نمی‌دهد}) \end{cases}$$

محاسبه $F_X(x)$ با استفاده از $f(x)$

برای محاسبه تابع احتمال $f(x)$ از روی تابع توزیع تجمعی $F_X(x)$ به صورت زیر عمل می‌کنیم:

$$\mathbf{F_X(x)}\;(\text{تابع توزیع تجمعی}) \xrightarrow[\text{در فواصل: } F'_X(x)\;(\text{مشتق تابع تجمعی})]{\text{درنقاط مرزی: } F(x^+)-F(x^-)\neq 0} \mathbf{f_X(x)}\;(\text{تابع احتمال})$$

نکته: هرگاه در یکی از نقاط مرزی تابع توزیع تجمعی رابطه $F_X(x^+) \neq F_X(x^-)$ برقرار باشد، آن‌گاه:

$$f(x) = F_X(x^+) - F_X(x^-)$$

البته فقط در توابع چندضابطه‌ای باید پیوسته بودن نقاط مرزی را بررسی کنیم.

مثال ۱ اگر تابع توزیع متغیر تصادفی X به صورت زیر باشد،تابع احتمال X را به دست آورید.

$$F_X(x)=\begin{cases}0 & , \quad x<0\\ \dfrac{x^2}{8} & , \quad 0\le x<2\\ 1 & , \quad x\ge 2\end{cases}$$

حل:

$$F_X(x)=\begin{cases}0 & , \quad x<0\\ \dfrac{x^2}{8} & , \quad 0\le x<2\\ 1 & , \quad x\ge 2\end{cases} \xrightarrow[F'(x)]{F(x^+)-F(x^-)} f_X(x)=\begin{cases}F(0^+)-F(0^-)=\dfrac{0^2}{8}-0=0 & , \quad x<0\\ \dfrac{x}{4} & , \quad 0\le x<2\\ F(2^+)-F(2^-)=1-\dfrac{2^2}{8}=\dfrac{1}{2} & , \quad x\ge 2\end{cases}$$

درنتیجه داریم:

$$f_X(x)=\begin{cases}\dfrac{x}{4} & , \quad 0<x<2\\ \dfrac{1}{2} & , \quad x=2\end{cases}$$

مثال ۲ اگر $F(x)$ یک تابع توزیع احتمال به صورت زیر باشد، تابع احتمال آن کدام است؟

$$F(x)=\begin{cases}0 & , \quad x<0\\ \dfrac{1}{2}x^2 & , \quad 0\le x\le 1\\ -1+2x-\dfrac{1}{2}x^2 & , \quad 1\le x<2\\ 1 & , \quad x\ge 2\end{cases}$$

حل:

$$F(x)=\begin{cases}0 & , \quad x<0\\ \dfrac{1}{2}x^2 & , \quad 0\le x\le 1\\ -1+2x-\dfrac{1}{2}x^2 & , \quad 1\le x<2\\ 1 & , \quad x\ge 2\end{cases} \xrightarrow[F'(x)]{F(x^+)-F(x^-)} f(x)=\begin{cases}0 & , \quad x=0\\ x & , \quad 0<x<1\\ 0 & , \quad x=1\\ 2-x & , \quad 1<x<2\\ 0 & , \quad x=2\end{cases}$$

تابع توزیع تجمعی $F(x)$ در تمام نقاط مرزی پیوسته است، زیرا:

$$F(0^+)-F(0^-)=0-\frac{0^2}{2}=0$$

$$F(1^+)-F(1^-)=\left(-1+2\times 1-\frac{1^2}{2}\right)-\frac{1^2}{2}=0$$

$$F(2^+)-F(2^-)=1-\left(-1+2\times 2-\frac{2^2}{2}\right)=0$$

و درنتیجه تابع چگالی احتمال به صورت زیر خواهد بود:

$$f(x)=\begin{cases}x & 0<x<1\\ 2-x & 1<x<2\end{cases}$$

نکته: اگر در تمام نقاط مرزی، $F(x^+) = F(x^-)$ باشد، کافی است فقط در فواصل از تابع مشتق بگیریم.

مثال ۳ تابع توزیع (تجمعی احتمال) متغیر تصادفی X به قرار زیر است، تابع چگالی احتمال X کدام است؟ (اقتصاد ـ ۸۳)

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{11}{25}x^2 - 2x & 0 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$

۱) $\frac{11}{25}x - 2$ ۲) $\frac{11}{75}x^3 - x^2$

۳) $\frac{11}{50}x + 2$ ۴) $\frac{22}{25}x - 2$

حل: گزینه ۴ درست است.

تابع توزیع تجمعی در نقاط مرزی $5, 0$ پیوسته است و کافی است فقط در فواصل از تابع مشتق بگیریم:

$$f(x) = F'(x)$$

$$f(x) = F'(x) = \frac{22}{25}x - 2 \qquad 0 \le x < 5$$

مثال ۴ تابع توزیع کمیت تصادفی X (زمان برطرف کردن نقص در تعمیر تلویزیون) به صورت زیر است:

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\frac{x}{100}} & x > 0 \end{cases}$$

تابع چگالی احتمال کمیت تصادفی X کدام است؟ (اقتصاد ـ ۷۲)

۱) $1 - 100e^{-\frac{x}{100}}$ ۲) $100e^{-\frac{x}{100}}$ ۳) $\frac{1}{100}e^{-\frac{x}{100}}$ ۴) $e^{-\frac{x}{100}}$

حل: گزینه ۳ درست است.

$$f(x) = F'_X(x)$$

$$f(x) = F'(x) = 0 + \frac{1}{100}e^{-\frac{x}{100}} = \frac{1}{100}e^{-\frac{x}{100}} \qquad x > 0$$

محاسبه احتمال با استفاده از $F_X(x)$

اگر $P(X \le x) = F_X(x)$، تابع توزیع تجمعی (تراکمی) متغیر تصادفی X با تابع چگالی احتمال $f_X(x)$ باشد، برای محاسبه احتمال در هر محدوده X می‌توانیم از قواعد زیر استفاده کنیم:

حالت اول

متغیر X در تمام نقاط مرزی، پیوسته باشد، یعنی $F_X(x^+) = F_X(x^-)$.

الف) اگر a و b در محدوده x باشند:

(احتمال در هر نقطه پیوسته $x = a$، دقیقاً برابر 0 است.)

$$1)\ P(X = a) = F_X(a^+) - F_X(a^-) = 0$$

$$2)\ P(X \le a) = P(X < a) = F_X(a)$$

$$3)\ P(X > a) = P(X \ge a) = 1 - F_X(a)$$

$$4)\ P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b) = F_X(b) - F_X(a)$$

ب) اگر a و b کمتر از حد پایین x باشند:

$$1)\ P(X = a) = 0$$

$$2)\ P(X < a) = P(X \le a) = 0$$

$$3)\ P(X>a)=P(X\geq a)=1$$
$$4)\ P(a\leq X\leq b)=P(a<X<b)=P(a\leq X<b)=P(a<X\leq b)=0$$

ج) اگر a و b بیشتر از حد بالای x باشند:

$$1)\ P(X=a)=0$$
$$2)\ P(X<a)=P(X\leq a)=1$$
$$3)\ P(X>a)=P(X\geq a)=0$$
$$4)\ P(a\leq X\leq b)=P(a<X<b)=P(a\leq X<b)=P(a<X\leq b)=0$$

مثال ۱ اگر $P(0\leq X\leq 4)=0.65$ و $F(0)=0.15$ باشد، مقدار $P(X\leq 4)=F(4)$ کدام است؟

۱) 0.8 ۲) 0.5 ۳) 0.45 ۴) 1

حل: گزینه ۱ درست است.

بنا بر قسمت (الف ـ ۴) داریم:

$$P(a\leq X\leq b)=F_X(b)-F_X(a)$$
$$P(0\leq X\leq 4)=F_X(4)-F_X(0)\ \rightarrow\ 0.65=F_X(4)-0.15\ \rightarrow\ F_X(4)=0.8$$

مثال ۲ تابع توزیع تجمعی متغیر تصادفی X به‌صورت زیر است.

$$F_X(x)=\begin{cases}0 & x<2\\ 1-\dfrac{4}{x^2} & 2\leq x<\infty\\ 1 & x\geq\infty\end{cases}$$

مطلوب است محاسبه:

الف) $P(X<5)$ ب) $P(1<X<1.5)$ ج) $P(X>4)$ د) $P(X=3)$

حل: تابع توزیع تجمعی نقطه مرزی $x=2$ پیوسته است، بنابراین:

$$F_X(2^+)=F_X(2^-)=0$$

الف) بنا بر قسمت (الف ـ ۲)

$$\begin{cases}P(X<5)=P(X\leq 5)=F_X(5)=1-\dfrac{4}{5^2}=1-\dfrac{4}{25}=\dfrac{21}{25}\\ F_X(x)=1-\dfrac{4}{x^2} \qquad 2\leq x<\infty\end{cases}$$

ب) بنا بر قسمت (ب ـ ۴)

$$\begin{cases}P(1<X<1.5)=0\\ F_X(x)=0 \qquad x<2\end{cases}$$

ج) بنا بر قسمت (الف ـ ۳)

$$\begin{cases}P(X>4)=1-F_x(4)=1-\left(1-\dfrac{4}{4^2}\right)=\dfrac{4}{16}=\dfrac{1}{4}\\ F_X(x)=1-\dfrac{4}{x^2} \qquad 2<x<\infty\end{cases}$$

د) بنا بر قسمت (الف ـ ۱)

$$\begin{cases} P(X=3)=0 \\ F_X(x)=1-\dfrac{4}{x^2} \qquad 2\le x<\infty \end{cases}$$

مثال ۳ تابع توزیع تجمعی متغیر تصادفی پیوسته X به صورت زیر است.

$$F_X(x)=\begin{cases} 0 & x<0 \\ \dfrac{x^2}{2} & 0\le x<1 \\ 2x-1-\dfrac{x^2}{2} & 1\le x<2 \\ 1 & x\ge 2 \end{cases}$$

مطلوب است محاسبه:

الف) $P\left(X\le\dfrac{1}{2}\right)$ ب) $P\left(\dfrac{1}{4}<X<\dfrac{1}{2}\right)$ ج) $P\left(\dfrac{1}{2}\le X<\dfrac{3}{2}\right)$ د) $P\left(X>\dfrac{3}{4}\right)$

حل: تابع توزیع تجمعی در نقاط مرزی $x=0,1,2$ پیوسته است، یعنی:

$$F_X\left(x^+\right)=F_X\left(x^-\right)$$

الف) بنا بر قسمت (الف ـ۲)

$$F_X(x)=\frac{x^2}{2}\ ;\ 0\le x<1\ \to\ P\left(X\le\frac{1}{2}\right)=F_X\left(\frac{1}{2}\right)=\frac{\left(\frac{1}{2}\right)^2}{2}=\frac{1}{8}$$

ب) بنا بر قسمت (الف ـ ۴)

$$F_X(x)=\frac{x^2}{2}\ ;\ 0\le x<1\ \to\ P\left(\frac{1}{4}<X<\frac{1}{2}\right)=F_X\left(\frac{1}{2}\right)-F_X\left(\frac{1}{4}\right)=\frac{\left(\frac{1}{2}\right)^2}{2}-\frac{\left(\frac{1}{4}\right)^2}{2}=\frac{1}{8}-\frac{1}{32}=\frac{3}{32}$$

ج) بنا بر قسمت (الف ـ ۴)

$$\begin{cases} F_X(x)=\dfrac{x^2}{2} & 0\le x<1 \\ F_X(x)=2x-1-\dfrac{x^2}{2} & 1\le x<2 \end{cases}$$

$$P\left(\frac{1}{2}\le X<\frac{3}{2}\right)=F_X\left(\frac{3}{2}\right)-F_X\left(\frac{1}{2}\right)=2\times\frac{3}{2}-1-\frac{\left(\frac{3}{2}\right)^2}{2}-\frac{\left(\frac{1}{2}\right)^2}{2}=\left(2-\frac{9}{8}\right)-\left(\frac{1}{8}\right)=\frac{6}{8}$$

د) بنا بر قسمت (الف ـ ۳)

$$F_X(x)=\frac{x^2}{2}\ ;\ 0\le x<1\ \to\ P\left(X>\frac{3}{4}\right)=1-F_X\left(\frac{3}{4}\right)=1-\frac{\left(\frac{3}{4}\right)^2}{2}=\frac{23}{32}$$

مثال ۴ اگر تابع توزیع کمیت تصادفی X به صورت زیر باشد:

$$F_X(x)=\begin{cases}0 & x\le 0\\ \frac{x}{10} & 0<x\le 10\\ 1 & x>10\end{cases}$$

احتمال $P(2<X<8)$ کدام است؟ (اقتصاد ـ ۷۲)

۱) 0.2 ۲) 0.4 ۳) 0.6 ۴) 0.8

حل: گزینه ۳ درست است.

تابع توزیع تجمعی در نقاط مرزی $x=0, 10$ پیوسته است.

$$F_X(x^+)=F_X(x^-)$$

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha)$$

$$P(2<X<8)=F(8)-F(2)=\frac{8}{10}-\frac{2}{10}=0.6$$

حالت دوم

متغیر X در بعضی نقاط مرزی گسسته باشد (پیوسته نباشد) یعنی $F_X(x^+)\neq F_X(x^-)$.

1) $P(X=a)=F_X(a^+)-F_X(a^-)$

2) $P(X<a)=F_X(a^-)$

3) $P(X\le a)=F_X(a)$

4) $P(a<X<b)=F_X(b^-)-F_X(a^+)$

✓ دقت کنید!

با توجه به پیوسته بودن $F_X(x)$ از راست، در روابط بالا می‌توان به‌جای $F_X(a^+)$ از $F_X(a)$ استفاده کرد.

با استفاده از قواعد بالا می‌توانیم احتمال‌های زیر را محاسبه کنیم:

1) $P(X\ge a)=1-P(X<a)=1-F(a^-)$

2) $P(X>a)=1-P(X\le a)=1-F(a)$

3) $P(a\le X\le b)=P(X=a)+P(a<X<b)+P(X=b)=F(b)-F(a^-)$

مثال ۵ تابع توزیع (پخش) متغیر تصادفی X به صورت زیر تعریف شده است:

$$F_X(x)=\begin{cases}0 & x<0\\ \frac{x^2}{4} & 0\le x<1\\ \frac{1}{2} & 1\le x<2\\ \frac{x}{3} & 2\le x<3\\ 1 & x\ge 3\end{cases}$$

مطلوب است محاسبه:

الف) $P(X=3)$, $P(X=2)$, $P(X=1)$, $P(X=0)$

ب) $P(X\le 2)$, $P(X<2)$

ج) $P(X\le 3)$, $P(X<3)$

د) $P(1\le X\le 3)$, $P(1<X<3)$

حل:

تابع توزیع تجمعی در نقاط $x=0,3$ پیوسته است $\left(F_X\left(x^+\right)=F_X\left(x^-\right)\right)$ اما در نقاط $x=1,2$ گسسته است $\left(F_X\left(x^+\right)\neq F_X\left(x^-\right)\right)$.

الف)

$$P(X=0)=F\left(0^+\right)-F\left(0^-\right)=\frac{0^2}{4}-0=0$$

$$P(X=1)=F\left(1^+\right)-F\left(1^-\right)=\frac{1}{2}-\frac{1^2}{4}=\frac{1}{4}$$

$$P(X=2)=F\left(2^+\right)-F\left(2^-\right)=\frac{2}{3}-\frac{1}{2}=\frac{1}{6}$$

$$P(X=3)=F\left(3^+\right)-F\left(3^-\right)=1-\frac{3}{3}=0$$

ب)

$$P(X<2)=F\left(2^-\right)=\frac{1}{2} \quad \rightarrow \quad P(X\geq 2)=1-P(X<2)=\frac{1}{2}$$

$$P(X\leq 2)=F(2)=\frac{2}{3} \quad \rightarrow \quad P(X>2)=1-P(X\leq 2)=\frac{1}{3}$$

ج)

$$P(X<3)=F\left(3^-\right)=\frac{3}{3}=1 \quad \rightarrow \quad P(X\geq 3)=1-P(X<3)=0$$

$$P(X\leq 3)=F(3)=1 \quad \rightarrow \quad P(X>3)=1-P(X\leq 3)=0$$

د)

$$P(1<X<3)=F\left(3^-\right)-F\left(1^+\right)=\frac{3}{3}-\frac{1}{2}=\frac{1}{2}$$

$$P(1\leq X\leq 3)=P(X=1)+P(1<X<3)+P(X=3)=\frac{1}{4}+\frac{1}{2}+0=\frac{3}{4}$$

محاسبه شاخص‌های مرکزی با استفاده از $F_X(x)$

اگر $F_X(x)$ تابع توزیع تجمعی (تراکمی) متغیر تصادفی پیوسته X باشد، آن‌گاه شاخص‌های مرکزی (نما، چندک، میانه و میانگین) می‌توانند به صورت زیر بررسی و محاسبه شوند.

محاسبه مد

طبق تعریف، مد یا نما (Mo) مقداری است که به ازای آن تابع چگالی $(f(x))$ ماکزیمم شود. همان‌طور که در مبحث محاسبه مد در تابع چگالی مطرح شد، یکی از روش‌های معمول آن است که قرار دهیم $f'(x)=0$. از آنجاکه رابطه $F'_X(x)=f_X(x)$ همواره برای متغیر تصادفی پیوسته X برقرار است، برای محاسبه مد می‌توان قرار داد $F''_X(x)=0$.

مثال تابع توزیع $F(x)=\frac{1}{3}\left(-\frac{x^3}{3}+3x^2-5\right)$ در دامنه $0\leq x\leq 3$ تعریف شده است. نما (مد) آن برابر است با: (اقتصاد ـ ۷۳)

۱) 0.6 ۲) 6 ۳) 3 ۴) 4

حل: گزینه ۳ درست است.

$$F''(Mo) = 0$$

$$F'(x) = \frac{1}{3}\left(-x^2 + 6x\right) \to F''(x) = \frac{1}{3}\left(-2x + 6\right) = 0 \to -2x + 6 = 0 \to x = Mo = 3$$

محاسبه چندک‌ها

اگر ثابت b یکی از شاخص‌های مرکزی (میانه، چارک، دهک و صدک) باشد، آن‌گاه:

$$P(X \le b) = \frac{1}{2} \qquad\qquad b = \text{میانه}$$

$$P(X \le b) = \frac{a}{4} \qquad a = 1,2,3 \qquad b = \text{چارک } a\text{ام}$$

$$P(X \le b) = \frac{a}{10} \qquad a = 1,2,\ldots,9 \qquad b = \text{دهک } a\text{ام}$$

$$P(X \le b) = \frac{a}{100} \qquad a = 1,2,\ldots,99 \qquad b = \text{صدک } a\text{ام}$$

ازآنجاکه $F_X(b) = P(X \le b)$ است، با توجه به روابط بالا برای محاسبه b (چندک مورد نظر) کافی است به صورت زیر عمل کنیم:

$$F_X(b) = \frac{1}{2} \qquad\qquad b = \text{میانه}$$

$$F_X(b) = \frac{a}{4} \qquad a = 1,2,3 \qquad b = \text{چارک } a\text{ام}$$

$$F_X(b) = \frac{a}{10} \qquad a = 1,2,\ldots,9 \qquad b = \text{دهک } a\text{ام}$$

$$F_X(b) = \frac{a}{100} \qquad a = 1,2,\ldots,99 \qquad b = \text{صدک } a\text{ام}$$

✓ **دقت کنید!**

روابط بالا وقتی همیشه صادق است که تابع $F_X(x)$ فقط یک ضابطه داشته باشد.

مثال ۱ اگر تابع توزیع تجمعی X به صورت زیر باشد، میانه توزیع کدام است؟

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \dfrac{2x+3}{18} & 0 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

۱) 6 ۲) 3
۳) 2 ۴) 1

حل: گزینه ۲ درست است.

$$F_X(b) = \frac{1}{2} \to \frac{2b+3}{18} = \frac{1}{2} \to 2b+3 = \frac{18}{2} \to b = 3$$

مثال ۲ تابع توزیع تجمعی X به صورت زیر بیان شده است، صدک هشتادم توزیع کدام است؟

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{10} & 0 \le x < 10 \\ 1 & x \ge 10 \end{cases}$$

۱) 6 ۲) 8
۳) 4 ۴) 0.8

حل: گزینه ۲ درست است.

$$F_X(b)=\frac{80}{100} \rightarrow \frac{b}{10}=\frac{80}{100} \rightarrow b=8$$

محاسبه امید ریاضی

برای محاسبه امید ریاضی (میانگین) متغیر تصادفی X با استفاده از تابع توزیع تجمعی به یکی از دو روش زیر عمل می‌کنیم:

روش اول:

ابتدا تابع احتمال $f(x)$ را محاسبه کرده، سپس به صورت زیر عمل می‌کنیم:

$$E(X)=\sum_{\forall x} x.f(x) \qquad (\text{X گسسته})$$

$$E(X)=\int x\cdot f(x)dx \qquad (\text{X پیوسته})$$

روش دوم:

از تابع توزیع تجمعی $F_X(x)$ مستقیماً برای محاسبه امید ریاضی (میانگین) **متغیر تصادفی پیوسته** X به صورت زیر استفاده می‌کنیم:

$$E(X)=\begin{cases}\int(1-F_X(x))dx & x\geq 0\\ \int(-F_X(x))dx & x<0\end{cases}$$

مثال ۱ اگر تابع توزیع تجمعی X به صورت زیر باشد، امید ریاضی (میانگین) توزیع کدام است؟

$$F_X(x)=\begin{cases}0 & x<0\\ x^2 & 0\leq x<1\\ 1 & x\geq 1\end{cases}$$

۱) $\frac{1}{3}$ ۲) $\frac{2}{3}$

۳) 2 ۴) 1

حل: گزینه ۲ درست است.

روش اول:

$$F'_X(x)=f(x)=2x \rightarrow E(x)=\int_0^1 x f(x)dx=\int_0^1 x\cdot 2xdx=\left[\frac{2x^3}{3}\right]_0^1=\frac{2}{3}$$

روش دوم:

ازآنجاکه مقادیر x مثبت است $(0<x<1)$ از رابطه زیر استفاده می‌کنیم:

$$E(X)=\int_0^1(1-F_X(x))dx=\int_0^1(1-x^2)dx=\left[x-\frac{x^3}{3}\right]_0^1=1-\frac{1}{3}=\frac{2}{3}$$

مثال ۲ اگر تابع توزیع متغیر تصادفی X به صورت باشد، متوسط X برابر است با:

$$F_X(x)=\begin{cases}0 & , & x<0\\ \frac{x^2}{8} & , & 0\leq x<2\\ 1 & , & x\geq 2\end{cases}$$

۱) $\frac{1}{9}$ ۲) $\frac{5}{3}$

۳) $\frac{2}{9}$ ۴) $\frac{4}{9}$

حل: گزینه ۲ درست است.

روش اول:

در مبحث محاسبه $f(x)$ از روی $F(x)$، تابع چگالی $F_X(x)$ را به صورت زیر به دست آوردیم:

$$f_X(x)=\begin{cases}\frac{x}{4} & 0<x<2\\ \frac{1}{2} & x=2\end{cases}$$

حال برای محاسبه متوسط X به صورت زیر عمل می‌کنیم:

$$E(X)=\int xf(x)dx+\sum xf(x)=\int_0^2 x\times\frac{x}{4}dx+2\times\frac{1}{2}=\frac{5}{3}$$

روش دوم:

ازآنجاکه مقادیر x مثبت است $(0<x<2)$ از رابطه زیر استفاده می‌کنیم:

$$E(X)=\int_0^2(1-F_X(x))dx=\int_0^2\left(1-\frac{x^2}{8}\right)dx=\left[x-\frac{x^3}{24}\right]_0^2=2-\frac{8}{24}=\frac{40}{24}=\frac{5}{3}$$

محاسبه $F_Y(y)$ با استفاده از $F_X(x)$

اگر $F_X(x)$ تابع توزیع تجمعی متغیر تصادفی پیوسته X باشد آن‌گاه برای محاسبه تابع توزیع تجمعی $F_Y(y)$ به صورت زیر عمل می‌کنیم:

$$F_Y(y)=P(Y\le y)=P(g(X)\le y)=\begin{cases}P\left(X\le g^{-1}(y)\right)=F_X\left(g^{-1}(y)\right)\\ P\left(X\ge g^{-1}(y)\right)=1-F_X\left(g^{-1}(y)\right)\end{cases}$$

بازه مربوط به y را به ترتیب با قرار دادن $x=a$ و $x=b$ در $Y=g(X)$ به دست می‌آوریم.

مثال تابع توزیع X به صورت زیر است:

$$F_X(x)=\begin{cases}0 & x<0\\ x^2 & 0\le x<1\\ 1 & x\ge 1\end{cases}$$

الف) تابع توزیع $Y=\frac{x+1}{3}$ را به دست آورید.

ب) تابع توزیع $Y=-\operatorname{Ln}x$ را به دست آورید.

حل:

الف)

$$F_Y(y)=P(Y\le y)=P\left(\frac{X+1}{3}\le y\right)=P(X\le 3y-1)=F_X(3y-1)=(3y-1)^2\quad;\quad\frac{1}{3}<y<\frac{2}{3}$$

ب)

$$F_Y(y)=P(Y\le y)=P(-\operatorname{Ln}X\le y)=P\left(X>e^{-y}\right)=1-F_X(e^{-y})=1-\left(e^{-y}\right)^2=1-e^{-2y}\quad;\ y>0$$

تابع مولد گشتاور (Moment Generating Function)

تعریف: تابع مولد گشتاور مرتبه t ام متغیر تصادفی X به صورت زیر تعریف می‌شود:

$$M_X(t) = E\left(e^{tX}\right) \qquad t \in R$$

تابع مولد گشتاور برای متغیر تصادفی گسسته و پیوسته X با تابع احتمال $f(x)$ به صورت زیر است:

X گسسته $\quad M_X(t) = E\left(e^{tX}\right) = \sum_x e^{tx} f(x)$

X پیوسته $\quad M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{+\infty} e^{tx} f(x)\,dx$

نکته:

۱- اگر $M_X(t)$ تابع مولد گشتاور متغیر تصادفی X باشد، مشتق r ام تابع مولد گشتاور در نقطه $t = 0$، $E\left(X^r\right)$ را تولید می‌کند که برابر گشتاور مرتبه r ام حول مبدأ است.

$$M_X^{(r)}(t=0) = \left.\frac{d^r M(t)}{dt^r}\right|_{t=0} = E\left(X^r\right)$$

مشتق اول تابع مولد گشتاور، امید ریاضی X است. $\quad E(X) = M'_X(t=0)$

مشتق دوم تابع مولد گشتاور، امید ریاضی X^2 است. $\quad E\left(X^2\right) = M''_X(t=0)$

$\vdots \qquad \vdots$

مشتق r ام تابع مولد گشتاور، امید ریاضی X^r است. $\quad E\left(X^r\right) = M_X^{(r)}(t=0)$

۲- محاسبه واریانس X با استفاده از تابع مولد گشتاور به صورت زیر است:

$$\sigma_X^2 = M''_X(t)\Big|_{t=0} - \left(M'_X(t)\Big|_{t=0}\right)^2$$

۳- همواره مقدار تابع مولد گشتاور در نقطه $t = 0$ برابر 1 است.

$$M_X(t=0) = 1$$

۴- تابع مولد گشتاور مجموع دو متغیر تصادفی مستقل، برابر است با حاصل‌ضرب تابع مولد گشتاور تک تک آن‌ها.

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

مثال ۱ تابع مولد گشتاور متغیر تصادفی X به صورت $M_X(t) = \frac{1}{(1-4t)^2}$ داده شده است. $E(X)$ برابر است با:

۱) 9 ۲) 8 ۳) 4 ۴) 2

حل: گزینه ۲ درست است.

برای محاسبه $E(X)$، ابتدا مشتق اول تابع مولد گشتاور را به دست می‌آوریم، سپس مقدار t را در آن برابر صفر قرار می‌دهیم.

$$E(X)=M'_X(t)=(-2)(-4)(1-4t)^{-3}\xrightarrow{t=0}E(X)=8$$

مثال ۲ تابع مولد گشتاور متغیر تصادفی X به صورت $M_X(t)=e^{-\lambda(1-e^t)}$ داده شده است؛ واریانس X برابر است با:

۱) λ ۲) λ^2 ۳) $\frac{1}{\lambda}$ ۴) $\frac{1}{\lambda^2}$

حل: گزینه ۱ درست است.

ابتدا مشتقات اول و دوم تابع مولد گشتاور را به دست آورده، مقدار t را در آن‌ها برابر صفر قرار می‌دهیم تا مقادیر $E(X)$ و $E(X^2)$ به دست آیند، سپس از رابطه $\sigma^2=E(X^2)-E(X)^2$، واریانس X را محاسبه می‌کنیم.

$$E(X)=M'_X(t)=\lambda e^t\cdot e^{-\lambda(1-e^t)}\xrightarrow{t=0}E(X)=\lambda$$

$$E(X^2)=M''_X(t)=\lambda e^t\cdot e^{-\lambda(1-e^t)}+\lambda^2e^{2t}\cdot e^{-\lambda(1-e^t)}\xrightarrow{t=0}E(X^2)=\lambda+\lambda^2$$

$$\sigma^2=E(X^2)-E(X)^2=\lambda+\lambda^2-\lambda^2=\lambda$$

تابع توزیع توأم (Joint Distribution Function)

گاهی در تحلیل پدیده‌های تصادفی دیده می‌شود که پدیده مورد نظر با مجموعه‌ای از متغیرهای تصادفی که با آن در ارتباط هستند، تفسیر می‌شود. برای تحلیل و تفسیر چنین پدیده‌هایی لازم است که تغییرات متغیرهای تصادفی در ارتباط با یکدیگر و به صورت توأم (مشترک) بررسی شود.

تابع توأم گسسته

مثال مفهومی: شرکتی در دو پروژه به صورت توأم (مشترک) شرکت کرده است، درصورتی‌که « X : میزان سود در پروژه اول» و « Y : میزان سود در پروژه دوم» باشد، تابع احتمال توأم برای X و Y به صورت زیر ارائه شده است:

$x \backslash y$	1	2	4
3	0.2	0.1	0.3
5	0.15	0.1	0.15

برای مثال، با احتمال 0.3، شرکت در پروژه اول $X = 3$ و در پروژه دوم $Y = 4$ سود می‌کند.

حال فرض کنید n متغیر از X و k متغیر از Y به صورت زیر با هم در ارتباط باشند:

$x \backslash y$	y_1	y_2	$\cdots$	y_k
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	$\cdots$	$f(x_1,y_k)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	$\cdots$	$f(x_2,y_k)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$f(x_n,y_1)$	$f(x_n,y_2)$	$\cdots$	$f(x_n,y_k)$

تعریف: $f(x,y)$ را تابع احتمال توأم دو متغیر تصادفی **گسسته** X و Y می‌نامند، هرگاه شرایط زیر را داشته باشد:

الف) $0 \le f(x,y) = P(x,y) \le 1$

به ازای تمام (x,y)ها، تابع $f(x,y)$ همان احتمال $P(x,y)$ است که مقداری بین صفر و یک دارد.

ب) $\sum_{\forall x}\sum_{\forall y} f(x,y) = \sum_{\forall y}\sum_{\forall x} f(x,y) = 1$

مجموع تمام مقادیر $f(x,y)$ برابر 1 است.

مثال ۱ مقدار k را طوری تعیین کنید که تابع $\begin{cases} f(x,y) = k(x+y) \\ X = 0,1,2 \ , \ Y = 0,1 \end{cases}$ را بتوان به عنوان تابع احتمال به کار برد؟

۱) $\frac{1}{6}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{2}$ ۴) $\frac{1}{9}$

حل: گزینه ۴ درست است.

راه حل اول: ازآنجاکه مجموع مقادیر $f(x,y)$ به ازای تمام زوج‌های (x,y) برابر با 1 است، داریم:

$$\sum_{y=0}^{1}\sum_{x=0}^{2} f(x,y)=1 \rightarrow f(0,0)+f(0,1)+f(1,0)+f(1,1)+f(2,0)+f(2,1)=1$$
$$\rightarrow k(0+0)+k(0+1)+k(1+0)+k(1+1)+k(2+0)+k(2+1)=1$$
$$\rightarrow 0+k+k+2k+2k+3k=1$$
$$\rightarrow 9k=1 \rightarrow k=\frac{1}{9}$$

راه حل دوم: ابتدا جدول تابع توأم $f(x,y)$ را برای تمام زوج‌های (x,y) به صورت زیر تشکیل می دهیم:

y \ x	0	1	2
0	$k(0+0)$	$k(1+0)$	$k(2+0)$
1	$k(0+1)$	$k(1+1)$	$k(2+1)$

حال با توجه به آنکه مجموع همه مقادیر $f(x,y)$ باید برابر با 1 باشد، داریم:

$$\sum\sum f(x,y)=1 \longrightarrow 0+k+k+2k+2k+3k=1 \rightarrow 9k=1 \rightarrow k=\frac{1}{9}$$

مثال ۲ دو متغیر تصادفی ناپیوسته (گسسته) X و Y با قانون توزیع (تابع احتمال) توأم $P(X=x, Y=y)=f(x,y)$ در نظر گرفته می‌شوند، کدام‌یک از روابط زیر برای توزیع فوق صادق است؟ (اقتصاد ـ ۷۱)

۱) $\sum_{x} f(x,y)=1$ ۲) $\sum_{x}\sum_{y} f(x,y)=1$ ۳) $\sum_{y} f(x,y)=1$ ۴) هیچ‌کدام

حل: گزینه ۲ درست است.

مثال ۳ تابع احتمال توأم زیر را درنظر بگیرید. مقدار α کدام است؟

۱) 0.2 ۲) 0.3
۳) 0.25 ۴) 0.15

x \ y	1	2	6
2	0.1	0.3	0.2
5	0.1	α	0.1

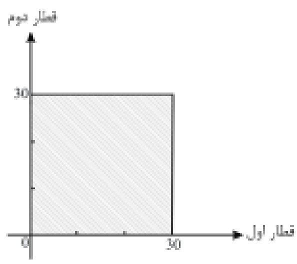
حل: گزینه ۱ درست است.

مجموع احتمال‌های تمام مقادیر تابع احتمال توأم $f(x,y)$ باید 1 شود:

$$\sum\sum f(x,y)=1 \rightarrow 0.1+0.3+0.2+0.1+\alpha+0.1=1 \rightarrow \alpha=0.2$$

تابع توأم پیوسته

مثال مفهومی: اگر دو قطار در فاصله زمانی ۸ تا ۸:۳۰ به ایستگاهی برسند، زمان رسیدن آن‌ها به ایستگاه متغیرهای تصادفی X و Y در بازه $(0,30)$ دقیقه است:



همان‌طورکه دیده می‌شود، فضای نمونه برای ورود دو قطار به صورت توأم پیوسته بوده و تشکیل سطحی به شکل مربع داده است.

تعریف: اگر X و Y دو متغیر تصادفی **پیوسته** با تابع چگالی توأم $f(x,y)$ باشند، آن‌گاه:

الف) $f(x,y) \geq 0$

به ازای تمام (x,y)ها، تابع $f(x,y)$ همان احتمال $P(x,y)$ است که مقداری بین صفر و یک دارد.

ب) $\iint f(x,y)\,dx\,dy = \iint f(x,y)\,dy\,dx = 1$

مجموع تمام مقادیر $f(x,y)$ برابر 1 است.

مثال ۱ تابع چگالی توأم $f(x,y) = cxy\ ;\ 0<x<1\ ,\ 0<y<2$ را در نظر بگیرید. ثابت c کدام است؟

۱) 1 ۲) $\frac{1}{2}$ ۳) 2 ۴) $\sqrt{2}$

حل: گزینه ۱ درست است.

بنا بر تعریف تابع چگالی توأم به صورت $\iint f(x,y)\,dx\,dy = \iint f(x,y)\,dy\,dx = 1$، داریم:

$$\int_0^2 \int_0^1 cxy\,dx\,dy = 1 \quad \rightarrow \quad \int_0^2 cy\left[\frac{x^2}{2}\right]_0^1 dy = \int_0^2 \frac{c}{2} y\,dy = \frac{c}{2}\left[\frac{y^2}{2}\right]_0^2 = 1 \rightarrow \quad c=1$$

بنابراین تابع چگالی توأم X و Y به صورت زیر است:

$$f(x,y) = xy \quad ; \quad 0<x<1\ ,\quad 0<y<2$$

مثال ۲ تابع توزیع توأم $f(x,y) = cx^2y\ ;\ 0<x<1\ ,\ 0<y<1$ را در نظر بگیرید. ثابت c کدام است؟

۱) 5 ۲) 4 ۳) 6 ۴) 2

حل: گزینه ۳ درست است.

بنا بر تعریف تابع چگالی توأم به صورت $\iint f(x,y)\,dx\,dy = \iint f(x,y)\,dy\,dx = 1$، داریم:

$$\int_0^1 \int_0^1 cx^2y\,dx\,dy = 1 \quad \rightarrow \quad \int_0^1 cy\left[\frac{x^3}{3}\right]_0^1 dy = \frac{c}{3}\left[\frac{y^2}{2}\right]_0^1 = 1 \quad \rightarrow \quad c=6$$

بنابراین تابع چگالی توأم X و Y به صورت زیر است:

$$f(x,y) = 6x^2y \quad ; \quad 0<x<1\ ,\quad 0<y<1$$

محاسبه احتمال در توابع توأم

توابع گسسته

اگر $f(x,y)$ تابع توأم (مشترک) متغیرهای تصادفی گسسته X و Y باشد، برای محاسبه احتمال روی ناحیه دلخواه A از رابطه زیر استفاده می‌کنیم:

$$P(A) = f(A) = \sum_A\sum f(x,y)$$

به عبارت دیگر، مجموع مقادیر $f(x,y)$ را به ازای ضابطه A محاسبه می‌کنیم.

مثال ۱ اگر مقادیر توزیع احتمال توأم X و Y به صورتی باشند که در جدول زیر داده شده‌اند.

x \ y	0	1	2	3
0	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{120}$
1	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{20}$	0
2	$\frac{1}{24}$	$\frac{1}{40}$	0	0

مطلوب است محاسبه:

الف) $P(X=1, Y=2)$

ب) $P(X=0, 1\le Y<3)$

ج) $P(X+Y\le 1)$

د) $P(X>Y)$

حل:

الف)
$$P(X=1,\ Y=2)=\frac{1}{20}$$

ب)
$$P(X=0\ ,1\le Y<3)=P(X=0\ ,Y=1)+P(X=0,\ Y=2)=\frac{1}{4}+\frac{1}{8}=\frac{3}{8}$$

ج)
$$P(X+Y\le 1)=P(X=0,Y=1)+P(X=1,Y=0)+P(X=0,Y=0)=\frac{1}{4}+\frac{1}{6}+\frac{1}{12}=\frac{6}{12}=\frac{1}{2}$$

د)
$$P(X>Y)=P(X=1,Y=0)+P(X=2,Y=0)+P(X=2,Y=1)=\frac{1}{6}+\frac{1}{24}+\frac{1}{40}=\frac{40+10+6}{240}=\frac{56}{240}=\frac{7}{30}$$

مثال ۲ تابع احتمال توأم زیر را درنظر بگیرید.

x \ y	1	2	6
2	0.1	0.3	0.2
5	0.1	0.2	0.1

مطلوب است محاسبه:

الف) $P(X=2,Y=1)$ ب) $P(X=2)$ ج) $P(Y=6)$ د) $P(X<Y)$

ه) $P(X>Y)$ و) $P(X=Y)$ ز) $P(X\le Y)$ ح) $P(X\ge Y)$

حل:

الف)
$$P(X=2,Y=1)=f(X=2,Y=1)=0.1$$

ب)
$$P(X=2)=P(X=2,Y=1)+P(X=2,Y=2)+P(X=2,Y=6)=0.1+0.3+0.2=0.6$$

x \ y	1	2	6	
2	0.1 +	0.3 +	0.2	$P(X=2)=0.6$
5	0.1	0.2	0.1	

ج) $P(Y=6)=P(X=2,Y=6)+P(X=5,Y=6)=0.2+0.1=0.3$

X \ Y	1	2	6
2	0.1	0.3	0.2
			+
5	0.1	0.2	0.1
			$P(Y=6)=0.3$

د) $P(X<Y)=f(X<Y)=\sum\sum_{x<y} f(x,y)=f(X=2,Y=6)+f(X=5,Y=6)=0.2+0.1=0.3$

ه) $P(X>Y)=f(X>Y)=\sum\sum_{x>y} f(x,y)=f(X=2,Y=1)+f(X=5,Y=1)+f(X=5,Y=2)=0.1+0.1+0.2=0.4$

و) $P(X=Y)=f(X=Y)=\sum\sum_{x=y} f(x,y)=f(X=2,Y=2)=0.3$

ز) $P(X\le Y)=P(X<Y)+P(X=Y)=0.3+0.3=0.6$

ح) $P(X\ge Y)=P(X>Y)+P(X=Y)=0.4+0.3=0.7$

توابع پیوسته

اگر $f(x,y)$ تابع توأم (مشترک) متغیرهای تصادفی پیوسته X و Y باشد، برای محاسبه احتمال روی ناحیه دلخواه A از رابطه زیر استفاده می‌کنیم:

$$P(A)=\iint_A f(x,y)\,dx\,dy=\iint_A f(x,y)\,dy\,dx$$

به عبارت دیگر از تابع توأم روی ناحیه A انتگرال می‌گیریم.

مثال تابع چگالی توأم $f(x,y)=xy\,;\,0<x<1\,,\,0<y<2$ را در نظر بگیرید. مطلوب است محاسبه:

الف) $P\left(X<\frac{1}{2}, Y>1\right)$ ب) $P\left(X<\frac{1}{2}\right)$ ج) $P\left(X=\frac{1}{3}, Y=1\right)$

حل:

الف) بنا بر تعریف، احتمال در هر بازه، سطح زیر منحنی (انتگرال) در آن بازه است، بنابراین:

$$P\left(X<\frac{1}{2}, Y>1\right)=\int_1^2\int_0^{\frac{1}{2}} xy\,dx\,dy=\int_1^2 y\left[\frac{x^2}{2}\right]_0^{\frac{1}{2}}dy=\int_1^2 \frac{y}{8}dy=\left[\frac{y^2}{16}\right]_1^2=\frac{3}{16}$$

ب) در محاسبه $P\left(x<\frac{1}{2}\right)$ حدود مربوط به y آزاد است و در تمام بازه $0<y<2$ در نظر گرفته می‌شود، بنابراین:

$$P\left(X<\frac{1}{2}\right)=P\left(0<X<\frac{1}{2},0<Y<2\right)=\int_0^2\int_0^{\frac{1}{2}}xydxdy=\int_0^2 y\left[\frac{x^2}{2}\right]_0^{\frac{1}{2}}dy=\int_0^2\frac{y}{8}dy=\left[\frac{y^2}{16}\right]_0^2=\frac{4}{16}=\frac{1}{4}$$

ج) تابع چگالی توأم X و Y پیوسته است، بنابراین: $P\left(X=\frac{1}{3},\ Y=1\right)=0$

تابع حاشیه‌ای (کناره‌ای) (Marginal Function)

با استفاده از تابع احتمال توأم $f(x,y)$ همواره می‌توان توابع احتمال جداگانه هر یک از متغیرهای تصادفی X و Y را پیدا کرد. به توابع $f(x)$ و $f(y)$ که از تابع احتمال توأم $f(x,y)$ به دست می‌آیند، توابع احتمال کناره‌ای (حاشیه‌ای) گفته می‌شود.

توابع حاشیه‌ای گسسته

تعریف: اگر X و Y متغیرهای تصادفی گسسته با تابع احتمال توأم $f(x,y)$ باشند، آن‌گاه توابع احتمال کناره‌ای (حاشیه‌ای) هریک از متغیرها به صورت زیر است:

$$f(x)=\sum_{\forall y}f(x,y)\quad,\quad f(y)=\sum_{\forall x}f(x,y)$$

تابع حاشیه‌ای $f(x)$

با توجه به جدول زیر دیده می‌شود که برای محاسبه تابع احتمال حاشیه‌ای $f(x)$ کافی است مجموع $\sum_{\forall y}f(x,y)$ را به ازای جمیع مقادیر Y محاسبه کنیم، در این وضعیت X ثابت در نظر گرفته شده است.

x \ y	y_1		y_2		$\dots$		y_k	$f(x)=P(x)$
x_1	$f(x_1,y_1)$	+	$f(x_1,y_2)$	+	$\dots$	+	$f(x_1,y_k)$	$f(x_1)$
x_2	$f(x_2,y_1)$	+	$f(x_2,y_2)$	+	$\dots$	+	$f(x_2,y_k)$	$f(x_2)$
$\vdots$	$\vdots$		$\vdots$				$\vdots$	$\vdots$
x_n	$f(x_n,y_1)$	+	$f(x_n,y_2)$	+	$\dots$	+	$f(x_n,y_k)$	$f(x_n)$

تابع حاشیه‌ای $f(y)$

با توجه به جدول زیر دیده می‌شود که برای محاسبه تابع احتمال حاشیه‌ای $f(y)$ کافی است مجموع $\sum_{\forall x}f(x,y)$ را به ازای جمیع مقادیر X محاسبه کنیم، در این وضعیت Y ثابت در نظر گرفته می‌شود.

$y \backslash x$	y_1	y_2	$\cdots$	y_k
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	$\cdots$	$f(x_1,y_k)$
	$+$	$+$		$+$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	$\cdots$	$f(x_2,y_k)$
	$+$	$+$		$+$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$+$	$+$		$+$
x_n	$f(x_n,y_1)$	$f(x_n,y_2)$	$\cdots$	$f(x_n,y_k)$
$f(y)=P(y)$	$f(y_1)$	$f(y_2)$		$f(y_k)$

مثال ۱ تابع احتمال توأم زیر را در نظر بگیرید:

$y \backslash x$	0	1	2
2	0.1	0.2	0.1
5	0.2	0.2	0.2

الف) تابع احتمال کناره‌ای (حاشیه‌ای) X را به دست آورید.

ب) تابع احتمال کناره‌ای (حاشیه‌ای) Y را به دست آورید.

حل:

الف)

$y \backslash x$	0	1	2	$f(x)=P(x)$
2	0.1 +	0.2 +	0.1	0.4
5	0.2 +	0.2 +	0.2	0.6

درنتیجه تابع کناره‌ای X به صورت زیر است:

x	2	5	
$P(x)=f(x)$	0.4	0.6	$\sum f(x)=1$

ب)

$y \backslash x$	0	1	2
2	0.1	0.2	0.1
	$+$	$+$	$+$
5	0.2	0.2	0.2
$P(y)=f(y)$	0.3	0.4	0.3

درنتیجه تابع کناره‌ای Y به صورت زیر است:

y	0	1	2	
$P(y)=f(y)$	0.3	0.4	0.3	$\sum f(y)=1$

توابع حاشیه‌ای پیوسته

تعریف: اگر X و Y متغیرهای تصادفی پیوسته با تابع چگالی احتمال توأم $f(x,y)$ باشند، آنگاه توابع احتمال کناره‌ای (حاشیه‌ای) هر کدام از متغیرها به صورت زیر است:

$$f(x)=\int_Y f(x,y)dy \quad , \quad f(y)=\int_X f(x,y)dx$$

مثال ۲ تابع توزیع توأم $f(x,y)=6x^2y\ ;\ 0<x<1\ ,\ 0<y<1$ را در نظر بگیرید:

الف) تابع کناره‌ای $f(x)$ را به دست آورید.

ب) تابع کناره‌ای $f(y)$ را به دست آورید.

حل:

الف)
$$f(x)=\int_y f(x,y)dy=\int_0^1 6x^2y\,dy=6x^2\left[\frac{y^2}{2}\right]_0^1=3x^2$$

ب)
$$f(y)=\int_x f(x,y)dx=\int_0^1 6x^2y\,dx=6y\left[\frac{x^3}{3}\right]_0^1=2y$$

محاسبه امید و واریانس در توابع توأم

ابتدا تابع کناره‌ای متغیر تصادفی مورد نظر را به دست می‌آوریم، سپس امید ریاضی را به همان روش توابع یک‌متغیره محاسبه می‌کنیم:

درصورتی‌که X و Y دو متغیر تصادفی **گسسته** باشند، آن‌گاه:

$$\begin{cases} E(X)=\sum_x xP(x) \quad ; \quad P(x)=\sum_y P(x,y) \ \text{چگالی کناری} \\ E(Y)=\sum_y yP(y) \quad ; \quad P(y)=\sum_x P(x,y) \ \text{چگالی کناری} \end{cases}$$

درصورتی‌که X و Y دو متغیر تصادفی **پیوسته** باشند، آن‌گاه:

$$\begin{cases} E(X)=\int_x x\cdot f(x)dx \quad ; \quad f(x)=\int_y f(x,y)dy \ \text{چگالی کناری} \\ E(Y)=\int_y y\cdot f(y)dy \quad ; \quad f(y)=\int_x f(x,y)dx \ \text{چگالی کناری} \end{cases}$$

امید ریاضی مجموع دو متغیر تصادفی

در شرایطی که c, b, a مقادیر ثابتی باشند (مثبت یا منفی)، با توجه به خطی بودن میانگین (امید) همواره رابطه زیر برقرار است:

$$\boxed{E\left[aX+bY+c\right]=aE(X)+bE(Y)+c}$$

رابطه بالا در هر شرایطی (X و Y مستقل یا وابسته) برقرار است.

مثال ۱ تابع احتمال توأم زیر را در نظر بگیرید.

x \ y	-1	0
1	0.1	0.2
2	0.3	0.4

مطلوب است محاسبه:

الف) $E(X)$

ب) $E(Y)$

ج) $E(X+Y)$

د) $E(2X-Y+1)$

حل:

الف) ابتدا تابع کناره‌ای X را به دست می‌آوریم:

x \ y	-1	0	$f(x)$
1	0.1 +	0.2	$f(X=1)=0.3$
2	0.3 +	0.4	$f(X=2)=0.7$

$$\Rightarrow \quad E(X)=\sum_{\forall x} x\cdot f(x)=1\times 0.3+2\times 0.7=1.7$$

ب) ابتدا تابع کناره‌ای Y را به دست می‌آوریم:

x \ y	-1	0
1	0.1	0.2
	+	+
2	0.3	0.4
$f(y)$	$f(Y=-1)=0.4$	$f(Y=0)=0.6$

$$\Rightarrow \quad E(Y)=\sum_{\forall y} y\cdot f(y)=-1\times 0.4+0\times 0.6=-0.4$$

ج)
$$E(X+Y)=E(X)+E(Y)=1.7+-0.4=1.3$$

د)
$$E(2X-Y+1)=2E(X)-E(Y)+1=2\times 1.7-(-0.4)+1=4.8$$

مثال ۲ تابع احتمال مشترک دو پروژه سرمایه‌گذاری و میزان سود حاصل به میلیون تومان در جدول زیر داده شده است که در آن X سود پروژه اول و Y سود پروژه دوم است. اگر شرکتی در هر دو پروژه سرمایه‌گذاری کند، امید ریاضی و کمیت انتظاری سود کل وی چقدر خواهد شد؟ (اقتصاد ـ ۷۳)

x \ y	1	2
-2	0.29	0.41
10	0.16	0.14

۱) 1.76

۲) 2.75

۳) 3.15

۴) 4.26

حل: گزینه ۳ درست است.

$x \backslash y$	1	2	$f(x)$
-2	0.29	0.41	0.7
10	0.16	0.14	0.3
$f(y)$	0.45	0.55	1

$$\Rightarrow \begin{cases} E(X)=\sum x f_X(x)=-2\times 0.7+10\times 0.3=1.6 \\ E(Y)=\sum y f_Y(y)=1\times 0.45+2\times 0.55=1.55 \\ E(X+Y)=E(X)+E(Y)=1.6+1.55=3.15 \end{cases}$$

توجه: سودکل = مجموع سود در هر دو پروژه

مثال ۳ تابع چگالی توأم $f(x,y)=\frac{3}{4}x^2y \; ; \; 0<x<2 \, , \; 0<y<1$ را در نظر بگیرید:

الف) مقدار $E(X)$ را به دست آورید.

ب) مقدار $E(Y)$ را به دست آورید.

حل:

الف) ابتدا تابع کناره‌ای X را به دست می‌آوریم:

$$f(x)=\int_Y f(x,y)dy=\int_0^1 \frac{3}{4}x^2y\,dy=\frac{3}{4}x^2\left[\frac{1}{2}y^2\right]_0^1=\frac{3}{8}x^2$$

$$E(X)=\int x f(x)dx=\int_0^2 x\,\frac{3}{8}x^2\,dx=\frac{3}{8}\left[\frac{1}{4}x^4\right]_0^2=\frac{3}{8}.4=\frac{3}{2}=1.5$$

ب) ابتدا تابع کناره‌ای Y را به دست می‌آوریم:

$$f(y)=\int_X f(x,y)dx=\int_0^2 \frac{3}{4}x^2y\,dx=\frac{3}{4}y\left[\frac{1}{3}x^3\right]_0^2=2y$$

$$E(Y)=\int y f(y)dy=\int_0^1 y.2y\,dy=2\left[\frac{1}{3}y^3\right]_0^1=\frac{2}{3}$$

محاسبه واریانس

$$\mathrm{Var}(Y)=E\left(Y^2\right)-\left(E(Y)\right)^2$$

$$\begin{cases} E\left(Y^2\right)=\int_Y y^2 f(y)dy \\ E(Y)=\int_Y y f(y)dy \end{cases}$$

$$f(y)=\int_X f(x,y)dx$$

$$\mathrm{Var}(X)=E\left(X^2\right)-\left(E(X)\right)^2$$

$$\begin{cases} E\left(X^2\right)=\int_X x^2 f(x)dx \\ E(X)=\int_X x f(x)dx \end{cases}$$

$$f(x)=\int_Y f(x,y)dy$$

مثال ۴ تابع چگالی توأم $f(x,y)=\frac{3}{4}x^2y \; ; \; 0<x<2 \, , \; 0<y<1$ را در نظر بگیرید:

الف) مقدار $\mathrm{Var}(X)$ را به دست آورید.

ب) مقدار $\mathrm{Var}(Y)$ را به دست آورید.

حل:

الف) ابتدا تابع کناره‌ای X را به دست می‌آوریم:

$$\begin{cases} Var(X)=E(X^2)-(E(X))^2=\frac{12}{5}-\left(\frac{3}{2}\right)^2=\frac{48-45}{20}=\frac{3}{20} \\ E(X^2)=\int_0^2 x^2\frac{3}{8}x^2dx=\frac{3}{8}\left[\frac{1}{5}x^5\right]_0^2=\frac{3}{8}\cdot\frac{32}{5}=\frac{12}{5} \\ E(X)=\frac{3}{2}\ ,\ f(x)=\frac{3}{8}x^2 \end{cases}$$

ب) ابتدا تابع کناره‌ای Y را به دست می‌آوریم:

$$\begin{cases} Var(Y)=E(Y^2)-(E(Y))^2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{9-8}{18}=\frac{1}{18} \\ E(Y^2)=\int_0^1 y^2.2y\,dy=2\left[\frac{1}{4}y^4\right]_0^1=\frac{1}{2} \\ E(Y)=\frac{2}{3}\ ,\ f(y)=2y \end{cases}$$

استقلال دو متغیر تصادفی

دو متغیر تصادفی X و Y را مستقل (ناوابسته) گویند، اگر و فقط اگر هیچ رابطه‌ای با هم نداشته باشند. در این صورت مقدار تابع توأم آن‌ها $(f(x,y))$ برابر با حاصل‌ضرب توابع کناره‌ای آن‌ها می‌شود؛ به عبارت دیگر:

X و Y مستقل (ناوابسته) $\rightleftarrows$ $f(x,y)=f(x)\cdot f(y)$

تبصره:

الف) اگر X و Y متغیرهای تصادفی گسسته باشند و $f(x)=P(x)$ و $f(y)=P(y)$ و $f(x,y)=P(x,y)$:

X و Y مستقل (ناوابسته) $\rightleftarrows$ $P(x\ ,\ y)=P(x)\cdot P(y)$

ب) اگر دو متغیر تصادفی X و Y مستقل باشند، آن‌گاه زوج متغیرهای تصادفی $\left(X,\frac{1}{Y}\right),\left(\frac{1}{X},Y\right),\left(\frac{1}{X},\frac{1}{Y}\right)$ نیز مستقل‌اند.

وابستگی دو متغیر تصادفی

دو متغیر تصادفی X و Y را وابسته گویند، اگر و فقط اگر با هم رابطه داشته باشند. در این صورت، مقدار تابع توأم آن‌ها $(f(x,y))$ برابر با حاصل‌ضرب توابع کناره‌ای آن‌ها نمی‌شود؛ به عبارت دیگر:

X و Y وابسته $\rightleftarrows$ $f(x,y)\neq f(x)\cdot f(y)$

بررسی استقلال دو متغیر گسسته

برای بررسی استقلال دو متغیر X و Y، ابتدا توابع کناره‌ای $f(x)$ و $f(y)$ را به دست می‌آوریم، سپس درستی رابطه $f(x_i, y_j) = f(x_i) \cdot f(y_j)$ را برای تمام زوج‌های (x_i, y_j) بررسی می‌کنیم.

$x \backslash y$	y_j	
x_i	$f(x_i, y_j) \rightarrow$ $\downarrow$	$f(x_i)$
	$f(y_j)$	

در این وضعیت دو حالت پیش می‌آید:

الف) درصورتی‌که به ازای تمام زوج‌های (x_i, y_j) رابطه $f(x_i, y_j) = f(x_i) \cdot f(y_j)$ برقرار باشد، آن‌گاه دو متغیر تصادفی مستقل‌اند.

ب) درصورتی‌که به ازای حداقل یک زوج (x_i, y_j) رابطه $f(x_i, y_j) \neq f(x_i) \cdot f(y_j)$ برقرار باشد، آن‌گاه دو متغیر تصادفی مستقل نبوده و وابسته‌اند.

تبصره:

ازآنجاکه همواره توابع کناره‌ای $f(y) \neq 0, f(x) \neq 0$ هستند، درصورتی‌که به ازای حداقل یک زوج (x, y)، داشته باشیم $f(x, y) = 0$، X و Y مستقل نیستند (وابسته هستند)، زیرا رابطه $f(x, y) = 0 = f(x \neq 0) \times f(y \neq 0)$ برقرار نیست، درنتیجه X و Y نمی‌توانند مستقل باشند.

مثال ۱ تابع توأم زیر را در نظر بگیرید. آیا X و Y مستقل‌اند؟

$x \backslash y$	2	3
-1	0.2	0.3
0	0.2	0.3

حل:

ابتدا توابع کناره‌ای X و Y را به دست می‌آوریم، سپس برقرار بودن یا نبودن رابطه $f(x, y) = f(x) \cdot f(y)$ را برای تمام زوج‌های (x, y) بررسی می‌کنیم.

$X \backslash Y$	2	3	$f(x)$
-1	0.2	0.3	0.5
0	0.2	0.3	0.5
$f(y)$	0.4	0.6	1

$\xrightarrow{f(x,y)=f(x) \cdot f(y)}$

$X \backslash Y$	2	3	$f(x)$
-1	$0.2 = 0.5 \times 0.4$	$0.3 = 0.5 \times 0.6$	0.5
0	$0.2 = 0.5 \times 0.4$	$0.3 = 0.5 \times 0.6$	0.5
$f(y)$	0.4	0.6	1

همان‌طورکه دیده می‌شود به ازای تمام زوج‌های (x, y) رابطه $f(x, y) = f(x) \cdot f(y)$ برقرار است، بنابراین X و Y مستقل‌اند.

مثال ۲ تابع احتمال توأم زیر را در نظر بگیرید. آیا X و Y مستقل‌اند؟

x \ y	2	3
−1	0.1	0.2
0	0.3	0.4

حل:
ابتدا توابع کناره‌ای X و Y را به دست می‌آوریم، سپس برقرار بودن یا نبودن رابطه $f(x,y)=f(x)\cdot f(y)$ را برای تمام زوج‌های (x,y) بررسی می‌کنیم.

x \ y	2	3	$f(x)$
−1	$\boxed{0.1}$	0.2	$\boxed{0.3}$
0	0.3	0.4	0.7
$f(y)$	$\boxed{0.4}$	0.6	1

$\xrightarrow{f(x,y)=f(x)\cdot f(y)}$

x \ y	2	3	$f(x)$
−1	$0.1 \neq 0.3\times 0.4$		0.3
0			
$f(y)$	0.4		

همان‌طورکه دیده می‌شود:

$$\underbrace{f(X=-1\,,Y=2)}_{0.1} \neq \underbrace{f(X=-1)}_{0.3}\times \underbrace{f(Y=2)}_{0.4}$$

درصورتی‌که حداقل یک زوج (x,y) وجود داشته باشد که $f(x,y)\neq f(x)\cdot f(y)$ باشد، آن‌گاه X و Y مستقل نیستند (وابسته‌اند).

مثال ۳ تابع توأم زیر را در نظر بگیرید، آیا X و Y مستقل‌اند؟

x \ y	−1	0	1
−1	0.1	0.6	0.1
1	0.1	0	0.1

حل:
$f(X=1,Y=0)=0$ است و با توجه به تبصره بیان‌شده، هرگاه به ازای حداقل یک زوج (x,y)، داشته باشیم $f(x,y)=0$، X و Y مستقل نیستند (وابسته‌اند).

مثال ۴ دو متغیر مستقل X و Y با تابع احتمال زیر داده شده‌اند. α کدام است؟ (حسابداری و مدیریت ـ ۸۵)

x \ y	1	2	3
0	0.12	0.2	0.08
2	0.18	α	β

۱) 0.12
۲) 0.2
۳) 0.25
۴) 0.3

حل: گزینه ۴ درست است.
ابتدا توابع کناره‌ای متغیرهای X و Y را به دست می‌آوریم:

$x \backslash y$	1	2	3	$f(x)$
0	0.12	$\boxed{0.2}$	0.08	$\boxed{0.4}$
2	0.18	α	β	0.6
$f(y)$	0.3	$\boxed{0.2+\alpha}$	$0.08+\beta$	1

حال با توجه به آنکه دو متغیر تصادفی مستقل هستند، باید رابطه $f(x\,,\,y)=f(x)f(y)$ به ازای تمام زوج‌های (x,y) از جمله $f(X=0\,,Y=2)$ برقرار باشد، بنابراین:

$$f(X=0\,,\,Y=2)=f(X=0)\times f(Y=2)\ \rightarrow\ 0.2=0.4\times(0.2+\alpha)\ \rightarrow\ \alpha=0.3$$

بررسی استقلال دو متغیر پیوسته

برای متغیرهای تصادفی **پیوسته** X و Y درصورتی‌که:

الف) حدود X و Y مستقل باشند،

ب) تابع چگالی توأم $f(x,y)$ را بتوان به صورت حاصل‌ضرب دو تابع مستقل بر حسب X و Y نوشت،

آن‌گاه X و Y مستقل هستند و در غیر این صورت وابسته خواهند بود.

مثال ۵ تابع چگالی توأم X و Y به صورت زیر داده شده است، آیا X و Y مستقل‌اند؟

$$f(x,y)=\begin{cases} xe^{-(x+y)} & ;\ x>0\ ,\ y>0 \\ 0 & ;\ \text{سایر مقادیر} \end{cases}$$

حل: با توجه به اینکه حدود X و Y به هم وابسته نیست و همچنین می‌توان تابع توأم X و Y را به صورت دو تابع مجزا از X و Y نوشت، نتیجه می‌گیریم که X و Y مستقل‌اند.

$$f(x,y)=xe^{-x}\cdot e^{-y}\ \ ;\ \ x>0\,,\ y>0$$

مثال ۶ تابع توزیع توأم $f(x,y)=6x^2y\ ;\ 0<x<1\,,\ 0<y<1$ را در نظر بگیرید، آیا X و Y مستقل‌اند؟

حل: با توجه به اینکه حدود X و Y به هم وابسته نیست و همچنین می‌توان تابع توأم X و Y را به صورت دو تابع مجزا از X و Y نوشت، نتیجه می‌گیریم که X و Y مستقل‌اند.

$$f(x,y)=6x^2\,.\,y\ \ ;\ \ 0<x<1\,,\ 0<y<1$$

به عبارت دیگر اگر توابع کناره‌ای X و Y را محاسبه کنیم:

$$\begin{cases} f(x)=\int_0^1 6x^2y\,dy=3x^2 \\ f(y)=\int_0^1 6x^2y\,dx=2y \end{cases} \rightarrow \begin{cases} f(x,y)=f(x)\cdot f(y) \\ 6x^2y=3x^2\cdot 2y \end{cases}$$

مثال ۷ تابع چگالی توأم $f(x,y)=\frac{1}{3}(x+y)\ ;\ 0<x<1\ ,\ 0<y<2$ را در نظر بگیرید، آیا X و Y مستقل هستند؟

حل:

X و Y مستقل نیستند.

$$\begin{cases} f(x)=\int_0^2 \frac{1}{3}(x+y)\,dy=\frac{1}{3}(2x+2) \\ f(y)=\int_0^1 \frac{1}{3}(x+y)\,dx=\frac{1}{3}\left(\frac{1}{2}+y\right) \end{cases} \rightarrow \begin{cases} f(x,y)\neq f(x)\cdot f(y) \\ \frac{1}{3}(x+y)\neq\frac{1}{3}(2x+2)\cdot\frac{1}{3}\left(\frac{1}{2}+y\right) \end{cases}$$

تابع توزیع تجمعی توأم (Cumulative Joint Distribution Function)

تعریف: تابع توزیع تجمعی توأم متغیرهای تصادفی X و Y به صورت زیر است:

$$F_{X,Y}(x,y)=P(X\le x\,,\,Y\le y)=\sum\sum f(x,y)$$

مثال تابع احتمال توأم زیر را در نظر بگیرید. مطلوب است محاسبه:

الف) $F(1,1)$

ب) $F(2,1)$

x \ y	0	1	2
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
1	$\frac{2}{9}$	$\frac{1}{6}$	0
2	$\frac{1}{36}$	0	0

حل:

الف)

$$F(1,1)=P(X\le 1,Y\le 1)=f(0,0)+f(0,1)+f(1,0)+f(1,1)=\frac{1}{6}+\frac{1}{3}+\frac{2}{9}+\frac{1}{6}=\frac{8}{9}$$

ب)

$$F(2,1)=P(X\le 2,Y\le 1)=f(0,0)+f(1,0)+f(2,0)+f(0,1)+f(1,1)+f(2,1)=\frac{1}{6}+\frac{2}{9}+\frac{1}{36}+\frac{1}{3}+\frac{1}{6}+0=\frac{33}{36}$$

محاسبه تابع چگالی توأم از روی تابع توزیع توأم پیوسته

برای به دست آوردن تابع چگالی توأم از روی تابع توزیع توأم، باید از هر متغیر مشتق جزیی بگیریم. البته فرقی نمی‌کند که ابتدا کدام متغیر را ثابت در نظر گرفته و نسبت به دیگری مشتق بگیریم.

$$f(x,y)=\frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x,y)=\frac{\partial}{\partial y}\frac{\partial}{\partial x}F(x,y)$$

مثال اگر تابع توزیع توأم دو متغیر تصادفی X و Y به صورت زیر باشد. چگالی احتمال توأم آن‌ها را بیابید.

$$F(x,y)=\begin{cases}\left(1-e^{-x}\right)\left(1-e^{-y}\right) & ;\quad x>0\ ,\ y>0\\ 0 & ;\quad \text{سایر جاها}\end{cases}$$

حل:

$$f(x,y)=\frac{\partial}{\partial y}\frac{\partial}{\partial x}F(x,y)=\frac{\partial}{\partial y}\left(e^{-x}\left(1-e^{-y}\right)\right)=e^{-y}.e^{-x}=e^{-(x+y)}$$

تابع احتمال شرطی (Conditional Probability Function)

اگر تابع احتمال توأم $f(x,y)$ و توابع کناره‌ای $f(x)$ و $f(y)$ مفروض باشند، داریم:

$$f(x|y)=\frac{f(x,y)}{f(y)} \quad , \quad f(y|x)=\frac{f(x,y)}{f(x)}$$

اگر متغیرهای تصادفی X و Y گسسته باشند، آن‌گاه $f(x,y)=P(x,y)$، $f(x)=P(x)$ و $f(y)=P(y)$ است و داریم:

$$P(x|y)=\frac{P(x,y)}{P(y)} \quad \text{و} \quad P(y|x)=\frac{P(x,y)}{P(x)}$$

اگر X و Y دو متغیر تصادفی مستقل از هم باشند، آن‌گاه $f(y|x)=f(y)$ و $f(x|y)=f(x)$ است و برعکس.

$$X \text{ و } Y \text{ مستقل} \quad \rightleftarrows \quad \begin{gathered} f(x|y)=f(x) \\ f(y|x)=f(y) \end{gathered}$$

یادآوری: اگر $f(x|y)=f(x)$ باشد، داریم:

$$f(x|y)=\frac{f(x,y)}{f(y)}=f(x) \quad \to \quad f(x,y)=f(x)\cdot f(y) \quad \to \quad X \text{ و } Y \text{ مستقل}$$

مثال ۱ تابع احتمال توأم زیر مفروض است. مطلوب است محاسبه:

الف) $P(X=-2|Y=0)$

ب) $P(Y=10|X=-2)$

x \ y	0	10
-2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

حل:

الف)

X \ Y	0	10
-2	$\boxed{\frac{1}{6}}$	$\frac{1}{3}$
	$+$	
2	$\frac{1}{3}$	$\frac{1}{6}$
$P(y)$	$P(Y=0)=\boxed{\frac{1}{2}}$	

$$\Rightarrow \quad P(X=-2|Y=0)=\frac{P(X=-2,Y=0)}{P(Y=0)}=\frac{\frac{1}{6}}{\frac{1}{2}}=\frac{1}{3}$$

ب)

X \ Y	0	10	$P(x)$
-2	$\frac{1}{6}$ +	$\boxed{\frac{1}{3}}$	$P(X=-2)=\boxed{\frac{1}{2}}$
2	$\frac{1}{3}$	$\frac{1}{6}$	

$$\Rightarrow \qquad P(Y=10\,|\,X=-2)=\frac{P(X=-2,Y=10)}{P(X=-2)}=\frac{\frac{1}{3}}{\frac{1}{2}}=\frac{2}{3}$$

مثال ۲ تابع چگالی توأم $f(x,y)=c(x+y)\ ;\ 0<x<1\,,0<y<2$ را در نظر بگیرید. مطلوب است محاسبه:

الف) ثابت c　　　ب) تابع کناره‌ای $f(x)$　　　ج) تابع کناره‌ای $f(y)$

د) تابع $f(x\,|\,y)$　　　ه) تابع $f(y\,|\,x)$

حل:

الف) بنا بر تعریف تابع چگالی توأم، رابطه $\iint f(x,y)\,dx\,dy=\iint f(x,y)\,dy\,dx=1$ برقرار است:

$$\int_0^2\int_0^1 c(x+y)\,dxdy=1 \;\rightarrow\; \int_0^2 c\left[\frac{x^2}{2}+xy\right]_0^1 dy=\int_0^2 c\left(\frac{1}{2}+y\right)dy=c\left[\frac{y}{2}+\frac{y^2}{2}\right]_0^2=c(1+2)=1 \;\rightarrow\; c=\frac{1}{3}$$

بنابراین تابع چگالی توأم X و Y برابر است با:

$$f(x,y)=\frac{1}{3}(x+y)\quad ;\quad 0<x<1\,,\quad 0<y<2$$

ب) برای محاسبه تابع کناره‌ای $f(x)$ باید تغییرات $f(x,y)$ را روی y در نظر بگیریم:

$$f(x)=\int_0^2\frac{1}{3}(x+y)\,dy=\frac{1}{3}\left[xy+\frac{y^2}{2}\right]_0^2=\frac{1}{3}(2x+2)$$

ج) برای محاسبه تابع کناره‌ای $f(y)$ باید تغییرات $f(x,y)$ را روی x در نظر بگیریم:

$$f(y)=\int_0^1\frac{1}{3}(x+y)\,dx=\frac{1}{3}\left[\frac{x^2}{2}+xy\right]_0^1=\frac{1}{3}\left(\frac{1}{2}+y\right)$$

د)

$$f(x\,|\,y)=\frac{f(x,y)}{f(y)}=\frac{\frac{1}{3}(x+y)}{\frac{1}{3}\left(\frac{1}{2}+y\right)}=\frac{x+y}{\frac{1}{2}+y}=\frac{2(x+y)}{2y+1}$$

ه)

$$f(y\,|\,x)=\frac{f(x,y)}{f(x)}=\frac{\frac{1}{3}(x+y)}{\frac{1}{3}(2x+2)}=\frac{(x+y)}{2(x+1)}$$

مثال ۳ X و Y متغیرهای تصادفی هستند و $1<x<3$ و $0<y<2$ است. درصورتی‌که $f(x,y)=\frac{1}{20}(x-y+4)$ و $f(x)=\frac{1}{20}(2x+6)$ باشد، تابع شرطی $f(y|x)$ کدام‌یک از گزینه‌های زیر است؟ (اقتصاد ـ ۷۵)

۱) $f(y|x)=\frac{2x+6}{x-y+4}$

۲) $f(y|x)=\frac{1}{400}(x-y+4)(2x+6)$

۳) $f(y|x)=\frac{x-y+4}{2x+6}$

۴) $f(x|y)=\frac{1}{400}(x-y+4)(2x+6)-2x^2+2xy$

حل: گزینه ۳ درست است.

$$f(y|x)=\frac{f(x,y)}{f(x)}=\frac{\frac{1}{20}(X-y+4)}{\frac{1}{20}(2x+6)}=\frac{x-y+4}{2x+6}$$

امید ریاضی شرطی

در شرایطی که:

۱- مقدار متوسط (ارزش انتظاری) متغیر تصادفی X به شرط وجود یک مقدار ثابت برای y مورد نظر باشد، آن‌گاه:

امید X به شرط y $=E(X|y)$

۲- مقدار متوسط (ارزش انتظاری) متغیر تصادفی Y به شرط وجود یک مقدار ثابت برای x مورد نظر باشد، آن‌گاه:

امید Y به شرط x $=E(Y|x)$

محاسبه امید شرطی درصورتی‌که X و Y دو متغیر تصادفی **گسسته** باشند، به صورت زیر است:

$$E(X|y)=\sum_{\forall x} x\cdot f(x|y)\quad,\quad E(Y|x)=\sum_{\forall y} y\cdot f(y|x)$$

و درصورتی‌که X و Y دو متغیر تصادفی **پیوسته** باشند، به صورت زیر است:

$$E(X|y)=\int x\cdot f(x|y)dx\quad,\quad E(Y|x)=\int y\cdot f(y|x)dy$$

مثال ۱ توزیع احتمال مشترک دو متغیر X و Y به صورت جدول زیر است، $E(Y|x=1)$ برابر است با: (اقتصاد ـ ۷۳)

x \ y	0	1	2
0	0.20	0.15	0.05
1	0.05	0.20	0.05
2	0.05	0.05	0.20

۱) 2

۲) 1

۳) 0.5

۴) 1.5

حل: گزینه ۲ درست است.

x \ y	0	1	2	$f(x)$
0	0.20	0.15	0.05	0.4
1	0.05	0.20	0.05	$f(x=1)=0.3$
2	0.05	0.05	0.20	0.3
$f(y)$	0.3	0.4	0.3	1

$$E(Y|x=1)=\frac{\sum yf\,X(x=1,y)}{f(X=1)}=\frac{0\times f(X=1,Y=0)+1\times f(X=1,Y=1)+2\times f(X=1,Y=2)}{f(X=1)}$$

$$=\frac{0\times 0.05+1\times 0.20+2\times 0.05}{0.3}=\frac{0.3}{0.3}=1$$

مثال ۲ جدول توزیع احتمال مشترک دو متغیر تصادفی X و Y به صورت زیر به دست آمده است:

y \ x	0	1	2
0	0.48	0.16	0.16
1	0.09	0.03	0.03
2	0.03	0.01	0.01

$E(X\,|\,y=0)$ برابر است با: (اقتصاد ـ ۷۳)

۱) 0.25 ۲) 0.15 ۳) 0.8 ۴) 0.6

حل: گزینه ۴ درست است.

y \ x	0	1	2	$f(y)$
0	0.48	0.16	0.16	$\boxed{0.8}$
1	0.09	0.03	0.03	0.15
2	0.03	0.01	0.01	0.05
$f(x)$	0.6	0.2	0.2	1

$$E(X\,|\,y=0)=\frac{\sum x\,f(X,Y=0)}{f(Y=0)}=\frac{0\times f(X=0,Y=0)+1\times f(X=1,Y=0)+2\times f(X=2,Y=0)}{0.8}$$

$$=\frac{0\times 0.48+1\times 0.16+2\times 0.16}{0.8}=0.6$$

مثال ۳ تابع چگالی توأم $f(x,y)=\frac{1}{3}(x+y)\ ;\ \ 0<x<1\ ,\ \ 0<y<2$ را در نظر بگیرید. مطلوب است محاسبه:

الف) $E(X\,|\,y=1)$ ب) $E(Y\,|\,x=0)$

حل:

الف)

$$\begin{cases} E(X\,|\,y=1)=\int_x x\cdot f(x\,|\,y)dx=\int_0^1 x\cdot\frac{2(x+1)}{2\times 1+1}dx=\int_0^1\frac{2}{3}\left(x^2+x\right)dx=\frac{2}{3}\left[\frac{1}{3}x^3+\frac{1}{2}x^2\right]_0^1=\frac{2}{3}\left(\frac{1}{3}+\frac{1}{2}\right)=\frac{5}{9} \\ f(x\,|\,y)=\frac{2(x+y)}{2y+1} \end{cases}$$

ب)

$$\begin{cases} E(Y|x=0)=\int_y y\cdot f(y|x)dy=\int_0^2 y\cdot\frac{0+y}{2(0+1)}dy=\int_0^2\frac{1}{2}y^2dy=\left[\frac{1}{6}y^3\right]_0^2=\frac{8}{6}=\frac{4}{3} \\ f(y|x)=\frac{x+y}{2(x+1)} \end{cases}$$

اگر X و Y متغیرهای تصادفی مستقل باشند، آن‌گاه:

X و Y مستقل	$\rightleftarrows$	$E(X\|y)=E(X)$ $E(Y\|x)=E(Y)$

امید ریاضی حاصل‌ضرب دو متغیر

میانگین (امید) حاصل‌ضرب دو متغیر تصادفی **گسسته** X و Y به صورت زیر محاسبه می‌شود:

$$\boxed{E(XY)=\sum\sum x_i\times y_j\times f(x_i,y_j)}$$

اگر X و Y متغیرهای تصادفی **پیوسته** باشند:

$$\boxed{E(XY)=\iint xyf(x,y)dx\,dy=\iint xy\,f(x,y)dy\,dx}$$

مثال ۱ تابع احتمال توأم زیر را در نظر بگیرید، مقدار $E(XY)$ کدام است؟

x \ y	2	3	5
-1	0.1	0.2	0.3
0	0.2	0.1	0.1

۱) -2.3

۲) 0.4

۳) 0.8

۴) -3.1

حل: گزینه ۱ درست است.

$$E(XY)=\sum\sum x_iy_jf(x_i\cdot y_j)=-1\times2\times0.1+(-1)\times3\times(0.2)+(-1)\times5\times0.3+0\times2\times0.2+0\times3\times0.1+0\times5\times0.1=-2.3$$

مثال ۲ تابع چگالی توأم زیر را در نظر بگیرید. $E(XY)$ کدام است؟

$$\begin{cases} f(x,y)=\frac{1}{3}(x+y) \\ 0<x<1\ ,\ 0<y<2 \end{cases}$$

۱) $\frac{1}{4}$

۲) $\frac{5}{6}$

۳) $\frac{2}{3}$

۴) 2

حل: گزینه ۳ درست است.

$$E(XY)=\int_0^2\int_0^1 x\cdot y\cdot\frac{1}{3}(x+y)dx\,dy=\int_0^2\int_0^1\frac{1}{3}(x^2y+xy^2)dx\,dy$$

$$=\frac{1}{3}\int_0^2\left[\frac{1}{3}x^3y+\frac{1}{2}x^2y^2\right]_0^1dy=\frac{1}{3}\int_0^2\left(\frac{1}{3}y+\frac{1}{2}y^2\right)dy=\frac{1}{3}\left[\frac{1}{6}y^2+\frac{1}{6}y^3\right]_0^2=\frac{1}{3}\left[\frac{4}{6}+\frac{8}{6}\right]=\frac{2}{3}$$

$E(XY)$ و استقلال متغیرهای تصادفی

$$E(XY)=E(X)\cdot E(Y) \quad \xrightleftharpoons[\not\leftarrow]{} \quad \text{X و Y مستقل (ناوابسته)}$$

این رابطه نشان می‌دهد، درصورتی‌که دو متغیر تصادفی X و Y مستقل (ناوابسته) باشند، آن‌گاه $E(XY)=E(X)\cdot E(Y)$ خواهد بود. **اما عکس این مطلب درست نیست**، یعنی ممکن است رابطه $E(XY)=E(X)\cdot E(Y)$ برقرار باشد، اما X و Y مستقل نباشند.

مثال اگر $E(X)=5$ و $E(Y)=2$ و X و Y مستقل باشند، آن‌گاه کدام عبارت درست است؟

۱) $E(X+Y)=7$ ۲) $E(X-Y)=3$ ۳) $E(XY)=10$ ۴) همه موارد

حل: گزینه ۴ درست است.

با توجه به خطی بودن میانگین (امید) همواره روابط زیر برقرار است (X و Y مستقل یا وابسته):

$$E(X+Y)=E(X)+E(Y)=5+2=7$$
$$E(X-Y)=E(X)-E(Y)=5-2=3$$

اما برای محاسبه $E(XY)$ با توجه به مستقل بودن X و Y داریم:

$$\text{X,Y مستقل} \rightarrow E(XY)=E(X)\cdot E(Y)=5\times 2=10$$

دقت کنید اگر در مثال بالا، شرط استقلال دو متغیر X و Y مطرح نمی‌شد، آن‌گاه لزوماً رابطه $E(XY)=E(X)\cdot E(Y)$ برقرار نبود و نمی‌توانستیم گزینه ۳ را بپذیریم. این در شرایطی است که روابط $E(X\pm Y)=E(X)\pm E(Y)$ همواره برقرار است و ارتباطی به مستقل بودن یا نبودن X و Y ندارد.

متغیرهای عمود

دو متغیر تصادفی X و Y بر هم عمودند اگر و فقط اگر امید ریاضی حاصل‌ضرب آن‌ها صفر باشد.

$$E(XY)=0 \quad \longleftrightarrow \quad \text{X و Y برهم عمودند.}$$

مثال اگر X و Y بر هم عمود باشند و $E(X)=2E(Y)=1$ باشد، حاصل عبارت $E[(X-E(X))Y]$ کدام است؟

۱) -0.5 ۲) 0 ۳) 0.5 ۴) -2

حل: گزینه ۱ درست است.

$$E[(X-E(X))Y]=E(XY-E(X)Y)=E(XY)-E(X)E(Y)=0-1\times\frac{1}{2}=-\frac{1}{2}$$

$$\text{X و Y بر هم عمودند} \rightarrow E(XY)=0\ ,\ E(X)=1\ ,E(Y)=\frac{1}{2}$$

امید ریاضی تقسیم دو متغیر

میانگین (امید) حاصل‌تقسیم دو متغیر تصادفی **گسسته** X و Y به صورت زیر محاسبه می‌شود:

$$E\left(\frac{X}{Y}\right)=\sum\sum\frac{x_i}{y_j}\times f(x_i,y_j)$$

اگر X و Y متغیرهای تصادفی **پیوسته** باشند:

$$\boxed{E\left(\frac{X}{Y}\right)=\iint \frac{x}{y} f(x,y)\,dx\,dy=\iint \frac{x}{y} f(x,y)\,dy\,dx}$$

برای محاسبه $E\left(\frac{Y}{X}\right)$ کافی است در روابط بالا جای X و Y را عوض کنیم.

بنابراین:

$$E\left(\frac{X}{Y}\right)\neq E\left(\frac{Y}{X}\right)$$

$E\left(\frac{X}{Y}\right)$ و استقلال متغیرهای تصادفی

همان‌طورکه می‌دانیم:

۱- اگر دو متغیر تصادفی X و Y مستقل باشند، متغیرهای X , $\frac{1}{Y}$ نیز مستقل هستند.

۲- حاصل تقسیم $\frac{X}{Y}$ را می‌توان به صورت ضرب $X\times\frac{1}{Y}$ نوشت.

۳- $E(XY)=E(X)E(Y) \;\underset{\not\Longleftarrow}{\Longrightarrow}\;$ X, Y مستقل

با توجه به موارد بالا داریم:

$$\boxed{E\left(\frac{X}{Y}\right)=E(X)E\left(\frac{1}{Y}\right) \;\underset{\not\Longleftarrow}{\Longrightarrow}\; \text{X و Y مستقل (ناوابسته)}}$$

✓ دقت کنید!

«امید تقسیم دو متغیر تصادفی» هیچ‌گاه نمی‌تواند با «تقسیم امید متغیرها» برابر باشد.

$$E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}$$

اثبات:

همان‌طورکه می‌دانیم همواره:

$$\begin{cases} E\left(\frac{1}{Y}\right)=\sum\sum \frac{1}{y} f(x,y) \\ \frac{1}{E(Y)}=\frac{1}{\sum\sum y f(x,y)} \end{cases} \longrightarrow E\left(\frac{1}{Y}\right)\neq\frac{1}{E(Y)}$$

درنتیجه داریم:

$$\begin{cases} E\left(\frac{X}{Y}\right)\overset{\text{X و Y مستقل}}{\overset{\uparrow}{=}}E(X)E\left(\frac{1}{Y}\right) \\ \frac{E(X)}{E(Y)}=E(X)\cdot\frac{1}{E(Y)} \end{cases} \xrightarrow{E\left(\frac{1}{Y}\right)\neq\frac{1}{E(Y)}} E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}$$

مثال در مورد دو متغیر تصادفی X و Y و صحت رابطه امید ریاضی $E\left(\frac{X}{Y}\right)=\frac{E(X)}{E(Y)}$ می‌توان گفت: (اقتصاد ـ ۸۱)

۱) وقتی صادق است که X و Y مستقل باشند.
۲) وقتی صادق است که X و Y کوواریانس صفر داشته باشند.
۳) وقتی صادق است که ناهمبسته باشند.
۴) هیچ وقت صادق نیست.

حل: گزینه ۴ درست است.

نتیجه:

۱- امید مجموع یا تفاضل دو متغیر تصادفی همواره برابر با مجموع یا تفاضل امید آن‌هاست.

$E(X\pm Y)=E(X)\pm E(Y)$ (X و Y مستقل یا وابسته)

۲- امید ضرب دو متغیر تصادفی فقط زمانی که X و Y مستقل باشند، برابر حاصل‌ضرب امید آن‌هاست.

$$\text{مستقل } X,Y \;\; \overset{\Longrightarrow}{\not\Longleftarrow} \;\; E(XY)=E(X)E(Y)$$

۳- امید تقسیم دو متغیر تصادفی برابر با تقسیم امید آن‌ها نیست.

$$E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}$$

البته زمانی که X و Y مستقل باشند:

$$E\left(\frac{X}{Y}\right)=E(X)\times E\left(\frac{1}{Y}\right)$$

واریانس حاصل‌ضرب دو متغیر

برای محاسبه واریانس حاصل‌ضرب دو متغیر تصادفی با توجه به تعریف واریانس داریم:

$$\mathrm{Var}(XY)=E\left(X^2Y^2\right)-E(XY)^2=E\left(X^2\right)E\left(Y^2\right)-E(X)^2E(Y)^2 \quad (\textbf{مستقل } Y \textbf{ و } X)$$

مثال فرض کنید X و Y دو متغیر تصادفی مستقل از هم باشند که در آن:

$E(X)=1\,,E(Y)=2\,,\mathrm{Var}(X)=2\,,\mathrm{Var}(Y)=3$

مقدار $\mathrm{Var}(XY)$ کدام است؟

۱) 10 ۲) 17 ۳) 21 ۴) 36

حل: گزینه ۲ درست است.

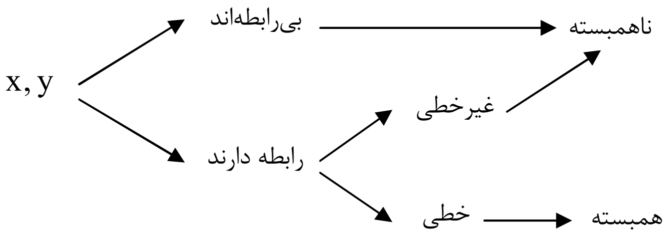
$$\mathrm{Var}(XY)=E\left(X^2Y^2\right)-E(XY)^2\overset{\text{مستقل } Y \text{ و } X}{=}E\left(X^2\right)E\left(Y^2\right)-E(X)^2E(Y)^2=3\times7-1^2\times2^2=21-4=17$$

$$\begin{cases}\mathrm{Var}(X)=E\left(X^2\right)-E(X)^2 \;\rightarrow\; 2=E\left(X^2\right)-1^2 \;\rightarrow\; E\left(X^2\right)=3\\ \mathrm{Var}(Y)=E\left(Y^2\right)-E(Y)^2 \;\rightarrow\; 3=E\left(Y^2\right)-2^2 \;\rightarrow\; E\left(Y^2\right)=7\end{cases}$$

همبستگی (Correlation)

می‌دانیم هر دو متغیر دلخواه X و Y به‌طور کلی یا وابسته‌اند (رابطه دارند) یا مستقل (بی‌رابطه‌اند). حال فرض کنید دو متغیر X و Y وابسته‌اند و این وابستگی، خطی است؛ در این صورت اصطلاحاً می‌گویند دو متغیر همبسته‌اند.

نمودار روابط



باتوجه به درستی روابط بالا، جدول زیر نتیجه می‌شود:

جدول همبستگی

وضعیت دو متغیر	رابطه دو متغیر
همبسته‌اند	رابطه خطی دارند.
ناهمبسته‌اند	یا مستقل‌اند (بی‌رابطه‌اند) یا رابطه غیرخطی دارند.

قضیه استثنا در ناهمبستگی:

اگر X و Y دارای تابع توزیع توأم نرمال و در عین حال ناهمبسته باشند، آن‌گاه X و Y حتماً مستقل‌اند.

عبارت‌های معادل برای همبسته و ناهمبسته

برای واژه‌های «همبسته» و «ناهمبسته» اصطلاحات دیگری نیز وجود دارد:

X و Y همبسته‌اند = وابستگی خطی دارند = استقلال خطی ندارند.

X و Y ناهمبسته‌اند = وابستگی خطی ندارند = استقلال خطی دارند.

در این بخش به تحلیل وضعیت همبستگی دو متغیر با استفاده از معیارهای «کوواریانس»، «ضریب همبستگی»، «ضریب تشخیص» و همچنین «معادله خط رگرسیون» می‌پردازیم.

کوواریانس

کوواریانس از نظر عددی برابر است با امید ریاضی حاصل‌ضرب انحرافات دو متغیر از میانگین آن‌ها که به صورت زیر محاسبه می‌شود:

$$\boxed{\operatorname{Cov}(X,Y)=E\big[(X-E(X))(Y-E(Y))\big]=E(XY)-E(X)E(Y)}$$

کوواریانس معیاری است که از طریق آن:

اولاً، می‌توان **نوع ارتباط** دو متغیر تصادفی X و Y را از نظر خطی بودن، مشخص کرد.

$$\begin{cases} \operatorname{Cov}(X,Y)\neq 0 & \text{(وابستگی خطی دارند = همبسته‌اند)} \\ \operatorname{Cov}(X,Y)=0 & \text{(وابستگی خطی ندارند = ناهمبسته‌اند)} \end{cases}$$

ثانیاً، در صورت خطی بودن ارتباط دو متغیر X و Y می‌توان **جهت ارتباط خطی** (همبستگی خطی) را تعیین کرد.

$$\begin{cases} Cov(X,Y) > 0 & \text{ارتباط خطی مستقیم} \\ Cov(X,Y) < 0 & \text{ارتباط خطی معکوس} \end{cases}$$

به عبارت دیگر، کوواریانس معیاری است که از طریق آن می‌توان نوع ارتباط خطی (همبسته یا ناهمبسته) و نیز جهت ارتباط خطی (مستقیم یا معکوس) بین دو متغیر را تعیین کرد.

تحلیل کوواریانس

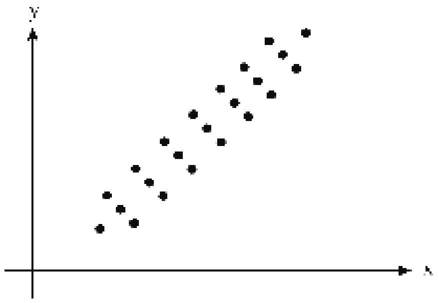
پس از محاسبه کوواریانس جامعه یا نمونه، با توجه به آنکه مقدار آن صفر یا مخالف صفر شده باشد می‌توان آن‌ها را به صورت زیر بررسی کرد:

$$\mathbf{Cov(X,Y) \neq 0}$$

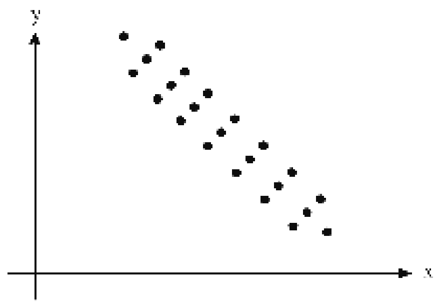
اگر کوواریانس دو متغیر تصادفی X و Y مخالف صفر باشد، آن دو متغیر **همبسته**‌اند و با یکدیگر وابستگی خطی دارند. در این شرایط رابطه بین X و Y را می‌توان با خطی به نام **خط رگرسیون** $(y = bx + a)$ مشخص کرد که:

الف) اگر $Cov(X,Y) > 0$: شیب خط مثبت است و ارتباط خطی مستقیم بین دو متغیر وجود دارد.

ب) اگر $Cov(X,Y) < 0$: شیب خط منفی است و ارتباط خطی معکوس بین دو متغیر وجود دارد.



$y = bx + a \quad \rightarrow \quad b > 0$



$y = bx + a \quad \rightarrow \quad b < 0$

مثال ۱ در صورتی‌که $Cov(X,Y) \neq 0$ باشد، آن‌گاه کدام عبارت درست است؟

۱) X و Y ناهمبسته‌اند.

۲) X و Y ارتباط خطی (همبستگی خطی) دارند.

۳) X و Y ارتباط غیرخطی دارند یا مستقل‌اند.

۴) X و Y لزوماً مستقل‌اند.

حل: گزینه ۲ درست است.

$$\mathbf{Cov(X,Y) = 0}$$

اگر کوواریانس دو متغیر تصادفی X و Y برابر با صفر شود، آن دو متغیر **ناهمبسته‌اند** و برعکس. در این وضعیت تمام مواردی که در «**نمودار روابط**» و «**جدول همبستگی**» برای ناهمبستگی دو متغیر بیان شد، برای $Cov(X,Y) = 0$ نیز برقرار است.

مثال ۲ اگر کوواریانس بین X و Y مساوی با صفر باشد، کدام یک از عبارات زیر صحیح است؟ (مدیریت ـ ۷۲)

۱) آن دو متغیر مستقل‌اند.

۲) دارای رابطه غیر خطی می‌باشند.

۳) موارد ۱ و ۲ هر دو صحیح است.

۴) یا دارای رابطه غیر خطی هستند یا هیچ رابطه‌ای با همدیگر ندارند.

حل: گزینه ۴ درست است.
وقتی کوواریانس دو متغیر برابر صفر می‌شود یعنی دو متغیر ناهمبسته‌اند درنتیجه با توجه به «جدول همبستگی»، دو متغیر یا مستقل‌اند یا رابطه غیرخطی دارند.

مثال ۳ اگر $Cov(X,Y)=0$ باشد، کدام بیان برای رابطه X و Y صحیح است؟ (حسابداری و مدیریت ـ ۸۵)

۱) رابطه خطی
۲) رابطه غیرخطی
۳) رابطه غیرخطی یا مستقل
۴) الزاماً مستقل

حل: گزینه ۳ درست است.
وقتی کوواریانس دو متغیر برابر صفر می‌شود یعنی دو متغیر ناهمبسته‌اند و طبق «جدول همبستگی»، دو متغیر یا مستقل‌اند یا رابطه غیرخطی دارند.

مثال ۴ فرض کنیم تابع چگالی احتمال‌های مشترک دو متغیر X و Y نرمال باشد، در این صورت اگر کوواریانس X و Y صفر باشد؟ (اقتصاد ـ ۸۰)

۱) ضریب همبستگی X و Y مثبت است.
۲) X و Y می‌توانند مستقل نباشند.
۳) X و Y مستقل از هم هستند.
۴) ضریب همبستگی X و Y منفی است.

حل: گزینه ۳ درست است.
با توجه به آنکه تابع چگالی مشترک دو متغیر نرمال و کوواریانس آن‌ها صفر است، طبق قضیه استثنا در ناهمبستگی، دو متغیر حتماً مستقل‌اند.
دقت کنید، اگر کلمه «نرمال» ذکر نمی‌شد، گزینه ۲ درست بود زیرا دو متغیر می‌توانستند رابطه غیرخطی داشته باشند.

مثال ۵ دو متغیر مستقل: (اقتصاد ـ ۷۵)

۱) استقلال خطی ندارد.
۲) کوواریانس آن‌ها مخالف صفر است.
۳) مانع‌الجمع‌اند.
۴) ناهمبسته‌اند.

حل: گزینه ۴ درست است.
گزینه ۱ نادرست است، زیرا وقتی دو متغیر مستقل‌اند بنا بر «نمودار روابط» ناهمبسته‌اند؛ در این حالت از عبارات معادل «وابستگی خطی ندارند» یا «استقلال خطی دارند» استفاده می‌شود.
گزینه ۲ نادرست است، زیرا وقتی دو متغیر مستقل‌اند بنابر «نمودار روابط» ناهمبسته‌اند، در نتیجه کوواریانس آن‌ها برابر صفر است.
گزینه ۳ نادرست است، زیرا «مانع‌الجمع» یعنی «ناسازگار» و دو متغیر ناسازگار همواره وابسته‌اند.
گزینه ۴ درست است، زیرا بنا بر «نمودار روابط»، دو متغیر مستقل همواره ناهمبسته‌اند.

مثال ۶ در رابطه با دو متغیر تصادفی X و Y کدام‌یک از گزینه‌های زیر غلط است؟ (اقتصاد ـ ۷۸)

۱) اگر $Cov(X,Y)=0$ باشد، دو متغیر تصادفی X و Y از لحاظ آماری مستقل از هم می‌باشند.
۲) اگر $P(X\cap Y)=P(X).P(Y)$ باشد، X و Y از لحاظ آماری مستقل از هم می‌باشند.
۳) اگر $Cov(X,Y)=0$ باشد، وابستگی خطی بین X و Y وجود ندارد.
۴) $Cov(X,Y)$ برابر $E(XY)-E(X)E(Y)$ است.

حل: گزینه ۱ درست است.
گزینه ۱ نادرست است، زیرا وقتی کوواریانس صفر است، دو متغیر ناهمبسته‌اند و بنا بر «جدول همبستگی» دو متغیر رابطه غیر خطی دارند (بنابراین لزوماً مستقل نیستند).

گزینه ۲ درست است، زیرا:

$$\text{مستقل } X,Y \rightleftarrows f(x,y)=f(x)\times f(y)$$

گزینه ۳ درست است، زیرا وقتی کوواریانس برابر با صفر است یعنی دو متغیر ناهمبسته‌اند. در این حالت از عبارات معادل «وابستگی خطی ندارند» یا «استقلال خطی دارند» استفاده می‌شود.

گزینه ۴ درست است، زیرا همان رابطه محاسبه کوواریانس است.

مثال ۷ اگر $E(X)=2$, $E(Y)=5$ و X و Y مستقل باشند، کدام مورد درست است؟

۱) $\mathrm{Cov}(X,Y)=0$ ۲) $E(XY)=10$ ۳) $E(X+Y)=7$ ۴) همه موارد

حل: گزینه ۴ درست است.

گزینه ۱ درست است، زیرا دو متغیر مستقل طبق «نمودار روابط»، حتماً ناهمبسته‌اند و درنتیجه کوواریانس آن‌ها صفر است.

گزینه ۲ درست است، زیرا:

$$\text{مستقل } X,Y \rightarrow E(XY)=E(X)E(Y)=10$$

گزینه ۳ درست است، زیرا رابطه $E(X\pm Y)=E(X)\pm E(Y)$ صرف‌نظر از مستقل یا وابسته بودن دو متغیر، همواره برقرار است، بنابراین:

$$E(X+Y)=E(X)+E(Y)=7$$

مثال ۸ هرگاه $\mathrm{Cov}(X,Y)=0$ باشد، آن‌گاه دو متغیر X و Y:

۱) دو متغیر مستقل‌اند. ۲) X و Y ناهمبسته‌اند (استقلال خطی دارند).

۳) $E(XY)=E(X)E(Y)$ ۴) موارد ۲ و ۳

حل: گزینه ۴ درست است.

اولاً، زمانی که $\mathrm{Cov}(X,Y)=0$ است، دو متغیر ناهمبسته‌اند، بنابراین یا مستقل‌اند یا رابطه غیرخطی دارند، پس لزوماً مستقل نیستند مگر آنکه بنا بر قضیه استثنا در ناهمبستگی، جامعه نرمال باشد.

$$\mathrm{Cov}(X,Y)=0 \rightarrow E(XY)-E(X)E(Y)=0 \rightarrow E(XY)=E(X)E(Y)$$

محاسبه کوواریانس

برای محاسبه کوواریانس دو متغیر تصادفی X و Y با توجه به روابط مطرح‌شده، بهتر است به صورت زیر عمل کنیم:

ابتدا استقلال دو متغیر تصادفی X و Y را بررسی می‌کنیم:

ـ اگر دو متغیر مستقل باشند، آن‌گاه ناهمبسته‌اند و درنتیجه کوواریانس صفر می‌شود.

$$\text{مستقل } X,Y \xrightarrow{f(x,y)=f(x)\times f(y)} \mathrm{Cov}(X,Y)=0$$

ـ اگر دو متغیر وابسته باشند، باید کوواریانس را محاسبه کنیم زیرا ممکن است مقدار آن، صفر (وابستگی غیرخطی) یا مخالف صفر (وابستگی خطی) شود.

$$\text{وابسته } X,Y \xrightarrow{f(x,y)\neq f(x)\times f(y)} \mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$$

مثال ۱ تابع احتمال توأم زیر را در نظر بگیرید. کوواریانس دو متغیر تصادفی X و Y کدام است؟

x \ y	2	3
−1	0.2	0.3
0	0.2	0.3

۱) 0 ۲) $-\frac{1}{2}$ ۳) $\frac{1}{4}$ ۴) 1

حل: گزینه ۱ درست است.

ابتدا مستقل بودن دو متغیر تصادفی X و Y را بررسی می‌کنیم، X و Y در جدول بالا مستقل هستند، درنتیجه:

X,Y مستقل (ناوابسته) $\rightarrow \quad \mathrm{Cov}(X,Y)=0$ (ناهمبسته)

یادآوری: برای بررسی استقلال متغیرهای تصادفی X و Y ابتدا توابع حاشیه‌ای $f(x)$ و $f(y)$ را محاسبه کرده، سپس برقرار بودن یا نبودن رابطه $f(x,y)=f(x)\cdot f(y)$ را برای تمام زوج‌های (x,y) بررسی می‌کنیم.

x \ y	2	3	f(x)
−1	0.2	0.3	0.5
0	0.2	0.3	0.5
f(y)	0.4	0.6	1

$\xrightarrow{f(x,y)=f(x)\cdot f(y)}$

x \ y	2	3	f(x)
−1	$0.2=0.5\times0.4$	$0.3=0.5\times0.6$	0.5
0	$0.2=0.5\times0.4$	$0.3=0.5\times0.6$	0.5
f(y)	0.4	0.6	1

مثال ۲ تابع احتمال توأم زیر را در نظر بگیرید، کوواریانس دو متغیر تصادفی X و Y کدام است؟

x \ y	−1	0	1
−1	0.1	0.6	0.1
1	0.1	0	0.1

۱) 0

۲) $-\frac{1}{2}$

۳) $\frac{1}{4}$

۴) $\frac{1}{8}$

حل: گزینه ۱ درست است.

ابتدا مستقل بودن دو متغیر تصادفی X و Y را بررسی می‌کنیم. با توجه به صفر بودن یکی از مقادیر $f(x,y)$، دومتغیر وابسته هستند و باید $\mathrm{Cov}(X,Y)$ را با استفاده از رابطه زیر محاسبه کنیم:

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)\cdot E(Y)$$

x \ y	−1	0	1	f(x)
−1	0.1	0.6	0.1	0.8
1	0.1	0	0.1	0.2
f(y)	0.2	0.6	0.2	

$$\begin{cases}\mathrm{Cov}(X,Y)=E(XY)-E(X)\cdot E(Y)=0-(-0.6)\times(0)=0\\ E(XY)=\sum\sum x_iy_j\,f(x_iy_j)=(-1)\times(-1)\times0.1+(-1)\times0\times0.6+(-1)\times1\times0.1+1\times(-1)\times0.1+1\times0\times0+1\times1\times0.1=0\\ E(X)=\sum_x x\,f(x)=-1\times0.8+1\times0.2=-0.6\\ E(Y)=\sum_y y\,f(y)=-1\times0.2+0\times0.6+1\times0.2=0\end{cases}$$

دقت کنید، همان‌طورکه در این مثال دیدیم، علی‌رغم اینکه X و Y مستقل نبوده‌اند (وابسته‌اند)، با این‌حال $\mathrm{Cov}(X,Y)=0$ است. بنابراین X و Y وابسته‌اند اما ارتباطشان غیرخطی است.

مثال ۳ تابع احتمال توأم زیر را در نظر بگیرید، کوواریانس دو متغیر تصادفی X و Y کدام است؟

x \ y	2	3
−1	0.1	0.2
0	0.3	0.4

۱) 0

۲) -0.02

۳) 0.25

۴) -0.4

حل: گزینه ۲ درست است.

ابتدا مستقل بودن دو متغیر تصادفی X و Y را بررسی می‌کنیم:

x \ y	2	3	f(x)
−1	0.1	0.2	0.3
0	0.3	0.4	0.7
f(y)	0.4	0.6	

$$\xrightarrow{f(x,y)=f(x)\cdot f(y)} \quad \underbrace{f(x=-1,y=2)}_{0.1} \underset{\neq}{\neq} \underbrace{f(x=-1)}_{0.3} \underset{\times}{\times} \underbrace{f(y=2)}_{0.4}$$

با توجه به آنکه X و Y مستقل نیستند (وابسته‌اند)، از رابطه زیر برای محاسبه کوواریانس استفاده می‌کنیم:

$$\begin{cases} Cov(X,Y)=E(XY)-E(X)E(Y)=-0.8-(-0.3)(2.6)=-0.02 \\ E(XY)=\sum\sum xyf(x,y)=(-1)\times 2\times 0.1+(-1)\times 3\times 0.2+0\times 2\times 0.3+0\times 3\times 0.4=-0.8 \\ E(X)=\sum x\cdot f(x)=(-1)\times 0.3+0\times 0.7=-0.3 \\ E(Y)=\sum y\cdot f(y)=2\times 0.4+3\times 0.6=2.6 \end{cases}$$

مثال ۴ فرض کنید متغیر تصادفی X در فاصله $(-1\,,1)$ دارای تابع چگالی $f(x)=\frac{1}{2}$ باشد. اگر $Y=X^2$ باشد، کوواریانس X و Y برابر است با: (اقتصاد ـ ۸۶)

۱) صفر ۲) $\frac{1}{12}$ ۳) $\frac{1}{4}$ ۴) 1

حل: گزینه ۱ درست است.

ازآنجاکه تابع احتمال توأم $f(x,y)$ وجود ندارد، نمی‌توانیم استقلال دو متغیر را بررسی کنیم؛ بنابراین به ناچار $Cov(X,Y)$ را محاسبه می‌کنیم:

$$\begin{cases} Cov(X,Y)=Cov\left(X,X^2\right)=E\left(XX^2\right)-E(X)E\left(X^2\right)=E\left(X^3\right)-E(X)E\left(X^2\right)=0-0\times E\left(X^2\right)=0 \\ E(X)=\int_{-1}^{1}\frac{1}{2}x\,dx=\left[\frac{1}{4}x^2\right]_{-1}^{1}=\frac{1}{4}\left(1-(-1)^2\right)=0 \\ E\left(X^3\right)=\int_{-1}^{1}\frac{1}{2}x^3dx=\left[\frac{1}{8}x^4\right]_{-1}^{1}=\frac{1}{8}\left(1-(-1)^4\right)=0 \end{cases}$$

محاسبه کوواریانس با استفاده از مشاهدات نمونه

با توجه به رابطه کوواریانس جامعه به صورت زیر:

$$Cov(X,Y)=E\left[(X-E(X))(Y-E(Y))\right]=\frac{1}{N}\left[\sum(X-E(X))(Y-E(Y))\right]=E(XY)-E(X)E(Y)$$
$$=\frac{\sum xy}{N}-\frac{\sum x}{N}\frac{\sum y}{N}=\frac{1}{N}\left[\sum xy-\frac{\sum x\sum y}{N}\right]$$

کوواریانس نمونه برابر است با:

$$\boxed{Cov(X,Y)=\frac{1}{n-1}\left[\sum(x-\bar{x})(y-\bar{y})\right]=\frac{1}{n-1}\left[\sum xy-\frac{\sum x\sum y}{n}\right]}$$

مثال ۵ در جدول داده‌های زیر کوواریانس X و Y کدام است؟

x	3	4	5
y	3	5	10

۱) صفر ۲) $\frac{7}{2}$

۳) $\frac{1}{3}$ ۴) 1

حل: گزینه ۲ درست است.

$$\begin{cases} Cov(X,Y)=\frac{1}{n-1}\left[\sum xy-\frac{\sum x\sum y}{n}\right]=\frac{1}{2}\left(79-\frac{12\times 18}{3}\right)=\frac{7}{2} \\ \sum xy=3\times 3+4\times 5+5\times 10=79 \\ \sum x=3+4+5=12 \\ \sum y=3+5+10=18 \end{cases}$$

مثال ۶ عبارت $\sum(x_i-\bar{x})(y_i-\bar{y})$ برابر است با: (اقتصاد ـ ۷۴)

۱) $\sum(x_i-\bar{x})\bar{y}$ ۲) $\sum(x_i-\bar{x})y_i$ ۳) $\sum(y_i-\bar{y})\bar{x}$ ۴) $\bar{x}\,\bar{y}$

حل: گزینه ۲ درست است.

$$\sum(x_i-\bar{x})(y_i-\bar{y})=\sum(x_i-\bar{x})y_i-\sum(x_i-\bar{x})\bar{y}=\sum(x_i-\bar{x})y_i-\bar{y}\sum(x_i-\bar{x})\ \underline{\underline{\sum(x_i-\bar{x})=0}}\ \sum(x_i-\bar{x})y_i$$

خواص کوواریانس

با توجه به رابطه $Cov(X,Y)=E(XY)-E(X)\cdot E(Y)$ و نیز مفهوم کوواریانس، خواص زیر برقرار است (X، Y، Z و T متغیرهای تصادفی و a، b، c و d ضرایب ثابت مثبت یا منفی هستند):

1) $Cov(X,Y)=Cov(Y,X)$
2) $Cov(X,X)=Var(X)=\sigma_X^2$
3) $Cov(X,a)=Cov(a,X)=Cov(a,b)=0$
4) $Cov(aX,bY)=ab\,Cov(X,Y)$
5) $Cov(X\pm c,Y\pm d)=Cov(X,Y)$
6) $Cov(aX\pm c,bY\pm d)=ab\,Cov(X,Y)$
7) $Cov(aX+cY,bZ+dT)=ab\,Cov(X,Z)+ad\,Cov(X,T)+cb\,Cov(Y,Z)+cd\,Cov(Y,T)$

توضیحات:

۱) کوواریانس بین X و Y همان کوواریانس بین Y و X است.

۲) کوواریانس یک متغیر با خودش برابر با واریانسش است.

۳) هیچ رابطه خطی بین دو ثابت (b,a) یا یک متغیر و یک ثابت (a,X) وجود ندارد، درنتیجه کوواریانس برابر صفر است.

۴) ضرایب متغیرها در کوواریانس X و Y با حفظ علامت خود، در کوواریانس دو متغیر X و Y ضرب می‌شود.

۵) افزودن یک مقدار ثابت به متغیرها یا کاستن یک مقدار ثابت از آن‌ها تأثیری در ارتباط خطی بین متغیرها ندارد.

۶) از ترکیب دو رابطه ۴ و ۵، رابطه ۶ نتیجه می‌شود.

۷) برای محاسبه کوواریانس دو عبارت که هر یک از چند متغیر استفاده می‌کنند، کافی است کوواریانس هر جمله از عبارت اول را با تک تک جملات عبارت دوم در نظر بگیریم و حاصل را با هم جمع کنیم.

برای مثال، کوواریانس هر یک از حالات زیر، با استفاده از خواص مطرح‌شده به دست می‌آید:

الف) $\text{Cov}(X,2)=\text{Cov}(2,X)=0$ — بنا بر خاصیت (۳)

ب) $\text{Cov}(-1,3)=0$ — بنا بر خاصیت (۳)

ج) $\text{Cov}(2X,-3Y)=(2)(-3)\text{Cov}(X,Y)=-6\text{Cov}(X,Y)$ — بنا بر خاصیت (۴)

د) $\text{Cov}(X+4,Y-10)=\text{Cov}(X,Y)$ — بنا بر خاصیت (۵)

ه) $\text{Cov}(-2X+3,5X+5)=(-2)(5)\text{Cov}(X,X)=-10\sigma_X^2$ — بنا بر خاصیت‌های (۲ و ۴)

و) $\text{Cov}(3X-Y,Z)=\text{Cov}(3X,Z)+\text{Cov}(-Y,Z)=3\text{Cov}(X,Z)-\text{Cov}(Y,Z)$ — بنا بر خاصیت (۷)

ز) $\text{Cov}(5X+2,Y+Z)=\text{Cov}(5X,Y)+\text{Cov}(5X,Z)=5\text{Cov}(X,Y)+5\text{Cov}(X,Z)$ — بنا بر خاصیت (۷)

ژ) $\text{Cov}(X+Y,X-Y)=\text{Cov}(X,X)+\text{Cov}(X,-Y)+\text{Cov}(Y,X)+\text{Cov}(Y,-Y)$ — بنا بر خاصیت (۷)

$$=\sigma_X^2-\text{Cov}(X,Y)+\text{Cov}(Y,X)-\sigma_Y^2=\sigma_X^2-\sigma_Y^2$$

مثال ۷ در صورتی‌که $Z=2X+4$ و $U=-5Y+1$ و $\text{Cov}(X,Y)=\frac{1}{5}$ باشد، آن‌گاه $\text{Cov}(U,Z)$ کدام است؟

۱) -1 ۲) -2 ۳) 0 ۴) 3

حل: گزینه ۲ درست است.

بنا بر خاصیت (۶) داریم:

$$\begin{cases}\text{Cov}(U,Z)=\text{Cov}(-5Y+\overset{0}{\cancel{1}},\,2X+\overset{0}{\cancel{4}})=(-5)(2)\text{Cov}(X,Y)=-10\times\frac{1}{5}=-2\\ \text{Cov}(X,Y)=\frac{1}{5}\end{cases}$$

مثال ۸ اگر $\sigma_Y^2=\frac{1}{2}$ و $\sigma_X^2=\frac{1}{5}$ و $\text{Cov}(X,Y)=1$ باشد، آن‌گاه مقدار $\text{Cov}(U,V)$ درصورتی‌که $U=2X+Y$ و $V=-5X+4Y+10$ هستند، کدام است؟

۱) 0 ۲) 3 ۳) 6 ۴) -2

حل: گزینه ۲ درست است.

بنا بر خاصیت (۷) داریم:

$$\text{Cov}(U,V)=\text{Cov}\left(2X+Y,-5X+4Y+\overset{0}{\cancel{10}}\right)$$

$$=(2)(-5)\text{Cov}(X,X)+(2)(4)\text{Cov}(X,Y)+(1)(-5)\text{Cov}(Y,X)+(1)(4)\text{Cov}(Y,Y)$$

$$=-10\sigma_X^2+8\text{Cov}(X,Y)-5\text{Cov}(Y,X)+4\sigma_Y^2=-10\times\frac{1}{5}+3\text{Cov}(X,Y)+4\times\frac{1}{2}=-2+3\times1+2=3$$

واریانس مجموع دو متغیر

اگر c,b,a مقادیر ثابتی باشند (مثبت یا منفی)، آن‌گاه:

$$\boxed{\text{Var}(aX+bY+c)=a^2\sigma_X^2+b^2\sigma_Y^2+2ab\,\text{Cov}(X,Y)}$$

یادآوری: $\sigma^2(c)=0$ (واریانس مقدار ثابت c برابر 0 است.)

نتیجه ۱: در صورتی‌که دو متغیر تصادفی X و Y مستقل باشند.

$$\text{مستقل } X,Y \xrightarrow{\text{Cov}(X,Y)=0} \sigma^2(aX+bY+c)=a^2\sigma_X^2+b^2\sigma_Y^2$$

نتیجه۲: در صورتی‌که X_1, X_2, ..., X_n، متغیرهای تصادفی مستقل باشند:

$$\text{مستقل } X_1,X_2,\ldots,X_n \xrightarrow[(i\neq j)]{Cov(X_i,X_j)=0} \sigma^2(a_1X_1+\ldots+a_nX_n)=a_1^2\sigma_{X_1}^2+\cdots+a_n^2\sigma_{X_n}^2$$

نتیجه ۳: در صورتی‌که X, Y, Z، سه متغیر تصادفی دلخواه باشند:

$$Var(aX+bY+cZ+d)=a^2\sigma_X^2+b^2\sigma_Y^2+c^2\sigma_Z^2+2ab\,Cov(X,Y)+2ac\,Cov(X,Z)+2bc\,Cov(Y,Z)$$

یادآوری: رابطه زیر همواره برقرار است:

$$E(aX+bY+c)=aE(X)+bE(Y)+c$$

مثال ۱ اگر $\sigma_X^2=\frac{1}{2}$ و $\sigma_Y^2=\frac{2}{3}$ و $\sigma^2(X+Y)=\frac{1}{2}$ باشد، مقدار کوواریانس X و Y کدام است؟

۱) $-\frac{1}{3}$ ۲) $\frac{1}{3}$ ۳) $\frac{2}{3}$ ۴) $\frac{1}{6}$

حل: گزینه ۱ درست است.

$$Var(X+Y)=\sigma_X^2+\sigma_Y^2+2Cov(X,Y) \rightarrow \frac{1}{2}=\frac{1}{2}+\frac{2}{3}+2Cov(X,Y) \rightarrow Cov(X,Y)=-\frac{1}{3}$$

مثال ۲ اگر $V(X)=8$ و $V(Y)=2$ و $Cov(X,Y)=-1$ باشد، واریانس $Z=\frac{1}{2}X-Y+5$ کدام است؟

۱) 8 ۲) 5 ۳) 10 ۴) 15

حل: گزینه ۲ درست است.

$$Var(Z)=Var\left(\frac{1}{2}X-Y+\overset{0}{\cancel{5}}\right)=\frac{1}{4}\sigma_X^2+(-1)^2\sigma_Y^2+2\left(\frac{1}{2}\right)(-1)Cov(X,Y)=\frac{1}{4}\times 8+2-(-1)=5$$

یادآوری: $V(X)=\sigma_X^2$, $V(Y)=\sigma_Y^2$

مثال ۳ اگر $\mu_X=2$ و $\mu_Y=3$ و $\mu_Z=5$ و واریانس‌ها به ترتیب $\sigma_X^2=1$ و $\sigma_Y^2=2$ و $\sigma_Z^2=4$ باشد و متغیرهای تصادفی X, Y, Z دو به دو مستقل باشند، آن‌گاه میانگین و واریانس $W=2X-Y-3Z$ کدام است؟ (چپ به راست)

۱) $-14, 42$ ۲) $2, 25$ ۳) $0, 40$ ۴) $1, 5$

حل: گزینه ۱ درست است.

$$E(2X-Y-3Z)=2E(X)-E(Y)-3E(Z)=2\times 2-3-3\times 5=-14$$

$$\sigma^2(2X-Y-3Z)=4\sigma_X^2+\sigma_Y^2+9\sigma_Z^2+2(2)(-1)Cov(X,Y)+2(2)(-3)Cov(X,Z)+2(-1)(-3)Cov(Y,Z)$$

$$\text{مستقل } X,Y,Z \rightarrow Cov(X,Y)=Cov(X,Z)=Cov(Y,Z)=0 \rightarrow \sigma^2(2X-Y-3Z)=4\times 1+2+9\times 4=42$$

مثال ۴ اگر $Var(X)=3$ و $Var(Y)=4$ و $Cov(X,Y)=-1$ باشند، آن‌گاه واریانس $Z=2X-Y+1$ چقدر است؟(اقتصاد ـ ۸۸)

۱) 12 ۲) 18 ۳) 20 ۴) 21

حل: گزینه ۳ درست است.

$$\begin{cases} Var(Z)=Var\left(2X-Y+\underset{0}{\cancel{1}}\right)=2^2Var(X)+(-1)^2Var(Y)+2(2)(-1)Cov(X,Y) \\ \qquad =4\times 3+1\times 4-4\times(-1)=12+4+4=20 \\ Var(X)=3 \ , \ Var(Y)=4 \ , \ Cov(X,Y)=-1 \end{cases}$$

مثال ۵ مقدار $\operatorname{Cov}(X-Y, X+Y)$ کدام است؟ (محیط زیست ـ ۸۷)

۱) $\operatorname{Var}(X)-\operatorname{Var}(Y)$
۲) $\operatorname{Var}(X+Y)-\operatorname{Var}(X-Y)$
۳) $\operatorname{Var}(X-Y)-\operatorname{Var}(X+Y)$
۴) $\operatorname{Var}(X)+\operatorname{Var}(Y)$

حل: گزینه ۱ درست است.

$$\operatorname{Cov}(X-Y, X+Y)=\operatorname{Cov}(X, X)\cancel{+\operatorname{Cov}(X, Y)-\operatorname{Cov}(Y, X)}-\operatorname{Cov}(Y, Y)=\operatorname{Var}(X)-\operatorname{Var}(Y)$$

$\mathbf{E(X \pm Y), V(X \pm Y)}$

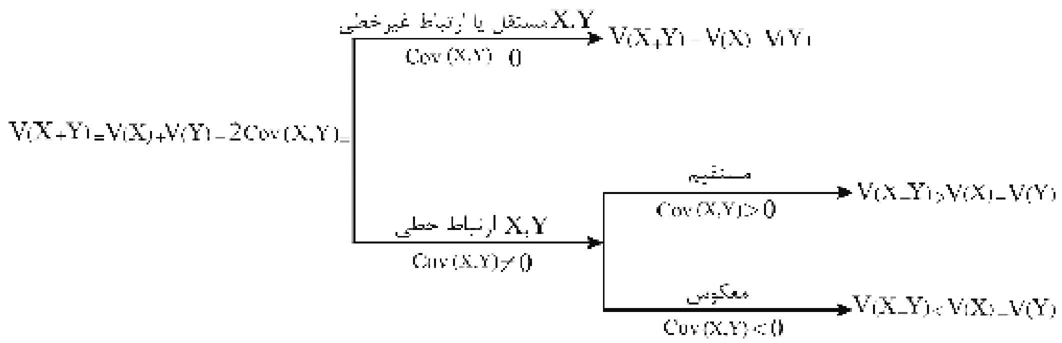
۱- بر اساس خاصیت خطی بودن میانگین، روابط زیر همواره برقرارند و با مستقل یا وابسته بودن متغیرهای تصادفی ارتباطی ندارند.

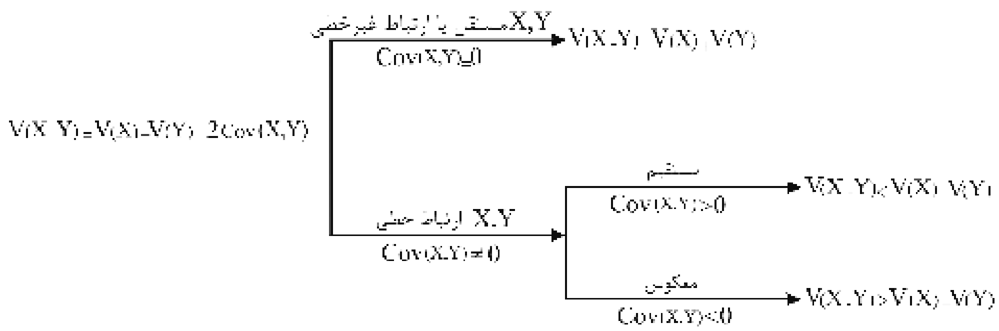
$$\boxed{\begin{aligned}E(X+Y)&=E(X)+E(Y)\\ E(X-Y)&=E(X)-E(Y)\end{aligned}}$$

۲- $V(X-Y), V(X+Y)$ با توجه به تعاریف بیان شده به صورت زیر محاسبه می شوند:

$$\boxed{\begin{aligned}V(X-Y)&=V(X)+V(Y)-2\operatorname{Cov}(X, Y)\\ V(X+Y)&=V(X)+V(Y)+2\operatorname{Cov}(X, Y)\end{aligned}}$$

که با در نظر گرفتن روابط متغیرهای X و Y به صورت زیر می‌توانند بررسی شوند:





نتیجه:

(X و Y یا مستقل‌اند یا وابسته‌اند با ارتباط غیرخطی) $\longrightarrow$

$V(X\pm Y)=V(X)+V(Y)$
$\mathrm{Cov}(X,Y)=0$
X و Y ناهمبسته‌اند.

به عبارت دیگر برقراری رابطه $V(X\pm Y)=V(X)+V(Y)$ تنها معادل ناهمبستگی یا $\mathrm{Cov}(X,Y)=0$ یا $E(XY)=E(X)\cdot E(Y)$ است و لزوماً نمی‌تواند مستقل بودن دو متغیر X و Y را نشان دهد، زیرا ممکن است دو متغیر وابسته با ارتباط غیرخطی باشند.

مثال ۱ اگر $V(X\pm Y)=V(X)+V(Y)$ باشد، کدام‌یک از روابط زیر برقرار است؟

۱) $\mathrm{Cov}(X,Y)=0$
۲) X و Y لزوماً مستقل‌اند.
۳) X و Y ناهمبسته‌اند.
۴) ۱ و ۳

حل: گزینه ۴ درست است.

مثال ۲ اگر $V(X)=2$ و $V(Y)=8$ و ارتباط دو متغیر، خطی و مستقیم باشد، کدام عبارت درست است؟

۱) $V(X+Y)=10$
۲) $V(X+Y)<10$
۳) $V(X+Y)>10$
۴) هیچ‌کدام

حل: گزینه ۳ درست است.

X,Y ارتباط خطی مستقیم دارند $\xrightarrow{\mathrm{Cov}(X,Y)>0} V(X+Y)=V(X)+V(Y)+2\mathrm{Cov}(X,Y)$

$$\to V(X+Y)=2+8+2\mathrm{Cov}(X,Y)=10+2\mathrm{Cov}(X,Y)\xrightarrow{\mathrm{Cov}(X,Y)>0} V(X+Y)>10$$

اگر در بین گزینه‌ها، گزینه $V(X-Y)<10$ وجود داشت، این عبارت هم می‌توانست درست باشد، زیرا:

$$V(X-Y)=V(X)+V(Y)-2\mathrm{Cov}(X,Y)\xrightarrow{\mathrm{Cov}(X,Y)>0} V(X-Y)<10$$

در شرایط زیر گزینه‌های ۱ و ۲ نیز می‌توانستند، درست باشند:

X,Y مستقل‌اند (ناوابسته) یا ارتباط غیرخطی دارند (گزینه ۱) $\xrightarrow{\mathrm{Cov}(X,Y)=0} V(X+Y)=V(X)+V(Y)=10$

X,Y ارتباط خطی معکوس دارند (گزینه ۲) $\xrightarrow{\mathrm{Cov}(X,Y)<0} V(X+Y)<V(X)+V(Y)=10$

ضریب همبستگی (Coefficient of Correlation)

همان‌طورکه گفته شد، کوواریانس معیاری است که جهت و نوع ارتباط خطی دو متغیر X و Y را مشخص می‌کند و با شیب خط رگرسیون هم‌علامت است. حال اگر بخواهیم علاوه بر تعیین جهت و نوع ارتباط خطی دو متغیر، **شدت ارتباط خطی** آن‌ها را از این نظر بررسی کنیم که آیا خط بیان‌کننده ارتباط خطی دو متغیر (خط رگرسیون) از تمام نقاط واقعی (x, y) عبور می‌کند (شدت کامل) یا عبور نمی‌کند (شدت ناقص)، باید از معیار دیگری به نام ضریب همبستگی به صورت زیر استفاده کنیم.

شدت کامل: خط رگرسیون از تمام نقاط واقعی عبور می‌کند (تمام زوج نقاط واقعی (x,y) در معادله خط رگرسیون صدق می‌کنند).

شدت ناقص: خط رگرسیون از تمام نقاط واقعی عبور نمی‌کند (حداقل یک زوج (x,y) واقعی وجود دارد که در معادله خط رگرسیون صدق نمی‌کند).

تعریف: ضریب همبستگی هر جامعه یک معیار نسبی بدون واحد در فاصله $-1 \le \rho_{X,Y} \le 1$ است که از رابطه زیر محاسبه می‌شود:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sqrt{E(X-\mu_X)^2}\ \sqrt{E(Y-\mu_Y)^2}} = \frac{\sum(x-\mu_X)(y-\mu_Y)}{\sqrt{\sum(x-\mu_X)^2}\ \sqrt{\sum(y-\mu_Y)^2}}$$

این معیار علاوه بر تعیین جهت و نوع ارتباط خطی دو متغیر، **شدت ارتباط خطی دو متغیر را از نظر کامل یا ناقص** بودن نشان می‌دهد.

درصورتی‌که بخواهیم با توجه به داده‌های حاصل از یک نمونه، ضریب همبستگی را بیابیم، از رابطه زیر استفاده می‌کنیم:

$$r_{X,Y} = \frac{Cov(X,Y)}{S_X .\ S_Y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

که در آن $Cov(X,Y)$، کوواریانس نمونه است.

قضیه: کوواریانس دو متغیر استاندارد، برابر با ضریب همبستگی جامعه اصلی است.

$$Cov\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = Cov\left(\frac{X}{\sigma_X} - \frac{\mu_X}{\sigma_X}, \frac{Y}{\sigma_Y} - \frac{\mu_Y}{\sigma_Y}\right) = Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) = \frac{Cov(X,Y)}{\sigma_X . \sigma_Y} = \rho_{X,Y}$$

تحلیل ضریب همبستگی

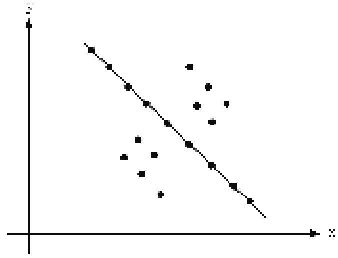
بعد از محاسبه ضریب همبستگی جامعه یا نمونه، با توجه به آنکه مقدار آن صفر یا مخالف صفر شده باشد می‌توان آن‌ها را به صورت زیر بررسی کرد:

$\rho_{X,Y} \neq 0$

اگر ضریب همبستگی دو متغیر تصادفی X و Y مخالف صفر باشد، آن‌گاه دو متغیر **همبسته‌اند** و با هم وابستگی خطی دارند. در این شرایط رابطه بین X و Y را می‌توان با خطی به نام خط رگرسیون مشخص کرد (در این وضعیت قطعاً $Cov(X,Y)$ و شیب خط رگرسیون نیز مخالف صفر هستند).

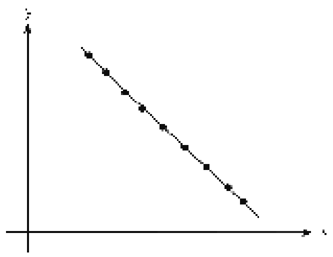
الف) مقدار $r_{X,Y}$ شدت ارتباط خطی را نشان می‌دهد (کامل یا ناقص).

ب) علامت $r_{X,Y}$ جهت ارتباط خطی را نشان می‌دهد (مانند کوواریانس و شیب خط رگرسیون).



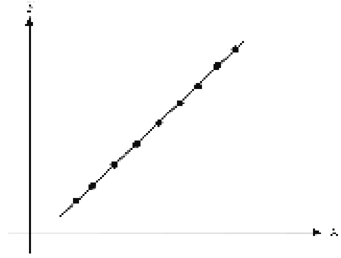
$-1 < \rho < 0$

(ارتباط خطی معکوس، شدت ناقص)



$\rho = -1$

(ارتباط خطی معکوس، شدت کامل)



$\rho = 1$

(ارتباط خطی مستقیم، شدت کامل)

$0 < \rho < 1$

(ارتباط خطی مستقیم، شدت ناقص)

مثال ۱ اگر $r_{X,Y} = -0.6$ باشد، کدام‌یک از عبارات زیر درست است؟

۱) ارتباط معکوس و شدت ناقص ـ شیب خط و کوواریانس منفی

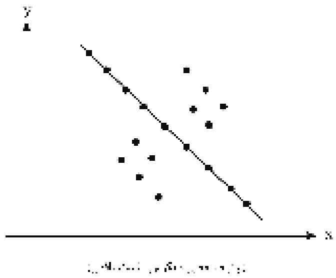
۲) ارتباط معکوس و شدت کامل ـ شیب خط و کوواریانس مثبت

۳) ارتباط مستقیم و شدت ناقص ـ شیب خط و کوواریانس منفی

۴) ارتباط معکوس و شدت ناقص ـ کوواریانس منفی ولی علامت شیب خط مشخص نیست.

حل: گزینه ۱ درست است.

ارتباط معکوس و شدت ناقص $\leftarrow$



$-1 < r_{X,Y} = -0.6 < 0 \rightarrow$

$r_{X,Y} = -0.6 < 0 \quad \longleftrightarrow \quad Cov(X,Y) < 0$ و شیب خط < 0

مثال ۲ معادله خط رگرسیون با توجه به نمونه 10تایی برآوردشده به صورت $\hat{y} = 2 - x$ است. کدام‌یک از این موارد نمی‌تواند ضریب همبستگی آن باشد؟

۱) -0.45 ۲) -0.90 ۳) 0.95 ۴) -0.85

حل: گزینه ۳ درست است.

شیب خط رگرسیون منفی است، بنابراین علامت ضریب همبستگی نیز قطعاً منفی است؛ پس تنها گزینه ۳ نمی‌تواند ضریب همبستگی باشد زیرا علامتش مثبت است.

مثال ۳ معادله رگرسیون $y = 400 - 20x$ را در نظر بگیرید. مقدار واقعی y به ازای $x = 15$ برابر با 150 است. ضریب همبستگی کدام است؟ (مدیریت ـ ۷۳)

۱) $0 < r < 1$ ۲) $-1 < r < 0$ ۳) $r = 1$ ۴) $r = -1$

حل: گزینه ۲ درست است.

شیب خط $y = 400 - 20x$ منفی است $(b = -20)$، بنابراین جهت ارتباط X و Y معکوس و ضریب همبستگی منفی است $(r < 0)$.

بنابراین با توجه به تحلیل ضریب همبستگی، زمانی که ضریب همبستگی منفی است دو وضعیت وجود دارد:

$$\begin{cases} r=-1 & \text{ارتباط معکوس و کامل} \\ -1<r<0 & \text{ارتباط معکوس و ناقص} \end{cases}$$

ازآنجاکه نقطه واقعی $(x=15\ ,\ y=150)$ در معادله خط رگرسیون $y=400-20x$ صدق نمی‌کند $(150\neq 400-20\times 15)$، ارتباط X و Y ناقص و $-1<r<0$ است.

توجه کنید که اگر نقطه در معادله خط رگرسیون صدق می‌کرد، باز هم نمی‌توانستیم نتیجه بگیریم که $r=-1$ است، زیرا باید تمام نقاط واقعی در معادله خط رگرسیون صدق کنند.

مثال ۴ معادله رگرسیون $y=2x$ را که از مجموعه نقاط زیر به دست آمده است، در نظر بگیرید. ضریب همبستگی کدام است؟

x	7	10	4	11
y	14	20	8	22

۱) $0<r<1$ ۲) $-1<r<0$

۳) $r=1$ ۴) $r=-1$

حل: گزینه ۳ درست است.

با توجه به آنکه شیب خط مثبت است و تمام نقاط در معادله خط صدق می‌کنند، ارتباط خطی، مستقیم و شدت، کامل و درنتیجه $r=1$ است.

$\rho_{X,Y}=0$

اگر ضریب همبستگی دو متغیر تصادفی X و Y برابر با صفر شود، حتماً کوواریانس صفر شده است، آن‌گاه دو متغیر **ناهمبسته‌اند** و برعکس. در این وضعیت تمام موارد که در **«نمودار روابط» و «جدول همبستگی»** برای ناهمبستگی دو متغیر بیان شد، برای $\rho_{X,Y}=0$ نیز برقرار است.

مثال ۵ هرگاه $\rho_{X,Y}=0$ باشد، آن‌گاه دو متغیر X و Y:

۱) $Cov(X,Y)=0$ ۲) X و Y ناهمبسته‌اند (استقلال خطی دارند)

۳) $E(XY)=E(X)E(Y)$ ۴) همه موارد

حل: گزینه ۴ درست است.

$$\rho_{X,Y}=0 \ \leftrightarrow\ Cov(X,Y)=0 \ \leftrightarrow\ E(XY)-E(X)E(Y)=0 \ \leftrightarrow\ E(XY)=E(X)E(Y)$$

محاسبه ضریب همبستگی

برای محاسبه ضریب همبستگی دو متغیر تصادفی X و Y با توجه به رابطه $\rho_{X,Y}=\dfrac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}$ ، بهتر است به صورت زیر عمل کنیم.

۱) اگر دو متغیر مستقل باشند، ناهمبسته‌اند، در نتیجه کوواریانس آن‌ها صفر است و ضریب همبستگی نیز صفر می‌شود.

$$X,Y \text{ مستقل } \xrightarrow{f(x,y)=f(x)\times f(y)} Cov(X,Y)=0 \ \rightarrow\ \rho_{X,Y}=0$$

۲) اگر دو متغیر وابسته باشند، باید کوواریانس آن‌ها را محاسبه کنیم، زیرا ممکن است مقدار آن باز هم صفرشود (وابستگی غیرخطی)، در این وضعیت باز هم ضریب پ صفر می شود.

$$X,Y \text{ وابسته } \xrightarrow{f(x,y)\neq f(x)\times f(y)} Cov(X,Y)=E(XY)-E(X)E(Y)$$

درصورتی‌که کوواریانس دو متغیر مخالف صفر باشد (وابستگی خطی)، ضریب همبستگی آن‌ها نیز مخالف صفر است و باید مقادیر σ_X,σ_Y را برای محاسبه ضریب همبستگی به دست آوریم.

مثال ۱ جدول توزیع احتمال مشترک زیر را در نظر بگیرید. ضریب همبستگی X و Y برابر است با: (اقتصاد ـ ۸۱)

y \ x	-1	0
-1	0.15	0.15
1	0.35	0.35

۱) -1　　۲) صفر

۳) 0.5　　۴) 1

حل: گزینه ۲ درست است.

y \ x	-1	0	$f(y)$
-1	0.15	0.15	0.30
1	0.35	0.35	0.70
$f(x)$	0.50	0.50	1

$\xrightarrow{f(x,y)=f(x)\cdot f(y)}$

y \ x	-1	0
-1	$0.15=0.3\times0.5$	$0.15=0.3\times0.5$
1	$0.35=0.7\times0.5$	$0.35=0.7\times0.5$

ازآنجاکه X و Y دو متغیر تصادفی مستقل‌اند، ضریب همبستگی آن‌ها برابر صفر است.

مثال ۲ تابع احتمال متغیر تصادفی X به صورت زیر است. متغیر $Y=X^2$ را در نظر بگیرید. مقدار ضریب همبستگی $\rho(X,Y)$ برابر است با: (اقتصاد ـ ۸۵)

x	-2	0	2
$P(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

۱) -1　　۲) 0

۳) $\frac{\sqrt{2}}{2}$　　۴) 1

حل: گزینه ۲ درست است.

ازآنجاکه تابع احتمال توأم $f(x,y)$ وجود ندارد، ابتدا $Cov(X,Y)$ را محاسبه می‌کنیم:

$$\begin{cases} Cov(X,X^2)=E(XX^2)-E(X)E(X^2)=E(X^3)-E(X)E(X^2)=0-0\times E(X^2)=0 \\ E(X)=\sum xP(x)=-2\times\frac{1}{4}+0\times\frac{1}{2}+2\times\frac{1}{4}=0 \\ E(X^3)=\sum x^3P(x)=(-2)^3\times\frac{1}{4}+0^3\times\frac{1}{2}+2^3\times\frac{1}{4}=0 \end{cases}$$

حال ازآنجاکه کوواریانس دو متغیر برابر با صفر شده است، ضریب همبستگی آن‌ها نیز صفر می‌شود:

$$\rho_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\ \rightarrow\ \rho_{X,X^2}=\frac{Cov(X,X^2)}{\sigma_X\sigma_{X^2}}=\frac{0}{\sigma_X\sigma_{X^2}}=0$$

محاسبه ضریب همبستگی نمونه

با توجه به رابطه ضریب همبستگی جامعه به صورت زیر:

$$\rho_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}$$

ضریب همبستگی نمونه برابر است با:

$$r_{X,Y}=\frac{Cov(X,Y)}{S_X.S_Y}=\frac{\frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}}{\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}\sqrt{\frac{\sum(y-\bar{y})^2}{n-1}}}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

از طرفی داریم:

$$r_{X,Y}=\frac{Cov(X,Y)}{S_X.S_Y}=\frac{\frac{1}{n-1}\left[\sum xy-\frac{\sum x\sum y}{n}\right]}{\sqrt{\frac{1}{n-1}\left[\sum x^2-\frac{\left(\sum x\right)^2}{n}\right]}\sqrt{\frac{1}{n-1}\left[\sum y^2-\frac{\left(\sum y\right)^2}{n}\right]}}=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\sum x^2-\frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2-\frac{\left(\sum y\right)^2}{n}}}$$

بنابراین، با توجه به داده‌های مسئله می‌توانیم از یکی از روابط زیر استفاده کنیم:

$$\boxed{r_{X,Y}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\sum x^2-\frac{\left(\sum x\right)^2}{n}}\sqrt{\sum y^2-\frac{\left(\sum y\right)^2}{n}}}}$$

و درصورتی‌که روابط زیر را در نظر بگیریم:

$$SP_{xy}=\sum(x-\bar{x})(y-\bar{y})=\sum xy-\frac{\sum x\sum y}{n}$$

$$S_{xx}=SS_x=\sum(x-\bar{x})^2=\sum x^2-\frac{\left(\sum x\right)^2}{n}$$

$$S_{yy}=SS_y=\sum(y-\bar{y})^2=\sum y^2-\frac{\left(\sum y\right)^2}{n}$$

آن‌گاه به رابطه زیر برای محاسبه ضریب همبستگی نمونه می‌رسیم:

$$\boxed{r_{X,Y}=\frac{SP_{xy}}{\sqrt{SS_x\times SS_y}}}$$

بیان داده‌ها برحسب انحراف از میانگین

گاهی داده‌ها برحسب انحراف از میانگین بیان می‌شوند. در این صورت داریم:

$$\begin{cases}\sum(x-\bar{x})^2=\sum x_t^2\\ \sum(y-\bar{y})^2=\sum y_t^2\\ \sum(x-\bar{x})(y-\bar{y})=\sum x_t y_t\end{cases}$$

بنابراین، ضریب همبستگی نمونه برابر است با:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}=\frac{\sum x_t y_t}{\sqrt{\sum x_t^2\sum y_t^2}}$$

مثال ۳ ضریب همبستگی جامعه دو متغیر وابسته X و Y برابر است با: (مدیریت ـ ۷۴)

x	7	10	4	11
y	14	20	8	22

۱) 2 ۲) 1

۳) $\frac{1}{2}$ ۴) -1

حل: گزینه ۲ درست است.

راه حل اول: با توجه به آنکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعدادی صحیح هستند،

$$\begin{cases} \bar{x}=\dfrac{\sum x_i}{n}=\dfrac{32}{4}=8 \\ \bar{y}=\dfrac{\sum y_i}{n}=\dfrac{64}{4}=16 \end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه، مناسب‌تر است:

x	y	$(x-\bar{x})$	$(y-\bar{y})$
7	14	−1	−2
10	20	2	4
4	8	−4	−8
11	22	3	6

$\longrightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
2	1	4
8	4	16
32	16	64
18	9	36
$\sum=60$	$\sum=30$	$\sum=120$

$$r_{X,Y}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}=\frac{60}{\sqrt{30\times120}}=1$$

راه حل دوم: با کمی دقت در هر زوج (x,y) متوجه می‌شویم که رابطه $y=2x$ بین تمام نقاط (x,y) برقرار است (مقدار هر y دو برابر x است).

حال ازآنجاکه شیب خط مثبت است و تمام نقاط در رابطه صدق می‌کنند، ارتباط خطی، مستقیم و شدت، کامل است، درنتیجه $r_{X,Y}=+1$ است.

مثال ۴ مقادیر عددی دو متغیر X و Y به شرح زیر است، ضریب همبستگی بین X و Y کدام است؟(برنامه ریزی شهری ـ ۸۴)

x	3	4	5	10
y	3	5	7	17

۱) $\frac{1}{2}$ ۲) $\frac{2}{3}$

۳) $\frac{3}{4}$ ۴) 1

حل: گزینه ۴ درست است.

راه حل اول:

x	y	xy	x^2	y^2
3	3	9	9	9
4	5	20	16	25
5	7	35	25	49
10	17	170	100	289
$\sum x=22$	$\sum y=32$	$\sum xy=234$	$\sum x^2=150$	$\sum y^2=372$

$$r=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\left(\sum x^2-\frac{(\sum x)^2}{n}\right)\left(\sum y^2-\frac{(\sum y)^2}{n}\right)}}=\frac{234-\frac{22\times32}{4}}{\sqrt{\left(150-\frac{(22)^2}{4}\right)\left(372-\frac{(32)^2}{4}\right)}}=\frac{234-176}{\sqrt{(150-121)\times(372-256)}}=\frac{58}{58}=1$$

راه حل دوم:

با کمی دقت در مقادیر X و Y می‌توان فهمید که رابطه $y = 2x - 3$ بین مقادیر X و Y برقرار است. حال ازآنجاکه این خط از تمام نقاط عبور کرده (شدت کامل) و شیب خط نیز مثبت است، می‌توان نتیجه گرفت $r = 1$ است.

مثال ۵ در جدول فراوانی داده‌های آماری دو صفت X و Y داریم:

$$\sum(x_i - \bar{x})^2 = 32 ،\sum(y_i - \bar{y})^2 = 50 ،\sum(x_i - \bar{x})(y_i - \bar{y}) = 24$$

ضریب همبستگی کدام است؟ (حسابداری ـ ۷۷)

۱) 0.4 ۲) 0.5 ۳) 0.6 ۴) 0.8

حل: گزینه ۳ درست است.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{24}{\sqrt{32 \times 50}} = \frac{24}{\sqrt{1600}} = \frac{24}{40} = 0.6$$

مثال ۶ برای بررسی همبستگی بین دو متغیر تصادفی X و Y، نمونه‌ای به حجم $n = 10$ از جامعه نرمال انتخاب کرده، بر اساس نتایج مشاهدات داریم: $\sum x_i = 30$ و $\sum y_i = 40$ و $\sum x_i^2 = 234$ و $\sum y_i^2 = 196$ و $\sum x_i y_i = 160$. تخمین ضریب همبستگی بین دو متغیر تصادفی X و Y کدام است؟ (اقتصاد ـ ۷۳)

۱) 0.55 ۲) 0.45 ۳) 0.85 ۴) 0.65

حل: گزینه ۱ درست است.

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} = \frac{160 - \frac{30 \times 40}{10}}{\sqrt{\left(234 - \frac{(30)^2}{10}\right)\left(196 - \frac{(40)^2}{10}\right)}} = \frac{160 - 120}{\sqrt{144 \times 36}} = \frac{40}{12 \times 6} = \frac{5}{9} = 0.55$$

مثال ۷ اگر معادله خط رگرسیون y نسبت به x به صورت $y = -2x + b$ باشد و $SS_y = 4SS_x$ و $SP_{xy} = -2SS_x$ باشد، ضریب همبستگی X و Y کدام است؟

۱) 1 ۲) 0.9 ۳) -1 ۴) -0.9

حل: گزینه ۳ درست است.

ازآنجاکه شیب خط منفی است، ضریب همبستگی نیز منفی خواهد بود و فعلاً گزینه‌های ۳ و ۴ می‌توانند درست باشند؛ حال مقدار ضریب همبستگی را محاسبه می‌کنیم:

$$\begin{cases} r_{X,Y} = \dfrac{SP_{xy}}{\sqrt{SS_x SS_y}} \\ SP_{xy} = -2SS_x \ , \ SS_y = 4SS_x \end{cases} \longrightarrow r_{X,Y} = \frac{-2SS_x}{\sqrt{SS_x \cdot 4SS_x}} = \frac{-2SS_x}{2\,SS_x} = -1$$

خواص ضریب همبستگی

با توجه به رابطه $-1 \le \rho_{X,Y} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y} \le 1$، خواص زیر برقرار است (a، b، c و d ضرایب ثابت مثبت یا منفی هستند):

1) $\rho_{X,Y} = \rho_{Y,X}$
2) $\rho_{X,a} = \rho_{a,X} = \rho_{a,b} = 0$
3) $\rho_{X,X} = \rho_{-X,-X} = 1$
4) $\rho_{X,-X} = \rho_{-X,X} = -1$
5) $\rho_{aX\pm b\,,\,cY\pm d} = \begin{cases} \rho_{X,Y} & (\text{c,a هم‌علامت}) \\ -\rho_{X,Y} & (\text{c,a مختلف‌العلامت}) \end{cases}$

توضیحات:

۱) ارتباط X با Y همان ارتباط Y با X است.

۲) هیچ ارتباط خطی بین دو ثابت یا یک متغیر و یک ثابت وجود ندارد و ضریب همبستگی بین آن‌ها صفر است.

۳) ارتباط یک متغیر با خودش، مستقیم و شدت، کامل است.

۴) ارتباط یک متغیر با منفی خودش، معکوس و شدت، کامل است.

۵) قدرمطلق ضریب همبستگی با تغییر مبدأ و مقیاس هیچ تغییری نمی‌کند. تغییرات خطی روی X و Y به صورت $aX \pm b$ و $cY \pm d$ هیچ تأثیری در مقدار ضریب همبستگی $(\rho_{X,Y})$ ندارد و تنها زمانی که ضرایب a و c مختلف‌العلامت باشند، علامت ضریب همبستگی برعکس می‌شود.

برای مثال، اگر $\rho_{X,Y} = 0.2$ باشد، ضریب همبستگی هر یک از حالات زیر، بر اساس خواص مطرح شده به دست می‌آید:

الف) $\rho_{2X+3,5Y+7} = \rho_{X,Y} = 0.2$ (بنا بر خاصیت ۵)

ب) $\rho_{-2X+3,5Y+7} = -\rho_{X,Y} = -0.2$ (بنا بر خاصیت ۵)

ج) $\rho_{-2X+4\,,\,-8Y+6} = \rho_{X,Y} = 0.2$ (بنا بر خاصیت ۵)

د) $\rho_{-2X+1,4X-5} = \rho_{-X,X} = -1$ (بنا بر خاصیت‌های ۵ و ۴)

ه) $\rho_{5X-3,8X+7} = \rho_{X,X} = 1$ (بنا بر خاصیت‌های ۵ و ۳)

و) $\rho_{2X,4X} = \rho_{X,X} = 1$ (بنا بر خاصیت ۵ و ۳)

ز) $\rho_{-2X,5Y} = -\rho_{X,Y} = -0.2$ (بنا بر خاصیت ۵)

ح) $\rho_{6,X} = \rho_{X,6} = 0$ (بنا بر خاصیت ۲)

ضریب تعیین (Coefficient of Determination)

اگر ضریب همبستگی دو متغیر تصادفی $r_{X,Y} = 0.8$ باشد، با توجه به آنچه در مبحث ضریب همبستگی بیان شد، می‌توان گفت:

اولاً، X و Y ارتباط خطی مستقیم دارند $(r_{X,Y} = 0.8 > 0)$.

ثانیاً، شدت ارتباط خطی دو متغیر تصادفی X و Y ناقص است $(0 < r_{X,Y} = 0.8 < 1)$، یعنی خط رگرسیون از تمام نقاط واقعی (x,y) عبور نمی‌کند.

اما نمی‌توان گفت:

۱- چند درصد از تغییرات Y توسط X (با استفاده از خط رگرسیون) قابل بیان است.

۲- چند درصد از تغییرات Y تحت تأثیر X است.

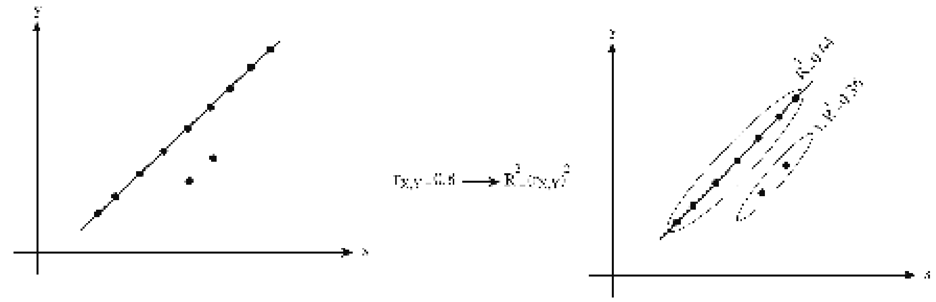
۳- شدت ارتباط ضریب همبستگی X و Y چند برابر قوی‌تر از شدت ارتباط دو متغیر دیگر است.

۴- چند درصد از نقاط واقعی Y توسط X از معادله خط رگرسیون $(y = ax + b)$ قابل پیش‌بینی است.

۵- چند درصد از نقاط واقعی (x, y) روی معادله خط رگرسیون قرار دارد.

تعریف: ضریب تعیین (R^2) معیاری است که از مجذور ضریب همبستگی $(r_{X,Y})$ به دست می‌آید و برای بیان درصد ارتباط خطی میان دو متغیر X و Y، ملاکی گویاتر از ضریب همبستگی است.

$$0 \le R^2 = (r_{X,Y})^2 \le 1$$



تعاریف زیر برای R^2 (ضریب تعیین) و $1-R^2$ بیان می‌شوند.

R^2

تعریف ۱: درصد تغییرات متغیر تابع یا وابسته Y که توسط خط رگرسیون $(y = bx + a)$ قابل بیان است.

تعریف ۲: درصد تغییرات متغیر تابع یا وابسته Y که توسط متغیر مستقل X قابل بیان (تعیین) است.

تعریف ۳: درصد ارتباط خطی میان دو متغیر X و Y که قابل بیان است.

تعریف ۴: درصد تغییرات متغیر تابع یا وابسته Y که تحت تأثیر متغیر مستقل(آزاد) X است.

$(1-R^2)$

تعریف ۱: درصد تغییرات متغیر تابع یا وابسته Y که توسط خط رگرسیون $(y = bx + a)$ قابل بیان نیست.

تعریف ۲: درصد تغییرات متغیر تابع یا وابسته Y که توسط متغیر مستقل X قابل بیان (تبیین) نیست.

تعریف ۳: درصد ارتباطی خطی میان دو متغیر X و Y که قابل بیان نیست.

تعریف ۴: درصد تغییرات متغیر تابع یا وابسته Y که تحت تأثیر متغیر مستقل (آزاد) X نیست.

مثال ۱ اگر $Cov(X,Y)=1$ و $\sigma_X^2=4,\sigma_Y^2=1$ باشد، ضریب تعیین چقدر است؟

۱) 0.25 ۲) 0.5 ۳) 0.75 ۴) 0.45

حل: گزینه ۱ درست است.

$$\begin{cases} r_{X,Y}=\dfrac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}=\dfrac{1}{2\times 1}=0.5 \quad\rightarrow\quad R^2=\left(r_{X,Y}\right)^2=0.25 \\ \sigma_X^2=4 \quad\rightarrow\quad \sigma_X=2 \\ \sigma_Y^2=1 \quad\rightarrow\quad \sigma_Y=1 \end{cases}$$

مثال ۲ اگر $Cov(X,Y)=-36$ و $\sigma_X^2=16$ و $\sigma_Y^2=100$ باشد، چند درصد تغییرات متغیر Y توسط X قابل بیان است؟

۱) 0.25 ۲) 0.9 ۳) 0.81 ۴) 0.19

حل: گزینه ۳ درست است.

$$r_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{-36}{4\times 10}=\frac{-9}{10} \quad\rightarrow\quad R^2=\left(r_{X,Y}\right)^2=0.81$$

مثال ۳ اگر $Cov(X,Y)=28$ و $\sigma_X^2=16$ و $\sigma_Y^2=100$ باشد، چند درصد تغییرات متغیر Y توسط X قابل بیان نیست؟

۱) 0.49 ۲) 0.70 ۳) 0.51 ۴) 0.23

حل: گزینه ۳ درست است.

$$r_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{28}{4\times 10}=\frac{7}{10} \quad\rightarrow\quad \begin{cases} R^2=\left(r_{X,Y}\right)^2=0.49 & \text{(درصد تغییرات بیان‌شده)} \\ 1-R^2=0.51 & \text{(درصد تغییرات بیان‌نشده)} \end{cases}$$

مثال ۴ اگر معادله خط رگرسیون برآوردی به صورت $\hat{y}=2.4-0.6x$ به دست آمده و ضریب تعیین 0.49 باشد، ضریب همبستگی کدام است؟

۱) $+0.7$ ۲) -0.49 ۳) ± 0.7 ۴) -0.7

حل: گزینه ۴ درست است.

ازآنجاکه علامت شیب خط رگرسیون منفی است (-0.6)، با توجه به هم‌علامت بودن شیب خط و ضریب همبستگی قطعاً ضریب همبستگی هم **منفی** خواهد بود $(r_{X,Y}<0)$، درنتیجه داریم:

$$r^2=0.49 \quad\Rightarrow\quad r=\sqrt{r^2}=\begin{cases} -0.7 & \text{قابل قبول} \\ +0.7 \end{cases}$$

قوی بودن ضریب همبستگی

ضریب تعیین $\left(R^2\right)$ بهترین معیار برای آن است که نشان دهیم ضریب همبستگی دو متغیر X و Y $\left(r_{X,Y}\right)$ چند برابر **قوی‌تر** از ضریب همبستگی دو متغیر T, Z $\left(r_{Z,T}\right)$ است. برای این کار از نسبت زیر استفاده می‌کنیم:

$$r_{Z,T} \text{ نسبت به } r_{X,Y} \text{ قوت } = \left(\frac{r_{X,Y}}{r_{Z,T}}\right)^2 = \frac{R^2_{X,Y}}{R^2_{Z,T}}$$

مثال ۱ اگر ضریب همبستگی بین دو متغیر 0.8 و ضریب همبستگی بین دو متغیر دیگر 0.2 باشد، می‌توان گفت همبستگی دو متغیر اول چند برابر قوی‌تر از دو متغیر دوم است؟

۱) دو برابر ۲) چهاربرابر ۳) شانزده برابر ۴) هشت برابر

حل: گزینه ۳ درست است.

$$\begin{cases} r_{Z,T} \text{ به } r_{X,Y} \text{ قوت } = \left(\dfrac{r_{X,Y}}{r_{Z,T}}\right)^2 = \left(\dfrac{0.8}{0.2}\right)^2 = 4^2 = 16 \text{ برابر} \\ r_{X,Y} = 0.8 \,,\, r_{Z,T} = 0.2 \end{cases}$$

مثال ۲ اگر ضریب تعیین دو متغیر 0.8 و ضریب تعیین دو متغیر دیگر 0.2 باشد، می‌توان گفت همبستگی دو متغیر اول چند برابر قوی‌تر از دو متغیر دوم است؟

۱) دو برابر ۲) چهار برابر ۳) شانزده برابر ۴) هشت برابر

حل: گزینه ۲ درست است.

$$\begin{cases} r_{Z,T} \text{ به } r_{X,Y} \text{ قوت } = \left(\dfrac{r_{X,Y}}{r_{Z,T}}\right)^2 = \dfrac{R^2_{X,Y}}{R^2_{Z,T}} = \dfrac{0.8}{0.2} = 4 \text{ برابر} \\ R^2_{X,Y} = 0.8 \,,\, R^2_{Z,T} = 0.2 \end{cases}$$

خط رگرسیون

یکی از اهداف تحقیقات آماری، پیش‌بینی تغییرات یک متغیر وابسته برحسب یک متغیر مستقل در جامعه است.
اگر X را متغیر مستقل و Y را متغیر وابسته در نظر بگیریم، برای مثال، پیش‌بینی « Y : میزان مخارج مصرفی خانوارها» برحسب « X : درآمد خانوار» یا « Y : میزان سرمایه‌گذاری» برحسب « X : نرخ بهره» از جمله مطالعاتی است که می‌تواند برای تعیین میزان ارتباط خطی X و Y انجام گیرد.

تعریف: برای تعیین میزان ارتباط و همبستگی میان متغیر مستقل X و متغیر وابسته Y باید همه نقاط x و y را روی صفحه مختصات مشخص کنیم. در این صورت از بین همه خطوطی که می‌توانند از بین نقاط عبور کنند تنها یک خط وجود دارد که از بیشتر نقاط از جمله نقطه میانگین عبور می‌کند و فاصله آن نسبت به بقیه نقاط مینیمم است؛ این خط، خط برازش یا خط رگرسیون نامیده می‌شود.

اگر $y = \beta x + \alpha$ معادله خط رگرسیون جامعه باشد، آن‌گاه ضرایب α و β به صورت زیر محاسبه می‌شوند:

شیب خط (β)

برای محاسبه شیب خط از رابطه زیر استفاده می‌شود:

$$\boxed{\beta = \frac{\operatorname{Cov}(X,Y)}{\sigma_X^2} = \rho_{X,Y} \times \frac{\sigma_Y}{\sigma_X}}$$

ثابت معادله یا عرض از مبدأ (α)

ازآنجاکه نقطه میانگین (μ_x , μ_y) همواره در معادله خط صدق می‌کند، مقدار ثابت معادله یا عرض از مبدأ از رابطه زیر به دست می‌آید:

$$\boxed{\alpha = \mu_y - \beta\mu_x}$$

می‌دانیم که پارامترهای جامعه همواره مجهول هستند، بنابراین باید معادله خط رگرسیون را با استفاده از نمونه برآورد کنیم.

برآورد خط رگرسیون

فرض کنید نمونه‌ای به حجم n به صورت (x_i , y_i) داشته باشیم که در آن به ازای هر مقدار x_i متعلق به x، مقدار متناظر آن یعنی y_i متعلق به y به دست آمده است.در این حالت برای $i = 1,2,\dots,n$ نقاط (x_1,y_1)، (x_2,y_2)، ...، (x_n,y_n) به دست می‌آید.

x	x_1	x_2	$\dots$	x_n
y	y_1	y_2	$\dots$	y_n

اگر $y = bx + a$ معادله خط رگرسیون نمونه (برآورد خط رگرسیون جامعه) باشد، ضرایب a و b به صورت زیر محاسبه می‌شوند:

برآورد شیب خط (b)

شیب خط رگرسیون از رابطه زیر برآورد می‌شود:

$$\boxed{b = r_{X,Y} \times \frac{S_Y}{S_X} = \frac{\operatorname{Cov}(X,Y)}{S_X^2} = \frac{SP_{xy}}{SS_x}}$$

درواقع این روابط به صورت زیر به دست آمده‌اند:

$$b = r_{X,Y} \times \frac{S_Y}{S_X} = \frac{Cov(X,Y)}{S_X . \cancel{S_Y}} \times \frac{\cancel{S_Y}}{S_X} = \frac{Cov(X,Y)}{S_X^2} = \frac{\frac{1}{n-1}\left[\sum(x-\bar{x})(y-\bar{y})\right]}{\frac{1}{n-1}\left[\sum(x-\bar{x})^2\right]} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{SP_{xy}}{SS_x}$$

یادآوری: در داده‌های آماری روابط زیر صادق هستند:

$$SP_{xy} = \sum(x-\bar{x})(y-\bar{y}) = \sum xy - \frac{\sum x \sum y}{n}$$

$$S_{xx} = SS_x = \sum(x-\bar{x})^2 = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

بنابراین، با توجه به داده‌های مسئله، می‌توان از روابط زیر برای محاسبه b استفاده کرد:

$$\boxed{b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}}$$

برآورد ثابت معادله یا عرض از مبدأ (a)

ازآنجاکه نقطه میانگین $(\bar{x},\bar{y})$ همواره در معادله خط صدق می‌کند، ثابت معادله یا عرض از مبدأ از رابطه زیر برآورد می‌شود:

$$\boxed{a = \bar{y} - b\bar{x}}$$

خط رگرسیون عبوری از مبدأ

اگر خط رگرسیون از مبدأ مختصات یعنی نقطه $(0,0)$ عبور کند، مقدار a برابر صفر و درنتیجه معادله این خط به صورت $y = bx$ خواهد بود. در این حالت مقدار b از رابطه زیر محاسبه می‌شود:

$$y = bx + a \xrightarrow{a=0} y = bx \xrightarrow{(\bar{x},\bar{y})} b = \frac{\bar{y}}{\bar{x}}$$

✓ **دقت کنید!**

در بعضی منابع:

ـ برآورد شیب خط رگرسیون را با $\hat{\beta}$ و برآورد عرض از مبدأ را با $\hat{\alpha}$ نشان می‌دهند.

ـ معادله خط رگرسیون به صورت $y = ax + b$ بیان می‌شود که در این حالت، a، شیب خط و b، عرض از مبدأ است.

مثال ۱ برای دو صفت X و Y در جدول روبه‌رو، معادله خط رگرسیون کدام است؟ (حسابداری و مدیریت ـ ۸۸)

x	2	3	5	6
y	3	2	4	3

۱) $\hat{y} = 0.2x + 2.2$

۲) $\hat{y} = -0.2x + 3.8$

۳) $\hat{y} = 0.3x + 1.8$

۴) $\hat{y} = -0.3x + 4.2$

حل: گزینه ۱ درست است.

ازآنجاکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعداد صحیحی هستند،

$$\begin{cases} \bar{x} = \frac{\sum x_i}{n} = \frac{16}{4} = 4 \\ \bar{y} = \frac{\sum y_i}{n} = \frac{12}{4} = 3 \end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه مناسب‌تر است:

x	y	$(x-\bar{x})$	$(y-\bar{y})$
2	3	-2	0
3	2	-1	-1
5	4	1	1
6	3	2	0

$\longrightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
0	4
1	1
1	1
0	4
$\sum = 2$	$\sum = 10$

$$\longrightarrow \quad b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{2}{10}$$

حال با توجه به این نکته که خط رگرسیون همواره از نقطه $(\bar{x},\bar{y})$ می‌گذرد، داریم:

$$\bar{y} = a + b\bar{x} \quad \rightarrow \quad \text{عرض از مبدأ} : a = \bar{y} - b\bar{x} \rightarrow a = 3 - 0.2 \times 4 = 3 - 0.8 = 2.2$$

مثال ۲ اگر شیب معادله رگرسیون -10 باشد، $\sum x = 100$ و $\bar{x} = 20$ و $\sum y = 20$ باشد، ثابت معادله کدام است؟ (مدیریت ـ ۷۴)

۱) 106 ۲) 110 ۳) 220 ۴) 204

حل: گزینه ۴ درست است.

$$\begin{cases} \bar{y} = b\bar{x} + a \quad \rightarrow \quad a = \bar{y} - b\bar{x} = 4 - (-10)(20) = 204 \\ \bar{x} = \frac{\sum x}{n} \quad \rightarrow \quad 20 = \frac{100}{n} \quad \rightarrow \quad n = 5 \\ \bar{y} = \frac{\sum y}{n} = \frac{20}{5} = 4 \text{ ، شیب خط} : b = -10 \end{cases}$$

مثال ۳ اگر $SP_{xy} = 20$ و $SS_x = 20$ و $SS_y = 20$ و $\bar{x} = 5$ و $\bar{y} = 4$ باشد. معادله خط رگرسیون برابر است با:(مدیریت ـ ۷۹)

۱) $y = -x + 1$ ۲) $y = x - 1$ ۳) $y = x + 1$ ۴) $y = \frac{x}{2} + 1$

حل: گزینه ۲ درست است.

$$\begin{cases} b = \frac{SP_{xy}}{SS_x} = \frac{20}{20} = 1 \\ \bar{y} = b\bar{x} + a \rightarrow a = 4 - 1 \times 5 = -1 \\ y = bx + a \longrightarrow y = x - 1 \end{cases}$$

راه حل تستی: با توجه به این نکته که نقطه $(\bar{x},\bar{y})$ همیشه از معادله خط رگرسیون برآوردشده می‌گذرد، با قرار دادن مقدار $\bar{x}$ و $\bar{y}$ در معادلات رگرسیون گزینه‌ها، به راحتی گزینه درست را می‌یابیم؛ درواقع گزینه‌ای که $\bar{x}$ و $\bar{y}$ در معادله آن صدق کند، پاسخ سؤال است.

مثال ۴ با استفاده از اطلاعات زیر معادله رگرسیون کدام است؟ (حسابداری ـ ۸۱)

$$\sum y = 50\ ,\ \sum x = 75\ ,\ n = 25\ ,\ \sum y^2 = 228\ ,\ \sum xy = 30\ ,\ \sum x^2 = 625$$

۱) $y = 2.9 - 0.3\,x$ ۲) $y = 2.9 - 0.15\,x$ ۳) $y = 5.8 - 0.3\,x$ ۴) $y = 8.7 - 0.6\,x$

حل: گزینه ۱ درست است.

$$\begin{cases} b = \dfrac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{(\sum x)^2}{n}} = \dfrac{30 - \dfrac{75 \times 50}{25}}{625 - \dfrac{(75)^2}{25}} = \dfrac{-120}{400} = \dfrac{-3}{10} = -0.3 \\ \bar{y} = b\bar{x} + a \quad \rightarrow \quad a = \bar{y} - b\bar{x} = \dfrac{\sum y}{n} - b \times \dfrac{\sum x}{n} = \dfrac{50}{25} - (-0.3) \times \dfrac{75}{25} = 2.9 \\ \hat{y} = a + bx = 2.9 - 0.3x \end{cases}$$

مثال ۵ فرض کنید $Cov(X,Y) = 12$ و $n = 10$ و $\sum x = \sum y = 50$ و $\sigma_X = 4$ و $\sigma_Y = 3$ است. معادله رگرسیون y برحسب x کدام است؟ (مدیریت ـ ۷۵، ۷۷)

۱) $y = 1.5 - 0.3\,x$ ۲) $y = 1.25 + 0.75\,x$ ۳) $y = 1.5 + 0.4\,x$ ۴) $y = 3 + 2.2\,x$

حل: گزینه ۲ درست است.

$$\begin{cases} \text{شیب خط} : b = \dfrac{Cov(X,Y)}{\sigma_X^2} = \dfrac{12}{(4)^2} = \dfrac{12}{16} = \dfrac{3}{4} = 0.75 \\ \text{عرض از مبدأ} : a = \bar{y} - b\bar{x} = \dfrac{\sum y}{n} - b \times \dfrac{\sum x}{n} = \dfrac{50}{10} - \dfrac{3}{4} \times \dfrac{50}{10} = 1.25 \\ \text{معادله خط رگرسیون} : \hat{y} = a + bx = 1.25 + 0.75x \end{cases}$$

پیش‌بینی مقدار y

یکی از کاربردهای معادله خط رگرسیون $(y = ax + b)$، پیش‌بینی مقدار متغیر وابسته y به ازای هر مقدار از متغیر مستقل x است. برای رسیدن به این هدف، معادله خط رگرسیون را به دست می‌آوریم، سپس مقدار x را در آن قرار می‌دهیم تا مقدار پیش‌بینی‌شده y $(\hat{y})$ به دست آید.

مثال ۱ رابطه بین میانگین X و Y خطی است و داده‌های نمونه به شرح زیر در دست است:

$$\sum x_i y_i = 1150\ ,\ \bar{y} = 10\ ,\ n = 20\ ,\ \sum x_i^2 = 550\ ,\ \bar{X} = 5$$

مقدار پیش‌بینی شده به ازای $x = 6$ برابر است با: (مدیریت ـ ۷۳)

۱) 10 ۲) 11 ۳) 12 ۴) 13

حل: گزینه ۴ درست است.

ابتدا معادله خط رگرسیون را به دست می‌آوریم، سپس $x = 6$ را در آن قرار می‌دهیم تا مقدار پیش‌بینی شده y $(\hat{y})$ به دست آید.

$$\begin{cases} b = \dfrac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{(\sum x)^2}{n}} = \dfrac{1150 - \dfrac{100 \times 200}{20}}{550 - \dfrac{100^2}{20}} = \dfrac{150}{50} = 3 \\ a = \bar{y} - b\bar{x} = 10 - 3 \times 5 = -5 \\ \hat{y} = a + bx = -5 + 3x \xrightarrow{x=6} \hat{y} = -5 + 3 \times 6 = 13 \end{cases}$$

اگر $x = 6$ را در معادله خط $y = 3x - 5$ قرار دهیم، مقدار پیش‌بینی شده y $(\hat{y})$ برابر 13 به دست می‌آید.

مثال ۲ اگر $Cov(X,Y) = -10$ و $\sigma_X = \sigma_Y = 2$ و $\bar{X} = \bar{Y} = 10$ باشد، مقدار پیش‌بینی شده y به ازای $x = 4$، چقدر است؟ (مدیریت ـ ۷۶)

۱) 25 ۲) 128.75 ۳) 32.25 ۴) 40

حل: گزینه ۱ درست است.

ابتدا معادله خط رگرسیون را به دست می‌آوریم، سپس $x = 4$ را در آن قرار می‌دهیم تا مقدار پیش‌بینی شده y $(\hat{y})$ به دست آید.

$$\begin{cases} b = \dfrac{Cov(X,Y)}{\sigma_X^2} = \dfrac{-10}{4} = -2.5 \\ a = \bar{y} - b\bar{x} = 10 - (-2.5) \times 10 = 35 \\ \hat{y} = a + bx = 35 - 2.5x \xrightarrow{x=4} \hat{y} = 35 - 2.5 \times 4 = 25 \end{cases}$$

R^2 و خط رگرسیون

اگر خط رگرسیون y روی x و خط رگرسیون x روی y به صورت زیر مفروض باشد:

$$\begin{cases} x = b'y + a' & (\text{خط x روی y}) \\ b' = \dfrac{Cov(x,y)}{\sigma_y^2} \end{cases} \qquad \begin{cases} y = bx + a & (\text{خط y روی x}) \\ b = \dfrac{Cov(x,y)}{\sigma_x^2} \end{cases}$$

آن‌گاه حاصلضرب شیب خطوط (bb') برابر R^2 (ضریب تعیین) یا مجذور $r_{X,Y}$ (ضریب همبستگی) است.

$$bb' = \frac{Cov^2(X,Y)}{\sigma_X^2 \sigma_Y^2} = \left(\frac{Cov(X,Y)}{\sigma_X \sigma_Y}\right)^2 = (r_{X,Y})^2 = R_{X,Y}^2$$

مثال در یک جامعه نرمال دو بعدی معادلات رگرسیون y برحسب x و x برحسب y به صورت زیر به دست آمده است. ضریب تعیین (R^2) کدام است؟

$y = 4.85 - 3.2x$ ، $x = 8.32 - 0.28y$

۱) -3.2 ۲) -2.28 ۳) 0.896 ۴) -0.896

حل: گزینه ۳ درست است.

$$r^2 = bb' = (-3.2)(-0.28) = 0.896$$

تست‌های طبقه‌بندی شده

تابع احتمال گسسته

۱. به ازای کدام مقدار a، تابع $P(X=x)=\dfrac{\binom{4}{x}}{3a+1}\ ;\ x=0,1,2,3,4$، یک تابع احتمال است؟

(حسابداری و مدیریت ـ ۸۷)

۱) 6　　۲) 5　　۳) 4　　۴) 3

۲. در یک تاس ناسالم احتمال آمدن هر شماره متناسب با وارون عدد آن است. با کدام احتمال در پرتاب این تاس عدد زوج ظاهر می‌شود؟ (حسابداری و مدیریت ـ ۸۸)

۱) $\frac{29}{60}$　　۲) $\frac{31}{60}$　　۳) $\frac{45}{147}$　　۴) $\frac{55}{147}$

۳. در یک تاس ناسالم احتمال آمدن هر عدد متناسب با خود آن عدد است. در پرتاب این تاس احتمال ظاهر شدن عدد زوج کدام است؟ (GIS ـ ۸۶)

۱) $\frac{7}{12}$　　۲) $\frac{5}{12}$　　۳) $\frac{4}{7}$　　۴) $\frac{3}{7}$

۴. اگر $P(X=x)=\dfrac{1}{x^2+x},\ x\in N$ تابع احتمال متغیر تصادفی X باشد، $P(2\le X\le 19)$ کدام است؟ (GIS ـ ۸۶)

۱) 0.36　　۲) 0.45　　۳) 0.54　　۴) 0.63

۵. چنانچه $n=20$، $\sum x_i^2=85$ و $\sum x_i=30$ باشد، حاصل عبارت $\sum(x_i-\bar{x})^2$ کدام است؟

(محیط زیست ـ ۸۷)

۱) 40　　۲) 45　　۳) 70　　۴) 82

تابع چگالی پیوسته

۶. تابع چگالی احتمال متغیر تصادفی X به صورت زیر است. مقدار $P(2\le X\le 6)$ برابر است با: (اقتصاد ـ ۸۷)

$$\begin{cases} f(x)=\frac{1}{4} & ;\quad 1<x<5 \\ f(x)=0 & ;\quad \text{برای سایر مقادیر } x \end{cases}$$

۱) 0.25　　۲) 0.5　　۳) 0.75　　۴) 1

۷. مقدار m در تابع $f(x)=\frac{m}{\sqrt{x}}$ برای $0\le x\le 1$ چقدر باشد تا $f(x)$ یک تابع چگالی احتمال باشد؟ **(اقتصاد ــ ۸۸)**

۱) $\frac{1}{2}$ ۲) 1 ۳) $\frac{3}{2}$ ۴) 2

۸. متغیر تصادفی X دارای تابع چگالی به صورت $f(x)=\begin{cases} ae^{-2x}; & x>0 \\ 0 & ; x\le 0\end{cases}$ است. a کدام است؟ **(GIS ــ ۸۷)**

۱) $\frac{2}{3}$ ۲) $\frac{1}{2}$ ۳) 1 ۴) 2

۹. در عبارت $f(x)=\begin{cases} a\left(\frac{1+x}{2}\right) & 1<x<4 \\ 0 & \text{سایر نقاط}\end{cases}$، مقدار a چقدر باشد تا $f(x)$ یک چگالی احتمال باشد؟

(محیط زیست ــ ۸۶)

۱) $\frac{21}{4}$ ۲) $\frac{21}{17}$ ۳) $\frac{17}{21}$ ۴) $\frac{4}{21}$

۱۰. تابع زیر مفروض است:

$$f(x)=\begin{cases} a+\frac{1}{4}x & -2\le x<0 \\ a-\frac{1}{4}x & 0\le x<2\end{cases}$$

مقدار a را چنان بیابید که $f(x)$ یک چگالی احتمال باشد؟ **(محیط زیست ــ ۸۸)**

۱) -2 ۲) $-\frac{1}{2}$ ۳) $\frac{1}{2}$ ۴) 2

تابع توزیع تجمعی

۱۱. تابع توزیع کمیت تصادفی پیوسته X (طول زمان کار دستگاه تا وقتی که از کار بیفتد) به قرار ذیل می‌باشد.
احتمال اینکه دستگاه در طول زمان $X\ge T$ از کار بیفتد چقدر است؟ **(اقتصاد ــ ۸۷)**

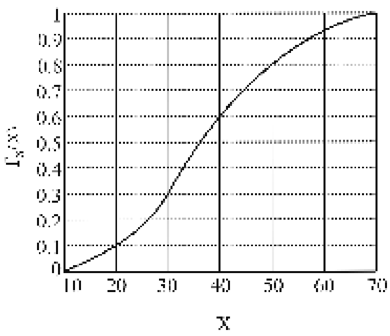
$$F(x)=1-\exp\left(-\frac{x}{T}\right) \qquad 0<x\le\infty$$

۱) e ۲) e^{-1} ۳) $1-e^{-1}$ ۴) $\frac{e^{-1}}{T}$

۱۲. تابع توزیع احتمال متغیر تصادفی X به صورت $F(x)=\begin{cases} 0 & x\le 2 \\ \frac{x^2-4}{96} & 2<x\le 10 \\ 1 & 10<x\end{cases}$ داده شده است. میانه متغیر تصادفی X کدام است؟ **(اقتصاد ــ ۸۸)**

۱) $\sqrt{52}$ ۲) $\sqrt{96}$ ۳) $\sqrt{48}$ ۴) $\sqrt{6}$

۱۳. تابع تجمعی توزیع متغیر پیوسته تصادفی X در شکل زیر نشان داده شده است. احتمال اینکه X دارای مقداری بیش از 30 باشد $P(X>30)$، حدوداً برابر است با : (محیط زیست ــ ۸۷)



۱) 0.3
۲) 0.7
۳) 0.9
۴) 1

امید ریاضی گسسته

۱۴. اگر $f(x)=\frac{2x-1}{25}\ ,\ x=1,2,3,4,5$ باشد، مقدار $E(X)$ کدام است؟ (GIS ــ ۸۶)

۱) 3.5　۲) 3.6　۳) 3.7　۴) 3.8

۱۵. در جدول مقابل، a چه مقدار باشد تا $E(X)=4$ باشد؟ (محیط زیست ــ ۸۸)

X	3	a	10	14
$P(X=x)$	$\frac{1}{10}$	$\frac{3}{10}$	b	$\frac{4}{10}$

۱) 6　۲) 8　۳) 13　۴) -13

امید ریاضی پیوسته

۱۶. امید ریاضی متغیر تصادفی X با تابع چگالی احتمال زیر چقدر است؟ (اقتصاد ــ ۸۶)

$$f(x)=\begin{cases}\dfrac{1}{2\sqrt{x}} & ; \quad 0\le x\le 1\\ 0 & ; \quad \text{برای سایر مقادیر } x\end{cases}$$

۱) $\frac{1}{2}$　۲) $\frac{4}{3}$　۳) $\frac{2}{3}$　۴) $\frac{1}{3}$

۱۷. امید ریاضی متغیر تصادفی X با تابع چگالی $f(x)=\begin{cases}\frac{1}{4}x & 1<x<3\\ 0 & \text{جای دیگر}\end{cases}$ کدام است؟ (برنامه‌ریزی شهری ــ ۸۷)

۱) 1.98　۲) 2　۳) 2.16　۴) 2.24

۱۸. به ازای مقداری از k تابع $f(x)=\begin{cases}kx^2 & ; \quad 0<x<4\\ 0 & \text{جای دیگر}\end{cases}$، یک تابع چگالی احتمال است. امید ریاضی X کدام است؟ (GIS ــ ۸۸)

۱) $\frac{3}{4}$　۲) $\frac{4}{3}$　۳) 2　۴) 3

خواص امید ریاضی

۱۹.در تابع

X	-1	0	1	2
$f(x)$	0.1	0.15	0.5	0.25

امید ریاضی $(X-1)^2$ کدام است ؟ (GIS ــ ۸۷)

۱) 0.65 ۲) 0.7 ۳) 0.75 ۴) 0.8

واریانس

۲۰. در جدول توزیع احتمال زیر، $Var\left(\frac{1}{2}X-2\right)$ کدام است؟ (حسابداری و مدیریت ــ ۸۸)

X	-2	0	2	4	5
$f(x)$	0.1	0.15	0.3	0.25	0.2

۱) 1.21 ۲) 1.25 ۳) 1.35 ۴) 1.41

۲۱.دو جعبه هر یک شامل 3 مهره است که از یک تا سه شماره‌گذاری شده‌اند. یک مهره به تصادف از هر جعبه انتخاب شده است. اگر متغیر تصادفی X نشان‌دهنده اختلاف بین اعداد دو مهره باشد، میانگین و واریانس X به ترتیب کدام است؟ (اقتصاد ــ ۸۶)

۱) $0, \frac{4}{81}$ ۲) $\frac{12}{9}, \frac{8}{9}$ ۳) $0, \frac{4}{9}$ ۴) $\frac{44}{81}, \frac{8}{9}$

۲۲. تابع احتمال متغیر تصادفی X به صورت زیر داده شده است. واریانس X کدام است؟ (برنامه‌ریزی شهری ــ ۸۷)

X	-1	1	3	5
$F(x)$	0.2	0.3	α	0.1

۱) 1.26 ۲) 2.46 ۳) 3.12 ۴) 3.36

خواص واریانس

۲۳.اگر $V(-2X+1)=5$ و $E(X)=1.5$ باشند، $E(X-2)^2$ کدام است؟ (برنامه‌ریزی شهری ــ ۸۸)

۱) 1.25 ۲) 1.5 ۳) 1.75 ۴) 2.25

۲۴.اگر $E(X)=4.5$ و $E(X^2)=24$ باشد، واریانس $(-2X+1)$ کدام است؟ (GIS ــ ۸۷)

۱) 15 ۲) 13.5 ۳) 12 ۴) 10.5

تابع چگالی توزیع توأم

۲۵.جدول زیر مفروض است. اگر $R=X^2+Y^2$ باشد، $\sqrt{E(R)}$ چقدر است؟ (محیط زیست ــ ۸۶)

Y \ X	0	1
0	$\frac{1}{10}$	$\frac{2}{10}$
1	$\frac{2}{10}$	$\frac{2}{10}$
2	$\frac{3}{10}$	0

۱) $\frac{\sqrt{2}}{2}$

۲) $\sqrt{2}$

۳) 2

۴) 4

محاسبه کوواریانس

۲۶. در توزیع احتمال توأم روبه‌رو، $Cov(X,Y)$ کدام است؟ **(حسابداری و مدیریت ـ ۸۶)**

Y \ X	0	1	2
1	0	0.1	0.2
3	0.3	0.4	0

۱) -0.56
۲) -0.46
۳) صفر
۴) 0.64

۲۷. در تابع احتمال توأم دو متغیر تصادفی X و Y، مقدار کوواریانس کدام است؟ **(حسابداری و مدیریت ـ ۸۷)**

X \ Y	1	4
0	$\frac{1}{6}$	$\frac{1}{3}$
5	$\frac{1}{3}$	$\frac{1}{6}$

۱) $-\frac{5}{4}$
۲) $\frac{5}{4}$
۳) $\frac{5}{3}$
۴) $-\frac{5}{3}$

۲۸. توزیع احتمال مشترک دو متغیر تصادفی X و Y به صورت زیر است. کوواریانس $(2X,3Y)$ کدام است؟
(اقتصاد ـ ۸۶)

X \ Y	0	1	2
0	0.2	0.3	0.1
1	0.1	0.2	0.1

۱) 0.04
۲) 0.12
۳) 0.24
۴) 0.36

۲۹. فرض کنید متغیر تصادفی X در فاصله $(-1,1)$ دارای تابع چگالی $f(x)=\frac{1}{2}$ باشد. اگر $Y=X^2$ باشد، کوواریانس X و Y برابر است با: **(اقتصاد ـ ۸۶)**

۱) صفر ۲) $\frac{1}{12}$ ۳) $\frac{1}{4}$ ۴) 1

۳۰. در توزیع توأم جدول مقابل، کوواریانس کدام است؟ **(برنامه‌ریزی شهری ـ ۸۸)**

Y \ X	−2	0	3
1	0.1	0.2	0.25
2	0.15	0.3	0

۱) -0.4125
۲) -0.0125
۳) -0.0145
۴) 0.2425

۳۱. از جدول تابع احتمال توأم دو متغیر X و Y مقدار $Cov(X,Y)$ کدام است؟ **(GIS ـ ۸۶)**

Y \ X	0	1	2
0	0	0.25	0.15
1	0.20	0.35	0.05

۱) -0.25
۲) -0.15
۳) 0.15
۴) 0.25

خواص کوواریانس

۳۲.اگر $Var(X)=3$ و $Var(Y)=4$ و $Cov(X,Y)=-1$ باشند، آن‌گاه واریانس $Z=2X-Y+1$ چقدر است؟
(اقتصاد ـ ۸۸)

۱) 12 ۲) 18 ۳) 20 ۴) 21

۳۳. مقدار $Cov(X-Y, X+Y)$ کدام است؟ **(محیط زیست ـ ۸۷)**

۱) $Var(X)-Var(Y)$ ۲) $Var(X+Y)-Var(X-Y)$
۳) $Var(X-Y)-Var(X+Y)$ ۴) $Var(X)+Var(Y)$

استقلال متغیرهای تصادفی

۳۴.اگر $V(X+Y)=V(X-Y)$ باشد، کدام بیان برای رابطه بین X و Y صحیح است؟ **(GIS ـ ۸۷)**

۱) استقلال ۲) رابطه غیر خطی
۳) رابطه غیر خطی یا استقلال ۴) رابطه‌ای وجود ندارد.

۳۵. جدول توزیع دومتغیری زیر مفروض است. در مورد وضعیت متغیرهای X و Y چه می‌توان گفت؟
(محیط زیست ـ ۸۷)

Y \ X	1	2	3
0	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{7}{20}$
1	0	$\frac{10}{20}$	0

۱) دو متغیر تصادفی X و Y به یکدیگر وابسته می‌باشند و همبستگی X و Y منفی است.
۲) دو متغیر تصادفی X و Y به یکدیگر وابسته می‌باشند و همبستگی X و Y مثبت است.
۳) دو متغیر تصادفی X و Y نسبت به یکدیگر مستقل هستند و همبستگی بین X و Y وجود ندارد.
۴) دو متغیر تصادفی X و Y مستقل از یکدیگرند اما همبستگی بین آن‌ها وجود دارد.

ضریب همبستگی

۳۶.ضریب همبستگی بین دو متغیر در جدول روبه‌رو، کدام است؟ **(حسابداری و مدیریت ـ ۸۶)**

x	2	3	4	5	6
y	3	5	1	4	2

۱) 0.3 ۲) -0.3
۳) -0.2 ۴) 0.2

۳۷. ضریب همبستگی بین دو صفت X و Y کدام است؟ **(GIS ـ ۸۸)**

x	2	5	8	11	14
y	12	10	8	6	4

۱) -1 ۲) -0.9
۳) -0.8 ۴) 0.2

۳۸. فرض کنید $Var(X)=8$ و $Var(Y)=2$ و ضریب همبستگی $\rho(X,Y)=\frac{1}{2}$ باشند، مقدار $Var(X-2Y)$ برابر است با: **(اقتصاد ـ ۸۷)**

۱) 2 ۲) 4 ۳) 8 ۴) 14

۳۹.فرض کنید مقدار واریانس X و Y به ترتیب 4 و 5 و ضریب همبستگی این دو متغیر نیز 0.8 باشد، در این صورت مقدار واریانس $Z = 2X + 4Y$ برابر است با: **(محیط زیست ـ ۸۶)**

۱) $96 + 12.8\sqrt{20}$ ۲) $96 + \sqrt{20}$ ۳) 96 ۴) 108.8

۴۰. دو متغیر تصادفی X و Y دارای ضریب همبستگی $\rho_{XY} = 0.25$ و $\mu_X = 100$، $\mu_Y = 80$، $\sigma_X = 10$ و $\sigma_Y = 8$ می‌باشد. واریانس $D = 2X - Y$ چقدر است؟ **(محیط زیست ـ ۸۷)**

۱) 384 ۲) 424 ۳) 464 ۴) 474

۴۱. هرگاه ضریب همبستگی بین دو متغیر -1 باشد، به این معنی است که ..... **(محیط زیست ـ ۸۷)**

۱) آن دو متغیر ارتباط خطی کامل دارند و در جهت عکس یکدیگر تغییر می‌کنند.
۲) آن دو متغیر ارتباط خطی نسبی دارند و در جهت عکس یکدیگر تغییر می‌کنند.
۳) آن دو متغیر ارتباط خطی کامل دارند و در یک جهت حرکت می‌کنند.
۴) هیچ ارتباط آماری بین این دو متغیر وجود ندارد.

رگرسیون

۴۲.برای دو صفت X و Y در جدول روبه‌رو، معادله خط رگرسیون کدام است؟ **(حسابداری و مدیریت ـ ۸۸)**

x	2	3	5	6
y	3	2	4	3

۱) $\hat{y} = 0.2x + 2.2$
۲) $\hat{y} = -0.2x + 3.8$
۳) $\hat{y} = 0.3x + 1.8$
۴) $\hat{y} = -0.3x + 4.2$

۴۳.معادله خط رگرسیون دو متغیر زیر کدام است؟ **(برنامه‌ریزی شهری ـ ۸۸)**

x	1	2	4	5
y	2	4	10	12

۱) $y = 2.6x - 0.8$
۲) $y = 2.4x - 0.2$
۳) $y = 2.3x + 0.1$
۴) $y = 2.2x - 0.4$

۴۴.ضریب همبستگی بین X و Y در یک نمونه تصادفی به حجم n برابر $r = 0.95$ می‌باشد. با فرض $S_x^2 = S_y^2$، شیب خط رگرسیون y بر روی x عبارت است از: **(اقتصاد ـ ۸۷)**

۱) 0.95 ۲) 0.90 ۳) 0.10 ۴) 0.05

۴۵.برای خط رگرسیون ساده مقیدی که از مبدأ مختصات می‌گذرد، اطلاعات $\sum xy = 480$، $\sum y = 16$ و $\sum x = 20$ بر اساس یک نمونه تصادفی 10تایی داده شده است. شیب خط رگرسیون کدام است؟ **(اقتصاد ـ ۸۸)**

۱) 2.75 ۲) 1.25
۳) 0.8 ۴) اطلاعات داده‌شده کافی نیست.

۴۶.فرض کنید برآورد خط رگرسیون y برحسب x برابر است با $y = 2x + 4$. اگر $x^* = 4x$ و $y^* = 2y$ باشد، در آن صورت ضریب همبستگی بین دو متغیر x^* و y^* کدام است؟ $\left(S_x^2 = 4, S_y^2 = 16\right)$ **(محیط زیست ـ ۸۸)**

۱) 1 ۲) $\frac{1}{2}$ ۳) 0 ۴) -1

پاسخ‌های تشریحی

تابع احتمال گسسته

۱. گزینه ۲ درست است.

مجموع احتمال‌ها در تابع احتمال گسسته برابر یک است، یعنی:

$$\sum P(x)=1 \rightarrow \sum_{x=0}^{4}\frac{\binom{4}{x}}{3a+1}=1 \rightarrow \frac{2^4}{3a+1}=1 \rightarrow 16=3a+1 \rightarrow 3a=15 \rightarrow a=5$$

یادآوری:

$$\sum_{x=0}^{n}\binom{n}{x}=2^n \rightarrow \sum_{x=0}^{4}\binom{4}{x}=2^4=16$$

۲. گزینه ۴ درست است.

X	1	2	3	4	5	6	
$P(x)$	$\frac{1}{1}x$	$\frac{1}{2}x$	$\frac{1}{3}x$	$\frac{1}{4}x$	$\frac{1}{5}x$	$\frac{1}{6}x$	$\sum P(x)=1$

$$\sum P(x)=1 \rightarrow x+\frac{1}{2}x+\frac{1}{3}x+\frac{1}{4}x+\frac{1}{5}x+\frac{1}{6}x=1 \rightarrow \frac{(60+30+20+15+12+10)x}{60}=1 \rightarrow x=\frac{60}{147}$$

$$P(\text{زوج } X)=P(X=2)+P(X=4)+P(X=6)=\frac{1}{2}x+\frac{1}{4}x+\frac{1}{6}x=\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}\right)x$$

$$=\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}\right)\times\frac{60}{147}=\frac{6+3+2}{12}\times\frac{60}{147}=\frac{55}{147}$$

توجه: چون احتمال هر شماره متناسب با وارون آن شماره است، داریم:

$$\text{احتمال هر عدد تاس}=\frac{1}{\text{عدد}}\times x$$

که ضریب ثابت x، تناسب را ایجاد می‌کند.

۳. گزینه ۳ درست است.

ابتدا جدول توزیع احتمال پرتاب تاس ناسالم را رسم می‌کنیم.

x	1	2	3	4	5	6	
$P(x)$	x	$2x$	$3x$	$4x$	$5x$	$6x$	$\sum P(x)=1$

$$\sum P(x)=1 \rightarrow x+2x+3x+4x+5x+6x=1 \rightarrow x=\frac{1}{21}$$

x	1	2	3	4	5	6
$P(x)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

$$P(\text{زوج } X)=P(X=2)+P(X=4)+P(X=6)=\frac{2}{21}+\frac{4}{21}+\frac{6}{21}=\frac{12}{21}=\frac{4}{7}$$

۴. گزینه ۲ درست است.

ابتدا تابع احتمال را کمی ساده‌تر می‌کنیم و سپس به محاسبه احتمال‌ها در نقطه می‌پردازیم.

$$P(X=x)=\frac{1}{x^2+x}=\frac{1}{x(x+1)}=\frac{1}{x}-\frac{1}{x+1}$$

با توجه به اینکه تابع احتمال متغیر گسسته است، داریم:

$$P(2\le X\le 19)=P(X=2)+P(X=3)+\dots+P(X=18)+P(X=19)$$

$$=\frac{1}{2}-\cancel{\frac{1}{3}}+\cancel{\frac{1}{3}}-\frac{1}{4}+\dots+\frac{1}{18}-\cancel{\frac{1}{19}}+\cancel{\frac{1}{19}}-\frac{1}{20}=\frac{1}{2}-\frac{1}{20}=\frac{9}{20}=0.45$$

۵. گزینه ۱ درست است.

$$\sum(x_i-\bar{x})^2=\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}=85-\frac{(30)^2}{20}=85-45=40$$

اثبات:

$$\sum(x_i-\bar{x})^2=\sum\left(x_i^2-2x_i\bar{x}+\bar{x}^2\right)=\sum x_i^2-2\bar{x}\underbrace{\sum x_i}_{n\bar{x}}+\sum\bar{x}^2=\sum x_i^2-2n\bar{x}^2+n\bar{x}^2=$$

$$=\sum x_i^2-n\bar{x}^2=\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}$$

یادآوری: قوانین $\sum$

$$\sum(ax+by)=a\sum x+b\sum y$$

$$\sum_{i=1}^{n}a=na$$

توجه: $\bar{x}$ یک عدد ثابت است و از $\sum$ بیرون می‌آید. همچنین می‌دانیم:

$$\bar{x}=\frac{\sum x_i}{n}\rightarrow\sum x_i=n\bar{x}$$

تابع چگالی پیوسته

۶. گزینه ۳ درست است.

یادآوری: احتمال در فاصله برابر انتگرال در فاصله است. $P(a<X<b)=\int_a^b f(x)dx$

دقت کنید حدود x در فاصله $(1,5)$ است و احتمال خواسته شده در فاصله $(2,6)$ است بنابراین باید فاصله را به $(2,5)$ تغییر دهیم.

$$P(2<X<6)=P(2<X<5)=\int_2^5 \frac{1}{4}dx=\left[\frac{x}{4}\right]_2^5=\frac{5-2}{4}=\frac{3}{4}=0.75$$

۷. گزینه ۱ درست است.

بنا بر قانون اول تابع چگالی احتمال پیوسته داریم:

$$\int_{\text{حد پایین}}^{\text{حد بالا}} f(x)dx=1$$

$$\int_0^1 f(x)dx=1\to\int_0^1\frac{m}{\sqrt{x}}dx=1\to\int_0^1 mx^{-\frac{1}{2}}dx=1\to m\left[2x^{\frac{1}{2}}\right]_0^1=1\to 2m=1\to m=\frac{1}{2}$$

۸. گزینه ۴ درست است.

یادآوری: مقدار انتگرال روی کل بازه تابع چگالی پیوسته برابر یک است. $\int_{-\infty}^{+\infty} f(x)dx=1$

$$\int_0^\infty ae^{-2x}dx=1\ \to\ a\left[-\frac{1}{2}e^{-2x}\right]_0^\infty=1\ \to\ \frac{1}{2}a=1\ \to\ a=2$$

راه حل تستی: در فصل بعد خواهیم دید که تابع چگالی دادهشده، تابع چگالی توزیع نمایی با $\lambda=2$ است $\left(f(x)=\lambda e^{-\lambda x}\ ;\ x>0\right)$، بنابراین مقدار $a=\lambda=2$ است.

۹. گزینه ۴ درست است.

بنا بر قانون اول تابع چگالی احتمال پیوسته داریم: $\int_{-\infty}^{+\infty} f(x)dx=1$

$$\int_1^4 a\left(\frac{1+x}{2}\right)dx=1\ \to\frac{a}{2}\left[x+\frac{x^2}{2}\right]_1^4=1\to\frac{a}{2}\left[4+\frac{16}{2}-1-\frac{1}{2}\right]=1\to\frac{a}{2}\times\frac{21}{2}=1\to a=\frac{4}{21}$$

۱۰. گزینه ۳ درست است.

یادآوری: مقدار انتگرال روی کل بازه تابع چگالی پیوسته برابر یک است. $\int_{-\infty}^{+\infty} f(x)dx=1$

بنابراین داریم:

$$\int_{-2}^0\left(a+\frac{1}{4}x\right)dx+\int_0^2\left(a-\frac{1}{4}x\right)dx=1\ \to\ \left[ax+\frac{1}{8}x^2\right]_{-2}^0+\left[ax-\frac{x^2}{8}\right]_0^2=1$$

$$\left[0-\left(-2a+\frac{(-2)^2}{8}\right)\right]+\left[2a-\frac{2^2}{8}-0\right]=1\ \to\ 2a-\frac{1}{2}+2a-\frac{1}{2}=1\ \to\ 4a=2\ \to\ a=\frac{1}{2}$$

تابع توزیع تجمعی

۱۱. گزینه ۲ درست است.

یادآوری: تابع توزیع تجمعی X عبارت است از:

$$\begin{cases} F_X(x) = P(X \le x) = \int_{\text{حد پایین}}^{x} f(x)dx \\ F'_X(x) = f(x) \end{cases}$$

راه حل اول:

$$P(X \ge T) = 1 - P(X < T) = 1 - F(T) = 1 - \left(1 - e^{-\frac{T}{T}}\right) = e^{-1}$$

راه حل دوم:

$$f(x) = F'(x) = \frac{1}{T}e^{-\frac{x}{T}} \quad ; \quad x > 0$$

احتمال در فاصله برابر انتگرال در فاصله است، پس داریم:

$$P(X \ge T) = \int_T^{\infty} \frac{1}{T}e^{-\frac{x}{T}}dx = \left[-e^{-\frac{x}{T}}\right]_T^{\infty} = \left(0 - \left(-e^{-\frac{T}{T}}\right)\right) = e^{-1}$$

۱۲. گزینه ۱ درست است.

یادآوری:

$$F(m) = \frac{1}{2} \quad ; \quad m : \text{میانه}$$

$$F(m) = \frac{m^2 - 4}{96} = \frac{1}{2} \rightarrow m^2 - 4 = \frac{96}{2} \rightarrow m^2 = 48 + 4 = 52 \rightarrow m = +\sqrt{52}$$

۱۳. گزینه ۲ درست است.

تابع توزیع تجمعی را می‌توان به صورت جدول توزیع احتمال نشان داد:

حدود دسته	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$	$60-70$
فراوانی تجمعی	کمتر از 0.1	0.3	0.6	0.8	بیشتر از 0.9	1

$$P(X > 30) = 1 - P(X \le 30) = 1 - F(30) = 1 - 0.3 = 0.7$$

امید ریاضی گسسته

۱۴. گزینه ۴ درست است.

با توجه به اینکه X متغیر تصادفی گسسته است، داریم:

راه حل اول:

$$E(X) = \sum xP(x) = \sum x \cdot \frac{2x-1}{25} = \frac{1}{25}\sum\left(2x^2 - x\right) = \frac{1}{25}\left(2\sum x^2 - \sum x\right)$$

$$=\frac{1}{25}\left(2\cdot\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)}{2}\right)\overset{n=5}{=\!=}\frac{1}{25}\left(\frac{5\times6\times11}{3}-\frac{5\times6}{2}\right)=\frac{95}{25}=3.8$$

راه حل دوم:

X_i	1	2	3	4	5	
$P(x_i)=\frac{2x_i-1}{25}$	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{5}{25}$	$\frac{7}{25}$	$\frac{9}{25}$	$\sum P(x_i)=1$

$$E(X)=\sum x.P(x)=1\times\frac{1}{25}+2\times\frac{3}{25}+3\times\frac{5}{25}+4\times\frac{7}{25}+5\times\frac{9}{25}=\frac{95}{25}=3.8$$

۱۵. گزینه ۴ درست است.

X	3	a	10	14	
$P(x)$	$\frac{1}{10}$	$\frac{3}{10}$	b	$\frac{4}{10}$	$\sum P(x)=1$

$$\sum P(x)=1 \rightarrow \frac{1}{10}+\frac{3}{10}+b+\frac{4}{10}=1 \rightarrow b=\frac{2}{10}$$

$$E(X)=4 \rightarrow \sum xP(x)=4 \rightarrow 3\times\frac{1}{10}+a\times\frac{3}{10}+10\times\frac{2}{10}+14\times\frac{4}{10}=4$$

$$\rightarrow \frac{3}{10}a=4-\frac{79}{10} \rightarrow \frac{3}{10}a=-\frac{39}{10} \rightarrow a=\frac{-39}{3}=-13$$

امید ریاضی پیوسته

۱۶. گزینه ۴ درست است.

$$E(X)=\int_0^1 xf(x)dx=\int_0^1 x.\frac{1}{2\sqrt{x}}dx=\int_0^1\frac{1}{2}x^{\frac{1}{2}}dx=\left[\frac{1}{2}\times\frac{2}{3}x^{\frac{3}{2}}\right]_0^1=\frac{1}{3}$$

۱۷. گزینه ۳ درست است.

$$E(X)=\int_{-\infty}^{+\infty}x\cdot f(x)dx=\int_1^3 x\cdot\frac{1}{4}x\,dx=\left[\frac{1}{12}x^3\right]_1^3=\frac{27-1}{12}=\frac{26}{12}=2.16$$

۱۸. گزینه ۴ درست است.

با توجه به قانون اول تابع چگالی احتمال متغیر تصادفی پیوسته داریم:

$$\int_{-\infty}^{+\infty}f(x)dx=1$$

$$\int_0^4 kx^2dx=1 \rightarrow k\left[\frac{1}{3}x^3\right]_0^4=1 \rightarrow \frac{64}{3}k=1 \rightarrow k=\frac{3}{64}$$

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^4 x.\frac{3}{64}x^2dx=\frac{3}{64}\left[\frac{1}{4}x^4\right]_0^4=\frac{3}{64}\times\frac{4^4}{4}=3$$

خواص امیدریاضی

۱۹.گزینه ۴ درست است.

راه حل اول:

X	−1	0	1	2	
f(x)	0.1	0.15	0.5	0.25	$\sum P(x)=1$

$$\begin{cases} E(X-1)^2 = E\left(X^2-2X+1\right) = E\left(X^2\right)-2E(X)+1 = 1.6-2(0.9)+1 = 0.8 \\ E\left(X^2\right) = \sum x^2 P(x) = (-1)^2\times 0.1+0^2\times 0.15+1^2\times 0.5+2^2\times 0.25 = 1.6 \\ E(X) = \sum x\,P(x) = (-1)\times 0.1+0\times 0.15+1\times 0.5+2\times 0.25 = 0.9 \end{cases}$$

راه حل دوم:

X−1	−2	−1	0	1	
f(x)	0.1	0.15	0.5	0.25	$\sum P(x)=1$

$$E(X-1)^2 = \sum (x-1)^2 f(x) = (-2)^2\times 0.1+(-1)^2\times 0.15+0^2\times 0.5+1^2\times 0.25 = 0.8$$

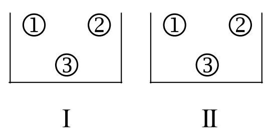
واریانس

۲۰. گزینه ۱ درست است.

X	−2	0	2	4	5	
f(x)	0.1	0.15	0.3	0.25	0.2	$\sum P(x)=1$

$$\begin{cases} Var\left(\frac{1}{2}X-\underset{0}{\cancel{2}}\right) = \left(\frac{1}{2}\right)^2 Var(X) = \frac{1}{4}Var(X) = \frac{1}{4}\times 4.84 = 1.21 \\ Var(X) = E\left(X^2\right)-E(X)^2 = 10.6-(2.4)^2 = 10.6-5.76 = 4.84 \\ E\left(X^2\right) = \sum x^2 P(x) = (-2)^2\times 0.1+0^2\times 0.15+2^2\times 0.3+4^2\times 0.25+5^2\times 0.2 = 0.4+0+1.2+4+5 = 10.6 \\ E(X) = \sum x\,P(x) = (-2)\times 0.1+0\times 0.15+2\times 0.3+4\times 0.25+5\times 0.2 = -0.2+0+0.6+1+1 = 2.4 \end{cases}$$

۲۱.گزینه ۴ درست است.



$$X: \text{اختلاف شماره دو مهره} \begin{cases} 0 & (1,1),(2,2),(3,3) \\ 1 & (2,1),(3,2),(1,2),(2,3) \\ 2 & (3,1),(1,3) \end{cases}$$

دقت کنید، $P(x)$ برای هر مقدار x از تقسیم تعداد حالات مساعد به کل حالات به دست می‌آید، یعنی $P(X=a)=\frac{n(a)}{n(S)}$.
مثلاً برای $P(X=0)$ سه حالت مساعد وجود دارد $(n(a)=3)$ و کل حالات نیز برابر 9 است $(n(S)=9)$، بنابراین احتمال آن $P(X=0)=\frac{3}{9}$ خواهد شد.

X	0	1	2	
$P(x)$	$\frac{3}{9}$	$\frac{4}{9}$	$\frac{2}{9}$	1

$$\rightarrow \begin{cases} \sigma_X^2 = E\left(X^2\right) - \left(E(X)\right)^2 = \frac{12}{9} - \left(\frac{8}{9}\right)^2 = \frac{12}{9} - \frac{64}{81} = \frac{44}{81} \\ E(X) = \sum x\,P(x) = 0 \times \frac{3}{9} + 1 \times \frac{4}{9} + 2 \times \frac{2}{9} = \frac{8}{9} \\ E\left(X^2\right) = \sum x^2 P(x) = 0^2 \times \frac{3}{9} + 1^2 \times \frac{4}{9} + 2^2 \times \frac{2}{9} = \frac{12}{9} \end{cases}$$

۲۲.گزینه ۴ درست است.

X	-1	1	3	5	
$P(x)=f(x)$	0.2	0.3	$\boxed{\alpha = 0.4}$	0.1	$\sum P(x)=1$

دقت کنید که به اشتباه در صورت سؤال به جای f از F استفاده شده است و می‌دانیم که در تابع احتمال گسسته $P(x)=f(x)$ و مجموع احتمال‌ها برابر یک است، بنابراین α برابر است با:

$$\sum P(x) = \sum f(x) = 1 \;\rightarrow\; 0.2+0.3+\alpha+0.1=1 \;\rightarrow\; \alpha = 0.4$$

$$\begin{cases} Var(X) = E\left(X^2\right) - E(X)^2 = 6.6 - (1.8)^2 = 3.36 \\ E\left(X^2\right) = \sum x^2\,P(x) = (-1)^2 \times 0.2 + 1^2 \times 0.3 + 3^2 \times 0.4 + 5^2 \times 0.1 = 6.6 \\ E(X) = \sum x\,P(x) = (-1) \times 0.2 + 1 \times 0.3 + 3 \times 0.4 + 5 \times 0.1 = 1.8 \end{cases}$$

خواص واریانس

۲۳.گزینه ۲ درست است.

$$\begin{cases} E(X-2)^2 = E\left(X^2 - 4X + 4\right) = E\left(X^2\right) - 4E(X) + 4 = 3.5 - 4 \times 1.5 + 4 = 1.5 \\ Var\left(-2X \overset{0}{\cancel{+1}}\right) = 5 \;\rightarrow\; (-2)^2\,Var(X) = 5 \;\rightarrow\; Var(X) = \frac{5}{4} \\ Var(X) = E\left(X^2\right) - \left(E(X)\right)^2 \;\rightarrow\; \frac{5}{4} = E\left(X^2\right) - (1.5)^2 \;\rightarrow\; E\left(X^2\right) = 1.25 + 2.25 = 3.5 \end{cases}$$

۲۴.گزینه ۱ درست است.

$$\begin{cases} Var\left(aX \underset{0}{\cancel{+b}}\right) = a^2\sigma_X^2 \rightarrow Var\left(-2X \underset{0}{\cancel{+1}}\right) = (-2)^2\,\sigma_X^2 = 4\sigma_X^2 = 4 \times 3.75 = 15 \\ \sigma_X^2 = E\left(X^2\right) - E(X)^2 = 24 - (4.5)^2 = 3.75 \end{cases}$$

تابع چگالی توزیع توأم

۲۵.گزینه ۲ درست است.

Y \ X	0	1	P(y)
0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$
1	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{4}{10}$
2	$\frac{3}{10}$	0	$\frac{3}{10}$
P(x)	$\frac{6}{10}$	$\frac{4}{10}$	1

$$\Rightarrow \begin{cases} E(X^2)=\sum x^2 P(x)=0^2\times\frac{6}{10}+1^2\times\frac{4}{10}=\frac{4}{10} \\ E(Y^2)=\sum y^2 P(y)=0^2\times\frac{3}{10}+1^2\times\frac{4}{10}+2^2\times\frac{3}{10}=\frac{16}{10} \end{cases}$$

$$E(R)=E(X^2+Y^2)=E(X^2)+E(Y^2)=\frac{4}{10}+\frac{16}{10}=\frac{20}{10}=2 \rightarrow \sqrt{E(R)}=\sqrt{2}$$

محاسبه کوواریانس

۲۶.گزینه ۲ درست است.

هرگاه در جدول توزیع احتمال توأم مقدار احتمال صفر وجود داشته باشد، کوواریانس X و Y صفر نیست.

Y \ X	0	1	2	f(y)
1	0	0.1	0.2	0.3
3	0.3	0.4	0	0.7
f(x)	0.3	0.5	0.2	1

$$\Rightarrow \begin{cases} E(XY)=\sum xyf(x,y)=0\times1\times0+0\times3\times0.3 \\ \quad +1\times1\times0.1+1\times3\times0.4+2\times1\times0.2+2\times3\times0=1.7 \\ E(X)=\sum xf(x)=0\times0.3+1\times0.5+2\times0.2=0.9 \\ E(Y)=\sum yf(y)=1\times0.3+3\times0.7=2.4 \end{cases}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=1.7-0.9\times2.4=-0.46$$

۲۷.گزینه ۱ درست است.

X \ Y	1	4	P(x)
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
5	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$
P(y)	$\frac{1}{2}$	$\frac{1}{2}$	1

$$\Rightarrow \begin{cases} E(XY)=\sum xyP(x,y)=5\times1\times\frac{1}{3}+5\times4\times\frac{1}{6}=\frac{15}{3}=5 \\ E(X)=\sum xP(x)=0\times\frac{1}{2}+5\times\frac{1}{2}=\frac{5}{2} \\ E(Y)=\sum yP(y)=1\times\frac{1}{2}+4\times\frac{1}{2}=\frac{5}{2} \end{cases}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=5-\frac{5}{2}\times\frac{5}{2}=-\frac{5}{4}$$

۲۸. گزینه ۳ درست است.

X \ Y	0	1	2	$f(x)$
0	0.2	0.3	0.1	0.6
1	0.1	0.2	0.1	0.4
$f(y)$	0.3	0.5	0.2	1

$$\Rightarrow \begin{cases} E(XY)=\sum xyf(x,y)=0\times0\times0.2+0\times1\times0.3+0\times2\times0.1 \\ \qquad +1\times0\times0.1+1\times1\times0.2+1\times2\times0.1=0.4 \\ E(X)=\sum xf(x)=0\times0.6+1\times0.4=0.4 \\ E(Y)=\sum yf(y)=0\times0.3+1\times0.5+2\times0.2=0.9 \end{cases}$$

$$Cov(2X,3Y)=2\times3Cov(X,Y)=6Cov(X,Y)=6\left[E(XY)-E(X)E(Y)\right]=6(0.4-0.4\times0.9)=0.24$$

۲۹. گزینه ۱ درست است.

$$\begin{cases} Cov(X,Y)=Cov\left(X,X^2\right)=E\left(XX^2\right)-E(X)E\left(X^2\right)=E\left(X^3\right)-E(X)E\left(X^2\right)=0-0\times E\left(X^2\right)=0 \\ E(X)=\int_{-1}^{1}\frac{1}{2}x\,dx=\left[\frac{1}{4}x^2\right]_{-1}^{1}=\frac{1}{4}\left(1-(-1)^2\right)=0 \\ E\left(X^3\right)=\int_{-1}^{1}\frac{1}{2}x^3dx=\left[\frac{1}{8}x^4\right]_{-1}^{1}=\frac{1}{8}\left(1-(-1)^4\right)=0 \end{cases}$$

نکته: اگر $f(x)$ تابعی فرد باشد، آن‌گاه $\int_{-a}^{a}f(x)dx=0$.

توجه: هرگاه تابع چگالی پیوسته یک عدد باشد، X دارای توزیع یکنواخت پیوسته است که در اینجا:

$$f(x)=\frac{1}{b-a}=\frac{1}{1-(-1)}=\frac{1}{2}\quad;\quad a=-1\ ,\ b=1\ \rightarrow\ E(X)=\frac{a+b}{2}=\frac{-1+1}{2}=0$$

۳۰. گزینه ۱ درست است.

Y \ X	−2	0	3	$f(y)$
1	0.1	0.2	0.25	0.55
2	0.15	0.3	0	0.45
$f(x)$	0.25	0.5	0.25	1

$$\Rightarrow \begin{cases} E(XY)=\sum_x\sum_y xyf(x,y)=1\times(-2)\times0.1+1\times0\times0.2+1\times3\times0.25 \\ \qquad +2\times(-2)\times0.15+2\times0\times0.3+2\times3\times0=-0.05 \\ E(X)=\sum_x xf(x)=(-2)(0.25)+0\times0.5+3\times0.25=0.25 \\ E(Y)=\sum_y yf(y)=1\times0.55+2\times0.45=1.45 \end{cases}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=-0.05-(0.25)(1.45)=-0.4125$$

۳۱. گزینه ۲ درست است.

Y \ X	0	1	2	P(y)
0	0	0.25	0.15	0.4
1	0.2	0.35	0.05	0.6
P(x)	0.2	0.6	0.2	1

$$\Rightarrow \begin{cases} E(XY)=\sum xyP(x,y)=1\times1\times0.35+1\times2\times0.05=0.45 \\ E(X)=\sum xP(x)=1\times0.6+2\times0.2=1 \\ E(Y)=\sum yP(x)=1\times0.6=0.6 \end{cases}$$

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=0.45-1\times0.6=-0.15$$

خواص کوواریانس

۳۲. گزینه ۳ درست است.

$$\mathrm{Var}(X)=3 \ , \ \mathrm{Var}(Y)=4 \ , \ \mathrm{Cov}(X,Y)=-1$$

$$\mathrm{Var}(Z)=\mathrm{Var}\Big(2X-Y+\underset{0}{\cancel{1}}\Big)=2^2\mathrm{Var}(X)+(-1)^2\mathrm{Var}(Y)+2(2)(-1)\mathrm{Cov}(X,Y)$$
$$=4\times3+1\times4-4\times(-1)=12+4+4=20$$

۳۳. گزینه ۱ درست است.

$$\mathrm{Cov}(X-Y,\,X+Y)=\mathrm{Cov}(X,X)\cancel{+\mathrm{Cov}(X,Y)-\mathrm{Cov}(Y,X)}-\mathrm{Cov}(Y,Y)=\mathrm{Var}(X)-\mathrm{Var}(Y)$$

استقلال متغیرهای تصادفی

۳۴. گزینه ۳ درست است.

$$\begin{cases} \mathrm{Var}(X+Y)=\sigma_X^2+\sigma_Y^2+2\mathrm{Cov}(X,Y) \\ \mathrm{Var}(X-Y)=\sigma_X^2+\sigma_Y^2-2\mathrm{Cov}(X,Y) \end{cases} \xrightarrow{\mathrm{Var}(X+Y)=\mathrm{Var}(X-Y)} +\mathrm{Cov}(X,Y)=-\mathrm{Cov}(X,Y)$$

تنها وقتی مثبت و منفی کوواریانس دو متغیر با هم برابر می‌شود که کوواریانس آن دو متغیر صفر باشد.

$\mathrm{Cov}(X,Y)=0 \rightarrow$ X و Y یا مستقل‌اند یا رابطه غیرخطی دارند.

۳۵. گزینه ۱ درست است.

یادآوری: با توجه به وجود احتمال توأم صفر در جدول احتمال توأم X و Y نتیجه می‌گیریم X و Y مستقل نیستند، پس دو گزینه ۳ و ۴ به راحتی حذف می‌شوند. حال باید کوواریانس X و Y را محاسبه کنیم تا گزینه درست مشخص شود.

Y \ X	1	2	3	P(y)
0	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{7}{20}$	$\frac{10}{20}$
1	0	$\frac{10}{20}$	0	$\frac{10}{20}$
P(x)	$\frac{1}{20}$	$\frac{12}{20}$	$\frac{7}{20}$	1

$$\Rightarrow \begin{cases} E(XY)=\sum\sum xy\,P(x,y)=1\times2\times\frac{10}{20}=1 \\ E(X)=\sum x\,P(x)=1\times\frac{1}{20}+2\times\frac{12}{20}+3\times\frac{7}{20}=\frac{46}{20} \\ E(Y)=\sum y\,P(y)=0\times\frac{10}{20}+1\times\frac{10}{20}=\frac{10}{20} \end{cases}$$

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=1-\frac{10}{20}\times\frac{46}{20}=-\frac{3}{20}$$

با توجه به منفی شدن کوواریانس X و Y نتیجه می‌گیریم که X و Y همبستگی خطی منفی (معکوس) دارند.

ضریب همبستگی

۳۶. گزینه ۲ درست است.

با توجه به آنکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعداد صحیح هستند:

$$\begin{cases} \bar{x} = \dfrac{\sum x_i}{n} = \dfrac{20}{5} = 4 \\ \bar{y} = \dfrac{\sum y_i}{n} = \dfrac{15}{5} = 3 \end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه مناسب‌تر است:

x	y	$(x-\bar{x})$	$(y-\bar{y})$
2	3	−2	0
3	5	−1	2
4	1	0	−2
5	4	1	1
6	2	2	−1

$\longrightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
0	4	0
−2	1	4
0	0	4
1	1	1
−2	4	1
$\sum = -3$	$\sum = 10$	$\sum = 10$

$$r_{XY} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{-3}{10}$$

۳۷. گزینه ۱ درست است.

با توجه به آنکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعداد صحیح هستند:

$$\begin{cases} \bar{x} = \dfrac{\sum x_i}{n} = \dfrac{40}{5} = 8 \\ \bar{y} = \dfrac{\sum y_i}{n} = \dfrac{40}{5} = 8 \end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه مناسب‌تر است:

x	y	$(x-\bar{x})$	$(y-\bar{y})$
2	12	−6	4
5	10	−3	2
8	8	0	0
11	6	3	−2
14	4	6	−4

$\longrightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
−24	36	16
−6	9	4
0	0	0
−6	9	4
−24	36	16
$\sum = -60$	$\sum = 90$	$\sum = 40$

$$r_{XY} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{-60}{3 \times 2 \times 10} = -1$$

۳۸. گزینه ۳ درست است.

یادآوری:

$$Var(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X,Y)$$

بنابراین باید ابتدا $Cov(X,Y)$ را محاسبه کنیم:

$$\begin{cases} Var(X-2Y)=\sigma_X^2+(-2)^2\sigma_Y^2+2(1)(-2)Cov(X,Y)=8+4\times 2-4\times 2=8 \\ \rho_{X,Y}=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y} \rightarrow \dfrac{1}{2}=\dfrac{Cov(X,Y)}{\sqrt{8}\sqrt{2}} \rightarrow Cov(X,Y)=2 \\ \sigma_X^2=8 \quad , \quad \sigma_Y^2=2 \end{cases}$$

۳۹.گزینه ۱ درست است.

$$Var(Z)=Var(2X+4Y)=2^2Var(X)+4^2Var(Y)+2(2)(4)Cov(X,Y)$$
$$=4\times 4+16\times 5+16\times 0.8\sqrt{20}=16+80+12.8\sqrt{20}=96+12.8\sqrt{20}$$

$$r_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y} \rightarrow 0.8=\frac{Cov(X,Y)}{\sqrt{4}\sqrt{5}} \rightarrow Cov(X,Y)=0.8\sqrt{20}$$

۴۰. گزینه ۱ درست است.

$$Var(D)=Var(2X-Y)=2^2Var(X)+(-1)^2Var(Y)+2(2)(-1)Cov(X,Y)$$
$$=4\times 10^2+1\times 8^2-4\times 20=400+64-80=384$$

$$\rho_{X,Y}=\frac{Cov(X,Y)}{\sigma_X.\sigma_Y} \rightarrow 0.25=\frac{Cov(X,Y)}{10\times 8} \rightarrow Cov(X,Y)=20$$

۴۱. گزینه ۱ درست است.

یادآوری:

$0<r<1 \rightarrow$ همبستگی خطی ناقص (نسبی) و مستقیم (در یک جهت)

$r=+1 \rightarrow$ همبستگی خطی کامل و مستقیم (در یک جهت)

$-1<r<0 \rightarrow$ همبستگی خطی ناقص و معکوس (در جهت عکس)

$r=-1 \rightarrow$ همبستگی خطی کامل و معکوس (در جهت عکس)

$r=0 \rightarrow$ رابطه خطی وجود ندارد.

رگرسیون

۴۲.گزینه ۱ درست است.

با توجه به آنکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعداد صحیح هستند:

$$\begin{cases} \bar{x}=\dfrac{\sum x_i}{n}=\dfrac{16}{4}=4 \\ \bar{y}=\dfrac{\sum y_i}{n}=\dfrac{12}{4}=3 \end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه مناسب‌تر است:

x	y	$(x-\bar{x})$	$(y-\bar{y})$
2	3	−2	0
3	2	−1	−1
5	4	1	1
6	3	2	0

$\Rightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
0	4
1	1
1	1
0	4
$\sum=2$	$\sum=10$

$$\Rightarrow b=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}=\frac{2}{10}$$

حال با توجه به این نکته که خط رگرسیون همواره از نقطه $(\bar{x},\bar{y})$ می‌گذرد، داریم:

$\bar{y}=a+b\bar{x} \quad\rightarrow$ عرض از مبدأ : $a=\bar{y}-b\bar{x}\rightarrow a=3-0.2\times4=3-0.8=2.2$

۴۳.گزینه ۱ درست است.

ازانجاکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعداد صحیح هستند:

$$\begin{cases}\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{12}{4}=3\\[2mm] \bar{y}=\dfrac{\sum y_i}{n}=\dfrac{28}{4}=7\end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه مناسب‌تر است:

x	y	$(x-\bar{x})$	$(y-\bar{y})$
1	2	−2	−5
2	4	−1	−3
4	10	1	3
5	12	2	5

$\Rightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
10	4
3	1
3	1
10	4
$\sum=26$	$\sum=10$

$$\Rightarrow b=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}=\frac{26}{10}=2.6$$

حال با توجه به این نکته که نقطه $(\bar{x},\bar{y})$ همواره از خط رگرسیون می‌گذرد، داریم:

$\bar{y}=a+b\bar{x} \quad\rightarrow$ عرض از مبدأ : $a=\bar{y}-b\bar{x}=7-(2.6)\times3=-0.8$

خط رگرسیون : $\hat{y}=a+bx \quad\rightarrow\quad \hat{y}=2.6x-0.8$

۴۴.گزینه ۱ درست است.

یادآوری: رابطه شیب خط و ضریب همبستگی

$$b=r\times\frac{S_y}{S_x}\xrightarrow[S_y=S_x]{S_y^2=S_x^2} b=r=0.95$$

۴۵.گزینه ۳ درست است.

توجه: در صورت سؤال گفته شده است که خط رگرسیون از مبدأ مختصات می‌گذرد و این به معنای صفر بودن عرض از مبدأ است $(\alpha = 0)$، در این صورت خط رگرسیون به صورت $y = \beta x$ خواهد بود.

یادآوری: خط رگرسیون برآوردشده همواره از نقطه $(\bar{x}, \bar{y})$ می‌گذرد.

$$\bar{y} = a + b\bar{x}$$

بنابراین:

$$\hat{y} = bx \xrightarrow{(\bar{x}, \bar{y})} \bar{y} = b\bar{x} \rightarrow \frac{16}{10} = b \times \frac{20}{10} \rightarrow b = \frac{1.6}{2} = 0.8$$

۴۶.گزینه ۱ درست است.

با توجه به خواص ضریب همبستگی داریم:

$$r_{X^*, Y^*} = r_{4X, 2Y} = (\text{ضرب علامت‌های X و Y}) r_{X,Y} = +1$$

$$r_{X,Y} = b\frac{S_x}{S_y} = 2 \times \frac{\sqrt{4}}{\sqrt{16}} = 1$$

خودآزمایی

تابع احتمال گسسته

۱. توزیع احتمال متغیر X به صورت جدول زیر است. احتمال اینکه حداقل X برابر 2 باشد، کدام است؟

X	0	1	2	3
$P(X=x)$	$\frac{1}{16}$	$\frac{1}{4}$	a	$\frac{1}{4}$

۱) $\frac{1}{8}$ ۲) $\frac{5}{8}$ ۳) $\frac{8}{11}$ ۴) $\frac{11}{16}$

تابع چگالی پیوسته

۲. هرگاه X یک متغیر تصادفی پیوسته با تابع چگالی احتمال $f(x)=\frac{3}{2}x^2+\frac{1}{2}k$ وقتی $0<x<1$ و در جای دیگر برابر صفر باشد، مقدار k چقدر است؟

۱) 1 ۲) 2 ۳) $\frac{2}{3}$ ۴) $\frac{3}{2}$

۳. تابع چگالی احتمال متغیر تصادفی X به صورت زیر است. مقدار $P\left(\frac{9}{4}\le X\le 4\right)$ کدام است؟

$$f(x)=\begin{cases}\frac{1}{4\sqrt{x}} & ;\ 1\le x\le 9\\ 0 & ;\ \text{سایر نقاط}\end{cases}$$

۱) $\frac{1}{2}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{8}$ ۴) $\frac{3}{8}$

۴. اگر متغیر پیوسته X تنها مقادیر غیرمنفی را با تابع چگالی احتمال $f(x)=e^{-x}$ اختیار کند، احتمال اینکه X مقداری بین 1 تا 3 را بگیرد، برابر است با:

۱) 0.2325 ۲) 0.3181 ۳) 0.1353 ۴) 0.4650

تابع توزیع تجمعی

۵. اگر متغیر تصادفی X دارای تابع توزیع تجمعی زیر باشد، مقدار $P(1 \le X < 3)$ کدام است؟

$$F(x)=\begin{cases}0 & x<1\\ \frac{1}{2} & 1\le x<2\\ \frac{5}{6} & 2\le x<3\\ 1 & x\ge 3\end{cases}$$

۱) $\frac{1}{3}$ ۲) $\frac{1}{6}$ ۳) $\frac{1}{4}$ ۴) $\frac{5}{6}$

امید ریاضی گسسته

۶. در پرتاب دو سکه سالم، رو شدن یک H دارای یک امتیاز و رو شدن دو H دارای دو امتیاز است. رو نشدن H چند امتیاز منفی داشته باشد تا امید ریاضی این بازی برابر صفر شود؟

۱) 3 ۲) 4 ۳) 5 ۴) 6

۷. جدول توزیع احتمال فروش نوعی کالا توسط یک شخص در روز با سود هر واحد معادل 5000 تومان به صورت

X	0	1	2	3
f(x)	0.1	0.4	0.3	0.2

و کل هزینه‌های فروش روزانه برابر مبلغ ثابت 2000 تومان است، امید ریاضی سود خالص فروش هر روز چند تومان است؟

۱) 8000 ۲) 7000 ۳) 6000 ۴) 5000

۸. برای توزیع احتمال متغیر تصادفی X به صورت زیر:

X	1	2	3	4
P(x)	0.15	b	0.25	c

امید ریاضی محاسبه‌شده برابر 2.7 است. بین b و c چه رابطه‌ای برقرار است؟

۱) $b \ne c$ ۲) $b < c$ ۳) $b = c$ ۴) $b > c$

۹. در یک بازی، تاسی (مکعب شش‌وجهی منتظم) را می‌ریزیم و معادل عددی که تاس نشان می‌دهد با واحد تومان جایزه می‌گیریم. برای پرتاب هر بار تاس چند تومان باید بپردازیم تا بازی عادلانه باشد؟ (جمع جبری امید ریاضی برد و باخت مساوی صفر شود.)

۱) 3.5 ۲) 3 ۳) 6 ۴) 7

امید ریاضی پیوسته

۱۰. اگر X با توزیع $f(x)=\left\{\frac{x}{6}+\frac{1}{12}:\ 0\le x\le 3\right.$ متغیر تصادفی پیوسته باشد، امید ریاضی X کدام است؟

۱) $\frac{9}{8}$ ۲) $\frac{3}{2}$ ۳) $\frac{15}{8}$ ۴) 1

خواص امید ریاضی

۱۱. اگر متغیر تصادفی X دارای توزیع احتمال به صورت زیر باشد، امید ریاضی متغیر تصادفی $Y = 5X - 2$ کدام است؟

X	0	1	2
$f(x)$	0.1	0.2	0.7

۱) 6　　۲) 5
۳) 5.6　　۴) 6.7

واریانس

۱۲. هرگاه X یک متغیر تصادفی باشد که در آن $P(X = c) = 1$ و c مقدار ثابت باشد، $E(X)$ و $Var(X)$ به ترتیب کدام است؟

۱) $0, 1$　　۲) $0, c$　　۳) $c, 1$　　۴) c^2, c

۱۳. فرض کنید X یک متغیر تصادفی باشد که مقادیر $2, 2^2, 2^3, \dots, 2^k, \dots$ را با احتمال $P(x_k) = P(2^k) = \frac{1}{2^k}$ اختیار کند. در این صورت کدام‌یک از گزاره‌های زیر درست است؟

۱) میانگین X، برابر با 1 است.　　۲) واریانس X، برابر با 1 است.
۳) میانگین قابل محاسبه نیست.　　۴) انحراف معیار برابر با $\sqrt{2}$ است.

۱۴. انحراف معیار توزیع احتمال زیر برابر است با:

X	0	1	2	3
$f_X(x)$	0.1	0.3	0.4	0.2

۱) 0.8　　۲) 0.9
۳) 1　　۴) 1.5

خواص واریانس

۱۵. اگر $E(X) = 3.4 , E(X^2) = 12$ باشد، $V\left(-\frac{1}{2}X + 3\right)$ چقدر است؟

۱) 0.11　　۲) 0.12　　۳) 0.13　　۴) 0.16

تابع چگالی توزیع توأم

۱۶. تابع احتمال توأم زیر را در نظر بگیرید. احتمالات حاشیه‌ای Y کدام است؟

X \ Y	0	10
−2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

۱)

Y	0	10
$f(y)$	$\frac{1}{6}$	$\frac{1}{3}$

۲)

Y	0	10
$f(y)$	$\frac{1}{3}$	$\frac{1}{6}$

۳)

Y	0	10
$f(y)$	$\frac{1}{2}$	$\frac{1}{2}$

۴) هیچ‌کدام

۱۷. در سؤال قبل، $P(Y>X)$ کدام است؟

۱) $\frac{1}{6}$ ۲) $\frac{1}{3}$ ۳) $\frac{1}{2}$ ۴) $\frac{2}{3}$

۱۸. تابع $f(x,y)=ax^2y$ که در آن $0\le x<1$ و $0\le y<1$ و در سایر نقاط $f(x,y)=0$. به ازای کدام مقدار a یک تابع چگالی است؟

۱) 2 ۲) 4 ۳) 6 ۴) 8

تابع احتمال شرطی

۱۹.تابع احتمال توأم زیر را درنظر بگیرید. $P(Y=10 \mid X=2)$ کدام است؟

X \ Y	0	10
–2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

۱) $\frac{1}{6}$ ۲) $\frac{1}{3}$

۳) $\frac{1}{2}$ ۴) 1

محاسبه کوواریانس

۲۰. تابع احتمال توأم زیر را درنظر بگیرید. $E(XY)$ کدام است؟

X \ Y	0	10
–2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

۱) $\frac{3}{10}$ ۲) صفر

۳) 10 ۴) $-\frac{10}{3}$

۲۱.در سؤال قبل، کوواریانس کدام است؟

۱) $\frac{3}{10}$ ۲) $\frac{10}{3}$ ۳) $-\frac{10}{3}$ ۴) صفر

۲۲. در تابع احتمال زیر، کدام رابطه زیر صادق است؟

X \ Y	0	10
–2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

۱) $E(X+Y)=E(X)$

۲) $E(X+Y)=E(Y)$

۳) $E(X+Y)=E(XY)$

۴) $E(X+Y)=E(X)-E(Y)$

۲۳.تابع احتمال توأم دو متغیر تصادفی X و Y به صورت جدول زیر است. کوواریانس بین دو متغیر کدام است؟

Y \ X	1	2	3
0	0.2	0.1	0.3
2	0.15	0.25	0

۱) 0.48
۲) 0.52
۳) -0.26
۴) -0.34

خواص کوواریانس

۲۴.اگر $V(X)=\frac{1}{2}$ و $V(Y)=\frac{2}{3}$ و $V(X+Y)=\frac{5}{6}$ باشد، آن‌گاه کدام عبارت درست است؟

۱) $Cov(X, Y)=\frac{1}{3}$
۲) $Cov(X, Y)=\frac{1}{2}$
۳) $Cov(X, Y)=-\frac{1}{6}$
۴) $Cov(X, Y)=-\frac{1}{4}$

۲۵.تابع احتمال توأم زیر را درنظر بگیرید. واریانس X و واریانس Y به ترتیب (از راست به چپ) کدام است؟

X \ Y	0	10
-2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

۱) 5 , 0
۲) 25 , 0
۳) 5 , 4
۴) 25 , 4

۲۶.در سؤال قبل، $V(X-Y)$ کدام است؟

۱) 29
۲) 21
۳) $\frac{77}{3}$
۴) $\frac{107}{3}$

استقلال متغیرهای تصادفی

۲۷.کدام‌یک از عبارات زیر نادرست است؟

۱) اگر دو متغیر تصادفی X و Y مستقل باشند، آن‌گاه X و Y ناهمبسته هستند.
۲) اگر دو پیشامد A و B مستقل باشند، آن‌گاه A و B' نیز مستقل‌اند.
۳) اگر A و B دو پیشامد جدا از هم باشند، آن‌گاه مستقل‌اند.
۴) اگر X و Y ناهمبسته بوده و متوسط X صفر باشد، آن‌گاه X و Y بر هم عمودند.

۲۸.اگر $r=0$ باشد، می‌توان گفت دو متغیر ....

۱) همبستگی با هم ندارند.
۲) همبستگی آن‌ها ضعیف است.
۳) همبستگی خطی با هم ندارند.
۴) همبستگی آن‌ها غیرخطی است.

ضریب همبستگی

۲۹.اگر معادله خط رگرسیون y نسبت به x به صورت $y=-2x+b$ و $SS_y=4SS_x$ باشد، ضریب همبستگی بین x و y کدام است؟

۱) 1
۲) 0.9
۳) -1
۴) -0.9

۳۰. اگر X_1, X_2, X_3 متغیرهای تصادفی با میانگین صفر و واریانس یک و دو به دو ناهمبسته باشند و $U=X_1+X_2+X_3$ و $V = X_1$، ضریب همبستگی بین U و V چقدر است؟

۱) $\frac{\sqrt{2}}{2}$ ۲) $-\frac{\sqrt{2}}{2}$ ۳) $\frac{\sqrt{3}}{3}$ ۴) $-\frac{\sqrt{3}}{3}$

۳۱. اگر در متغیرهای تصادفی $U=X+Y$ و $V=aX-Y$ که $\sigma_X^2=\sigma_Y^2=1$ بوده، ضریب همبستگی X و Y برابر 0.8 باشد و U و V ناهمبسته باشند، مقدار a برابر است با:

۱) 1 ۲) $-\frac{1}{2}$ ۳) $\frac{1}{2}$ ۴) -1

۳۲. معادله خط رگرسیون که باتوجه به نمونه‌ای 8 تایی برآوردشده، به صورت $\hat{y} = -5 + 2x$ است. کدام گزینه نمی‌تواند ضریب همبستگی آن باشد؟

۱) 0.45 ۲) -0.90 ۳) 0.95 ۴) 0.85

۳۳. ضریب همبستگی داده‌های زیر با کدام‌یک از این موارد برابر است؟

x	5	7	9
y	20	15	13

۱) $-\frac{7}{2\sqrt{10}}$ ۲) $-\frac{7}{2\sqrt{13}}$

۳) $\frac{7}{2\sqrt{10}}$ ۴) $\frac{7}{2\sqrt{13}}$

ضریب تعیین

۳۴. اگر ضریب همبستگی بین دو متغیر 0.6 و بین دو متغیر دیگر 0.3 باشد، می‌توان گفت همبستگی دو متغیر اول چند برابر قوی‌تر از دو متغیر دوم است؟

۱) دو برابر ۲) سه برابر ۳) چهار برابر ۴) نه برابر

رگرسیون

۳۵. از یک نمونه‌گیری اطلاعات زیر در دست است:

$$r_{XY} = 0.9 \quad , \quad \sigma_X = \sqrt{2} \quad , \quad \sigma_Y = \sqrt{8} \quad , \quad \bar{y} = 6 \quad , \quad \bar{x} = 3$$

معادله رگرسیون کدام است؟

۱) $\hat{y} = 0.6 + 1.8x$ ۲) $\hat{y} = 1.2 + 1.8x$ ۳) $\hat{y} = -0.6 + 1.8x$ ۴) $\hat{y} = 1.2 - 1.8x$

۳۶. در یک آزمایش از 10 زوج مقادیر X و Y، رابطه $\sum x_i = \sum y_i = 30$ برقرار بوده و معادله خط رگرسیون به صورت $y = -9 + \beta x$ است. مقدار شیب این خط، چقدر است؟

۱) 4 ۲) 3 ۳) -4 ۴) -3

۳۷. شیب خط رگرسیون با کدام‌یک از این موارد برابر است؟

x	5	7	9
y	20	15	13

۱) -1.00 ۲) -1.75

۳) -2.00 ۴) -2.75

۳۸. در سؤال قبل، عرض از مبدأ (a) کدام است؟

۱) 10.25 ۲) 18.25 ۳) 28.25 ۴) 38.25

پاسخنامه

۱	۱	۲	۳	●
۲	●	۲	۳	۴
۳	۱	●	۳	۴
۴	۱	●	۳	۴
۵	۱	۲	۳	●
۶	۱	●	۳	۴
۷	۱	۲	●	۴
۸	۱	۲	●	۴
۹	●	۲	۳	۴
۱۰	۱	۲	●	۴
۱۱	●	۲	۳	۴
۱۲	۱	●	۳	۴
۱۳	۱	۲	●	۴
۱۴	۱	●	۳	۴
۱۵	●	۲	۳	۴
۱۶	۱	۲	●	۴
۱۷	۱	۲	۳	●
۱۸	۱	۲	●	۴
۱۹	۱	●	۳	۴
۲۰	۱	۲	۳	●

۲۱	۱	۲	●	۴
۲۲	۱	●	۳	۴
۲۳	۱	۲	●	۴
۲۴	۱	۲	●	۴
۲۵	۱	۲	۳	●
۲۶	۱	۲	۳	●
۲۷	۱	۲	●	۴
۲۸	۱	۲	●	۴
۲۹	۱	۲	●	۴
۳۰	۱	۲	●	۴
۳۱	●	۲	۳	۴
۳۲	۱	●	۳	۴
۳۳	۱	●	۳	۴
۳۴	۱	۲	●	۴
۳۵	●	۲	۳	۴
۳۶	●	۲	۳	۴
۳۷	۱	●	۳	۴
۳۸	۱	۲	●	۴
۳۹	۱	۲	۳	۴
۴۰	۱	۲	۳	۴

برای دریافت پاسخ تشریحی سؤالات به سایت www.Tourani.ir مراجعه کنید.

فصل ۴ توزیع‌های گسسته و پیوسته

مقدمه

هر آزمایش آماری با فضای نمونه **گسسته** یا **پیوسته** وضعیتی را ایجاد می‌کند که می‌توانیم متغیر تصادفی (X) و تابع احتمال $(f(x))$ را بر اساس آن تعریف کنیم.

در این فصل سعی داریم ضمن بررسی آزمایش‌های مختلف گسسته و پیوسته، آن‌ها را دسته‌بندی کرده و برای هر کدام یک تابع احتمال معرفی کنیم و به تحلیل آن و یا محاسبه امید ریاضی و واریانس و تقریب و … بپردازیم.

توابع گسسته	
۱) یکنواخت	$DU(x_1,x_2,\dots,x_N)$
۲) برنولی	$Bin(1,p)$ یا $B(1,p)$
۳) دوجمله‌ای	$Bin(n,p)$ یا $B(n,p)$
۴) چندجمله‌ای	$MN(n_1,n_2,\dots,p_1,p_2,\dots)$
۵) دوجمله‌ای منفی	$NB(r,p)$
۶) هندسی	$Ge(p)$ یا $G(p)$
۷) فوق هندسی	$HG(N,k,n)$
۸) پواسون	$P(\lambda)$

توابع پیوسته	
۱) یکنواخت	$U(a,b)$
۲) نمایی	$E(\lambda)$ یا $Exp(\lambda)$
۳) گاما	$G(r,\lambda)$ یا $\Gamma(r,\lambda)$
۴) نرمال	$N(\mu,\sigma^2)$
۵) (کای - دو یا خی دو)	$\chi^2_{(n)}$
۶) t ـ استیودنت	$t_{(n)}$
۷) کوشی	$C(\mu,\sigma)$
۸) F (فیشر)	F_{n_1,n_2}

توزیع یکنواخت گسسته (Discrete Uniform Distribution)

مسئله (درک مطلب): اگر متغیر تصادفی X نتیجه پرتاب یک تاس به صورت زیر باشد، امید ریاضی و واریانس X کدام است؟

$$f(x)=\frac{1}{6} \quad ; \quad x=1,2,\dots,6$$

مقدمه: اگر نتیجه انجام آزمایشی مانند x بتواند به یکی از N وضعیت هم‌شانس $x_1,\dots,x_N$ با احتمال یکسان منجر شود، آن‌گاه احتمال وقوع هر یک از وضعیت‌های x_i به صورت زیر برابر $\frac{1}{N}$ است:

$$\sum_{i=1}^{N} P(x_i)=1 \rightarrow P(x_1)+P(x_2)+\dots+P(x_N)=1 \xrightarrow{P(x_1)=\dots=P(x_N)} P(x_i)=\frac{1}{N} \; ; \; i=1,2,\dots,N$$

تعریف: درصورتی‌که متغیر تصادفی یکنواخت گسسته X بتواند N مقدار مختلف $(x_N,\dots,x_2,x_1)$ را با احتمال‌های برابر $\frac{1}{N}$ اختیار کند، آن‌گاه X دارای توزیع **یکنواخت گسسته** است.

X	x_1	x_2	$\dots$	x_N
$f(x)$	$\frac{1}{N}$	$\frac{1}{N}$	$\dots$	$\frac{1}{N}$

$$\longrightarrow f(x)=\frac{1}{N} \; ; \; x=x_1,x_2,\dots,x_N$$

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع یکنواخت گسسته با N مقدار از X باشد، پارامتر آن N است و به صورت $X \sim DU(x_1,x_2,\dots,x_N)$ نمایش داده می‌شود.

تابع احتمال	$P(x)=\frac{1}{N} \quad ; \quad x=x_1,\dots,x_N$	احتمال وقوع هر یک از مقادیر x_i
امید ریاضی (میانگین)	$\mu_X=E(X)=\frac{\sum x_i}{N}$	متوسط مقادیر x_i
واریانس	$\sigma_X^2=\frac{\sum(x_i-\mu)^2}{N}=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2$	واریانس مقادیر x_i
انحراف معیار	$\sigma_X=\sqrt{\frac{\sum(x_i-\mu)^2}{N}}$	
تابع مولد گشتاور	$M_X(t)=\frac{e^t(1-e^{Nt})}{N(1-e^t)}$	

حالات خاص

الف) اگر متغیر تصادفی یکنواخت گسسته $x = x_1, \dots, x_N$ به صورت $x = 1, 2, \dots, N$ در نظر گرفته شود، آن‌گاه:

تابع احتمال	امید ریاضی (میانگین)	واریانس	انحراف معیار
$P(x) = \frac{1}{N}$ $x = 1, 2, \dots, N$	$\mu_X = E(X) = \frac{N+1}{2}$	$\sigma_X^2 = \frac{N^2-1}{12}$	$\sigma_X = \sqrt{\frac{N^2-1}{12}}$

ب) اگر متغیر تصادفی یکنواخت گسسته $x = x_1, x_2, \dots, x_N$ به صورت یک تصاعد حسابی (با قدرنسبت d) به صورت $x = x_1, x_1 + d, \dots, x_1 + (N-1)d$ مطرح شود، آن‌گاه:

تابع احتمال	امید ریاضی (میانگین)	واریانس	انحراف معیار
$P(x) = \frac{1}{N}$ $x = x_1, x_1 + d, .., x_1 + (N-1)d$	$\mu_X = E(X) = \frac{x_1 + x_N}{2}$	$\sigma_X^2 = d^2 \cdot \frac{N^2-1}{12}$	$\sigma_X = \sqrt{d^2 \cdot \frac{N^2-1}{12}}$

حل مسئله (درک مطلب):

با توجه به حالت خاص (الف) خواهیم داشت:

$$\mu = \frac{N+1}{2} = \frac{6+1}{2} = 3.5$$

$$\sigma^2 = \frac{N^2-1}{12} = \frac{6^2-1}{12} = \frac{35}{12}$$

مثال ۱ اگر کمیت تصادفی X به صورت مقابل توزیع شده باشد:

$$f(x) = \begin{cases} \frac{1}{9} & ; \quad x = 1,2,3,4,5,6,7,8,9 \\ 0 & ; \quad \text{سایر نقاط} \end{cases}$$

میانگین و واریانس آن برابر است با: (اقتصاد ـ ۸۴)

۱) $V(X) = \frac{81}{12}, E(X) = \frac{9}{2}$

۲) $V(X) = \frac{20}{3}, E(X) = 5$

۳) $V(X) = \frac{12}{3}, E(X) = 5$

۴) $V(X) = \frac{9}{12}, E(X) = \frac{71}{10}$

حل: گزینه ۲ درست است.

با توجه به حالت خاص (الف) داریم:

$$\begin{cases} E(X) = \frac{N+1}{2} = \frac{9+1}{2} = 5 \\ V(X) = \frac{N^2-1}{12} = \frac{9^2-1}{12} = \frac{80}{12} = \frac{20}{3} \end{cases}$$

مثال ۲ اگر تابع یکنواخت متغیر تصادفی گسسته X به صورت زیر باشد:

$$\begin{cases} f(x) = \frac{1}{5} \\ x = 2, 6, 10, 14, 18 \end{cases}$$

میانگین و واریانس X کدام است؟

۱) $5, 10$ ۲) $32, 10$ ۳) $20, 64$ ۴) $16, 8$

حل: گزینه ۲ درست است.

همان‌طورکه دیده می‌شود مشاهدات، یک تصاعد حسابی با $x_1 = 2, x_N = 18, N = 5, d = 4$ را تشکیل می‌دهند، بنابراین با توجه به حالت خاص (ب) داریم:

$$\mu = \frac{x_1 + x_N}{2} = \frac{2+18}{2} = 10$$

$$\sigma^2 = d^2 \cdot \frac{N^2 - 1}{12} = 4^2 \cdot \frac{5^2 - 1}{12} = 32$$

آزمایش برنولی

اگر هر بار انجام آزمایش، مستقل از نتایج به‌دست‌آمده از دفعات قبلی، فقط و فقط به یکی از دو پیشامد «موفقیت» یا «شکست» با احتمال ثابت (p) یا $(q=1-p)$ منجر شود، آن‌گاه به هر یک از این آزمایش‌ها «آزمایش برنولی» و به متغیر تصادفی X، **تعداد موفقیت** در هر بار انجام آزمایش $(x=0,1)$، «توزیع برنولی» نسبت داده می‌شود.

توزیع‌های حاصل از آزمایش برنولی

هرگاه یک آزمایش برنولی با احتمال موفقیت p را انجام دهیم یا آن را با توجه به هدف مورد نظر به صورت مستقل تکرار کنیم، یکی از چهار توزیع زیر حاصل می‌شود:

۱- توزیع برنولی (دونقطه‌ای)

آزمایش برنولی را یک بار انجام می‌دهیم. هدف، بررسی تعداد موفقیت‌های حاصل است که می‌تواند 0 یا 1 باشد.
مانند، تعداد خط در 1 بار پرتاب سکه.

۲- توزیع دوجمله‌ای (باینم)

آزمایش برنولی را n بار به صورت مستقل تکرار می‌کنیم. هدف، بررسی تعداد موفقیت‌های حاصل است که می‌تواند $0, 1, \dots, n$ باشد.
مانند، تعداد خط در n بار پرتاب سکه.
* توزیع برنولی حالت خاصی از توزیع دوجمله‌ای است، وقتی $n=1$ باشد.

۳- توزیع دوجمله‌ای منفی (پاسکال)

آزمایش برنولی را به صورت مستقل آن‌قدر تکرار می‌کنیم تا به rامین موفقیت برسیم. هدف، بررسی تعداد آزمایش لازم است که می‌تواند $r, r+1, \dots$ باشد.
مانند، تعداد پرتاب سکه برای رسیدن به rامین خط.

۴- توزیع هندسی

آزمایش برنولی را به صورت مستقل آن‌قدر تکرار می‌کنیم تا به اولین موفقیت برسیم. هدف، بررسی تعداد آزمایش لازم است که می‌تواند $1, 2, \dots$ باشد.
مانند، تعداد پرتاب سکه برای رسیدن به اولین خط.
* توزیع هندسی حالت خاصی از توزیع دوجمله‌ای منفی است، وقتی $r=1$ باشد.

نکته: «موفقیت»: وقوع پیشامد مطلوب (مورد نظر سؤال) با احتمال ثابت p

«شکست»: عدم وقوع پیشامد مطلوب (مورد نظر سؤال) با احتمال ثابت $q=1-p$

نکته: آزمایش‌هایی که فقط دو پیامد دارند، مانند (برد و باخت)، (پسر و دختر)، (شیر و خط)، (زوج و فرد)، (معیوب و سالم) با توجه به پیشامد مطلوب (مورد نظر سؤال) ممکن است در هر زوج یکی «موفقیت» و دیگری «شکست» باشد. به عبارت دیگر:

«موفقیت» = (برد یا باخت)، (پسر یا دختر)، (شیر یا خط)، ...

«شکست» = (باخت یا برد)، (دختر یا پسر)، (خط یا شیر)، ...

احتمال ثابت در هر آزمایش

شرط لازم برای هر آزمایش برنولی آن است که وقوع پیشامد موفقیت یا شکست مستقل از نتیجه آزمایش‌های قبلی، ثابت باشد. در هر یک از شرایط زیر، احتمال وقوع پیشامد ثابت است و می‌توانند به عنوان یک آزمایش برنولی در نظر گرفته شوند.

۱- هرگاه احتمال یا نسبت ثابتی داده شود (نمونه‌گیری با جایگذاری یا بدون جایگذاری).

برای مثال:

ـ 0.20 کالاهای یک کارخانه معیوب است.

ـ از هر 100 کالا، 20 کالا معیوب است.

ـ با احتمال 0.20 پرتاب‌های یک بازیکن به هدف اصابت می‌کند.

۲- هرگاه از یک جامعه **محدود** N **تایی**، نمونه‌گیری **با جایگذاری** انجام شود.

برای مثال، از یک جامعه با 4 کالای سالم و 6 کالای معیوب، نمونه‌ای با جایگذاری انتخاب می‌کنیم (بعد از هر بار انتخاب نمونه، آن را دوباره به جعبه برمی‌گردانیم).

احتمال معیوب بودن در هر بار نمونه‌گیری $= \frac{6}{10}$ (ثابت)

احتمال سالم بودن در هر بار نمونه‌گیری $= \frac{4}{10}$ (ثابت)

✓ دقت کنید!

اولاً، در سؤالات به طور پیش‌فرض انتخاب بدون جایگذاری است، بنابراین برای جامعه محدود باید در صورت سؤال به‌گونه‌ای «باجایگذاری بودن انتخاب» ذکر شود، در غیر این صورت احتمال ثابت نخواهد بود.

ثانیاً، اگر از جامعه محدود مانند مثال بالا نمونه‌گیری بدون جایگذاری انجام شود، احتمال معیوب یا سالم بودن در هر بار آزمایش دیگر ثابت نیست و وابسته به آزمایش‌های قبلی است و دیگر نمی‌تواند به عنوان آزمایش برنولی در نظر گرفته شود.

انتخاب در بار اول (بدون جایگذاری)		انتخاب در بار دوم (بدون جایگذاری)
احتمال معیوب بودن $= \frac{6}{10}$	$\xrightarrow[\text{در انتخاب اول}]{\text{خارج کردن معیوب}}$	$\frac{4}{9} =$ سالم و $\frac{5}{9} =$ معیوب
احتمال سالم بودن $= \frac{4}{10}$	$\xrightarrow[\text{در انتخاب اول}]{\text{خارج کردن سالم}}$	$\frac{3}{9} =$ سالم و $\frac{6}{9} =$ معیوب

۳- در هر بار پرتاب سکه یا تاس، احتمال ثابتی داریم؛ برای مثال:

$\frac{1}{2} =$ خط آمدن سکه $\quad \frac{1}{2} =$ شیر آمدن سکه

$\frac{3}{6} =$ فرد آمدن تاس $\quad \frac{3}{6} =$ زوج آمدن تاس

۴- پیشامد به دنیا آمدن پسر یا دختر در هر بار زایمان مادر با احتمال ثابتی اتفاق می‌افتد.

(دختر و پسر هم‌شانس) $\quad \frac{1}{2} =$ دختر $\quad \frac{1}{2} =$ پسر $\quad$: پیش‌فرض

(پیشامد پسر 3 برابر دختر) $\quad \frac{1}{4} =$ دختر $\quad \frac{3}{4} =$ پسر

توزیع برنولی (Bernoulli Distribution)

تعریف: هرگاه آزمایش برنولی با احتمال موفقیت p را یک بار انجام دهیم، آن‌گاه متغیر تصادفی « X : **تعداد موفقیت** در انجام 1 **بار** آزمایش برنولی» $(x=0,1)$ دارای توزیع **برنولی** است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع برنولی باشد، پارامتر آن p است و به صورت $X \sim Bin(1,p)$ یا $X \sim B(1,p)$ نمایش داده می‌شود.

نکته: ازآنجاکه در توزیع برنولی، تعداد موفقیت (X) در انجام 1 بار آزمایش برنولی فقط می‌تواند یکی از دو وضعیت «صفر موفقیت» یا « 1 موفقیت» را داشته باشد، به آن توزیع دونقطه‌ای نیز می‌گویند:

X : تعداد موفقیت در 1 بار انجام آزمایش برنولی	0	1
$f(x)=P(x)$	$q=1-p$	p

تابع احتمال	$f(x)=P(x)=p^x q^{1-x} \quad ; \quad x=0,1$	احتمال x موفقیت در یک آزمایش برنولی
امید ریاضی (میانگین)	$E(X)=\mu=p$	میانگین (امید) تعداد موفقیت در یک آزمایش برنولی
واریانس	$\sigma_X^2=pq$	واریانس تعداد موفقیت در یک آزمایش برنولی
انحراف معیار	$\sigma_X=\sqrt{pq}$	
تابع مولد گشتاور	$M_X(t)=pe^t+q$	

✓ **دقت کنید!**

۱- ازآنجاکه $p+q=1$ است، می‌توان $f(x)=p^x(1-p)^{1-x}$ را در نظر گرفت.

۲- در بعضی منابع، احتمال موفقیت را با نماد θ نمایش می‌دهند، درنتیجه داریم: $f(x)=\theta^x(1-\theta)^{1-x}$.

مثال اگر کمیت تصادفی X بر طبق قانون دونقطه‌ای با تابع احتمال زیر توزیع شده باشد:

$$P_X(x)=p^x q^{1-x} \quad , \quad x=0,1 \quad ; \quad q=1-p$$

امید ریاضی آن کدام است؟ (اقتصاد ـ ۷۲)

۱) $\frac{pq}{n}$ ۲) $p(1-p)$ ۳) p ۴) np

حل: گزینه ۳ درست است.

توزیع دوجمله‌ای (Binominal Distribution)

مسئله ۱ (درک مطلب): 0.20 از محصولات کارخانه‌ای معیوب است. (از هر 100 کالا، 20 کالا معیوب). احتمال آنکه از 3 کالای خریداری‌شده از این کارخانه یک کالا سالم باشد، چقدر است؟

مسئله۲ (درک مطلب): در خانواده‌ای با 3 فرزند، احتمال آنکه حداقل 1 فرزند پسر باشد، چقدر است؟

مسئله۳ (درک مطلب): در 3 بار پرتاب یک تاس سالم، احتمال آنکه حداکثر 1 بار عدد 4 یا 6 ظاهر شود، چقدر است؟

مسئله۴ (درک مطلب): از جعبه‌ای با 8 کالای سالم و 2 کالای معیوب، 3 کالا **با جایگذاری** انتخاب می‌کنیم. احتمال آنکه 1 کالا معیوب باشد، چقدر است؟

مقدمه: در هنگام انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، اگر به دنبال محاسبه احتمال x **موفقیت** در n **بار تکرار** (نمونه) آزمایش باشیم، باید به صورت زیر عمل کنیم:

برای مثال، اگر بخواهیم احتمال وقوع $x = 2$ موفقیت را در $n = 3$ بار تکرار آزمایش مستقل برنولی بررسی کنیم:

آزمایش / وضعیت‌های ممکن	اول		دوم		سوم	
وضعیت اول	p	×	p	×	q	$= p^2q^1$
						+
وضعیت دوم	p	×	q	×	p	$= p^2q^1$
						+
وضعیت سوم	q	×	p	×	p	$= p^2q^1$

$$\left.\begin{matrix} p^2q^1 \\ + \\ p^2q^1 \\ + \\ p^2q^1 \end{matrix}\right\} \rightarrow 3p^2q^1$$

همان‌طور که دیده می‌شود:

۱- وقوع 2 موفقیت در 3 آزمایش، 3 حالت دارد که باید آن را محاسبه کنیم، برای این کار می‌توانیم از $\binom{n}{x} = \binom{3}{2} = 3$، که همان ترکیب x از n است، استفاده کنیم تا تعداد حالات وقوع x موفقیت در n آزمایش را به دست آوریم.

۲- احتمال وقوع $x = 2$ موفقیت در هر حالت برابر است با $p^xq^{n-x} = p^2q^1$ که برای محاسبه احتمال کل باید در تعداد حالات ضرب شود، بنابراین:

$$P(\text{موفقیت در } n = 3 \text{ آزمایش } X = 2) = \binom{n}{x}p^xq^{n-x} = \binom{3}{2}p^2q^1 = 3p^2q$$

> **تعریف:** هرگاه یک آزمایش مستقل برنولی را n **بار تکرار** کنیم، (یک نمونه مستقل nتایی از آزمایش برنولی انتخاب کنیم) آن‌گاه متغیر تصادفی « X : **تعداد موفقیت در n بار آزمایش مستقل برنولی**» $(x = 0,1,\ldots,n)$ دارای توزیع **دوجمله‌ای (باینم)** خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع دوجمله‌ای باشد، پارامترهای آن n و p است و به صورت $X \sim Bin(n,p)$ یا $X \sim B(n,p)$ نمایش داده می‌شود.

تابع احتمال	$P(x)=\binom{n}{x}p^x q^{n-x}$ $x=0\,,1,2,\dots,\,n$	احتمال x بار موفقیت در n بار آزمایش برنولی
امید ریاضی (میانگین)	$E(X)=np$	میانگین تعداد موفقیت مورد انتظار در n بار آزمایش مستقل برنولی
واریانس	$\sigma_X^2=npq$	واریانس تعداد موفقیت در n بار آزمایش مستقل برنولی
انحراف معیار	$\sigma_X=\sqrt{npq}$	
تابع مولد گشتاور	$M_X(t)=\left(pe^t+q\right)^n$	

نکته: همان‌طورکه در ابتدا مطرح شد، $C_n^x=\binom{n}{x}$ حالات وقوع x موفقیت در n آزمایش را محاسبه می‌کند که برابر است با:

$$C_n^x=\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

✓ دقت کنید!

احتمال وقوع دقیقاً x موفقیت (هر موفقیت با احتمال p) در n بار آزمایش مستقل برنولی به مفهوم داشتن $n-x$ شکست (هر شکست با احتمال q) نیز است، به طوری که:

$$P(x)=\binom{n}{x}p^x q^{n-x}=\frac{n!}{x!(n-x)!}p^x q^{n-x}$$

مثال ۱ در یک توزیع دوجمله‌ای خاص با $p=0.4$، مقدار $\frac{7!}{3!4!}(0.4)^3(0.6)^4$ نشان‌دهنده احتمال کدام‌یک از این حالات است؟

۱) دقیقاً 4 موفقیت در 7 آزمایش
۲) 3 موفقیت یا بیشتر در 7 آزمایش
۳) دقیقاً 3 موفقیت در 7 آزمایش
۴) 4 موفقیت یا کمتر در 7 آزمایش

حل: گزینه ۳ درست است.

$$\begin{cases} P(X=3)=\binom{7}{3}(0.4)^3(0.6)^4=\frac{7!}{3!4!}(0.4)^3(0.6)^4 \\ n=7\,,x=3\,,\ p=0.4\,,\,q=0.6 \end{cases}$$

این احتمال نشان‌دهنده احتمال دقیقاً 3 موفقیت یا دقیقاً 4 شکست در 7 بار تکرار آزمایش است.

محاسبه احتمال

برای محاسبه احتمال x موفقیت در n بار آزمایش مستقل برنولی (احتمال موفقیت، p و احتمال شکست، q) به صورت زیر عمل می‌کنیم:

الف) n، p و x را مشخص می‌کنیم:

n : تعداد کل آزمایش

p : احتمال وقوع وضعیت مطلوب در هر آزمایش (موفقیت)

x : تعداد مورد نظر وقوع برای وضعیت مطلوب در n بار آزمایش

ب) با استفاده از رابطه زیر (بسط نیوتن) احتمال را محاسبه می‌کنیم:

$$P(x)=(p+q)^n=\binom{n}{x}p^x q^{n-x}$$

با توجه به آنکه همواره $\sum P(x)=1$ (مجموع مقادیر تابع احتمال متغیر تصادفی گسسته برابر با 1 است)، برای تابع احتمال دوجمله‌ای، رابطه $\sum_{x=0}^{n}P(x)=\sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}=1$ برقرار است و داریم:

$$\underbrace{\binom{n}{0}p^0q^n}_{P(X=0)}+\underbrace{\binom{n}{1}p^1q^{n-1}}_{P(X=1)}+\underbrace{\binom{n}{2}p^2q^{n-2}}_{P(X=2)}+\cdots+\underbrace{\binom{n}{n}p^nq^0}_{P(X=n)}=1$$

درنتیجه برای مثال:

(احتمال عدم موفقیت در n بار آزمایش) $P(X=0)=\binom{n}{0}p^0q^{n-0}=q^n$

(احتمال وقوع 1 موفقیت در n بار آزمایش) $P(X=1)=\binom{n}{1}p^1q^{n-1}=npq^{n-1}$

(احتمال وقوع 2 بار موفقیت در n بار آزمایش) $P(X=2)=\binom{n}{2}p^2q^{n-2}$

(احتمال وقوع حداکثر 1 موفقیت در n بار آزمایش) $P(X\le 1)=P(X=0)+P(X=1)$

(احتمال وقوع بیش از 1 موفقیت در n بار آزمایش) $P(X>1)=1-P(X\le 1)=1-P(X=0)-P(X=1)$

(احتمال وقوع حداقل 1 موفقیت در n بار آزمایش) $P(X\ge 1)=1-P(X<1)=1-P(X=0)=1-q^n$

حل مسئله ۱ (درک مطلب):

در این سؤال به ترتیب $n=3$، $p=0.80$ (سالم) و $x=1$ است و داریم:

$$P(X=1)=\binom{n}{x}p^xq^{n-x}=\binom{3}{1}(0.8)^1(0.2)^2=0.096$$

حل مسئله ۲ (درک مطلب):

در این سؤال به ترتیب $n=3$، $p=\frac{1}{2}$ (پسر) و $x\ge 1$ است و داریم:

$$P(X\ge 1)=1-P(X=0)=1-\binom{n}{0}p^0q^{n-0}=1-\binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3=1-\frac{1}{8}=\frac{7}{8}$$

حل مسئله ۳ (درک مطلب):

در این سؤال به ترتیب $n=3$، $p=\frac{2}{6}=\frac{1}{3}$ و $x\le 1$ است و داریم:

$$P(X\le 1)=P(X=0)+P(X=1)=\binom{n}{0}p^0q^{n-0}+\binom{n}{1}p^1q^{n-1}=\binom{3}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3+\binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2=\frac{8}{27}+\frac{12}{27}=\frac{20}{27}$$

حل مسئله ۴ (درک مطلب):

با توجه به نوع انتخاب (با جایگذاری) احتمال وقوع سالم یا معیوب بودن کالا در هر بار انتخاب ثابت و به ترتیب برابر $\frac{8}{10}$ و $\frac{2}{10}$ است، بنابراین شرایط آزمایش برنولی وجود دارد و داریم:

$$\begin{cases} P(X=1)=\binom{3}{1}\left(\frac{2}{10}\right)^1\left(\frac{8}{10}\right)^2=0.384 \\ p=P(\text{معیوب})=\frac{2}{10}\ ,\ q=\frac{8}{10}\ ,\ n=3 \end{cases}$$

مثال ۲ امتحانی شامل 4 سؤال 3 گزینه‌ای است. فردی به صورت تصادفی به همه سؤالات پاسخ می‌دهد. اگر X تعداد جواب‌های درست باشد، مطلوب است:

الف) تابع احتمال X

ب) امید ریاضی و واریانس X

ج) اگر معلمی مقیاس نمره را طبق تبدیل $Y = 22.5X + 10$ عوض کند، امید ریاضی و واریانس Y چقدر است؟

حل:

در یک سؤال 3 گزینه‌ای، احتمال درست جواب دادن، $\frac{1}{3}$ و احتمال غلط جواب دادن، $\frac{2}{3}$ است. حال اگر به $n = 4$ سؤال پاسخ داده شود، متغیر تصادفی X یعنی تعداد پاسخ درست در 4 سؤال، دارای توزیع دوجمله‌ای با پارامترهای $n = 4$ و $p = \frac{1}{3}$ است.

الف)

$$\begin{cases} P(x) = \binom{n}{x} p^x q^{n-x} = \binom{4}{x}\left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{4-x} \\ p = \frac{1}{3}\ ,\ q = \frac{2}{3}\ ,\ n = 4 \end{cases}$$

ب)

$$\begin{cases} E(X) = np = 4 \times \frac{1}{3} = \frac{4}{3} \\ \sigma_X^2 = npq = 4 \times \frac{1}{3} \times \frac{2}{3} = \frac{8}{9} \\ p = \frac{1}{3}\ ,\ q = \frac{2}{3}\ ,\ n = 4 \end{cases}$$

ج)

$$\begin{cases} E(Y) = E(22.5X + 10) = 22.5E(X) + 10 = 22.5 \times \frac{4}{3} + 10 = 40 \\ \sigma_Y^2 = \sigma^2(22.5X \overset{0}{\cancel{+10}}) = (22.5)^2 \sigma_X^2 = (22.5)^2 \times \frac{8}{9} = 450 \end{cases}$$

مثال ۳ 0.40 از کارمندان یک کارخانه، قراردادی و بقیه رسمی هستند. درصورتی‌که یک نمونه $n = 100$ نفری از کارمندان انتخاب کرده باشیم و « X : تعداد کارمندان قراردادی در نمونه» باشد، تابع احتمال X کدام است؟

۱) $(0.4)^x (0.6)^{100-x}$

۲) $C_{100}^{x} (0.4)^x (0.6)^{100-x}$

۳) $C_{100}^{x} (0.6)^{100-x} (0.4)$

۴) $(0.6)^x (0.4)^{100-x}$

حل: گزینه ۲ درست است.

در این سؤال به ترتیب احتمال موفقیت $p = 0.4$ (کارمند قراردادی) و تعداد نمونه $n = 100$ است:

$$\begin{cases} P(x) = \binom{n}{x} p^x q^{n-x} = \binom{100}{x} (0.4)^x (0.6)^{100-x} = C_{100}^{x} (0.4)^x (0.6)^{100-x} \\ n = 100\ ,\ p = P(\text{قراردادی}) = 0.40\ ,\ q = P(\text{رسمی}) = 0.60 \end{cases}$$

مثال ۴ 0.80 از محصولات کارخانه‌ای سالم است. احتمال اینکه از 4 کالای خریداری‌شده از این کارخانه یک کالا سالم باشد، چقدر است؟

۱) $\frac{12}{625}$ ۲) $\frac{16}{625}$ ۳) $\frac{24}{625}$ ۴) $\frac{5}{16}$

حل: گزینه ۲ درست است.

در این سؤال به ترتیب $n=4$، $p=0.8$ (سالم بودن کالا) و $x=1$ انتخاب یک کالای سالم است.

$$P(X=1)=\binom{n}{x}p^x q^{n-x}=\binom{4}{1}(0.8)^1(0.2)^3=0.0256=\frac{16}{625}$$

مثال ۵ امید و واریانس تعداد 4 یا 5 آمدن در 120 بار پرتاب یک تاس سالم چقدر است؟

۱) $\mu_X=40\ ,\ \sigma_X^2=\frac{80}{3}$ ۲) $\mu_X=20\ ,\ \sigma_X^2=\frac{50}{3}$

۳) $\mu_X=20\ ,\ \sigma_X^2=\frac{80}{3}$ ۴) $\mu_X=40\ ,\ \sigma_X^2=\frac{50}{3}$

حل: گزینه ۱ درست است.

ازآنجاکه آزمایش برنولی با دو نتیجه (ظهور یا عدم ظهور 4 یا 5) به تعداد 120 بار تکرار می‌شود، توزیع احتمال دوجمله‌ای است و درنتیجه:

$$\begin{cases} E(X)=np=120\times\frac{1}{3}=40 \\ \sigma_X^2=npq=120\times\frac{1}{3}\times\frac{2}{3}=\frac{80}{3} \\ n=120\ ,\ p=P(X=4 \text{ or } 5)=\frac{2}{6}=\frac{1}{3}\ ,\ q=1-p=\frac{2}{3} \end{cases}$$

مثال ۶ اگر در یک توزیع دوجمله‌ای، میانگین 4 و انحراف معیار $\sqrt{2}$ باشد، مقدار $P(X>0)$ کدام است؟

۱) 1 ۲) $1-\left(\frac{1}{2}\right)^8$ ۳) $\left(\frac{1}{2}\right)^8$ ۴) $1-\left(\frac{1}{3}\right)^4$

حل: گزینه ۲ درست است.

اولاً، اگر $\sigma=\sqrt{2}$ باشد، آن‌گاه $\sigma^2=2$ است. ثانیاً، در توزیع دوجمله‌ای $\mu=np$ و $\sigma^2=npq$ است، بنابراین:

$$\sigma^2=npq=2 \xrightarrow{\mu=np=4} 4q=2 \rightarrow q=\frac{1}{2}\ ,\ p=\frac{1}{2}$$

$$\mu=np=4 \xrightarrow{p=\frac{1}{2}} n\times\frac{1}{2}=4 \rightarrow n=8$$

همواره در توزیع دوجمله‌ای داریم:

$$P(X=0)+P(X=1)+\ldots+P(X=n)=1$$

بنابراین:

$$P(X>0)=P(X\geq 1)=1-P(X=0)=1-\binom{n}{0}p^0q^n=1-q^n=1-\left(\frac{1}{2}\right)^8$$

دقت کنید اگر در صورت مسئله احتمال $P(X \geq 0)$ مد نظر بود، آن‌گاه مقدار احتمال برابر 1 و گزینه ۱ درست بود.

مثال ۷ تابع چگالی احتمال $f(x)$ به صورت $f(x)=2x \; ; \; 0<x<1$ است. اگر 3 بار نقطه‌ای از فاصله $0<x<1$ انتخاب کنیم، احتمال آنکه 2 بار نقطه در فاصله $\left(0<X<\frac{1}{2}\right)$ باشد کدام است؟

۱) $\frac{9}{64}$ ۲) $\frac{3}{16}$ ۳) $\frac{1}{64}$ ۴) $\frac{1}{3}$

حل: گزینه ۱ درست است.

در این سؤال $n=3$ و $x=2$ و داریم:

$$p=P\left(0<X<\frac{1}{2}\right)=\int_0^{\frac{1}{2}} 2x\,dx=\left[x^2\right]_0^{\frac{1}{2}}=\frac{1}{4} \quad \rightarrow \quad q=\frac{3}{4}$$

درنتیجه:

$$P(X=2)=\binom{3}{2}p^2q^1=3\times\left(\frac{1}{4}\right)^2\times\frac{3}{4}=\frac{9}{64}$$

مثال ۸ احتمال اینکه در 2 بار پرتاب، حداقل یک پرتاب به هدف اصابت کند، مساوی 0.36 است. احتمال اصابت در هر بار پرتاب چقدر است؟

۱) 0.2 ۲) 0.8 ۳) 0.5 ۴) 0.3

حل: گزینه ۱ درست است.

اولاً، ازآنجاکه احتمال اصابت در هر پرتاب (p) خواسته شده، احتمال ثابت و آزمایش برنولی است.

ثانیاً، « X : تعداد اصابت در 2 بار پرتاب» دارای توزیع دوجمله‌ای و $x=0,1,2$ است.

حال با توجه به آنکه احتمال حداقل یک اصابت در 2 بار پرتاب 0.36 شده است، داریم:

$$P(X\geq 1)=1-P(X=0)=1-\binom{2}{0}p^0q^2=0.36 \quad \rightarrow \quad q^2=0.64 \quad \rightarrow \quad q=0.8 \;,\; p=0.2$$

نمایش خاص توزیع دوجمله‌ای

هرگاه متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای n و $p=q=\frac{1}{2}$ باشد، تابع احتمال آن به صورت زیر است:

$$\boxed{f(x)=\binom{n}{x}\left(\frac{1}{2}\right)^n=\frac{\binom{n}{x}}{2^n} \quad ; \quad x=0,1,2,\dots,n}$$

اثبات:

$$f(x)=\binom{n}{x}p^xq^{n-x} \xlongequal{p=q=\frac{1}{2}} \binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x}=\binom{n}{x}\left(\frac{1}{2}\right)^n=\frac{\binom{n}{x}}{2^n}$$

مثال ۹ شانه تخم‌مرغ‌های 6 تایی به قیمت 600 تومان و شانه‌های تاریخ‌گذشته آن به قیمت 400 تومان عرضه می‌شود. اگر تابع احتمال تخم‌مرغ‌های سالم در شانه تاریخ‌گذشته به صورت زیر باشد، احتمال آنکه خرید آن مقرون به صرفه باشد، چقدر است؟ (اقتصاد ـ ۸۴)

$$f(x)=\binom{6}{x}\frac{1}{64} \quad ; \quad x=0,1,\dots,6$$

۱) $\frac{7}{64}$ ۲) $\frac{16}{64}$

$$f(x)=0 \quad ; \quad \text{برای سایر مقادیر } x$$

۳) $\frac{28}{64}$ ۴) $\frac{39}{64}$

حل: گزینه ۱ درست است.

دقت کنید که اگر هر شانه 6 تایی سالم را 600 تومان بخریم، برای هر عدد تخم مرغ، 100 تومان هزینه کرده‌ایم. حال اگر یک شانه 6 تایی تاریخ‌گذشته را 400 تومان بخریم، درصورتی‌که 4 تخم مرغ سالم داشته باشد، برای هر عدد تخم مرغ، 100 تومان هزینه کرده‌ایم که این هزینه فرقی با هزینه خرید یک شانه سالم ندارد و مقرون به صرفه نیست.

پس برای آنکه خرید یک شانه تاریخ گذشته مقرون به صرفه باشد، باید بیش از 4 تخم مرغ سالم (حداقل 5 سالم) داشته باشد:

$$\begin{cases} P(\text{مقرون به صرفه بودن})=P(X\geq 5)=P(X=5)+P(X=6)=\binom{6}{5}\frac{1}{64}+\binom{6}{6}\frac{1}{64}=\frac{6}{64}+\frac{1}{64}=\frac{7}{64} \\ X: \text{تعداد تخم مرغ سالم در هر شانه تاریخ‌گذشته} \end{cases}$$

با کمی دقت متوجه می‌شوید که تابع $f(x)$ همان نمایش خاص از توزیع دوجمله‌ای با پارامترهای $n=6$ و $p=\frac{1}{2}$ است که:

$$\begin{cases} f(x)=\binom{6}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{6-x}=\binom{6}{x}\left(\frac{1}{2}\right)^6=\binom{6}{x}\frac{1}{64} \quad ; \quad x=0,1,2,\dots,6 \\ X: \text{تعداد تخم مرغ سالم در هر شانه تاریخ‌گذشته} \end{cases}$$

مثال ۱۰ اگر $x=0,1,\dots,10$; $f(x)=\frac{\binom{10}{x}}{2^{10}}$ باشد، ضریب تغییرات کمیت X چند درصد است؟

۱) 5 ۲) $\frac{\sqrt{5}}{2}$ ۳) $20\sqrt{2}$ ۴) $20\sqrt{\frac{5}{2}}$

حل: گزینه ۴ درست است.

با کمی دقت متوجه می‌شویم که تابع $f(x)$ همان نمایش خاص از توزیع دوجمله‌ای با پارامترهای $p=\frac{1}{2}$ و $n=10$ است.

حال برای محاسبه ضریب پراکندگی توزیع به صورت زیر عمل می‌کنیم:

$$\begin{cases} CV_X\times 100=\frac{\sigma}{\mu}\times 100=\frac{\sqrt{\frac{5}{2}}}{5}\times 100=20\sqrt{\frac{5}{2}} \\ \mu=np=10\times\frac{1}{2}=5 \\ \sigma=\sqrt{npq}=\sqrt{10\times\frac{1}{2}\times\frac{1}{2}}=\sqrt{\frac{5}{2}} \end{cases}$$

احتمال موفقیت و نمودار توزیع

درصورتی‌که احتمال وقوع موفقیت در هر بار آزمایش دوجمله‌ای برابر p باشد، آن‌گاه تأثیر این احتمال بر نمودار توزیع دوجمله‌ای به صورت زیر است:

الف) $p>0.5$: اگر احتمال موفقیت در هر آزمایش **بیشتر از** 0.5 باشد، نمودار توزیع، **چوله به چپ** است.

ب) $p=0.5$: اگر احتمال وقوع موفقیت در هر آزمایش **برابر با** 0.5 باشد، نمودار توزیع، **متقارن** است.

ج) $p<0.5$: اگر احتمال وقوع موفقیت در هر آزمایش **کمتر از** 0.5 باشد، نمودار توزیع، **چوله به راست** است.

مثال ۱۱ اگر در توزیع دوجمله‌ای $p=0.3$ باشد، نمودار احتمال آن چه شکلی خواهد داشت؟ (اقتصاد ـ ۸۱)

۱) متقارن ۲) چوله به چپ ۳) چوله به راست ۴) به n بستگی دارد.

حل: گزینه ۳ درست است.

ازآنجاکه $p=0.3<0.5$ است، نمودار توزیع احتمال دوجمله‌ای، چوله به راست خواهد بود.

محتمل‌ترین پیشامد

پیشامدی را که احتمال وقوع آن از سایر پیشامدها بیشتر باشد، محتمل‌ترین پیشامد گویند. حال در n بار تکرار آزمایش برنولی با احتمال ثابت p، می‌خواهیم بدانیم کدام‌یک از مقادیر X (تعداد موفقیت)، بزرگ‌ترین احتمال را دارند، برای این کار از رابطه زیر استفاده می‌کنیم:

$$\boxed{np-q\le x\le np+p}$$

الف) اگر مقدار $np+p$، عدد صحیح شود، آن‌گاه $np-q$ نیز عدد صحیح خواهد شد. در این حالت دو عدد برای مقادیر X وجود دارد که بزرگ‌ترین احتمال را دارند.

ب) اگر مقدار $np+p$، عدد اعشاری شود، آن‌گاه $np-q$ نیز عدد اعشاری خواهد شد. در این حالت تنها عدد صحیح بین این دو عدد اعشاری، بزرگ‌ترین احتمال را دارد.

نکته: محتمل‌ترین تعداد برای وقوع پیشامد X (موفقیت) $[p(n+1)]$ است، یعنی جزء صحیح (حد پایین) عدد اعشاری $p(n+1)$.

مثال در یک توزیع دوجمله‌ای احتمال وقوع S، سه برابر احتمال وقوع F است. اگر در این توزیع $n=10$ باشد، محتمل‌ترین تعداد دفعاتی که S رخ می‌دهد برابر است با:

۱) 4 ۲) 8 ۳) 5 ۴) 2

حل: گزینه ۲ درست است.

راه حل اول:

$$\begin{cases} np-q\le x\le np+p \;\rightarrow\; 10\times\frac{3}{4}-\frac{1}{4}\le x\le 10\times\frac{3}{4}+\frac{3}{4} \;\rightarrow\; \frac{29}{4}\le x\le\frac{33}{4} \\ n=10\;,\; p=\frac{3}{4}\;,\; q=\frac{1}{4} \end{cases}$$

بنا بر قسمت «ب» هر دو عدد $np+p$ و $np-q$ اعشاری هستند و تنها عدد صحیح بین آن دو $x=8$ است، بنابراین $x=8$ محتمل‌ترین وقوع پیشامد S (پیروزی) است.

راه حل دوم:

$$\text{محتمل‌ترین تعداد } x=\left[p(n+1)\right]=\left[\frac{3}{4}(10+1)\right]=\left[\frac{33}{4}\right]=[8.25]=8$$

توزیع چندجمله‌ای (Multinomial Distribution)

مسئله (درک مطلب): نتیجه هر مسابقه برای یک تیم به ترتیب با **احتمال‌های** 0.1، 0.4 **و** 0.5 می‌تواند **برد، باخت و مساوی** باشد، احتمال آنکه این تیم در 5 مسابقه دارای 2 برد، 2 باخت و 1 مساوی شود کدام است؟

مقدمه: در توزیع دوجمله‌ای انجام هر آزمایش منجر به یکی از دو نتیجه «موفقیت» یا «شکست» می‌شد (برد یا باخت، شیر یا خط، پسر یا دختر). هرگاه انجام آزمایشی بتواند بیش از دو نتیجه ممکن داشته باشد، به طوری که احتمال هر یک از نتایج در هر آزمایش ثابت باشد و آزمایش‌ها از هم مستقل باشند، توزیع مربوط به این آزمایش، چندجمله‌ای خواهد بود (برد، باخت یا مساوی؛ خوب، عالی، متوسط یا ضعیف؛ شاغل، بازنشسته یا بیکار).

تعریف: هرگاه انجام آزمایشی منجر به یکی از k نتیجه مجزا با احتمال‌های ثابت $p_1, p_2, \dots, p_k$ شود، به‌گونه‌ای که $p_1 + p_2 + \dots + p_k = 1$ باشد، اگر آن آزمایش را n بار به طور مستقل انجام دهیم، احتمال وقوع x_1 بار از نتیجه اول و x_2 بار از نتیجه دوم و ... و x_k بار از نتیجه kام $(x_1 + x_2 + \dots + x_k = n)$ دارای توزیع **چندجمله‌ای** است و به صورت زیر محاسبه می‌شود:

$$P(x_1, x_2, \dots, x_k) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع چندجمله‌ای باشد، پارامترهای آن $n_1, n_2, \dots$ و $p_1, p_2, \dots$ هستند و به صورت $X \sim MN(n_1, n_2, \dots, p_1, p_2, \dots)$ نمایش داده می‌شود.

میانگین و واریانس در توزیع چندجمله‌ای

برای محاسبه میانگین و واریانس **«تعداد دفعات تکرار مربوط به نتیجه iام در آزمایش چندجمله‌ای»** از **توزیع دوجمله‌ای** استفاده می‌کنیم.

تابع احتمال	$P(x_1, \dots, x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$	x_1 بار از نتیجه اول و ... و x_k بار از نتیجه kام $(x_1 + \dots + x_k = n)$
امید ریاضی (میانگین)	$E(X) = np$	میانگین نتیجه iام در توزیع چندجمله‌ای
واریانس	$\sigma_X^2 = npq$	واریانس نتیجه iام در توزیع چندجمله‌ای
انحراف معیار	$\sigma_X = \sqrt{npq}$	
	$p = p_i \ , \ q = 1 - p_i$	

حل مسئله (درک مطلب):

در این مسئله با توجه به شرایط توزیع چندجمله‌ای، احتمال مورد نظر به صورت زیر محاسبه می‌شود:

$$\begin{cases} P(X_1 = 2 \ , \ X_2 = 2 \ , \ X_3 = 1) = \frac{5!}{2!2!1!}(0.1)^2 (0.4)^2 (0.5)^1 \\ p_1 = 0.1 \ , \ p_2 = 0.4 , \ p_3 = 0.5 \\ x_1 = 2 \ , \ x_2 = 2 \ , x_3 = 1 \\ n = 5 \end{cases}$$

مثال ۱ در یک شهر بزرگ به ترتیب 0.40، 0.30 و 0.30 از مردم، شاغل، بیکار و بازنشسته هستند. درصورتی‌که 6 نفر از مردم این شهر را انتخاب کنیم، احتمال اینکه به ترتیب 2 نفر شاغل، 2 نفر بیکار و 2 نفر بازنشسته باشند، چقدر است؟

۱) 0.6 ۲) 0.4 ۳) 0.1166 ۴) 0.2

حل: گزینه ۳ درست است.

$$P(X_1=2\ ,\ X_2=2\ ,\ X_3=2)=\frac{6!}{2!2!2!}(0.4)^2(0.3)^2(0.3)^2=0.1166$$

مثال ۲ از دانش‌آموزان یک مدرسه 0.20 کلاس اول، 0.30 کلاس دوم و 0.50 کلاس سوم هستند. اگر 4 نفر از دانش‌آموزان مدرسه انتخاب شده باشند، مطلوب است احتمال آنکه:

الف) 1 نفر کلاس اول، 2 نفر کلاس دوم و 1 نفر کلاس سوم باشند.

ب) 1 نفر کلاس اول و 3 نفر کلاس دوم باشند.

ج) 1 نفر کلاس اول و بقیه کلاس دوم یا سوم باشند.

حل:

الف)

$$\begin{cases}P(X_1=1\ ,\ X_2=2\ ,X_3=1)=\dfrac{4!}{1!2!1!}(0.2)^1(0.3)^2(0.5)^1=0.108\\ x_1=1,\ x_2=2,x_3=1\end{cases}$$

ب)

$$\begin{cases}P(X_1=1,X_2=3,X_3=0)=\dfrac{4!}{1!3!0!}(0.2)^1(0.3)^3(0.5)^0=0.0216\\ x_1=1,x_2=3,x_3=0\end{cases}$$

ج)

$$\begin{cases}P(X_1=1,X_2=3)=\dfrac{4!}{1!3!}(0.2)^1(0.8)^3=0.4096\\ \text{کلاس اول}:p_1=0.2\quad,\quad \text{کلاس دوم یا سوم}:p_2=0.3+0.5=0.8\\ x_1=1\quad,\quad x_2=3\end{cases}$$

مثال ۳ نظر حسابرسان درباره حساب‌های شرکتی ممکن است قبول، مردود، عدم اظهار نظر یا اظهار نظر مشروط باشد. در سال پیش 0.20، 0.15، 0.40 و 0.25 نظرها به ترتیب قبول، مردود، عدم اظهار نظر و اظهار نظر مشروط بوده است. اگر فرض کنیم 6 شرکت به صورت تصادفی انتخاب شده‌اند، احتمال اینکه حساب‌های 2 شرکت قبول و حساب‌های 2 شرکت مردود باشد، چقدر است؟

۱) 0.00038 ۲) 0.082 ۳) 0.0342 ۴) 0.2737

حل: گزینه ۳ درست است.

در این مثال، با توجه به شرایط توزیع چندجمله‌ای، احتمال مورد نظر به صورت زیر محاسبه می‌شود:

$$\begin{cases}P(X_1=2,X_2=2,X_3=2)=\dfrac{6!}{2!2!2!}(0.2)^2(0.15)^2(0.4+0.25)^2=0.0342\\ \text{قبول}:p_1=0.2\quad,\quad \text{مردود}:p_2=0.15\quad,\quad p_3=1-(0.2+0.15)=0.65\\ x_1=2\ ,\ x_2=2\ ,\ x_3=6-(2+2)=2\end{cases}$$

توجه: در این مثال فقط احتمال 2 قبول و 2 مردود خواسته شده و در مورد 2 شرکت باقی‌مانده صحبتی نشده است، بنابراین 2 شرکت باقی‌مانده را یا مشروط و یا عدم اظهار نظر می‌دانیم و احتمال مربوط به 2 شرکت باقی‌مانده برابر است با $0.4+0.25=0.65$.

مثال ۴ در یک کارخانه به ترتیب 0.2، 0.4 و 0.4 کالاها در شیفت‌های صبح، عصر و شب تولید می‌شوند. اگر 100 کالا انتخاب کنیم، امید و واریانس تعداد کالاهای شیفت صبح در این نمونه کدام است؟

۱) 20 , 16 ۲) 8 , 64 ۳) 4 , 16 ۴) 9 , 25

حل: گزینه ۱ درست است.

با در نظر گرفتن $p=0.2$ برای کالاهای شیفت صبح و با استفاده از توزیع دوجمله‌ای داریم:

$$\begin{cases} E(X)=np=100\times 0.2=20 \\ \sigma_X^2=npq=100\times 0.2\times 0.8=16 \\ n=100, p=0.2, q=1-p=0.8 \end{cases}$$

مثال ۵ یک اظهارنامه مالیاتی به ترتیب با احتمال‌های 0.4، 0.3، 0.2 و 0.1 می‌تواند ۱ـ درست پر شود، ۲ـ شامل خطاهایی به نفع دولت باشد، ۳ـ شامل خطاهایی به نفع مالیات‌دهنده باشد، ۴ـ شامل خطاهایی از هر دو نوع باشد. درصورتی‌که 10 اظهارنامه را جمع‌آوری کرده باشیم:

الف) امید و واریانس تعداد اظهارنامه‌هایی که درست پر شده باشند چقدر است؟

ب) امید و وایانس تعداد اظهارنامه‌هایی که شامل خطاهایی به نفع دولت یا مالیات‌دهنده باشند، چقدر است؟

ج) امید و واریانس تعداد اظهارنامه‌هایی که شامل هر دو نوع خطا هستند، چقدر است؟

حل:

الف) با در نظر گرفتن $p=0.4$ برای اظهارنامه‌هایی که درست پر شده‌اند و با استفاده از توزیع دوجمله‌ای داریم:

$$\begin{cases} E(X)=np=10\times 0.4=4 \\ \sigma_X^2=npq=10\times 0.4\times 0.6=2.4 \\ p=0.4\ ,\ q=1-p=0.6\ ,\ n=10 \end{cases}$$

ب) با در نظر گرفتن $p=0.3+0.2=0.5$ برای اظهارنامه‌هایی که شامل خطا به نفع دولت یا خطا به نفع مالیات‌دهنده هستند و با استفاده از توزیع دوجمله‌ای داریم:

$$\begin{cases} E(X)=np=10\times 0.5=5 \\ \sigma_X^2=npq=10\times 0.5\times 0.5=2.5 \\ p=0.2+0.3=0.5,\ q=1-p=0.5\ ,\ n=10 \end{cases}$$

ج) با در نظر گرفتن $p=0.1$ برای اظهارنامه‌هایی که شامل هر دو نوع خطا هستند و با استفاده از توزیع دوجمله‌ای داریم:

$$\begin{cases} E(X)=np=10\times 0.1=1 \\ \sigma_X^2=npq=10\times 0.1\times 0.9=0.9 \\ p=0.1\ ,\ q=1-p=0.9\ ,\ n=10 \end{cases}$$

توزیع دوجمله‌ای منفی (پاسکال) (Negative Binomial Distribution)

مسئله ۱ (درک مطلب): احتمال آنکه هر پرتاب بازیکنی به هدف بخورد، $\frac{2}{3}$ است. احتمال اینکه سومین پرتابی که به هدف می‌خورد، پنجمین پرتاب وی باشد، چقدر است؟

مسئله ۲ (درک مطلب): در یک ظرف، 10 توپ سفید و 5 توپ سیاه داریم. یک توپ به طور تصادفی با جایگذاری انتخاب می‌کنیم. این عمل را آن‌قدر تکرار می‌کنیم تا چهارمین توپ سیاه انتخاب شود. متوسط تعداد انتخاب (تکرار) چقدر است؟

مقدمه: در هنگام انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، محاسبه احتمال وقوع x موفقیت در n بار تکرار آزمایش برنولی با استفاده از توزیع دوجمله‌ای برابر است با $\binom{n}{x}p^xq^{n-x}$.

حال اگر «**پیشامد وقوع r امین موفقیت در x امین آزمایش**» مورد نظر باشد $(x \ge r)$، برای محاسبه احتمال این پیشامد:

اولاً، باید در $x-1$ آزمایش قبلی دقیقاً به $r-1$ موفقیت برسیم: $\binom{x-1}{r-1}p^{r-1}q^{x-r}$

ثانیاً، در آزمایش x ام به موفقیت r ام برسیم: p

به عبارت دیگر:

$$\underbrace{P(x)}_{\text{r امین موفقیت در x امین آزمایش}} = \overbrace{\underbrace{\binom{x-1}{r-1}p^{r-1}q^{x-r}}_{r-1\ \text{موفقیت}}}^{x-1\ \text{آزمایش قبلی}} \cdot \overbrace{\underbrace{p}_{\text{موفقیت r ام}}}^{\text{آزمایش x ام}} = \binom{x-1}{r-1}p^rq^{x-r}$$

تعریف: هرگاه یک آزمایش برنولی با احتمال موفقیت p را به طور مستقل، آن‌قدر تکرار کنیم تا به **r امین موفقیت** برسیم، آن‌گاه متغیر تصادفی «X : تعداد آزمایشات لازم برای رسیدن به r امین موفقیت» $(x = r, r+1, \ldots)$ دارای توزیع **دوجمله‌ای منفی (پاسکال)** یا توزیع «زمان انتظار» خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع دوجمله‌ای منفی باشد، پارامترهای آن r و p هستند و به صورت $X \sim NB(r,p)$ نمایش داده می‌شود.

احتمال وقوع r امین موفقیت در x امین آزمایش (احتمال وقوع موفقیت r ام در آزمایش x ام)	$P(x) = \binom{x-1}{r-1}p^rq^{x-r}$ $x = r, r+1, \ldots$	**تابع احتمال**
میانگین تعداد آزمایش مورد انتظار برای رسیدن به r امین موفقیت	$E(X) = \frac{r}{p}$	**امید ریاضی (میانگین)**
واریانس تعداد آزمایش لازم برای رسیدن به r امین موفقیت	$\sigma_X^2 = \frac{rq}{p^2}$	**واریانس**
	$\sigma_X = \sqrt{\frac{rq}{p^2}}$	**انحراف معیار**
	$M_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$	**تابع مولد گشتاور**

محاسبه احتمال

الف) ابتدا وضعیتی را که به دنبال rامین وقوع آن هستیم، به عنوان موفقیت (با احتمال p) در نظر می‌گیریم.

ب) $\binom{x-1}{r-1}p^r q^{x-r}$؛ احتمال وقوع rامین موفقیت در xامین آزمایش است.

حل مسئله ۱ (درک مطلب):

الف) در این مسئله $p=\frac{2}{3}$ (احتمال به هدف خوردن) است.

ب)

$$\begin{cases} P(\text{سومین هدف، پنجمین پرتاب}) = \binom{x-1}{r-1}p^r q^{x-r} = \binom{5-1}{3-1}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 = \binom{4}{2}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 \\ r=3\ ,\ x=5 \\ p=\frac{2}{3}\ ,\ q=\frac{1}{3} \end{cases}$$

حل مسئله ۲ (درک مطلب):

با توجه به اینکه انتخاب با جایگذاری است، احتمال موفقیت (سیاه بودن) ثابت بوده و توزیع تعداد تکرار برای رسیدن به چهارمین موفقیت، دوجمله‌ای منفی است.

$$\begin{cases} E(X)=\frac{r}{p}=\frac{4}{\frac{5}{15}}=12 \\ p=P(\text{سیاه})=\frac{5}{15}\quad ,\quad q=P(\text{سفید})=\frac{10}{15} \end{cases}$$

مثال ۱ 0.20 تولیدات کارخانه‌ای معیوب است. احتمال آنکه سومین کالای انتخاب‌شده دومین کالای معیوب باشد، چقدر است؟

۱) 0.08 ۲) 0.064 ۳) 0.128 ۴) 0.032

حل: گزینه ۲ درست است.

الف) $p=0.2$ (معیوب)

ب)

$$\begin{cases} P(\text{دومین معیوب، سومین کالا}) = \binom{x-1}{r-1}p^r q^{x-r} = \binom{3-1}{2-1}(0.2)^2(0.8)^1 = 0.064 \\ r=2\ ,\ x=3\ ,\ p=0.2\ ,\ q=0.8 \end{cases}$$

مثال ۲ در یک توزیع پاسکال احتمال شکست 0.8 است. اگر X تعداد آزمایش برنولی در این توزیع باشد، $E(X)$ برای پیشامد بیستمین موفقیت کدام است؟

۱) 100 ۲) 25 ۳) 30 ۴) 45

حل: گزینه ۱ درست است.

در این مثال $p=0.2$ (موفقیت) و $r=20$ (بیستمین موفقیت)، بنابراین:

$$E(X)=\frac{r}{p}=\frac{20}{0.2}=\frac{200}{2}=100$$

دقت کنید که اگر در صورت سؤال پیشامد بیستمین شکست مطرح شده بود، آن‌گاه $p = 0.8$ و $r = 20$ را در نظر می‌گرفتیم و گزینه ۲ درست بود.

$$E(X) = \frac{r}{p} = \frac{20}{0.8} = 25$$

مثال ۳ 0.20 تولیدات یک کارخانه معیوب است. متوسط و واریانس تعداد کالاهای انتخاب‌شده برای رسیدن به دهمین کالای معیوب کدام است؟

۱) $\mu = 0.02 , \sigma^2 = 200$ ۲) $\mu = 50 , \sigma^2 = 4$ ۳) $\mu = 0.02 , \sigma^2 = 4$ ۴) $\mu = 50 , \sigma^2 = 200$

حل: گزینه ۴ درست است.

با توجه به $p = 0.2$ و $r = 10$ (دهمین کالای معیوب) داریم:

$$\begin{cases} E(X) = \dfrac{r}{p} = \dfrac{10}{0.2} = 50 \\ \sigma_X^2 = \dfrac{rq}{p^2} = \dfrac{10 \times 0.8}{(0.2)^2} = 200 \\ r = 10 \quad , \quad p = 0.2 \quad , \quad q = 0.8 \end{cases}$$

توزیع هندسی (Geometric Distribution)

مسئله۱ (درک مطلب): احتمال اصابت موشکی به یک جنگنده 0.4 است. با اصابت یک موشک، جنگنده سقوط می‌کند. احتمال آنکه **جنگنده در پرتاب پنجمین موشک سقوط کند** چقدر است؟

مسئله۲ (درک مطلب): 10 درصد تولیدات کارخانه‌ای معیوب‌اند، احتمال اینکه **سومین کالای** کنترل‌شده، **اولین کالای معیوب باشد**، چقدر است؟

مسئله۳ (درک مطلب): در یک ظرف، 10 توپ سفید و 5 توپ سیاه داریم. یک توپ به طور تصادفی با جایگذاری انتخاب می‌کنیم. این عمل را **آن‌قدر تکرار می‌کنیم تا توپ سیاه انتخاب شود**. متوسط تعداد انتخاب (تکرار) کدام است؟

مقدمه: در انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، اگر «پیشامد وقوع اولین موفقیت در xامین آزمایش» مورد نظر باشد $(x \ge 1)$، برای محاسبه احتمال این پیشامد:

$$\begin{cases} \text{اولاً، باید در تمام } x-1 \text{ آزمایش قبلی با شکست مواجه شده باشیم: } q^{x-1} \\ \text{ثانیاً، در آزمایش } x \text{ام به موفقیت برسیم (اولین موفقیت): } p \end{cases}$$

به عبارت دیگر:

$$\text{احتمال وقوع اولین موفقیت در } x\text{امین آزمایش} = \underbrace{\overbrace{\boxed{q^{x-1}}}^{x-1 \text{ آزمایش قبلی}}}_{x-1 \text{ شکست}} \cdot \underbrace{\overbrace{\boxed{p}}^{\text{آزمایش } x\text{ام}}}_{\text{اولین موفقیت}} = q^{x-1}p$$

تعریف: هرگاه یک آزمایش برنولی با احتمال موفقیت p را به طور مستقل آن‌قدر تکرار کنیم تا به **اولین موفقیت** برسیم (بالاخره موفق شویم)، آن‌گاه متغیر تصادفی «X : **تعداد آزمایشات** لازم برای رسیدن به **اولین موفقیت**» $(x = 1, 2, \ldots)$ دارای توزیع **هندسی** خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع هندسی باشد، پارامتر آن p است و به صورت $X \sim G(p)$ یا $X \sim Ge(p)$ نمایش داده می‌شود.

تابع احتمال	$P(x) = q^{x-1} \cdot p$ $x = 1, 2, \ldots$	احتمال وقوع اولین موفقیت در xامین آزمایش (احتمال آنکه در آزمایش xام بالاخره موفق شویم)
امید ریاضی (میانگین)	$E(X) = \frac{1}{p}$	میانگین تعداد آزمایش مورد انتظار برای رسیدن به اولین موفقیت
واریانس	$\sigma_X^2 = \frac{q}{p^2}$	واریانس تعداد آزمایش لازم برای رسیدن به اولین موفقیت
انحراف معیار	$\sigma_X = \sqrt{\frac{q}{p^2}}$	
تابع مولد گشتاور	$M_X(t) = \frac{pe^t}{1 - qe^t}$	

نکته: توزیع دوجمله‌ای منفی در حالت خاص $(r = 1)$ همان توزیع هندسی است، بنابراین توزیع هندسی حالت خاصی از توزیع دوجمله‌ای منفی است زمانی که به دنبال اولین موفقیت $(r = 1)$ هستیم.

محاسبه احتمال (هندسی)

$$\begin{cases} \text{الف) ابتدا وضعیتی را که به دنبال اولین وقوع آن هستیم، به عنوان موفقیت (با احتمال } p \text{) در نظر می‌گیریم.} \\ \text{ب) } P(x)=q^{x-1}p \text{، احتمال وقوع اولین موفقیت در } x \text{امین آزمایش است.} \end{cases}$$

با توجه به آنکه همواره $\sum P(x)=1$ (مجموع مقادیر تابع احتمال متغیر تصادفی گسسته برابر با 1 است)، برای تابع احتمال هندسی رابطه $\sum_{x=1}^{\infty} P(x)=\sum_{x=1}^{\infty} p\,q^{x-1}=1$ برقرار است و داریم:

$$\underbrace{p}_{P(X=1)}+\underbrace{qp}_{P(X=2)}+\underbrace{q^2p}_{P(X=3)}+\underbrace{q^3p}_{P(X=4)}+\cdots=1$$

برای مثال:

۱) سه آزمایش برای اولین موفقیت لازم باشد (سومین آزمایش، اولین موفقیت). $P(X=3)=q^2p$

۲) حداکثر 2 آزمایش برای اولین موفقیت لازم باشد. $P(X\le 2)=P(X=1)+P(X=2)=p+qp$

۳) حداقل 2 آزمایش برای اولین موفقیت لازم باشد.

روش اول:

$$P(X\ge 2)=1-P(X=1)=1-p=q$$

روش دوم:

$$P(X\ge 2)=P(X=2)+P(X=3)+\ldots=qp+q^2p+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{qp}{1-q}=q$$

۴) بیش از 2 آزمایش برای اولین موفقیت لازم باشد.

$$P(X>2)=P(X=3)+P(X=4)+\ldots=q^2p+q^3p+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{q^2p}{1-q}=q^2$$

۵) تعداد آزمایش لازم برای رسیدن به اولین موفقیت، فرد باشد.

$$P(\text{فرد } X)=P(X=1)+P(X=3)+P(X=5)+\ldots=p+q^2p+q^4p+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{p}{1-q^2}$$

یادآوری:

هر تصاعد هندسی به صورت زیر است:

$$\begin{cases} a_1, a_2, \ldots, a_n \rightarrow a_1, a_1q, a_1q^2, \ldots a_1q^{n-1} \\ q: \text{قدرنسبت} \end{cases}$$

و مجموع هر تصاعد هندسی را می‌توان به فرم زیر در نظر گرفت:

$$a_1+a_1q+a_1q^2+\ldots+a_1q^{n-1}=\frac{a_1\left(1-q^n\right)}{1-q}\xrightarrow[0<q<1]{n\to\infty}=\frac{a_1}{1-q}=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}$$

حل مسئله ۱ (درک مطلب):

الف) در این مسئله $p = 0.4$ (به هدف خوردن)

ب)

$$\begin{cases} P(x) = q^{x-1}p = (0.6)^4(0.4)^1 = 0.05184 \\ x = 5\ ,\ \ p = 0.4\ ,\ q = 0.6 \end{cases}$$

حل مسئله ۲ (درک مطلب):

الف) در این مسئله $p = 0.1$ (معیوب بودن)

ب)

$$\begin{cases} P(x) = q^{x-1}p = (0.9)^2(0.1)^1 = 0.081 \\ x = 3\ ,\ \ p = 0.1\ ,\ q = 0.9 \end{cases}$$

حل مسئله ۳ (درک مطلب):

با توجه به اینکه انتخاب با جایگذاری است، احتمال موفقیت (سیاه بودن) ثابت بوده و توزیع تعداد تکرار برای رسیدن به اولین موفقیت، هندسی است.

$$\begin{cases} E(X) = \dfrac{1}{p} = \dfrac{1}{\frac{5}{15}} = 3 \\ p = P(\text{سیاه}) = \dfrac{5}{15}\ \ ,\ \ q = P(\text{سفید}) = \dfrac{10}{15} \end{cases}$$

مثال ۱ احتمال آنکه دانشجویی یک تست را درست پاسخ دهد $\frac{1}{4}$ است. احتمال آنکه چهارمین تستی که دانشجو پاسخ می‌دهد، اولین تست درست او باشد، چقدر است؟

حل:

الف) در این مسئله $p = \frac{1}{4}$ (پاسخ درست دادن)

ب)

$$\begin{cases} P(x) = q^{x-1}p = \left(\dfrac{3}{4}\right)^3\left(\dfrac{1}{4}\right)^1 = \dfrac{27}{256} \\ x = 4\ ,\ \ p = \dfrac{1}{4}\ ,\ q = \dfrac{3}{4} \end{cases}$$

مثال ۲ احتمال آنکه یک راننده از چراغ قرمز عبور کند و پلیس متوجه نشود $\frac{1}{3}$ است. احتمال آنکه در سومین چهارراه برای اولین بار پلیس متوجه شود، چقدر است؟

حل:

الف) در این مسئله $p = \frac{2}{3}$ (متوجه شدن)

ب)

$$\begin{cases} P(x) = q^{x-1}p = \left(\dfrac{1}{3}\right)^2\left(\dfrac{2}{3}\right)^1 = \dfrac{2}{27} \\ x = 3\ ,\ \ p = \dfrac{2}{3}\ ,\ q = \dfrac{1}{3} \end{cases}$$

مثال ۳ هنگام تهیه یک فیلم سینمایی، احتمال اینکه بازیگری نقش خود را در هر دور فیلمبرداری درست بازی کند 0.4 است. احتمال اینکه در چهارمین دور فیلمبرداری برای اولین بار نقش خود را درست بازی کند، چقدر است؟

حل:

الف) در این مسئله $p = 0.4$ (درست بازی کردن)

ب)

$$\begin{cases} P(x) = q^{x-1}p = (0.6)^3(0.4) = 0.0864 \\ x = 4\ ,\ p = 0.4\ ,\ q = 0.6 \end{cases}$$

مثال ۴ 0.20 محصولات کارخانه‌ای معیوب است. اگر X تعداد آزمایش مورد نیاز برای رسیدن به اولین کالای معیوب باشد، متوسط (امید) تعداد انتخاب کدام است؟

۱) 5　　۲) 4　　۳) 2.5　　۴) 3

حل: گزینه ۱ درست است.

تعداد انتخاب برای رسیدن به اولین کالای معیوب دارای توزیع هندسی با پارامتر $p = 0.2$ بوده و متوسط تعداد آزمایش مورد نیاز برابر است با:

$$E(X) = \frac{1}{p} = \frac{1}{0.2} = 5$$

مثال ۵ در یک جعبه با 5 مهره قرمز و 15 مهره آبی، مهره‌ها را با جایگذاری خارج می‌کنیم تا به اولین مهره قرمز برسیم. امید و واریانس تعداد مهره‌های انتخابی کدام است؟

۱) $3\ ,\ 4$　　۲) $12\ ,\ 4$　　۳) $3\ ,\ \frac{4}{3}$　　۴) $12\ ,\ \frac{4}{3}$

حل: گزینه ۲ درست است.

ازآنجاکه نمونه‌گیری با جایگذاری از جامعه محدود انجام می‌شود، شرایط قانون برنولی (داشتن احتمال ثابت) برقرار است ($\frac{5}{20}$= احتمال قرمز بودن و $\frac{15}{20}$= احتمال آبی بودن).

از طرفی تعداد انتخاب برای رسیدن به اولین مهره قرمز دارای توزیع هندسی با پارامتر $p = \frac{5}{20}$ است.

$$\begin{cases} E(X) = \dfrac{1}{p} = \dfrac{1}{\frac{1}{4}} = 4 \\ \sigma_X^2 = \dfrac{q}{p^2} = \dfrac{\frac{3}{4}}{\left(\frac{1}{4}\right)^2} = 12 \\ p = P(\text{قرمز}) = \dfrac{5}{20} = \dfrac{1}{4} = 0.25\ ,\quad q = P(\text{آبی}) = \dfrac{15}{20} = \dfrac{3}{4} = 0.75 \end{cases}$$

مثال ۶ یک تاس سالم را آن‌قدر پرتاب می‌کنیم تا برای اولین بار نتیجه زوج ظاهر شود $(6,4,2)$. مطلوب است احتمال آنکه:

الف) حداقل 2 پرتاب برای رسیدن به اولین نتیجه زوج لازم باشد.

ب) تعداد پرتاب‌های لازم برای رسیدن به اولین نتیجه زوج مضربی از 3 باشد.

ج) تعداد پرتاب‌های لازم برای رسیدن به اولین نتیجه زوج، عددی فرد باشد.

حل:

در این مثال « X : تعداد پرتاب لازم برای رسیدن به اولین نتیجه زوج» دارای توزیع هندسی است و داریم:

$$\begin{cases} f(x)=P(x)=q^{x-1}p \quad ; \quad x=1,2,\ldots \\ p=P(\text{نتیجه زوج})=\frac{3}{6}=\frac{1}{2}\ ,\ q=\frac{1}{2} \end{cases}$$

الف)

$$P(X\ge 2)=P(X=2)+P(X=3)+\ldots=qp+q^2p+\ldots=\frac{qp}{1-q}=\frac{qp}{p}=q=\frac{1}{2}$$

ب)

$$P(3\ \text{مضرب}\ X)=P(X=3)+P(X=6)+P(X=9)+\ldots=q^2p+q^5p+q^8p+\ldots=\frac{q^2p}{1-q^3}=\frac{\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)}{1-\left(\frac{1}{2}\right)^3}=\frac{1}{7}$$

ج)

$$P(\text{فرد}\ X)=P(X=1)+P(X=3)+P(X=5)+\ldots=\underbrace{\cancel{q^0}}_{1}p+q^2p+q^4p+\ldots=\frac{p}{1-q^2}=\frac{\frac{1}{2}}{1-\left(\frac{1}{2}\right)^2}=\frac{2}{3}$$

توزیع فوق هندسی (Hypergeometric Distribution)

مسئله (درک مطلب): از جعبه‌ای با 12 کالا که 4 تای آن سالم و 8 تا معیوب است، 3 کالا به‌تصادف انتخاب شده است. احتمال آنکه 2 کالا سالم باشد چقدر است؟

مقدمه: جامعه محدود N تایی را در نظر بگیرید که k تای آن موفقیت و $N-k$ تای دیگر شکست تلقی شود. حال اگر یک نمونه n تایی را از جامعه انتخاب کنیم، توزیع «تعداد موفقیت در نمونه» بسته به آنکه نمونه‌گیری «بدون جایگذاری» یا «با جایگذاری» باشد با هم متفاوت است.

۱- با جایگذاری: در این وضعیت احتمال موفقیت در هر بار نمونه‌گیری مستقل، ثابت و برابر با $p=\frac{k}{N}$ است، درنتیجه آزمایش، برنولی و توزیع «تعداد موفقیت در نمونه» همان‌طورکه قبلاً بررسی شده بود، دوجمله‌ای است.

$$\text{احتمال موفقیت} = \underbrace{\overset{\text{بار اول}}{\frac{k}{N}}}_{p} \times \underbrace{\overset{\text{بار دوم}}{\frac{k}{N}}}_{p} \times \cdots$$

۲- بدون جایگذاری (پیش‌فرض): در این وضعیت احتمال موفقیت در هر بار نمونه‌گیری، وابسته به دفعات قبل است، زیرا در هر بار انتخاب، حجم جامعه 1 واحد کم می‌شود و احتمال موفقیت، دیگر ثابت نیست.

$$\text{احتمال موفقیت} = \overset{\text{بار اول}}{\frac{k}{N}} \times \overset{\text{بار دوم}}{\frac{k \text{ یا } k-1}{N-1}} \times \cdots$$

در چنین شرایطی برای محاسبه احتمال از همان رابطه کلاسیک فصل احتمال یعنی:

$$\text{احتمال} = \frac{\text{حالات مساعد}}{\text{حالات ممکن}}$$

استفاده می‌کنیم. درعین‌حال برای محاسبه حالات مساعد و ممکن از ترکیب استفاده می‌کنیم.

یادآوری:

۱- در صورت عدم بیان، نوع انتخاب به طور **پیش‌فرض بدون جایگذاری** در نظر گرفته می‌شود.

۲- ترکیب $\binom{N}{n}$ حالات انتخاب یک نمونه n تایی (بدون جایگذاری) از یک جامعه N تایی است.

برای مثال در مسئله بالا حالات ممکن برای انتخاب 3 کالا از بین 12 کالا برابر است با $\binom{12}{3}$ که با توجه به 4 کالای سالم و 8 کالای معیوب، وضعیت‌های متفاوتی برای انتخاب به شرح زیر وجود دارد:

$$\underset{\text{حالات ممکن}}{\binom{12}{3}} = \underset{\text{هر 3 کالا سالم}}{\binom{4}{3}\binom{8}{0}} + \underset{\text{2 سالم و 1 معیوب}}{\binom{4}{2}\binom{8}{1}} + \underset{\text{1 سالم و 2 معیوب}}{\binom{4}{1}\binom{8}{2}} + \underset{\text{هر 3 کالا معیوب}}{\binom{4}{0}\binom{8}{3}}$$

در این وضعیت احتمال 2 کالای سالم در نمونه به صورت زیر است:

حل مسئله (درک مطلب):

$$P(X=2)=\frac{\binom{4}{2}\binom{8}{1}}{\binom{12}{3}}$$

تعریف: هرگاه از یک **جامعه محدود** N تایی که K تای آن موفقیت و $N-K$ تای آن شکست است، یک نمونه n تایی **بدون جایگذاری** انتخاب کنیم، آن‌گاه « X : تعداد موفقیت در نمونه n تایی» دارای توزیع **فوق هندسی** است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع فوق هندسی باشد، پارامترهای آن N، k و n هستند و به صورت $X \sim HG(N, k, n)$ نمایش داده می‌شود.

احتمال آنکه در یک نمونه n تایی (بدون جایگذاری)، x تا متعلق به مجموعه k (موفقیت) باشد.	$P(x)=\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ $x=0,1,2,....,\min(n,k)$	**تابع احتمال**
متوسط تعداد موفقیت در نمونه n تایی با احتمال موفقیت $p=\dfrac{K}{N}$	$E(X)=n\cdot p$	**امید ریاضی(میانگین)**
واریانس تعداد موفقیت در نمونه n تایی با احتمال موفقیت $p=\dfrac{K}{N}$ ضریب تصحیح واریانس $=\dfrac{N-n}{N-1}$	$\sigma_X^2=\dfrac{N-n}{N-1}\cdot n\cdot p\cdot q$	**واریانس**
	$\sigma_X=\sqrt{\dfrac{N-n}{N-1}\cdot n\cdot p\cdot q}$	**انحراف معیار**

مثال ۱ می‌خواهیم در یک مدرسه با 5 معلم مرد و 4 معلم زن یک کمیته 3 نفری انتخاب کنیم. اگر متغیر تصادفی X تعداد معلم زن در کمیته مورد نظر باشد، تابع توزیع احتمال آن کدام است؟

۱) $f(x)=\binom{9}{x}\left(\frac{4}{9}\right)^x\left(\frac{5}{9}\right)^{9-x}$

۲) $f(x)=\dfrac{\binom{4}{x}\binom{5}{3-x}}{\binom{9}{3}}$

۳) $f(x)=\dfrac{\binom{4}{x}\binom{5}{3}}{\binom{9}{3}}$

۴) هیچ‌کدام

حل: گزینه ۲ درست است.

انتخاب نمونه (n تایی) بدون جایگذاری از جامعه محدود (N تایی) دارای توزیع فوق هندسی است و داریم:

$$\begin{cases} f(x)=\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}=\dfrac{\binom{4}{x}\binom{5}{3-x}}{\binom{9}{3}} \\ N=9\ ,\ n=3\ ,\ k=4\ ,\ N-k=5 \end{cases}$$

مثال ۲ از 9 عدد کالای یکسان موجود در یک کارتن، 3 عدد معیوب است. 4 کالا به طور تصادفی از بین آن‌ها برداشته می‌شود. با کدام احتمال لااقل سه کالای برداشته‌شده، سالم است؟ (حسابداری ـ ۸۴)

۱) $\frac{8}{21}$ ۲) $\frac{13}{21}$ ۳) $\frac{17}{42}$ ۴) $\frac{25}{42}$

حل: گزینه ۴ درست است.

در این مثال جامعه محدود است $(N=9)$ و داریم $k=6$ (کالای سالم) و $n=4$ (نمونه پیش‌فرض بدون جایگذاری)، درنتیجه توزیع فوق هندسی است و داریم:

$$\begin{cases} P(X \geq 3) = P(X=3) + P(X=4) = \dfrac{\binom{6}{3}\binom{3}{1}}{\binom{9}{4}} + \dfrac{\binom{6}{4}\binom{3}{0}}{\binom{9}{4}} = \dfrac{60+15}{126} = \dfrac{25}{42} \\ N=9 , n=4 , k=6 \end{cases}$$

مثال ۳ از جوراب‌های بسته‌بندی‌شده در جعبه‌ای 9 عدد سالم و 3 عدد معیوب است. یک مشتری به طور تصادفی 4 عدد را خریداری می‌کند. میانگین و واریانس تعداد جوراب‌های معیوب در این خرید به ترتیب از چپ به راست چقدر است؟ (اقتصاد ـ ۷۳)

۱) $1 , \frac{6}{11}$ ۲) $1 , \frac{3}{4}$ ۳) $1 , \frac{27}{12}$ ۴) $2 , \frac{9}{12}$

حل: گزینه ۱ درست است.

در این مثال جامعه محدود است $(N=12)$ و داریم $k=3$ (کالای معیوب) و $n=4$ (نمونه پیش‌فرض بدون جایگذاری)، درنتیجه توزیع فوق هندسی است و داریم:

$$\begin{cases} E(X) = np = 4 \times \dfrac{3}{12} = 1 \\ Var(X) = npq\left(\dfrac{N-n}{N-1}\right) = 4 \times \dfrac{3}{12} \times \dfrac{9}{12} \times \dfrac{12-4}{12-1} = \dfrac{6}{11} \\ N=12 \quad , \quad n=4 \quad , \quad k=3 \quad , \quad p = \dfrac{k}{N} = \dfrac{3}{12} \end{cases}$$

مثال ۴ از 10 محصول تولیدی به وسیله یک ماشین، 3 واحد آن معیوب است. یک نمونه 2 تایی از محصولات این ماشین، انتخاب شده است. احتمال اینکه هیچ‌کدام سالم نباشد، چقدر است؟ (مدیریت ـ ۷۹)

۱) صفر ۲) $\frac{49}{90}$ ۳) $\frac{9}{100}$ ۴) $\frac{1}{15}$

حل: گزینه ۴ درست است.

در این مثال جامعه محدود است $(N=10)$ و داریم $k=3$ (کالای معیوب) و $n=2$ (نمونه پیش‌فرض بدون جایگذاری)، درنتیجه توزیع فوق هندسی است و داریم:

$$\begin{cases} P(X=2) = \dfrac{\binom{3}{2}\binom{7}{0}}{\binom{10}{2}} = \dfrac{3}{45} = \dfrac{1}{15} \\ N=10 , n=2 , k=3 \end{cases}$$

مثال ۵ از یک مجموعه شامل 3 کالای معیوب و 7 کالای سالم، یک نمونه 4 تایی انتخاب کرده‌ایم. احتمال آنکه حداقل 3 کالا سالم باشد، چقدر است؟

حل:
دقت کنید کل حالات انتخاب به شرح زیر است:

$$\binom{10}{4}=\underbrace{\binom{7}{0}\cancel{\binom{3}{4}}}_{\text{غیرممکن}}+\binom{7}{1}\binom{3}{3}+\binom{7}{2}\binom{3}{2}+\underbrace{\binom{7}{3}\binom{3}{1}+\binom{7}{4}\binom{3}{0}}_{\text{حداقل 3 سالم}}$$

همان‌طورکه مشاهده می‌شود به هیچ وجه امکان انتخاب $\binom{7}{0}\binom{3}{4}$ وجود ندارد و جواب سؤال $\dfrac{\binom{7}{3}\binom{3}{1}+\binom{7}{4}\binom{3}{0}}{\binom{10}{4}}$ است

ضریب تصحیح واریانس

کمیت $\dfrac{N-n}{N-1}$ به عنوان ضریب تصحیح برای واریانس توزیع فوق هندسی به کار برده می‌شود و همان‌طور که در ادامه خواهیم دید، در بعضی شرایط $\left(\dfrac{n}{N}\le 0.05\right)$، از این ضریب چشم‌پوشی می‌شود.

تقریب توزیع فوق هندسی به دوجمله‌ای

زمانی که N (حجم جامعه) بزرگ و n (حجم نمونه) کوچک شود (بنا بر قانون سرانگشتی n از 5 درصد N تجاوز نکند)، تفاوت چندانی بین نمونه‌گیری بدون جایگذاری و نمونه‌گیری با جایگذاری وجود ندارد؛ در این وضعیت می‌توان برای تقریب احتمال‌های توزیع فوق هندسی از توزیع دوجمله‌ای پارامترهای $n\,,\,p=\dfrac{K}{N}$ استفاده کرد. بدیهی است در این شرایط از ضریب تصحیح $\dfrac{N-n}{N-1}$ برای واریانس نیز چشم‌پوشی می‌شود.

قضیه: هرگاه در یک توزیع فوق هندسی، حجم نمونه (n)، را کاهش و حجم جامعه (N) را افزایش دهیم به طوری که $\dfrac{n}{N}\le 0.05$ آن‌گاه توزیع دوجمله‌ای تقریب مناسبی برای توزیع فوق هندسی است و از ضریب تصحیح واریانس $\dfrac{N-n}{N-1}$ چشم‌پوشی می‌کنیم. واضح است در این شرایط روابط زیر برقرار است:

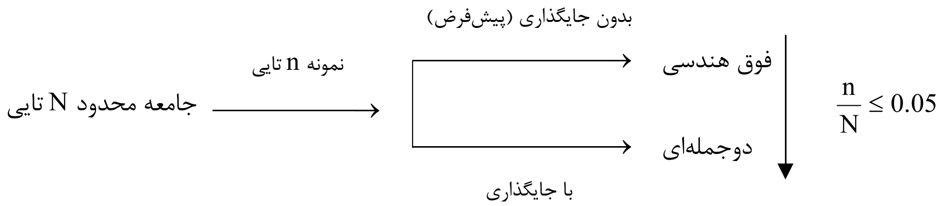
$$\frac{n}{N}\le\frac{5}{100}\;\rightarrow\;\begin{cases} n\le \%5N & \text{نمونه از \%5 جامعه تجاوز نمی‌کند.}\\ 20n\le N & \text{جامعه حداقل 20 برابر نمونه است.}\end{cases}$$

✓ **دقت کنید!**

هرگاه $\dfrac{n}{N}>0.05$ یا $n>0.05N$ باشد، تقریب دوجمله‌ای دیگر مناسب نیست و از ضریب تصحیح $\dfrac{N-n}{N-1}$ در واریانس چشم‌پوشی نمی‌شود.

نتیجه:

برای تعیین توزیع «**تعداد موفقیت در نمونه**» همواره:



مثال ۱ در کدام‌یک از موارد زیر می‌توان از توزیع دوجمله‌ای برای تقریب توزیع فوق هندسی استفاده کرد؟

۱) K و N هر دو بزرگ باشند.
۲) K بزرگ ولی N کوچک باشد.
۳) N بزرگ و n کوچک باشد.
۴) K و N هر دو کوچک باشند.

حل: گزینه ۳ درست است.

مثال ۲ از یک جامعه 4000 نفره یک نمونه تصادفی 40 تایی انتخاب شده است. در این حالت تابع احتمال متغیر تصادفی X، کدام است؟ (حسابداری ـ ۸۱)

۱) هندسی
۲) دوجمله‌ای
۳) فوق هندسی
۴) هم فوق هندسی و هم دوجمله‌ای

حل: گزینه ۴ درست است.

اولاً، ازآنجاکه جامعه محدود است $(N = 4000)$ و نمونه $(n = 40)$ به طور پیش‌فرض بدون‌جایگذاری انتخاب شده، توزیع فوق هندسی است.

ثانیاً، با توجه به آنکه $\frac{n}{N} = \frac{40}{4000} = 0.01 \le 0.05$ است، توزیع دوجمله‌ای تقریب مناسبی برای توزیع فوق هندسی است.

ثالثاً، اگر گزینه ۴ وجود نداشت گزینه ۲ را انتخاب می‌کردیم زیرا دوجمله‌ای تقریب فوق هندسی است و در شرایط تقریب، محاسبه تقریب توزیع آسان‌تر است.

دقت کنید که اگر $\frac{n}{N} > 0.05$ بود، آن‌گاه تقریب دوجمله‌ای مناسب نبود و گزینه ۳ پاسخ درست بود.

مثال ۳ از یک جعبه محتوی 4 کالای سالم و 6 کالای معیوب، یک نمونه 3 تایی (بدون جایگذاری) خارج می‌کنیم. در این حالت توزیع X، تعداد کالاهای سالم در نمونه، کدام است؟

۱) دوجمله‌ای
۲) فوق هندسی
۳) هندسی
۴) دوجمله‌ای منفی

حل: گزینه ۲ درست است.

انتخاب نمونه (n تایی) بدون جایگذاری از جامعه محدود (N تایی) دارای توزیع فوق هندسی است. (البته اگر در سؤال کلمه بدون جایگذاری عنوان نمی‌شد باز هم پاسخ سؤال، توضیح فوق هندسی بود چون پیش‌فرض در این سؤالات انتخاب بدون جایگذاری است.)

دقت کنید اگر نمونه‌گیری با جایگذاری انجام می‌شد، گزینه ۱ پاسخ درست بود.

توزیع پواسون (Poisson Distribution)

مسئله ۱ (درک مطلب): به طور متوسط در هر ساعت، 12 اتومبیل به یک پمپ بنزین مراجعه می‌کنند. احتمال اینکه در 15 دقیقه 3 اتومبیل مراجعه کنند، چقدر است؟

مسئله ۲ (درک مطلب): به طور متوسط در هر 200 متر پارچه، 3 زدگی وجود دارد. احتمال آنکه در یک بسته 600 متری 1 زدگی وجود داشته باشد، چقدر است؟

مقدمه: گاهی **تعداد اتفاقات** (رویداد، رخداد) در یک بازه **مکانی یا زمانی** مطرح است، مانند:

تعداد اتفاقات در واحد زمان:
- **تعداد** تماس‌های تلفنی در **یک ساعت**
- **تعداد** مشتریان در **یک روز**
- **تعداد** خرابی‌های اتومبیل در **یک سال**
- **تعداد** مسافران هواپیما در **شش ماه**

تعداد اتفاقات در واحد مکان:
- **تعداد** اتومبیل‌ها در 2 **کیلومتر**
- **تعداد** زدگی‌ها در 100 **متر** پارچه
- **تعداد** غلط‌های املایی در 10 **صفحه**
- **تعداد** نقطه‌ها در 100 **سانتی‌متر**

در این وضعیت با استفاده از توزیعی به نام پواسون به تشریح احتمال مسایل می‌پردازیم.

تعریف: فرض کنید تعداد اتفاقات در یک فاصله مشخص از زمان یا مکان مورد نظر باشد؛ در این صورت «X : **تعداد اتفاقات** در یک بازه **زمانی** یا **مکانی**» $(x = 0,1,2,\dots)$ دارای **توزیع پواسون** است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع پواسون باشد، پارامتر آن λ است و به صورت $X \sim P(\lambda)$ نمایش داده می‌شود.

احتمال x اتفاق در یک بازه زمانی یا مکانی با متوسط λ	$P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $x = 0,1,2,\dots$	**تابع احتمال**
متوسط تعداد اتفاقات در بازه زمانی یا مکانی	$E(X) = \lambda$	**امید ریاضی (میانگین)**
واریانس تعداد اتفاقات در بازه زمانی یا مکانی	$\sigma_X^2 = \lambda$	**واریانس**
	$\sigma_X = \sqrt{\lambda}$	**انحراف معیار**
	$M_X(t) = e^{-\lambda(1-e^t)}$	**تابع مولد گشتاور**

توجه: توزیع پواسون تنها توزیعی است که در آن میانگین و واریانس توزیع با هم برابرند و هر دو مساوی پارامتر توزیع (λ) هستند.

پارامتر پواسون

متوسط تعداد اتفاقات در هر **بازه زمانی یا مکانی** به عنوان پارامتر توزیع پواسون شناخته شده و با نماد λ نمایش داده می‌شود.

اگر $0 \le \lambda \le 10$ باشد، امکان استفاده از توزیع پواسون برای حل مسایل مناسب است، در غیر این صورت $(\lambda > 10)$ بهتر است از تقریب نرمال که بعداً در توزیع نرمال بررسی می‌شود، استفاده شود.

محاسبه احتمال در پواسون

الف) مقدار λ (پارامتر پواسون) را با توجه به زمان یا مکان مشخص می‌کنیم؛ درصورتی‌که زمان یا مکان تغییر کند، با استفاده از تناسب مقدار λ را به دست می‌آوریم.

برای مثال، اگر $\lambda = 2$ مشتری در دقیقه باشد، در 20 ثانیه داریم:

زمان	متوسط تعداد مشتری (λ)
1 دقیقه = 60 ثانیه	2
20 ثانیه	?

$$\Rightarrow \lambda = \frac{2\times 20}{60} = \frac{2}{3}$$

و در 5 دقیقه خواهیم داشت:

زمان	متوسط تعداد مشتری (λ)
1 دقیقه	2
5 دقیقه	?

$$\Rightarrow \lambda = \frac{2\times 5}{1} = 10$$

ب) $\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \\ x = 0,1,2,\dots \end{cases}$، احتمال وقوع x اتفاق در واحد زمان یا مکان است.

با توجه به آنکه همواره $\sum P(x) = 1$ (مجموع مقادیر تابع احتمال متغیر تصادفی گسسته برابر با 1 است)، برای تابع احتمال پواسون رابطه $\sum_{x=0}^{\infty} P(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = 1$ برقرار است و داریم:

$$\underbrace{\frac{e^{-\lambda}\lambda^0}{0!}}_{P(X=0)} + \underbrace{\frac{e^{-\lambda}\lambda^1}{1!}}_{P(X=1)} + \underbrace{\frac{e^{-\lambda}\lambda^2}{2!}}_{P(X=2)} + \cdots = 1$$

$$\begin{cases} \text{احتمال عدم وقوع اتفاق} = P(X=0) = \dfrac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} \\ \text{احتمال وقوع 1 اتفاق} = P(X=1) = \dfrac{e^{-\lambda}\lambda^1}{1!} = \lambda e^{-\lambda} \\ \text{احتمال وقوع حداکثر 1 اتفاق} = P(X\le 1) = P(X=0) + P(X=1) = (\lambda+1)e^{-\lambda} \\ \text{احتمال وقوع حداقل 1 اتفاق} = P(X\ge 1) = 1 - P(X=0) = 1 - e^{-\lambda} \\ \text{احتمال وقوع بیش از 1 اتفاق} = P(X>1) = 1 - P(X=0) - P(X=1) = 1 - (\lambda+1)e^{-\lambda} \\ \vdots \end{cases}$$

حل مسئله ۱ (درک مطلب):

به طور متوسط $\lambda = 12$ اتومبیل در هر ساعت به پمپ بنزین مراجعه می‌کنند، ولی احتمال در 15 دقیقه خواسته شده است؛ بنابراین:

زمان	متوسط تعداد اتومبیل (λ)
1 ساعت = 60 دقیقه	12
15 دقیقه	?

$$\Rightarrow \lambda = \frac{12\times 15}{60} = 3$$

یعنی به طور متوسط $\lambda = 3$ اتومبیل در 15 دقیقه به پمپ بنزین مراجعه می‌کنند؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \xrightarrow[\lambda=3]{x=3} P(X=3) = \dfrac{e^{-3}3^{3}}{3!} = \dfrac{27e^{-3}}{6} = \dfrac{9}{2}e^{-3} = 4.5e^{-3} \\ X : \text{تعداد اتومبیل‌هایی که در 15 دقیقه به پمپ بنزین مراجعه می‌کنند} \quad , \quad \lambda = 3 \end{cases}$$

حل مسئله ۲ (درک مطلب):

به طور متوسط $\lambda = 3$ زدگی در 200 متر پارچه دیده می‌شود، ولی احتمال در بسته‌های 600 متری خواسته شده است؛ بنابراین:

زمان	متوسط تعداد زدگی (λ)
200 متر	3
600 متر	?

$$\Rightarrow \lambda = \frac{3\times 600}{200} = 9$$

یعنی به طور متوسط $\lambda = 9$ زدگی در 600 متر پارچه خواهد بود؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \xrightarrow[\lambda=9]{x=1} P(X=1) = \dfrac{e^{-9}9^{1}}{1!} = 9e^{-9} \\ X : \text{تعداد زدگی‌ها در یک بسته 600 متری} \quad , \quad \lambda = 9 \end{cases}$$

مثال ۱ تعداد مشتریان یک فروشگاه در ساعت خاصی از روز دارای توزیع پواسون با متوسط 4 نفر در ساعت است.مطلوب است محاسبه:

الف) احتمال آنکه در یک ساعت هیچ مشتری وارد نشود.

ب) احتمال آنکه در نیم ساعت 3 مشتری وارد فروشگاه شود.

ج) احتمال آنکه در 5 ساعت حداکثر یک مشتری وارد فروشگاه شود.

د) احتمال آنکه در یک ربع ساعت حداقل یک مشتری وارد فروشگاه شود.

ه) تابع احتمال تعداد مشتریان در 10 ساعت.

حل:

الف) $\lambda = 4$ نفر در ساعت است و با توجه به آنکه احتمال در ساعت خواسته شده، λ تغییری نمی‌کند؛ بنابراین:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \xrightarrow[\lambda=4]{x=0} P(X=0) = \dfrac{e^{-4}4^{0}}{0!} = e^{-4} \\ X : \text{تعداد مشتری‌هایی که در یک ساعت وارد می‌شوند} \quad , \quad \lambda = 4 \end{cases}$$

ب) $\lambda = 4$ نفر در ساعت است ولی احتمال در نیم ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
$\frac{1}{2}$ ساعت	?

$$\Rightarrow \lambda = \frac{4\times\frac{1}{2}}{1} = 2$$

یعنی به طور متوسط $\lambda = 2$ نفر در نیم‌ساعت وارد می‌شوند؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \xrightarrow[\lambda=2]{x=3} P(X=3) = \dfrac{e^{-2}2^{3}}{3!} = \dfrac{8}{6}e^{-2} = \dfrac{4}{3}e^{-2} \\ X : \text{تعداد مشتری‌هایی که در نیم‌ساعت وارد می‌شوند} \quad , \quad \lambda = 2 \end{cases}$$

ج) $\lambda = 4$ نفر در ساعت است اما احتمال در 5 ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
5 ساعت	?

$$\Rightarrow \lambda = \frac{4\times 5}{1} = 20$$

یعنی به طور متوسط $\lambda = 20$ نفر در 5 ساعت وارد می‌شود؛ درنتیجه:

$$\begin{cases} P(X \le 1) = P(X=0) + P(X=1) = \dfrac{e^{-\lambda}\lambda^0}{0!} + \dfrac{e^{-\lambda}\lambda^1}{1!} = e^{-20} + 20e^{-20} = 21e^{-20} \\ X: \text{تعداد مشتری‌هایی که در 5 ساعت وارد می‌شوند} \quad , \ \lambda = 20 \end{cases}$$

د) $\lambda = 4$ نفر در ساعت است اما احتمال در یک ربع ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
$\frac{1}{4}$ ساعت	?

$$\Rightarrow \ \lambda = \frac{4 \times \frac{1}{4}}{1} = 1$$

یعنی به طور متوسط $\lambda = 1$ نفر در یک ربع ساعت وارد می‌شود؛ در نتیجه:

$$\begin{cases} P(X \ge 1) = 1 - P(X=0) = 1 - \dfrac{e^{-\lambda}\lambda^0}{0!} = 1 - e^{-1} \\ X: \text{تعداد مشتری‌هایی که در یک ربع ساعت وارد می‌شوند} \quad , \ \lambda = 1 \end{cases}$$

ه) $\lambda = 4$ نفر در ساعت است و تابع احتمال در 10 ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
10 ساعت	?

$$\Rightarrow \ \lambda = \frac{4 \times 10}{1} = 40$$

یعنی به طور متوسط $\lambda = 40$ نفر در 10 ساعت وارد می‌شود؛ در نتیجه:

$$\begin{cases} f(x) = P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-40}40^x}{x!} \\ X: \text{تعداد مشتری‌هایی که در 10 ساعت وارد می‌شوند} \quad , \ \lambda = 40 \end{cases}$$

مثال ۲ تعداد اشتباهات یک حروف‌چین در حروف‌چینی یک متن 500 سطری 2 اشتباه است. احتمال اینکه در 100 سطر اول اشتباهی وجود نداشته باشد، چقدر است؟

۱) e^{-2} ۲) $e^{-\frac{2}{5}}$ ۳) $1 - e^{-\frac{2}{5}}$ ۴) $e^{-\frac{5}{2}}$

حل: گزینه ۲ درست است.

$\lambda = 2$ اشتباه در 500 سطر است و احتمال در 100 سطر خواسته شده؛ بنابراین:

مکان	متوسط تعداد اشتباهات (λ)
500 سطر	2
100 سطر	?

$$\Rightarrow \ \lambda = \frac{2 \times 100}{500} = \frac{2}{5}$$

یعنی به طور متوسط $\lambda = \frac{2}{5}$ اشتباه در 100 سطر وجود دارد؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda = \frac{2}{5}]{x=0} P(X=0) = \dfrac{e^{-\frac{2}{5}} \left(\frac{2}{5}\right)^0}{0!} = e^{-\frac{2}{5}} \\ X: \text{تعداد اشتباهات حروف‌چینی در 100 سطر} \quad , \ \lambda = \frac{2}{5} \end{cases}$$

مثال ۳ در یک توزیع پواسون $P(X=1)=P(X=2)$ برقرار است. ضریب تغییرات متغیر تصادفی X چند درصد است؟

۱) 50 ۲) $50\sqrt{2}$ ۳) $200\sqrt{2}$ ۴) 100

حل: گزینه ۲ درست است.

ابتدا با توجه به تابع احتمال پواسون $P(x)=\frac{e^{-\lambda}\lambda^x}{x!}$ داریم:

$$P(X=1)=P(X=2) \rightarrow \frac{\cancel{e^{-\lambda}}\lambda^1}{1!}=\frac{\cancel{e^{-\lambda}}\lambda^2}{2!} \rightarrow \lambda=2$$

حال برای محاسبه درصد ضریب تغییرات (پراکندگی) به صورت زیر عمل می‌کنیم:

$$\begin{cases} CV\times 100=\frac{\sigma}{\mu}\times 100=\frac{\sqrt{2}}{2}\times 100=50\sqrt{2} \\ \mu=\lambda=2 \\ \sigma=\sqrt{\lambda}=\sqrt{2} \end{cases}$$

تقریب توزیع دوجمله‌ای به پواسون

گاهی در یک توزیع دوجمله‌ای با پارامترهای n و p شرایطی پیش می‌آید که در آن تعداد تکرار آزمایش برنولی (n) زیاد و احتمال موفقیت در هر آزمایش (p) کم است، به طوری که یکی از حالات زیر اتفاق می‌افتد:

$$\begin{cases} \text{I)}\quad n\ge 20\ ,\ p\le 0.05 \\ \text{II)}\quad n\ge 100\ ,\ np\le 10 \end{cases}$$

بدیهی است در این شرایط محاسبه احتمال مشکل است. در این وضعیت می‌توانیم به‌جای استفاده از توزیع دوجمله‌ای از توزیع پواسون با پارامتر $\lambda=np$ بهره گیریم تا محاسبات ساده‌تر انجام شود.

قضیه: هرگاه در یک توزیع دوجمله‌ای با پارامترهای n و p، یکی از شرایط زیر برقرار باشد آن‌گاه توزیع پواسون با پارامتر $\lambda=np$ تقریب مناسبی برای توزیع دوجمله‌ای است.

$$X\sim Bin(n,p)\ \xrightarrow[(\text{II}):\,n\ge 100\,,\,np\le 10]{(\text{I}):\,n\ge 20,\,p\le 0.05}\ X\sim P(\lambda=np)$$

مثال ۱ در کدام‌یک از موارد زیر توزیع پواسون تقریب خوبی برای توزیع دوجمله‌ای است؟

۱) $p=0.32,\ n=40$ ۲) $p=0.68\ ,\ n=40$

۳) $p=0.98\ ,\ n=200$ ۴) $p=0.03\ ,\ n=20$

حل: گزینه ۴ درست است.

در گزینه ۱ با توجه به شرط اول $p=0.32\not\le 0.05$ و $n=40\ge 20$ تقریب مناسب نیست.

در گزینه ۲ با توجه به شرط اول $p=0.68\not\le 0.05$ و $n=40\ge 20$ تقریب مناسب نیست.

در گزینه ۳ با توجه به شرط دوم $np=196\not\le 10$ و $n=200\ge 100$ تقریب مناسب نیست.

در گزینه ۴ با توجه به شرط اول $p=0.03\le 0.05$ و $n=20\ge 20$ تقریب مناسب است.

مثال ۲ اگر یک توزیع دو جمله‌ای 100 تکرار داشته باشد و احتمال موفقیت در هر تکرار 0.01 باشد، بهترین توزیع برای تقریب احتمال‌های آن کدام است؟

۱) نمایی ۲) نرمال ۳) پواسون ۴) یکنواخت

حل: گزینه ۳ درست است.

با توجه به شرط دوم $n=100\ge 100\ ,\ np=100\times 0.01=1\le 10$ تقریب دوجمله‌ای به پواسون مناسب است.

مثال ۳ نسبت خرابی کالا در یک کارخانه برابر 0.01 است. احتمال آنکه در 100 کالا حداکثر یک کالای خراب وجود داشته باشد، چقدر است؟ (مدیریت ـ ۷۹)

۱) e^2 ۲) e^{-2} ۳) $2e$ ۴) $2e^{-1}$

حل: گزینه ۴ درست است.

با توجه به آنکه احتمال معیوب بودن در هر انتخاب $p = 0.01$ ثابت است و نمونه 100تایی از جامعه انتخاب شده است، توزیع تعداد کالای معیوب در نمونه دوجمله‌ای است.

اما شرایط تقریب پواسون با توجه به شرط دوم برقرار است، بنابراین محاسبه احتمال از طریق آن ساده‌تر است.

$$n = 100 \ge 100 \ , \ np = 100 \times 0.01 = 1 \le 10 \ \rightarrow \ \text{تقریب پواسون مناسب است} \ \rightarrow \ \lambda = np = 100 \times 0.01 = 1$$

$$P(X \le 1) = P(X = 0) + P(X = 1) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1} \times 1^0}{0!} + \frac{e^{-1} \times 1^1}{1!} = 2e^{-1} = 0.736$$

در این مسئله اگر به طور مستقیم برای محاسبه احتمال از توزیع دوجمله‌ای استفاده می‌کردیم، محاسبه بسیار مشکل‌تر بود زیرا:

$$P(X \le 1) = P(X = 0) + P(X = 1) = \binom{100}{0}(0.01)^0(0.99)^{100} + \binom{100}{1}(0.01)^1(0.99)^{99} = 0.736$$

با این حال پاسخ هر دوی آن‌ها یکسان است.

مثال ۴ طبق آمار سالیانه‌ای که اداره راهنمایی و رانندگی منتشر کرده است، از هر 100 هزار نفر به طور متوسط 3 نفر در اثر حوادث رانندگی کشته می‌شوند. در شهری با 200 هزار نفر جمعیت مطلوب است محاسبه احتمال آنکه:

الف) 4 نفر کشته شوند.

ب) کمتر از 3 نفر کشته شوند.

حل:

با توجه به شرط دوم:

$$n = 200000 \ge 100 \ , \ np = 200000 \times 0.00003 = 6 \le 10 \ \rightarrow \ \text{تقریب پواسون مناسب است.} \rightarrow \lambda = np = 6$$

الف) $P(X = 4) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-6}6^4}{4!}$

ب) $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \displaystyle\sum_{x=0}^{2} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-6} + 6e^{-6} + \frac{6^2 e^{-6}}{2!} = 25e^{-6}$

در تمام توزیع‌های گسسته $P(X = a) = f(X = a) \ge 0$ است.

توزیع یکنواخت پیوسته (Continuous Uniform Distribution)

مسئله (درک مطلب): درصورتی‌که تابع چگالی روبه‌رو در نظر گرفته شود: $f(x)=\frac{2}{3} \quad ; \quad -1 \le x < \frac{1}{2}$

الف) احتمال آنکه X در فاصله 0 تا $\frac{1}{2}$ باشد کدام است؟

ب) امید ریاضی و واریانس X کدام است؟

مقدمه: فاصله پیوسته‌ای را به صورت $\alpha < x < \beta$ در نظر بگیرید که وقوع احتمال در هر نقطه از آن با هر نقطه دیگر در آن برابر باشد؛ در این صورت احتمال وقوع در هر نقطه یا فاصله را می‌توانیم به صورت زیر نمایش دهیم:

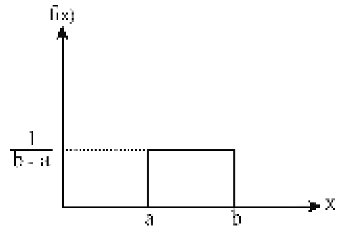
$$P(x)=\frac{\text{طول بازه } x}{\beta-\alpha}$$

استفاده از رابطه بالا به آن علت است که فاصله $\alpha < x < \beta$ به طور یکنواخت بین تمام نقاط توزیع شده است تا احتمال در تمام نقاط با هم برابر باشد در چنین حالتی طبیعی است که:

$$P(c<X<d)=\frac{d-c}{\beta-\alpha}$$

$$P(X=e)=\frac{e-e}{\beta-\alpha}=0$$

تعریف: متغیر تصادفی پیوسته X در بازه a تا b دارای **چگالی یکنواخت** است اگر و تنها اگر تابع چگالی آن به ازای هر مقدار از x برابر با عدد ثابت $\frac{1}{b-a}$ باشد.



پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع یکنواخت پیوسته باشد، پارامترهای آن a و b است و به صورت $X \sim u(a,b)$ نمایش داده می‌شود.

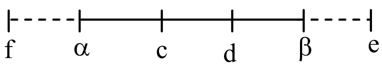
تابع احتمال	$f(x)=\frac{1}{b-a} \quad ; \quad a<x<b$	مقدار ثابت $\frac{1}{b-a}$ به ازای هر مقداری از x
امید ریاضی (میانگین)	$E(X)=\mu=\frac{a+b}{2}$	متوسط مقادیر x در بازه a تا b
واریانس	$\sigma_X^2=\frac{(b-a)^2}{12}$	واریانس مقادیر x در بازه a تا b
انحراف معیار	$\sigma_X=\sqrt{\frac{(b-a)^2}{12}}$	
تابع مولد گشتاور	$M_X(t)=\frac{e^{tb}-e^{ta}}{t(b-a)}$	

نکته: هرگاه تابع چگالی $f(x)$ در بازه $\alpha < x < \beta$ به صورت «$f(x) =$ ثابت» باشد، آن‌گاه حتماً X دارای توزیع یکنواخت پیوسته با تابع چگالی $f(x) = \dfrac{1}{\beta - \alpha}$ است و برعکس؛ به عبارت دیگر:

$$\begin{cases} f(x) = \text{ثابت} \\ \alpha < x < \beta \end{cases} \Longleftrightarrow \begin{cases} f(x) = \dfrac{1}{\beta - \alpha} \\ \alpha < x < \beta \end{cases}$$

محاسبه احتمال

هرگاه $\alpha < x < \beta$ دارای توزیع چگالی یکنواخت باشد، برای محاسبه احتمال در بازه دلخواه $c < x < d$، درصورتی‌که در فاصله مورد نظر باشد، به صورت زیر عمل می‌کنیم:



$$P(c < X < d) = \int_c^d f(x)\,dx = \int_c^d \frac{1}{\beta - \alpha}\,dx = \left[\frac{x}{\beta - \alpha}\right]_c^d = \frac{d - c}{\beta - \alpha}$$

بنابراین:

$$P(\text{بازه}) = \frac{\text{طول بازه}}{\beta - \alpha}$$

$$\begin{cases} P(c < X < d) = \dfrac{d - c}{\beta - \alpha} \quad , \quad P(X < c) = \dfrac{c - \alpha}{\beta - \alpha} \\ P(X > d) = \dfrac{\beta - d}{\beta - \alpha} \quad , \quad P(X = c) = \dfrac{c - c}{\beta - \alpha} = 0 \end{cases}$$

✓ دقت کنید!

اگر بازه احتمال خارج از بازه $(\alpha < x < \beta)$ یعنی بازه تابع یکنواخت پیوسته باشد (مانند نقاط f, e)، آن‌گاه:

$$\begin{cases} P(f < X < \alpha) = 0 \\ P(f < X < c) = P(f < X < \alpha) + P(\alpha < X < c) = \dfrac{c - \alpha}{\beta - \alpha} \\ P(d < X < e) = P(d < X < \beta) + P(\beta < X < e) = \dfrac{\beta - d}{\beta - \alpha} \end{cases}$$

حل مسئله (درک مطلب):

با توجه به نکته، ازآنجاکه (ثابت) $f(x) = \dfrac{2}{3}$ است، حتماً تابع یکنواخت پیوسته است، یعنی:

$$f(x) = \frac{1}{\beta - \alpha} = \frac{1}{\frac{1}{2} - (-1)} = \frac{1}{\frac{3}{2}} = \frac{2}{3}$$

الف)

$$P\left(0 < X < \frac{1}{2}\right) = \frac{\frac{1}{2} - 0}{\frac{1}{2} - (-1)} = \frac{1}{3}$$

ب)

$$\begin{cases} E(X)=\dfrac{\alpha+\beta}{2}=\dfrac{-1+\frac{1}{2}}{2}=-\dfrac{1}{4} \\ \sigma_X^2=\dfrac{(\beta-\alpha)^2}{12}=\dfrac{\left(\frac{1}{2}-(-1)\right)^2}{12}=\dfrac{\frac{9}{4}}{12}=\dfrac{3}{16} \end{cases}$$

مثال ۱ متغیر تصادفی X با چگالی یکنواخت زیر را در نظر بگیرید:

$$f(x)=\begin{cases} \dfrac{1}{8} & ; \quad \alpha<x<\beta \\ 0 & ; \quad \text{در غیر این صورت} \end{cases}$$

اگر امید ریاضی متغیر تصادفی X برابر 50 باشد، آن‌گاه مقادیر β, α کدام است؟

۱) $\beta=56 \ , \ \alpha=44$ ۲) $\beta=54 \ , \alpha=46$ ۳) $\beta=40 \ , \alpha=22.5$ ۴) $\beta=50 \ , \alpha=46$

حل: گزینه ۲ درست است.

$$\begin{cases} E(X)=\dfrac{\alpha+\beta}{2}=50 \ \rightarrow \ \alpha+\beta=100 \\ f(x)=\dfrac{1}{\beta-\alpha}=\dfrac{1}{8} \ \rightarrow \ \beta-\alpha=8 \end{cases} \rightarrow \begin{array}{l} \beta=54 \\ \alpha=46 \end{array}$$

مثال ۲ تابع چگالی یکنواخت زیر را در نظر بگیرید:

$$f(x)=\begin{cases} \dfrac{1}{9} & ; \quad 2\le x\le 11 \\ 0 & ; \quad \text{در غیر این‌صورت} \end{cases}$$

مقدار میانگین (μ) و واریانس $\left(\sigma^2\right)$ به ترتیب کدام است؟

۱) $\sigma^2=\dfrac{27}{4} \ , \ \mu=6.5$ ۲) $\sigma^2=\dfrac{9}{2} \ , \ \mu=6.5$ ۳) $\sigma^2=\dfrac{9}{2} \ , \ \mu=8$ ۴) $\sigma^2=\dfrac{27}{2} \ , \ \mu=7$

حل: گزینه ۱ درست است.

$$\begin{cases} \mu=\dfrac{\alpha+\beta}{2}=\dfrac{2+11}{2}=6.5 \\ \sigma^2=\dfrac{(\beta-\alpha)^2}{12}=\dfrac{(11-2)^2}{12}=\dfrac{81}{12}=\dfrac{27}{4} \\ a=2 \ , \ b=11 \end{cases}$$

مثال ۳ تابع چگالی یکنواخت زیر را در نظر بگیرید:

$$f(x)=\begin{cases} \dfrac{1}{8} & ; \quad 2\le x\le 10 \\ 0 & ; \quad \text{در غیر این‌صورت} \end{cases}$$

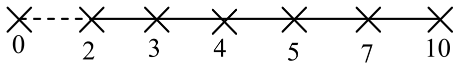
احتمال مربوط به هر یک از حالات زیر کدام است؟

الف) $P(3<X<7)$ ب) $P(X>4)$ ج) $P(X<5)$

د) $P(X=3)$ ه) $P(3<X<10)$ و) $P(0<X<3)$

حل:

روش اول:



احتمال در هر بازه برابر با انتگرال (سطح زیر منحنی) در بازه مورد نظر است:

الف) $P(3<X<7)=\int_3^7 \frac{1}{8}dx=\left[\frac{x}{8}\right]_3^7=\frac{7-3}{8}=\frac{4}{8}=\frac{1}{2}$

ب) $P(X>4)=\int_4^{10} \frac{1}{8}dx=\left[\frac{x}{8}\right]_4^{10}=\frac{10-4}{8}=\frac{6}{8}$

ج) $P(X<5)=\int_2^5 \frac{1}{8}dx=\left[\frac{x}{8}\right]_2^5=\frac{5-2}{8}=\frac{3}{8}$

د) $P(X=3)=\int_3^3 \frac{1}{8}dx=0$

ه) $P(3<X<10)=\int_3^{10} \frac{1}{8}dx=\left[\frac{x}{8}\right]_3^{10}=\frac{10-3}{8}=\frac{7}{8}$

و) $P(0<X<3)=\int_2^3 \frac{1}{8}dx=\left[\frac{x}{8}\right]_2^3=\frac{3-2}{8}=\frac{1}{8}$

روش دوم:

الف) $P(3<X<7)=\frac{7-3}{10-2}=\frac{4}{8}$

ب) $P(X>4)=\frac{10-4}{10-2}=\frac{6}{8}$

ج) $P(X<5)=\frac{5-2}{10-2}=\frac{3}{8}$

د) $P(X=3)=\frac{3-3}{10-2}=0$

ه) $P(3<X<10)=\frac{10-3}{10-2}=\frac{7}{8}$

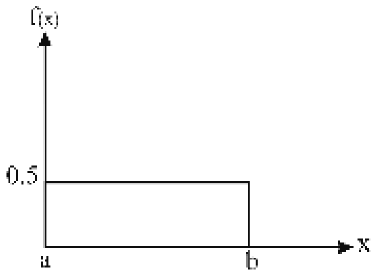
و) $P(0<X<3)=P(2<X<3)=\frac{3-2}{10-2}=\frac{1}{8}$

مثال ۴ تابع چگالی یکنواخت پیوسته زیر را در نظر بگیرید:

الف) تابع چگالی $f(x)$ کدام است؟

ب) درصورتی‌که $P(X<C)=0.15$ باشد، مقدار C کدام است؟

ج) امید ریاضی (میانگین) و واریانس X کدام است؟



حل:

الف) با توجه به آنکه a در مبدأ مختصات قرار دارد، $a = 0$ است:

$$f(x)=\frac{1}{b-a}\xrightarrow[f(x)=\frac{1}{2}]{a=0}\frac{1}{2}=\frac{1}{b-0}\rightarrow b=2$$

در نتیجه:

$$f(x)=\frac{1}{2}\quad ;\quad 0\le x\le 2$$

ب) **روش اول:**

$$P(X<C)=0.15\rightarrow\int_0^C\frac{1}{2}dx=0.15\rightarrow\left[\frac{x}{2}\right]_0^C=0.15\rightarrow\frac{C}{2}=0.15\rightarrow C=0.3$$

روش دوم:

$$P(X<C)=0.15\rightarrow\frac{C-\alpha}{\beta-\alpha}=0.15\rightarrow\frac{C-0}{2-0}=0.15\rightarrow C=0.3$$

ج)

$$\begin{cases}E(X)=\dfrac{a+b}{2}=\dfrac{0+2}{2}=1\\ \sigma_X^2=\dfrac{(b-a)^2}{12}=\dfrac{(2-0)^2}{12}=\dfrac{1}{3}\end{cases}$$

مثال ۵ تابع توزیع تجمعی مربوط به تابع چگالی یکنواخت پیوسته $f(x)=\frac{1}{\beta-\alpha}\ ;\ \alpha\le x\le\beta$ کدام است؟

۱) $\frac{x-\alpha}{\beta-\alpha}$ ۲) $\frac{\alpha}{\beta-\alpha}$ ۳) $\frac{x}{\beta-\alpha}$ ۴) هیچ‌کدام

حل: گزینه ۱ درست است.

$$F_X(x)=\int_{\text{حد پایین}}^{x}f(x)dx=\int_\alpha^x\frac{1}{\beta-\alpha}dx=\left[\frac{x}{\beta-\alpha}\right]_\alpha^x=\frac{x-\alpha}{\beta-\alpha}$$

مثال ۶ سود شرکتی دارای توزیع یکنواخت بین -3 تا 5 میلیون ریال است.

الف) تابع چگالی احتمال را به دست آورید.

ب) احتمال آنکه سود شرکت بین 0 تا 1.2 باشد، چقدر است؟

ج) متوسط سود مورد انتظار شرکت چقدر است؟

د) واریانس سود شرکت چقدر است؟

حل:

الف)

$$\begin{cases}f(x)=\dfrac{1}{\beta-\alpha}\\ \alpha<x<\beta\end{cases}\rightarrow\begin{cases}f(x)=\dfrac{1}{5-(-3)}=\dfrac{1}{8}\\ -3<x<5\end{cases}$$

ب) **روش اول:**

$$P(0<X<1.2)=\int_0^{1.2}f(x)dx=\int_0^{1.2}\frac{1}{8}dx=\left[\frac{x}{8}\right]_0^{1.2}=\frac{1.2}{8}=0.15$$

روش دوم:

$$P(0<X<1.2)=\frac{1.2-0}{5-(-3)}=\frac{1.2}{8}=0.15$$

ج) $$E(X)=\frac{\alpha+\beta}{2}=\frac{-3+5}{2}=1\ (\text{میلیون})$$

د) $$\sigma_X^2=\frac{(\beta-\alpha)^2}{12}=\frac{\left(5-(-3)\right)^2}{12}=\frac{64}{12}=\frac{16}{3}$$

مثال ۷ سود مورد انتظار شرکتی دارای توزیع یکنواخت با میانگین 130 میلیون تومان و واریانس 300 است. احتمال آنکه سود این شرکت بیشتر از 120 میلیون تومان شود، چقدر است؟

۱) 0.33 ۲) 0.667 ۳) 0.5 ۴) 0.2

حل: گزینه ۲ درست است.

$$\left.\begin{array}{l} E(X)=\dfrac{\alpha+\beta}{2}=130 \\ \sigma_X^2=\dfrac{(\beta-\alpha)^2}{12}=300 \end{array}\right. \rightarrow \begin{array}{l} \alpha+\beta=260 \\ \beta-\alpha=60 \end{array} \rightarrow \alpha=100\ ,\beta=160$$

$$\begin{cases} P(X>a)=\dfrac{\beta-a}{\beta-\alpha} \\ \alpha<x<\beta \end{cases} \rightarrow \begin{cases} P(X>120)=\dfrac{160-120}{160-100}=\dfrac{40}{60}=\dfrac{2}{3}=0.667 \\ 100<x<160 \end{cases}$$

توزیع نمایی (Exponential Distribution)

مسئله ۱ (درک مطلب): تعداد مشتریان یک فروشگاه در ساعت خاصی از روز دارای توزیع پواسون با میانگین 4 نفر است. متوسط زمان بین ورود دو مشتری یا ورود اولین مشتری کدام است؟

مسئله ۲ (درک مطلب): زمان ورود مشتریان به یک بانک دارای توزیع نمایی با میانگین $\frac{1}{2}$ ساعت است. مطلوب است احتمال آنکه:

الف) مشتری بعدی دقیقاً رأس 1 ساعت وارد شود.

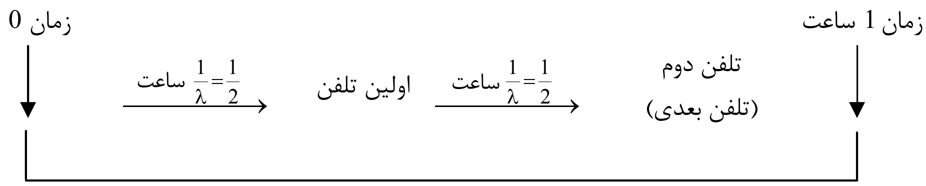
ب) مشتری بعدی تا 3 ساعت بعد وارد شود.

ج) صندوق‌دار بیش از 20 دقیقه منتظر اولین مشتری شود.

د) مشتری بعدی بین 2 تا 3 ساعت وارد شود.

مقدمه: در بخش توزیع‌های گسسته دیدیم که تعداد وقایعی که در یک «فاصله زمانی یا مکانی» رخ می‌دهند، دارای توزیع پواسون هستند. حال اگر بخواهیم «فاصله زمانی بین دو اتفاق متوالی یا زمان لازم برای وقوع اولین اتفاق» در توزیع پواسون را بررسی کنیم، با یک متغیر تصادفی پیوسته (زمان) روبه‌رو می‌شویم که به آن **نمایی** گفته می‌شود. به مثال زیر توجه کنید:

برای مثال، فرض کنید تعداد تماس‌های تلفنی در ساعت به یک بانک دارای توزیع پواسون با متوسط $\lambda = 2$ تلفن در ساعت باشد، آن‌گاه:



متوسط $\lambda = 2$ تلفن در ساعت

درنتیجه، اگر به طور متوسط 2 تماس تلفنی در ساعت برقرار شود، انتظار داریم در هر $\frac{1}{2}$ ساعت، 1 تماس داشته باشیم.

تعریف: درصورتی‌که تعداد اتفاقات در واحد زمان دارای توزیع پواسون با میانگین λ اتفاق باشد، آن‌گاه متغیر تصادفی« X : **زمان** بین دو اتفاق متوالی یا **زمان** لازم برای اولین اتفاق» $(x \geq 0)$ دارای **توزیع نمایی** با میانگین زمان $\frac{1}{\lambda}$ است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع نمایی باشد، پارامتر آن λ است و به صورت $X \sim E(\lambda)$ نمایش داده می‌شود.

X : زمان لازم برای وقوع اتفاق بعدی زمان لازم برای وقوع اولین اتفاق زمان لازم بین دو اتفاق متوالی	$f(x) = \lambda e^{-\lambda x} \quad ; \quad 0 \le x < \infty$	**تابع احتمال**
متوسط زمان لازم برای وقوع اولین اتفاق یا دو اتفاق متوالی	$E(X) = \mu = \frac{1}{\lambda}$	**امید ریاضی (میانگین)**
واریانس زمان لازم برای وقوع اولین اتفاق یا دو اتفاق متوالی	$\sigma_X^2 = \frac{1}{\lambda^2}$	**واریانس**
	$\sigma_X = \frac{1}{\lambda}$	**انحراف معیار**
	$M_X(t) = \frac{\lambda}{\lambda - t}$	**تابع مولد گشتاور**

توجه: توزیع نمایی تنها توزیعی است که در آن میانگین توزیع با انحراف معیار آن برابر است و درنتیجه همواره ضریب تغییرات (ضریب پراکندگی) آن یک است.

رابطه بین λ و $\frac{1}{\lambda}$

$\frac{1}{\lambda} =$ متوسط زمان لازم برای وقوع اولین اتفاق یا دو اتفاق متوالی $\longleftrightarrow$ $\lambda =$ متوسط تعداد اتفاقات در واحد زمان

برای محاسبه احتمال وقوع اتفاق بعدی یا اولین اتفاق در زمان مشخص به صورت زیر عمل می‌کنیم:

ابتدا λ (متوسط تعداد اتفاقات در واحد زمان) را محاسبه می‌کنیم و بر اساس آن داریم:

الف) وقوع اتفاق بعدی یا اولین اتفاق دقیقاً در زمان a: $P(X = a) = 0$

ب) وقوع اتفاق بعدی یا اولین اتفاق قبل از زمان a (حداکثر زمان a): $P(X < a) = 1 - e^{-\lambda a}$

ج) وقوع اتفاق بعدی یا اولین اتفاق بعد از زمان a (حداقل زمان a): $P(X > a) = e^{-\lambda a}$

د) وقوع اتفاق بعدی یا اولین اتفاق بین زمان a تا b: $P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}$

برای اثبات روابط بالا با توجه به تابع چگالی نمایی:

$$f(x) = \lambda e^{-\lambda x} \qquad 0 \le x < \infty$$

الف) $P(X = a) = 0$ (در تابع چگالی احتمال پیوسته)

ب) $P(X < a) = \int_0^a \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_0^a = 1 - e^{-\lambda a}$

ج) $P(X > a) = \int_a^\infty \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_a^\infty = e^{-\lambda a}$

د) $P(a < X < b) = \int_a^b \lambda e^{-\lambda x}dx = \left[-e^{-\lambda x}\right]_a^b = e^{-\lambda a} - e^{-\lambda b}$

حل مسئله ۱ (درک مطلب):

با توجه به رابطه λ و $\frac{1}{\lambda}$ داریم:

$\frac{1}{\lambda}=\frac{1}{4}$ متوسط ساعت (15 دقیقه) برای ورود مشتری بعدی یا اولین مشتری $\longrightarrow$ $(\lambda=4)$ متوسط نفر در ساعت

حل مسئله ۲ (درک مطلب):

$\frac{1}{\lambda}=\frac{1}{2}$ متوسط ساعت برای ورود مشتری بعدی یا اولین مشتری $\longrightarrow$ $\lambda=2$ متوسط نفر در ساعت

حال بنا بر نکته و $\lambda=2$ داریم:

الف) $P(X=a)=0 \quad \to \quad P(X=1)=0$

ب) $P(X<a)=1-e^{-\lambda a} \quad \to \quad P(X<3)=1-e^{-2\times 3}=1-e^{-6}$

ج) $P(X>a)=e^{-\lambda a} \xrightarrow{\text{20 دقیقه معادل } \frac{1}{3} \text{ ساعت}} P\left(X>\frac{1}{3}\right)=e^{-2\times\frac{1}{3}}=e^{-\frac{2}{3}}$

د) $P(a<X<b)=e^{-\lambda a}-e^{-\lambda b} \quad \to \quad P(2<X<3)=e^{-2\times 2}-e^{-2\times 3}=e^{-4}-e^{-6}$

مثال ۱ تابع چگالی نمایی زیر را در نظر بگیرید:

$$f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}} & ; \quad x>0\\ 0 & ; \quad \text{در غیر این صورت}\end{cases}$$

الف) میانگین و واریانس X چقدر است؟

ب) احتمال اینکه X مقداری بین $\frac{1}{2}$ و $\frac{3}{2}$ را اختیار کند، چقدر است؟

حل:

الف) با توجه به آنکه در توزیع نمایی $E(X)=\frac{1}{\lambda}$ و $\sigma^2(X)=\frac{1}{\lambda^2}$ است، داریم:

$$f(x)=\lambda e^{-\lambda x}=\frac{1}{2}e^{-\frac{x}{2}} \to \lambda=\frac{1}{2} \to \begin{cases}E(X)=\frac{1}{\lambda}=\frac{1}{\frac{1}{2}}=2\\ \sigma^2(X)=\frac{1}{\lambda^2}=\frac{1}{\frac{1}{4}}=4\end{cases}$$

ب) **راه حل اول:**

بنا بر نکته (د):

$$P(a<X<b)=e^{-\lambda a}-e^{-\lambda b} \xrightarrow{\lambda=\frac{1}{2}} P\left(\frac{1}{2}<X<\frac{3}{2}\right)=e^{-\frac{1}{4}}-e^{-\frac{3}{4}}$$

راه حل دوم: بنابر روش معمول محاسبه احتمال در توابع چگالی پیوسته، احتمال در هر فاصله برابر با انتگرال روی تابع چگالی $f(x)$ در فاصله مورد نظر است؛ درنتیجه به صورت زیر عمل می‌کنیم:

$$P\left(\frac{1}{2}<X<\frac{3}{2}\right)=\int_{\frac{1}{2}}^{\frac{3}{2}}\frac{1}{2}e^{-\frac{x}{2}}dx=\left[-e^{-\frac{x}{2}}\right]_{\frac{1}{2}}^{\frac{3}{2}}=-e^{-\frac{3}{4}}-\left(-e^{-\frac{1}{4}}\right)=e^{-\frac{1}{4}}-e^{-\frac{3}{4}}$$

مثال ۲ تعداد خرابی‌های ماشین در ماه (30 روز) دارای توزیع پواسون با میانگین 3 خرابی است.

الف) متوسط زمان بین دو خرابی چند روز است؟

ب) احتمال اینکه در 10 روز اول پس از سرویس، ماشین خراب شود، چقدر است؟

حل:

الف) با توجه به رابطه λ و $\frac{1}{\lambda}$:

$$\lambda = 3 \text{ متوسط خرابی در ماه} \longrightarrow \frac{1}{\lambda} = \frac{1}{3} \text{ متوسط ماه بین دو خرابی یا برای اولین خرابی}$$

بنابراین به طور متوسط $\frac{1}{\lambda} = \frac{1}{3}$ ماه (10 روز) بین دو خرابی یا برای اولین خرابی زمان لازم است.

ب) ازآنجاکه به طور متوسط $\lambda = 3$ خرابی در ماه است، داریم:

راه حل اول: بنا بر نکته (ب) احتمال آنکه قبل از 10 روز ($\frac{1}{3}$ ماه) پس از سرویس، ماشین خراب شود برابر است با:

$$P(X < a) = 1 - e^{-\lambda a} \quad \rightarrow \quad P\left(X < \frac{1}{3}\right) = 1 - e^{-3.\frac{1}{3}} = 1 - e^{-1}$$

راه حل دوم: با توجه به روش معمول محاسبه احتمال در توابع چگالی پیوسته، احتمال در هر فاصله برابر است با انتگرال روی تابع چگالی $f(x)$ در فاصله مورد نظر؛ بنابراین ابتدا تابع چگالی نمایی را با توجه به $\lambda = 3$ به صورت زیر به دست می‌آوریم:

$$f(x) = \lambda e^{-\lambda x} = 3e^{-3x} \quad , \quad 0 < x < \infty$$

حال احتمال مورد نظر را محاسبه می‌کنیم:

$$P\left(X < \frac{1}{3}\right) = \int_0^{\frac{1}{3}} 3e^{-3x}\, dx = \left[-e^{-3x}\right]_0^{\frac{1}{3}} = 1 - e^{-1}$$

دقت کنید که حتماً باید واحد زمانی احتمال خواسته‌شده، منطبق بر واحد زمانی λ باشد. در این سؤال ازآنجاکه واحد زمانی ($\lambda = 3$ خرابی) ماه است، بنابراین واحد زمانی 10 روز برحسب ماه $\frac{10}{30} = \frac{1}{3}$ ماه خواهد بود.

مثال ۳ اگر متغیر تصادفی X دارای توزیع نمایی با پارامتر λ باشد، تابع توزیع تجمعی آن کدام است؟

۱) $e^{-\lambda x}$ ۲) $\lambda e^{-\lambda x}$ ۳) $1 - e^{-\lambda x}$ ۴) $-e^{-\lambda x}$

حل: گزینه ۳ درست است.

راه حل اول:

$$F_X(x) = P(X \le x) = 1 - e^{-\lambda x}$$

راه حل دوم:

$$F_X(x) = P(X \le x) = \int_0^x \lambda e^{-\lambda x}\, dx = \left[-e^{-\lambda x}\right]_0^x = 1 - e^{-\lambda x}$$

مثال ۴ مدت‌زمان تعمیر ماشینی دارای توزیع نمایی با میانگین 3 ساعت است. مطلوب است احتمال آنکه:

الف) ماشین کمتر از 1 ساعت تعمیر شود.

ب) مدت تعمیر بین 1 تا 3 ساعت طول بکشد.

حل:

$$(\text{متوسط در ساعت})\ \frac{1}{\lambda}=3 \longrightarrow \lambda=\frac{1}{3}\ (\text{متوسط}\ \frac{1}{3}\ \text{ماشین در ساعت})$$

الف) بنا بر نکته (ب)

$$P(X<a)=1-e^{-\lambda a} \;\to\; P(X<1)=1-e^{-\frac{1}{3}\times 1}=1-e^{-\frac{1}{3}}$$

ب) بنا بر نکته (د)

$$P(a<X<b)=e^{-\lambda a}-e^{-\lambda b} \;\to\; P(1<X<3)=e^{-\frac{1}{3}\times 1}-e^{-\frac{1}{3}\times 3}=e^{-\frac{1}{3}}-e^{-1}$$

مثال ۵ به طور متوسط هر 0.5 دقیقه 2 مشتری با توزیع پواسون به گیشه پرداخت بانکی مراجعه می‌کند. احتمال اینکه اولین مشتری بعد از 2 دقیقه وارد شود چقدر است؟

حل:

$$(\text{مشتری در}\ \frac{1}{2}\ \text{دقیقه})\lambda=2 \longrightarrow \lambda=2\times 2=4\ (\text{مشتری در 1 دقیقه})$$

با توجه به نکته (ج):

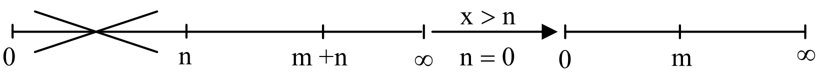
$$P(X>a)=e^{-\lambda a} \;\to\; P(X>2)=e^{-4\times 2}=e^{-8}$$

خاصیت عدم حافظه (بی‌حافظه بودن)

یکی از مهم‌ترین خصوصیات توزیع نمایی، خاصیت «بی‌حافظه بودن» است:

$$P(X>m+n \mid x>n)=P(X>m)$$

بی‌حافظه بودن به آن مفهوم است که اگر متغیر X با توزیع نمایی تا زمان n اتفاق نیفتاده باشد $(X>n)$، آن‌گاه وقوع آن در m واحد زمان بعدی مستقل از n است،؛به عبارت دیگر وقوع X بعد از m واحد زمانی را می‌توان بدون در نظر گرفتن n، یک توزیع نمایی مستقل دانست که از زمان صفر در نظر گرفته می‌شود:



کاربرد توزیع نمایی

یکی از کاربردهای مهم توزیع نمایی با استفاده از خاصیت «بی‌حافظه بودن»، محاسبه طول عمر قطعات برقی است، زیرا یک دستگاه برقی هر چقدر هم که عمر کرده باشد، طول عمر باقی‌مانده آن ربطی به مدت‌زمان کارکرد قبلی نداشته و دوباره از 0 در نظر گرفته می‌شود؛ به عبارت دیگر هر دستگاه برقی ممکن است هر لحظه خراب شود، از این رو معمولاً طول عمر قطعات برقی شامل گارانتی نمی‌شود!

مثال اگر طول عمر یک لامپ دارای توزیع نمایی با میانگین 20 ساعت باشد و این لامپ تا به حال 100 ساعت کار کرده باشد، احتمال آنکه بعد از 15 ساعت خراب شود چقدر است؟

۱) e^{-4} ۲) $e^{-\frac{3}{4}}$ ۳) $1-e^{-\frac{3}{4}}$ ۴) $e^{-\frac{1}{4}}$

حل: گزینه ۲ درست است.

با توجه به خاصیت عدم حافظه داریم:

$$\begin{cases} P(X>15+100\,|\,x>100)=P(X>15)=e^{-\frac{1}{20}\times 15}=e^{-\frac{15}{20}}=e^{-\frac{3}{4}} \\ \frac{1}{\lambda}=20 \quad \to \quad \lambda=\frac{1}{20} \end{cases}$$

یادآوری: برای محاسبه $P(X>15)$ از نکته (ج) در توزیع نمایی استفاده کرده‌ایم:

$$P(X>a)=e^{-\lambda a}$$

توزیع گاما (Gamma Distribution)

اگر $X_1, X_2, \dots, X_r$، r متغیر مستقل از هم و دارای توزیع نمایی با پارامتر λ باشند، آن‌گاه توزیع مجموع این متغیرها گاما با پارامترهای (r,λ) است.

$$X_i \sim \lambda \ \text{نمایی با پارامتر} \to \left(\sum_{i=1}^{r} x_i\right) \sim (r,\lambda) \ \text{گاما با پارامتر}$$

تعریف: توزیع گاما به عنوان مدت‌زمان انتظار برای r اتفاق در نظر گرفته می‌شود، به این صورت که اگر اتفاقات به طور تصادفی در طول زمان رخ دهند، آن‌گاه مدت‌زمانی که فرد باید منتظر باشد تا r اتفاق رخ دهد، دارای توزیع گاما با پارامترهای (r,λ) است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع گاما باشد، پارامترهای آن r,λ است و به صورت $X \sim G(r,\lambda)$ یا $X \sim \Gamma(r,\lambda)$ نمایش داده می‌شود.

X: مدت‌زمان انتظار برای r اتفاق	$f(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} \quad ; \quad 0<x<\infty \quad , \quad r\geq 1$ $\Gamma(r)=(r-1)!$	**تابع احتمال**
متوسط مدت‌زمان انتظار برای r اتفاق	$E(X)=\frac{r}{\lambda}$	**امید ریاضی (میانگین)**
واریانس مدت‌زمان انتظار برای r اتفاق	$\sigma_X^2=\frac{r}{\lambda^2}$	**واریانس**
	$\sigma_X=\frac{\sqrt{r}}{\lambda}$	**انحراف معیار**
	$M_X(t)=\left(\frac{\lambda}{\lambda-t}\right)^r$	**تابع مولد گشتاور**

حالت خاص

هرگاه در توزیع گاما قرار دهیم $r=1$، توزیع متغیر تصادفی X دقیقاً نمایی با پارامتر λ است؛ به عبارت دیگر:

$$f(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} \xrightarrow{r=1} f(x)=\frac{\lambda}{\Gamma(1)}x^{1-1}e^{-\lambda x}=\lambda e^{-\lambda x}$$

یادآوری: $\Gamma(n)=(n-1)!$

توزیع نرمال (Normal Distribution)

مقدمه: توزیع نرمال یا توزیع زنگی (بهنجار) به عنوان یک توزیع متقارن، مهم‌ترین توزیع پیوسته است و دارای کاربردهای فراوانی از جمله موارد زیر است:

۱- بسیاری از پدیده‌های طبیعی دارای توزیع نرمال هستند.

۲- بسیاری از توزیع‌ها در شکل حدی دارای تقریب نرمال هستند (دوجمله‌ای، پواسون، کای دو، t–استیودنت و فیشر).

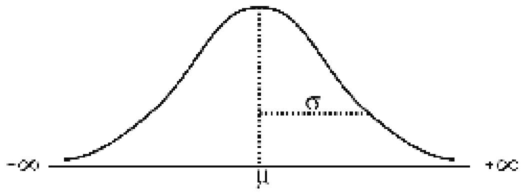
تعریف: اگر متغیر تصادفی پیوسته X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، تابع چگالی و تابع مولد گشتاور آن به صورت زیر است:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad -\infty < x < +\infty$$

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

پارامترهای توزیع

اگر متغیر تصادفی X دارای توزیع نرمال باشد، پارامترهای آن μ (میانگین) و σ^2 (واریانس) است و از شکل نمادین زیر برای نمایش آن استفاده می‌شود:



$$\boxed{X \sim N(\mu, \sigma^2)}$$

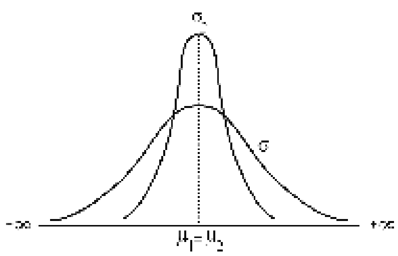
تأثیرات μ و σ^2 روی منحنی نرمال

در صورت معلوم بودن پارامترهای μ و σ^2، به راحتی می‌توانیم توزیع را مشخص و منحنی آن را ترسیم کنیم. وضعیت‌های زیر تأثیرات μ و σ^2 را روی منحنی نرمال نشان می‌دهند.

دو منحنی نرمال به صورت $N(\mu_1, \sigma_1^2)$ و $N(\mu_2, \sigma_2^2)$ را در نظر می‌گیریم.

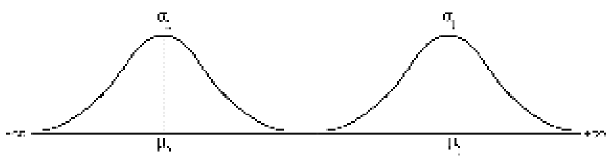
الف) $\mu_1 = \mu_2 \quad , \quad \sigma_1 > \sigma_2$

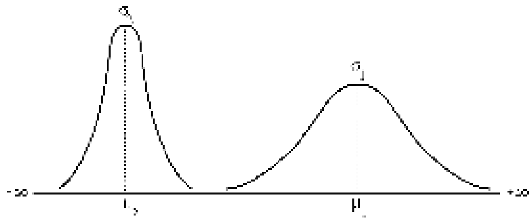
(میانگین‌ها برابر و انحراف معیارها متفاوت)



ب) $\mu_1 > \mu_2 \quad , \quad \sigma_1 = \sigma_2$

(میانگین‌ها متفاوت و انحراف معیارها برابر)





ج) $\sigma_1 > \sigma_2 \quad , \quad \mu_1 > \mu_2$

(میانگین‌ها متفاوت و انحراف معیارها متفاوت)

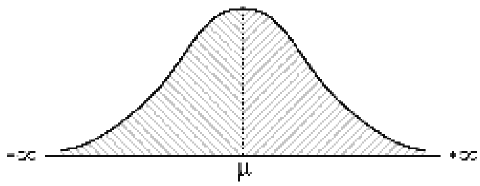
نتیجه:

۱) افزایش میانگین، منحنی را به سمت راست و کاهش میانگین، منحنی را به سمت چپ انتقال می‌دهد.

۲) افزایش انحراف معیار، ارتفاع منحنی را کوتاه‌تر (پخ‌تر)، پراکندگی را بیشتر و تمرکز حول میانگین را کمتر می‌کند و کاهش انحراف معیار، ارتفاع منحنی را بلندتر (کشیده‌تر)، پراکندگی را کمتر و تمرکز حول میانگین را بیشتر می‌کند.

خصوصیات توزیع نرمال

۱- سطح زیر منحنی نرمال با توجه به تعریف تابع چگالی (پیوسته)، برابر 1 است.



$$\boxed{\int_{-\infty}^{+\infty} f(x)\,dx = 1}$$

۲- پارامترهای میانگین (μ)، میانه (Md) و مد (نما) (Mo) در توزیع نرمال با هم برابر هستند.

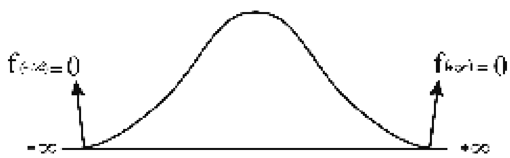
$$\boxed{\mu = Md = Mo}$$

۳- با توجه به برابری $\mu = Mo$ حداکثر مقدار تابع در نقطه $x = \mu$ به دست می‌آید؛ به عبارت دیگر:

$$\boxed{f'_X(\mu) = 0 \quad \rightarrow \quad f(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{\mu-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\,\sigma}}$$

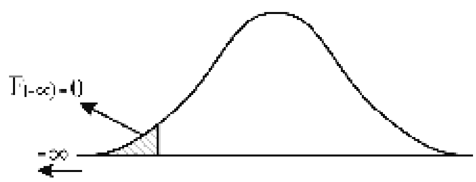
۴- در منحنی توزیع نرمال، با فاصله گرفتن از میانگین (μ) در هر دو سمت منحنی، به محور x ها نزدیک می‌شویم به طوری که:

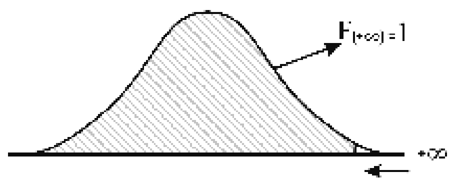
$$\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x) = 0 \;\rightarrow\; f(-\infty) = f(+\infty) = 0$$



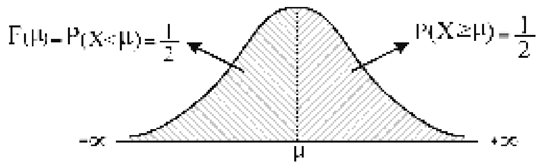
نتیجه:

$$\begin{cases} \lim\limits_{x \to +\infty} F_X(x) = P(X \le +\infty) = 1 \\ \lim\limits_{x \to -\infty} F_X(x) = P(X \le -\infty) = 0 \end{cases} \rightarrow \begin{cases} F(+\infty) = 1 \\ F(-\infty) = 0 \end{cases}$$





۵- خط $x = \mu$ محور تقارن منحنی است (با توجه به برابری $Md = \mu$) و درنتیجه:

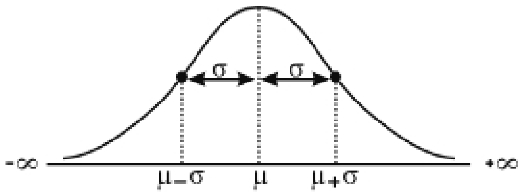


$$P(X \le \mu) = P(X \ge \mu) = \frac{1}{2} \rightarrow F(\mu) = \frac{1}{2}$$

یادآوری: با توجه به پیوسته بودن توزیع نرمال $P(X = \mu) = 0$ است، بنابراین:

$$P(X \ge \mu) = P(X > \mu) \quad , \quad P(X \le \mu) = P(X < \mu)$$

۶- نقاط $\mu \pm \sigma$ تنها دو نقطه عطف منحنی نرمال هستند به طوری که $f''(\mu \pm \sigma) = 0$.



مثال ۱ اگر متغیر تصادفی X دارای توزیع نرمال با میانگین 50 باشد، $P(X \le 50)$ کدام است؟

۱) 0 ۲) $\frac{3}{4}$ ۳) $\frac{1}{2}$ ۴) 1

حل: گزینه ۳ درست است.

$$\begin{cases} P(X \le \mu) = P(X \ge \mu) = \frac{1}{2} \\ \mu = 50 \end{cases}$$

مثال ۲ اگر متغیر تصادفی X دارای توزیع نرمال به فرم $X \sim N(60, 25)$ باشد، مقدار مد توزیع کدام است؟

۱) 60 ۲) 30 ۳) 25 ۴) 5

حل: گزینه ۱ درست است.

در توزیع نرمال به صورت $X \sim N(60, 25)$، مقدار $\mu = 60$ و $\sigma^2 = 25$ است، بنابراین چون در توزیع نرمال میانگین، میانه و مد بر هم منطبق هستند، مقدار مد نیز برابر با 60 است.

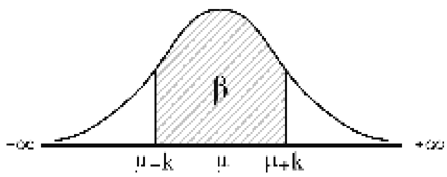
مثال ۳ اگر متغیر تصادفی X، دارای توزیع نرمال باشد، حداکثر مقدار تابع کدام است؟

۱) μ ۲) $\frac{1}{\sqrt{2\pi}\,\sigma}$ ۳) $\frac{1}{\sqrt{2\pi\sigma}}$ ۴) σ

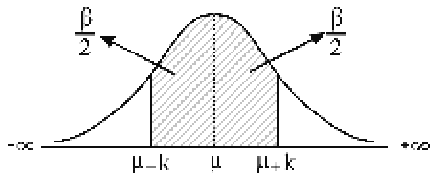
حل: گزینه ۲ درست است.

تقارن و سطح زیرمنحنی

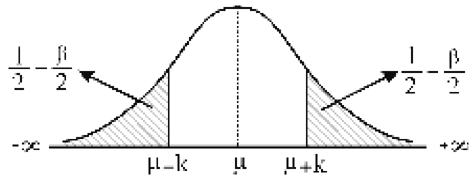
به ازای هر نقطه دلخواه k در روابط زیر، به علت تقارن منحنی نسبت به خط $x = \mu$ داریم:



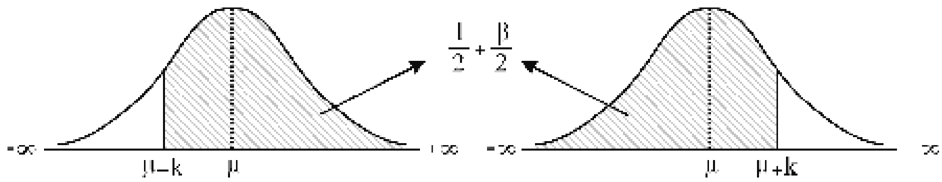
$$P(\mu - k \le X \le \mu + k) = \beta$$



$$1)\ P(\mu-k\le X\le\mu)=P(\mu\le X\le\mu+k)=\frac{\beta}{2}$$



$$2)\ P(X\le\mu-k)=P(X\ge\mu+k)=\frac{1}{2}-\frac{\beta}{2}$$

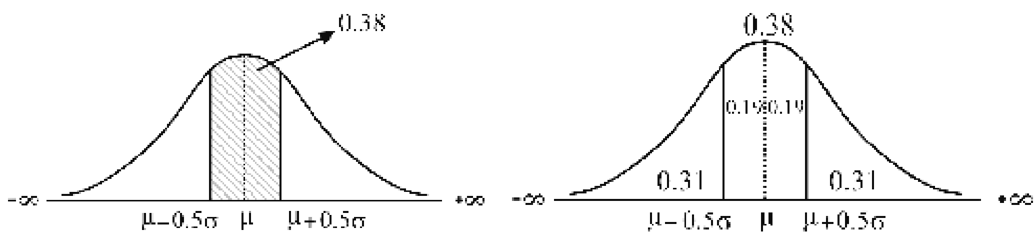


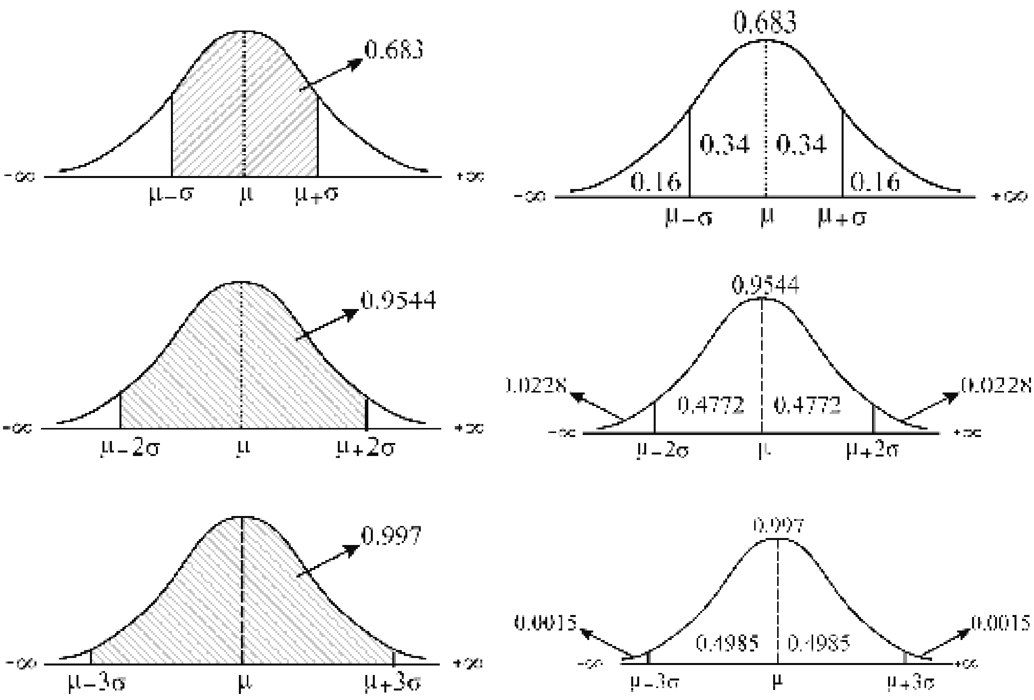
$$3)\ P(X\ge\mu-k)=P(X\le\mu+k)=\frac{1}{2}+\frac{\beta}{2}$$

انحرافات حول میانگین

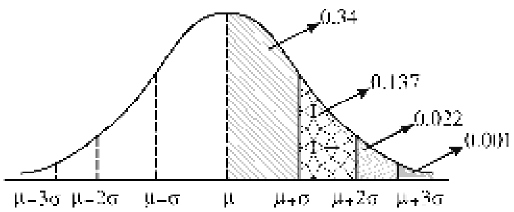
در توزیع نرمال، مقدار احتمال برای 0.5 ,1 ,2 و 3 انحراف معیار حول میانگین به صورت زیر است:

احتمال در فاصله 0.5 انحراف معیار (0.5σ) حول میانگین:	$P(\mu-0.5\sigma\le X\le\mu+0.5\sigma)\simeq 0.38$
احتمال در فاصله 1 انحراف معیار (σ) حول میانگین:	$P(\mu-\sigma\le X\le\mu+\sigma)=0.683\simeq 0.68$
احتمال در فاصله 2 انحراف معیار (2σ) حول میانگین:	$P(\mu-2\sigma\le X\le\mu+2\sigma)=0.9544\simeq 0.95$
احتمال در فاصله 3 انحراف معیار (3σ) حول میانگین:	$P(\mu-3\sigma\le X\le\mu+3\sigma)=0.997$





در حالت کلی مقدار احتمال در فواصل 1, 2 و 3 انحراف معیار بالای میانگین به طور تقریبی به صورت زیر است:

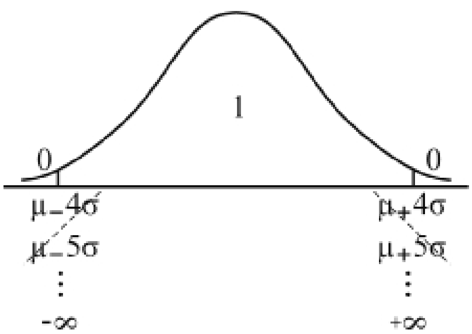


نکته:

۱ـ با توجه به احتمال مربوط به 3 انحراف معیار حول میانگین (0.997) و با در نظر گرفتن این موضوع که سطح کل زیر منحنی نرمال 1 است، می‌توان نتیجه گرفت که:

اولاً: احتمال یا سطح زیر منحنی خارج از فاصله $(\mu - 3\sigma , \mu + 3\sigma)$ تقریباً برابر 0 است.

ثانیاً: احتمال یا سطح زیر منحنی، درون فواصل بیش از 3 انحراف معیار یعنی $(\mu \pm 4\sigma), (\mu \pm 5\sigma), \ldots$ تقریباً برابر 1 است.



۲ـ برای هر بازه به صورت $(\mu - k\sigma , \mu + k\sigma)$ همواره می‌توان μ,σ و k را با استفاده از روابط زیر محاسبه کرد:

$$\Big(\underbrace{\mu - k\sigma}_{a} , \underbrace{\mu + k\sigma}_{b}\Big) \longrightarrow \begin{cases} \mu = \dfrac{a+b}{2} \\ k\sigma = \dfrac{b-a}{2} \end{cases}$$

مثال ۱ درآمد حدود 95% از رانندگان تاکسی در روز بین 1000 تا 5000 تومان است. با فرض نرمال بودن توزیع درآمد، انحراف معیار درآمد این صنف کدام است؟

۱) 1000 ۲) 1500 ۳) 2000 ۴) 666.6

حل: گزینه ۱ درست است.

می‌دانیم در توزیع نرمال: $P(\mu-2\sigma < X < \mu+2\sigma) = 0.9544 \simeq 0.95 \quad \to \quad k=2$

بنابراین با توجه به نکته ۲ داریم:

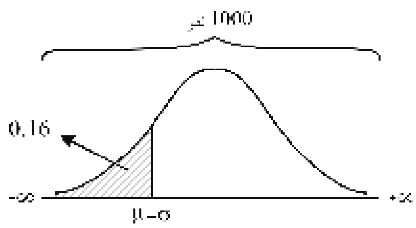
$$\Big(\underbrace{1000}_{a} , \underbrace{5000}_{b}\Big) \longrightarrow \begin{cases} \mu = \dfrac{a+b}{2} = \dfrac{1000+5000}{2} = 3000 \\ k\sigma = \dfrac{b-a}{2} \longrightarrow 2\sigma = \dfrac{5000-1000}{2} \longrightarrow \sigma = 1000 \end{cases}$$

مثال ۲ برای بررسی وضعیت تحصیلی 1000 دانش‌آموز، نمرات آن‌ها را بر روی محور نرمال برده‌اند. برای این‌کار ابتدا میانگین (μ) و انحراف معیار (σ) نمرات را به دست آورده و سپس نمرات را روی منحنی نرمال در نظر گرفته‌اند. مشخص کنید چند نفر از دانش‌آموزان در محدوده نمرات زیر قرار دارند؟

حل:

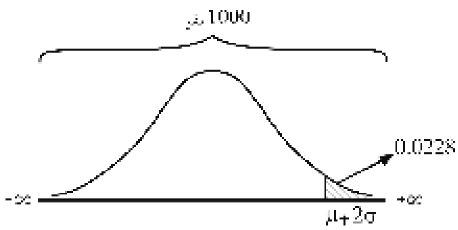
محدوده نمرات

$x < \mu - \sigma$



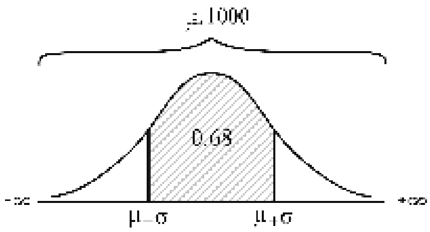
نفر $\to = 1000 \times 0.16 = 160$

$x > \mu + 2\sigma$



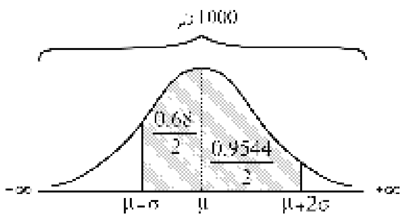
نفر $\to = 1000 \times 0.0228 = 22.8 \simeq 23$

$\mu-\sigma<x<\mu+\sigma$



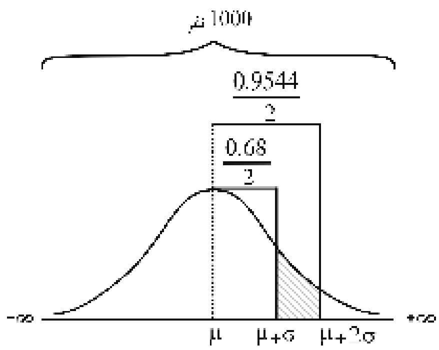
$\rightarrow = 1000\times(0.68) = 680$ نفر

$\mu-\sigma<x<\mu+2\sigma$



$\rightarrow = 1000\times\left(\frac{0.9544}{2}+\frac{0.68}{2}\right)$
$= 1000\times 0.8172 = 817.2 \simeq 818$ نفر

$\mu+\sigma<x<\mu+2\sigma$



$\rightarrow = 1000\times\left(\frac{0.9544}{2}-\frac{0.68}{2}\right)$
$= 1000\times 0.1372 = 137.2 \simeq 138$ نفر

ترکیب خطی از توزیع نرمال

هرگاه X دارای توزیع نرمال $X\sim N(\mu,\sigma^2)$ باشد، هر ترکیب خطی از آن به صورت $Y = bX + a$ نیز نرمال است؛ در این وضعیت کافی است امید و واریانس Y را محاسبه کنیم.

مثال متغیر تصادفی X بر طبق قانون نرمال با امید ریاضی 150 و واریانس 64 توزیع شده است. اگر متغیر تصادفی Y براساس معادله $Y = \frac{1}{2}X + 25$ از X پیروی کند، آن‌گاه تابع چگالی احتمال متغیر تصادفی Y عبارت است از: (اقتصاد ـ ۸۳)

۱) $f(y) = \frac{1}{4\sqrt{2\pi}}e^{-\frac{(y-100)^2}{32}}$

۲) $f(y) = \frac{1}{4\sqrt{2\pi}}e^{-\frac{(y-100)^2}{16}}$

۳) $f(y) = \frac{1}{16\sqrt{2\pi}}e^{-\frac{(y-150)^2}{64}}$

۴) $f(y) = \frac{1}{16\sqrt{2\pi}}e^{-\frac{(y-75)^2}{32}}$

حل: گزینه ۱ درست است.

$$\begin{cases} X\sim N\left(\mu_X = 150\,,\,\sigma_X^2 = 64\right) \\ Y = \frac{1}{2}X + 25 \rightarrow \begin{cases} E(Y) = \frac{1}{2}E(X) + 25 = 100 \rightarrow \mu_Y = 100 \\ \sigma^2(Y) = \frac{1}{4}\sigma_X^2 = 16 \rightarrow \sigma_Y = 4 \end{cases} \end{cases}$$

$$f(y)=\frac{1}{\sqrt{2\pi}\ \sigma_y}e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma}\right)^2}=\frac{1}{\sqrt{2\pi}\ 4}e^{-\frac{1}{2}\left(\frac{y-100}{4}\right)^2}=\frac{1}{4\sqrt{2\pi}}e^{-\frac{(y-100)^2}{32}}$$

توزیع $\sum_{i=1}^{n} x_i$

اگر $X_1, X_2, \dots, X_n$ مستقل از هم و دارای توزیع نرمال $X_i \sim N\left(\mu, \sigma^2\right)$ باشند، آن‌گاه $\sum_{i=1}^{n} x_i$ دارای توزیع نرمال $N\left(n\mu, n\sigma^2\right)$ است:

$$\sum_{i=1}^{n} x_i \sim N\left(n\mu, n\sigma^2\right)$$

نتیجه:

مجموع متغیرهای مستقل نرمال، همواره نرمال است و میانگین و واریانس مجموع متغیرهای مستقل نرمال برابر است با مجموع میانگین‌ها و مجموع واریانس‌ها.

توزیع نرمال استاندارد (Standard Normal Distribution)

ازآنجاکه انتگرال‌گیری از تابع چگالی نرمال برای محاسبه احتمال در فاصله محدود غیرممکن است، با استفاده از روش‌های عددی، جدولی به دست آمده است که مقادیر احتمال مربوط به توزیع نرمال با میانگین 0 و واریانس 1 در آن قابل محاسبه است؛ بنابراین برای محاسبه احتمال در هر توزیع نرمال به صورت $X \sim N\left(\mu, \sigma^2\right)$ ابتدا باید توزیع نرمال را به صورت $N(0,1)$ تبدیل کنیم، سپس از روی جدول نرمال استاندارد مقدار عددی احتمال را به دست آوریم.

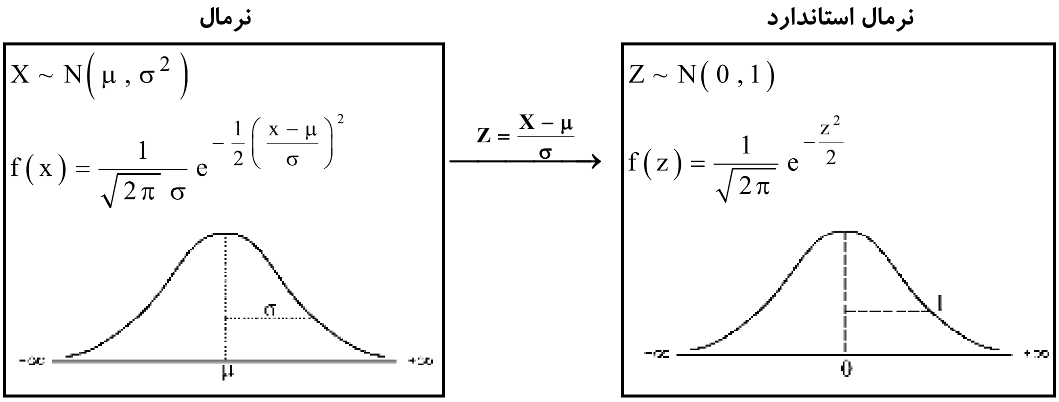
تبدیل نرمال $N\left(\mu, \sigma^2\right)$ به نرمال استاندارد $N(0,1)$

با استفاده از تغییر متغیر $\mathbf{Z = \frac{X - \mu}{\sigma}}$، تابع نرمال $X \sim N\left(\mu, \sigma^2\right)$ به صورت زیر به تابع نرمال استاندارد $Z \sim N(0,1)$ تبدیل می‌شود:

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = E\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{E(X)}{\sigma} - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$\sigma^2(Z) = \sigma^2\left(\frac{X-\mu}{\sigma}\right) = \sigma^2\left(\frac{X}{\sigma} - \underbrace{\cancel{\frac{\mu}{\sigma}}}_{0}\right) = \left(\frac{1}{\sigma}\right)^2 \sigma_X^2 = \frac{\sigma^2}{\sigma^2} = 1$$

درنتیجه:

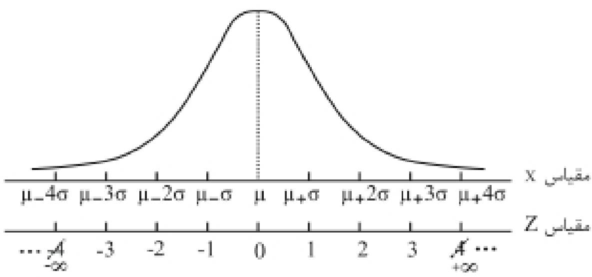


$$X \sim N\left(\mu, \sigma^2\right) \xrightarrow{\mathbf{Z = \frac{X - \mu}{\sigma}}} Z \sim N(0,1)$$

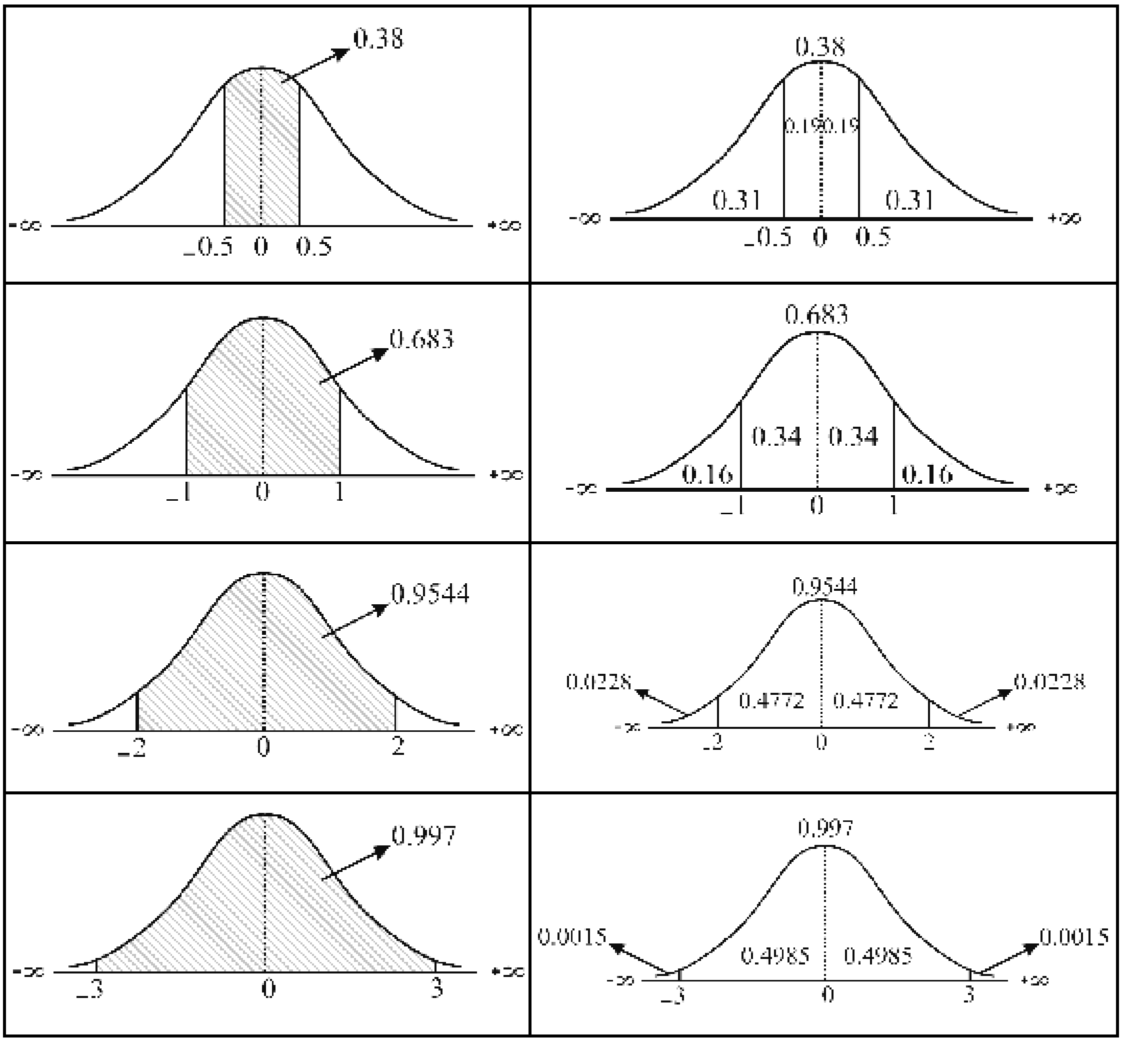
تغییر مقیاس از $X \sim N\left(\mu, \sigma^2\right)$ به $Z \sim N(0,1)$

با تبدیل متغیر $Z = \frac{X-\mu}{\sigma}$، مقیاس از X به Z به صورت زیر تغییر می‌کند:

با توجه به سطوح مطرح‌شده در توزیع نرمال:



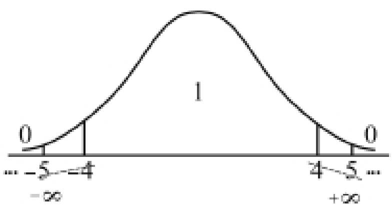
و با در نظر گرفتن محور تقارن $Z = 0$ در نرمال استاندارد خواهیم داشت:



نکته: با توجه به احتمال مربوط به $P(-3 < Z < 3) = 0.997$ و با در نظر گرفتن این موضوع که سطح کل زیرمنحنی نرمال استاندارد 1 است، می‌توان نتیجه گرفت که:

اولاً: احتمال یا سطح زیرمنحنی خارج از فاصله $(-3 < Z < 3)$ تقریباً برابر صفر است.

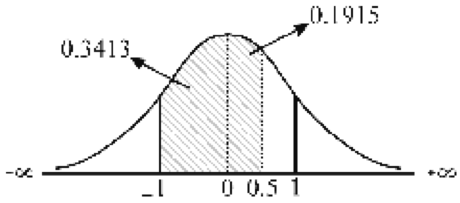
ثانیاً: احتمال یا سطح زیرمنحنی، درون فواصل بیش از $Z = \pm 3$ یعنی $(-4 < Z < 4)$ یا $(-5 < Z < 5)$ و ... تقریباً برابر 1 است.



مثال ۱ اگر $P(0 \le Z \le 0.5) = 0.1915$ و $P(0 \le Z \le 1) = 0.3413$ باشد، $P(-1 \le Z \le 0.5)$ کدام است؟

۱) 0.1498 ۲) 0.4672 ۳) 0.5328 ۴) 0.8502

حل: گزینه ۳ درست است.



ازآنجاکه در منحنی نرمال استاندارد، $z=0$ محور تقارن است، با توجه به شکل، رابطه زیر برقرار است

$$P(-1<Z<0)=P(0<Z<1)=0.3413$$

و داریم:

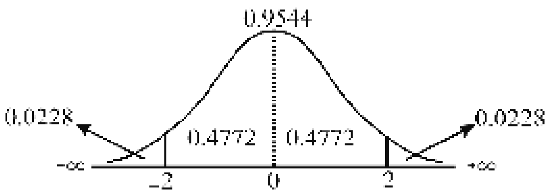
$$P(-1<Z<0.5)=P(-1<Z<0)+P(0<Z<0.5)=0.3413+0.1915=0.5328$$

مثال ۲ فرض کنید Z متغیر استاندارد و $\int_0^2 f(z)dz=0.4772$ است. $P(Z>2)$ کدام است؟

۱) 0.0228 ۲) 0.9772 ۳) 0.4772 ۴) 0.5228

حل: گزینه ۱ درست است.

با توجه به شکل:

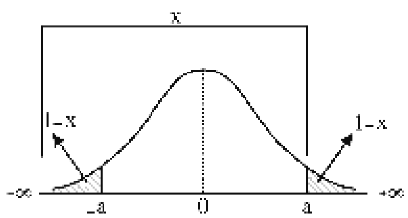


$$\int_0^2 f(z)dz=0.4772 \rightarrow P(0<Z<2)=0.4772 \rightarrow P(Z>2)=0.5-0.4772=0.0228$$

مثال ۳ اگر در یک توزیع نرمال استاندارد $\int_{-\infty}^{a}=x$ باشد، آن‌گاه $\int_{-a}^{+a}$ کدام است؟

۱) $x-0.5$ ۲) $2-2x$ ۳) $2x-1$ ۴) $2x-0.5$

حل: گزینه ۳ درست است.



با توجه به محور تقارن $z=0$، داریم $P(Z>a)=P(Z<-a)$، بنابراین:

$$\int_{-\infty}^{a}=P(-\infty<Z<a)=P(Z<a)=x \quad \rightarrow \quad \begin{cases} P(Z>a)=1-x \\ P(Z<-a)=1-x \end{cases}$$

حال با توجه به شکل ازآنجاکه سطح کل زیرمنحنی برابر 1 است، داریم:

$$P(-a<Z<a)=1-\left[\underbrace{P(Z>a)}_{1-x}+\underbrace{P(Z<-a)}_{1-x}\right]=1-2(1-x)=2x-1$$

یا

$$P(-a<Z<a)=\underbrace{P(Z<a)}_{x}-\underbrace{P(Z<-a)}_{1-x}=2x-1$$

توزیع $\sum_{i=1}^{n} z_i$

درصورتی‌که $Z_1, Z_2, \ldots, Z_n$ مستقل از هم و دارای توزیع نرمال استاندارد $Z_i \sim N(0,1)$ باشند، آن‌گاه $\sum_{i=1}^{n} z_i$ دارای توزیع نرمال $N(0,n)$ است و دیگر نرمال استاندارد نیست:

$$\boxed{\sum_{i=1}^{n} z_i \sim N(0,n)}$$

نتیجه: مجموع متغیرهای مستقل نرمال استاندارد، دارای توزیع نرمال است ولی استاندارد نیست.

مثال ۱ درصورتی‌که Z_1, Z_2 مستقل و دارای توزیع نرمال استاندارد باشند، آن‌گاه $Z_1 + Z_2$ دارای توزیع ....

۱) نرمال استاندارد است. ۲) $N(0,2)$ است. ۳) نرمال نیست. ۴) $N(0,3)$ است.

حل: گزینه ۲ درست است.

$$Z_1 + Z_2 \sim N(0,2)$$

نکته: هرگاه X دارای توزیع نرمال استاندارد $Z \sim N(0,1)$ باشد، دو قاعده زیر برقرار است:

1) $E\left(Z^{2m+1}\right) = 0 \quad ; \; m = 0,1,2,\ldots \rightarrow$ امید توان فرد Z برابر صفر است.

2) $E\left(Z^{2m}\right) = \dfrac{(2m)!}{2^m m!} \quad ; \; m = 0,1,2,\ldots$

مثال ۲ متغیر تصادفی X دارای توزیع نرمال با میانگین صفر و واریانس 2 است. $E\left(X^4\right)$ برابر است با:

۱) 18 ۲) 16 ۳) 12 ۴) 6

حل: گزینه ۳ درست است.

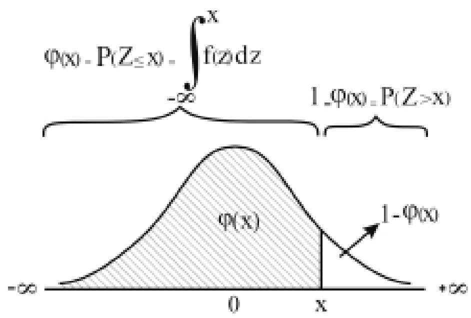
ابتدا X را به Z تبدیل می‌کنیم، سپس با توجه به قاعده دوم داریم:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-0}{\sqrt{2}} \rightarrow Z^4 = \frac{X^4}{4} \rightarrow X^4 = 4Z^4 \rightarrow E\left(X^4\right) = 4E\left(Z^4\right) \xlongequal{m=2} 4\frac{(2\times 2)!}{2!2^2} = 3\times 4 = 12$$

تابع توزیع تجمعی (Cumulative Distribution Function)

اگر $Z \sim N(0,1)$ متغیر نرمال استاندارد باشد، تابع توزیع تجمعی آن $F_Z(x) = P(Z \le x)$ به صورت زیر خواهد بود:

$$P(Z \le x) = \int_{-\infty}^{x} f(z)dz = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \varphi(x)$$



درنتیجه:

$$\boxed{\begin{array}{l} P(Z \le x) = \varphi(x) \\ P(Z > x) = 1 - \varphi(x) \end{array}}$$

با توجه به تقارن توزیع نرمال استاندارد داریم:

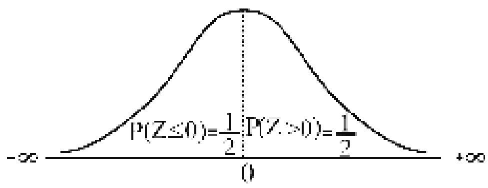
$$P(Z \le -x) = P(Z \ge +x) \quad \rightarrow \quad \boxed{\varphi(-x) = 1 - \varphi(x)}$$

یادآوری ۱: با استفاده از جدول نرمال، مقدار $\varphi(x)$ قابل محاسبه است.

یادآوری ۲: با توجه به پیوسته بودن توزیع نرمال، $P(Z = x) = 0$ است.

تابع توزیع تجمعی $\varphi(0)$

اگر $Z \sim N(0,1)$ استاندارد باشد، آن‌گاه:



$$\begin{cases} P(Z \le 0) = \varphi(0) = \dfrac{1}{2} \\ P(Z > 0) = 1 - \varphi(0) = \dfrac{1}{2} \end{cases}$$

اگر k مقدار ثابتی باشد (مثبت یا منفی)، آن‌گاه همواره:

$$P(kZ \le 0) = P(kZ \ge 0) = \frac{1}{2}$$

مثال اگر $Z \sim N(0,1)$ استاندارد باشد، مقدار $P(-6Z \le 0)$ کدام است؟

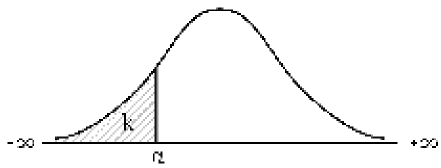
۱) 0.5 ۲) 0.25 ۳) $\frac{1}{6}$ ۴) 0

حل: گزینه ۱ درست است.

$$P(-6Z \le 0) = P\left(\frac{-6Z}{-6} \ge \frac{0}{-6}\right) = P(Z \ge 0) = \frac{1}{2}$$

نمایش سطح در نرمال استاندارد

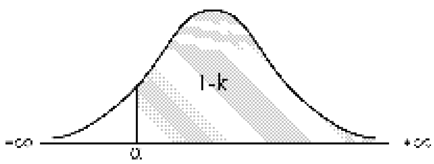
اگر $Z \sim N(0,1)$ باشد، سطح زیر منحنی به صورت‌های زیر نمایش داده می‌شود:



نمایش اول: $\int_{-\infty}^{\alpha} = k$

نمایش دوم: $F(\alpha) = \varphi(\alpha) = P(Z \le \alpha) = k$

نمایش سوم: $Z_k = \alpha$



نمایش اول: $\int_{\alpha}^{+\infty} = 1 - k$

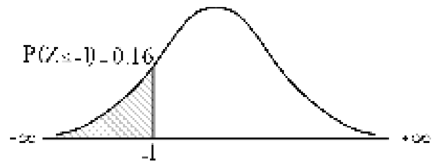
نمایش دوم: $P(Z > \alpha) = 1 - k$

نمایش سوم: $Z_{1-k} = \alpha$

نمایش اول : $\int_{-\infty}^{-1} = 0.16$

نمایش دوم : $F(-1) = \varphi(-1) = P(Z \le -1) = 0.16$

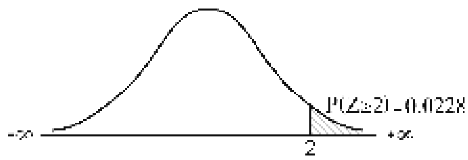
نمایش سوم : $Z_{0.16} = -1$



نمایش اول : $\int_{2}^{+\infty} = 0.0228$

نمایش دوم : $P(Z > 2) = 0.0228$

نمایش سوم : $Z_{0.0228} = 2$



انواع مسایل نرمال

۱- تبدیل نرمال به نرمال استاندارد

بین هر متغیر X با توزیع نرمال $X \sim N(\mu, \sigma^2)$ و متغیر استاندارد Z با توزیع $Z \sim N(0,1)$ همواره تبدیل متغیر زیر وجود دارد:

$$Z = \frac{X - \mu}{\sigma}$$

مثال ۱ اگر کمیت $X \sim N(9,16)$ توزیع شده باشد، مقدار استانداردشده $z = 1.5$ با کدام مقدار X برابر است؟

۱) 20 ۲) 15 ۳) 30 ۴) 7

حل: گزینه ۲ درست است.

$$Z = \frac{X - \mu}{\sigma} \xrightarrow[\mu=9,\ \sigma=4]{z=1.5} 1.5 = \frac{x-9}{4} \to x = 15$$

مثال ۲ نمرات دو داوطلب در آزمون درس آمار، نرمال و برابر با 14 و 12 است. این نمرات برحسب واحدهای استاندارد به ترتیب 0.25 و -0.25 شده، میانگین و واریانس نمرات آزمون چقدر است؟

۱) 16 , 13 ۲) 1 , 0 ۳) 4 , 12 ۴) 4 , 13

حل: گزینه ۱ درست است.

با توجه به رابطه $Z_i = \frac{X_i - \mu}{\sigma}$ که بین یک متغیر (X) نرمال و یک متغیر (Z) نرمال استاندارد وجود دارد و با توجه به مقادیر $x_1 = 14$, $x_2 = 12$ و $z_1 = 0.25$, $z_2 = -0.25$ داریم:

$$\begin{cases} Z_1 = \dfrac{X_1 - \mu}{\sigma} \\ Z_2 = \dfrac{X_2 - \mu}{\sigma} \end{cases} \to \begin{cases} 0.25 = \dfrac{14 - \mu}{\sigma} \\ -0.25 = \dfrac{12 - \mu}{\sigma} \end{cases} \to \begin{cases} 0.25\sigma = 14 - \mu \\ -0.25\sigma = 12 - \mu \end{cases} \to 0 = 26 - 2\mu \to \mu = 13$$

حال با توجه به $\mu = 13$، مقدار σ را می‌توانیم محاسبه کنیم:

$$0.25 = \frac{14 - \mu}{\sigma} \xrightarrow{\mu=13} \frac{1}{4} = \frac{14-13}{\sigma} \to \sigma = 4 \to \sigma^2 = 16$$

مثال ۳ اندازه قد دانش‌آموزان کلاس اول دارای توزیع نرمال با میانگین 100 و انحراف معیار 5 سانتی‌متر است. متغیر نرمال استاندارد برای قد دانش‌آموزان بین 110 تا 115 سانتی‌متر در کدام فاصله است؟

۱) $1 < z < 2$ ۲) $2 < z < 3$ ۳) $1.5 < z < 2$ ۴) $2 < z < 2.5$

حل: گزینه ۲ درست است.

$$\begin{cases} 110 < x < 115 \\ \mu = 100 \ , \ \sigma = 5 \end{cases} \xrightarrow{Z=\frac{X-\mu}{\sigma}} \frac{110-100}{5} < \frac{x-\mu}{\sigma} < \frac{115-100}{5} \rightarrow 2 < z < 3$$

۲- محاسبه احتمال در توزیع نرمال

اگر متغیر تصادفی X دارای توزیع نرمال به صورت $X \sim N(\mu, \sigma^2)$ باشد، برای محاسبه احتمال در هر بازه دلخواه به صورت زیر عمل می‌کنیم:

الف) با استفاده از رابطه $Z = \frac{X-\mu}{\sigma}$، احتمال مورد نظر را به فرم استاندارد تبدیل می‌کنیم.

ب) به کمک جدول نرمال استاندارد، مقدار احتمال مورد نظر را به دست می‌آوریم:

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right) = \varphi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X > a) = P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = P\left(Z > \frac{a-\mu}{\sigma}\right) = 1 - \varphi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \varphi\left(\frac{b-\mu}{\sigma}\right) - \varphi\left(\frac{a-\mu}{\sigma}\right)$$

مثال ۱ اگر X دارای توزیع نرمال با $\mu = 20$ و $\sigma^2 = 9$ باشد، احتمال آنکه کمیت تصادفی 2X مقادیری کمتر از 42 اختیار کند $(P(2X < 42))$، برابر است با:

۱) $\varphi(0.33)$ ۲) $\varphi(0.2)$ ۳) $\varphi(2)$ ۴) $\varphi(0.5)$

حل: گزینه ۱ درست است.

$$\begin{cases} P(2X < 42) = P(X < 21) = P\left(\frac{X-\mu}{\sigma} < \frac{21-20}{3}\right) = P\left(Z < \frac{1}{3}\right) = \varphi\left(\frac{1}{3}\right) = \varphi(0.33) \\ \mu = 20 \\ \sigma^2 = 9 \rightarrow \sigma = 3 \end{cases}$$

مثال ۲ در یک توزیع نرمال با میانگین 32 و واریانس 4، تقریباً چند درصد داده‌ها بین دو عدد 38 و 26 قرار می‌گیرند؟ $\left(S_{-\infty}^{-3} = 0.0013\right)$ (حسابداری ومدیریت ـ ۸۵)

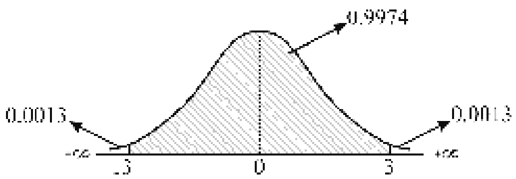
۱) 89.6 ۲) 92.3 ۳) 95.4 ۴) 99.7

حل: گزینه ۴ درست است.

$$\begin{cases} P(26 < X < 38) = P\left(\frac{26-32}{\sqrt{4}} < \frac{X-\mu}{\sigma} < \frac{38-32}{\sqrt{4}}\right) = P(-3 < Z < 3) = 0.9974 \rightarrow 99.7 \text{ درصد} \\ \mu = 32 \ , \ \sigma^2 = 4 \end{cases}$$

$$S_{-\infty}^{-3} = P(Z < -3) = P(Z > 3) = 0.0013$$

$$P(-3 < Z < 3) = 1 - 2 \times 0.0013 = 0.9974$$



مثال ۳ وزن خالص قوطی‌های روغن کارخانه‌ای بر روی قوطی، 500 گرم نوشته شده است، اما درواقع وزن قوطی‌های روغن دارای توزیع نرمال با میانگین 509.8 گرم و انحراف معیار 5 گرم است. چه نسبتی از قوطی‌ها کمتر از 500 گرم وزن دارند؟
(اقتصاد ـ ۷۶)

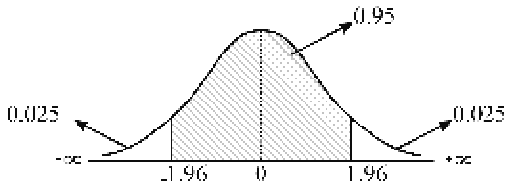
۱) 0.02 ۲) 0.0228 ۳) 0.025 ۴) 0.05

حل: گزینه ۳ درست است.

$$X \sim N\left(\mu = 509.8 \,,\, \sigma^2 = 5^2\right)$$

$$P\,(X < 500) = P\left(\frac{X-\mu}{\sigma} < \frac{500-509.8}{5}\right)$$

$$= P\left(Z < -\frac{9.8}{5}\right) = P\,(Z < -1.96) = 0.025$$



مثال ۴ متغیر تصادفی X بر طبق قانون نرمال با $\mu = 20$ و $\sigma^2 = 25$ توزیع شده است. احتمال اینکه کمیت تصادفی X مقادیری بیش از 10 اختیار کند، چقدر است؟ (با فرض اینکه $F_Z\,(-2) = 0.0228$ باشد.) (اقتصاد ـ ۷۵)

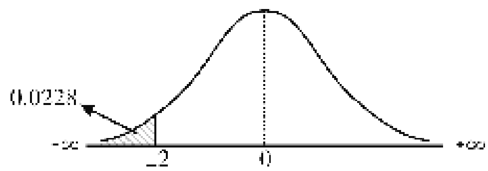
۱) 0.0228 ۲) 0.23413 ۳) 0.6587 ۴) 0.9772

حل: گزینه ۴ درست است.

$$\begin{cases} P\,(X > 10) = P\left(\dfrac{X-\mu}{\sigma} > \dfrac{10-20}{\sqrt{25}}\right) = P\left(Z > -2\right) = 0.9772 \\ \mu = 20 \;,\; \sigma^2 = 25 \;\;\rightarrow\;\; \sigma = 5 \end{cases}$$

$$F_Z\,(-2) = P\left(Z \le -2\right) = 0.0228$$

$$\rightarrow \quad P\left(Z > -2\right) = 1 - 0.0228 = 0.9772$$



مثال ۵ در یک بررسی از یک توزیع نرمال، $\mu = 100$ و $\sigma = 15$ مشاهده شده است. چه نسبتی از افراد 130 یا بالاتر از 130 هستند؟ $\left(\int_{-2}^{2} f\,(z)\,dz = 0.9544\right)$

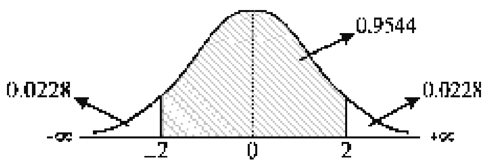
۱) 0.0228 ۲) 0.0456 ۳) 0.0912 ۴) 0.4772

حل: گزینه ۱ درست است.

$$\begin{cases} P\left(X \ge 130\right) = P\left(\dfrac{X-\mu}{\sigma} \ge \dfrac{130-100}{15}\right) = P\left(Z > 2\right) = 0.0228 \\ \mu = 100,\, \sigma = 15 \end{cases}$$

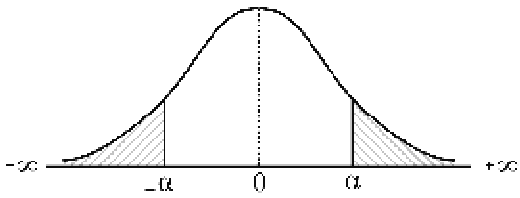
$$\int_{-2}^{2} f\left(z\right)dz = P\left(-2 < Z < 2\right) = 0.9544 \rightarrow$$

$$P\left(Z > 2\right) = P\left(Z < -2\right) = \frac{1-0.9544}{2} = 0.0228$$



۳- استفاده معکوس از جدول نرمال استاندارد

درصورتی‌که α مقدار ثابتی در بازه $-\infty < z < +\infty$ باشد، با توجه به شکل زیر داریم:



$$\rightarrow \boxed{\begin{array}{l} P(Z>\alpha)=P(Z<-\alpha) \\ P(Z<\alpha)=P(Z>-\alpha) \end{array}}$$

به عبارت دیگر اگر A و B مقادیر ثابتی در بازه $-\infty < z < +\infty$ باشند، آن‌گاه:

$$\boxed{\begin{array}{ll} \begin{array}{l} P(Z>A)=P(Z<B) \\ P(Z<A)=P(Z>B) \end{array} & \rightarrow \quad A=-B \\ \begin{array}{l} P(Z>A)=P(Z>B) \\ P(Z<A)=P(Z<B) \end{array} & \rightarrow \quad A=B \end{array}}$$

فرض کنید:

۰) X دارای توزیع نرمال به صورت $N(\mu,\sigma^2)$ است (σ^2 و μ معلوم یا مجهول).

۱) احتمال $P(X<a)$ یا $P(X>a)$ (a معلوم یا مجهول) مشخص و برابر با مقدار معلوم t باشد.

۲) با توجه به جدول نرمال استاندارد $P(Z<b)=t$ یا $P(Z>b)=t$ نیز مشخص است (b معلوم).

با توجه به این مفروضات، اگر یکی از مقادیر a، μ و σ مجهول باشند، برای محاسبه مقدار آن کافی است با استفاده از رابطه $Z=\frac{X-\mu}{\sigma}$ ابتدا مقدار استانداردشده قسمت (1) را به دست آوریم، سپس به کمک قسمت (2) مقدار مجهول را محاسبه کنیم؛ به عبارت دیگر:

$$(1)\ P(X<a)=P\left(Z<\frac{a-\mu}{\sigma}\right)=t \rightarrow \begin{cases} \xrightarrow{(2)\ P(Z<b)=t} \boxed{\frac{a-\mu}{\sigma}=b} \\ \xrightarrow[(2)\ P(Z>b)=t]{} \boxed{\frac{a-\mu}{\sigma}=-b} \end{cases}$$

یا

$$(1)\ P(X>a)=P\left(Z>\frac{a-\mu}{\sigma}\right)=t \rightarrow \begin{cases} \xrightarrow{(2)\ P(Z<b)=t} \boxed{\frac{a-\mu}{\sigma}=-b} \\ \xrightarrow[(2)\ P(Z>b)=t]{} \boxed{\frac{a-\mu}{\sigma}=b} \end{cases}$$

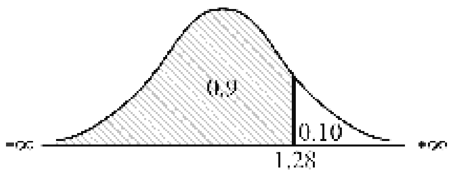
مثال ۱ در یک آزمون بزرگ، میانگین نمرات 60 و انحراف معیار نمرات 20 است. اگر 10% از شرکت‌کنندگان بتوانند نمره قبولی بگیرند، حداقل نمره قبولی چقدر است؟ $\left(S_{-\infty}^{1.28}=0.9\right)$

۱) 75.6 ۲) 75 ۳) 85 ۴) 85.6

حل: گزینه ۴ درست است.

$$(1)\begin{cases} P(X>\alpha)=0.10 \\ \mu=60\ ,\ \sigma=20 \end{cases} \rightarrow P\left(\frac{X-\mu}{\sigma}>\frac{\alpha-60}{20}\right)=0.10 \rightarrow P\left(Z>\frac{\alpha-60}{20}\right)=0.10$$

$$(2)\ S_{-\infty}^{1.28} = P(Z<1.28) = 0.90 \rightarrow P(Z>1.28) = 0.1$$



حال با در نظر گرفتن روابط (۱) و (۲) داریم:

$$\begin{aligned}&(1)\ P\left(Z>\frac{\alpha-60}{20}\right)=0.10\\&(2)\ \ P(Z>1.28)=0.1\end{aligned} \quad \rightarrow \quad 1.28=\frac{\alpha-60}{20} \quad \rightarrow \quad \alpha=85.6$$

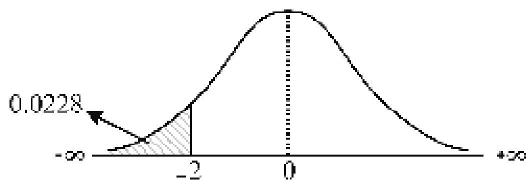
مثال ۲ اگر $P(Z\le -2)=0.0228$ و X نیز دارای توزیع نرمال با میانگین 15 و $P(X\le 35)=0.9772$ باشد، انحراف معیار X کدام است؟

۱) صفر   ۲) 5   ۳) 10   ۴) 15

حل: گزینه ۳ درست است.

$$(1)\ P(X\le 35)=P\left(\frac{X-\mu}{\sigma}\le\frac{35-\mu}{\sigma}\right)=P\left(Z\le\frac{35-15}{\sigma}\right)=0.9772$$

$$\begin{aligned}&(2)\ P(Z\le -2)=0.0228\\&\rightarrow\ P(Z\ge -2)=1-0.0228=0.9772\end{aligned}$$



$$\begin{aligned}&(1):P\left(Z\le\frac{20}{\sigma}\right)=0.9772\\&(2):P(Z\ge -2)=0.9772\end{aligned} \quad \rightarrow \quad -2=-\left(\frac{20}{\sigma}\right) \quad \rightarrow \quad \sigma=10$$

مثال ۳ فرض کنید $X\sim N(20,\sigma^2)$ است. اگر $P(X\le 22+\sigma)=0.9772$ باشد، مقدار انحراف معیار (σ) کدام است؟ $\left(S_2^5=0.0228\right)$

۱) 1   ۲) 2   ۳) 3   ۴) 4

حل: گزینه ۲ درست است.

$X\sim N(20,\sigma^2)$ و $\mu=20$ است، بنابراین:

$$(1):P(X\le 22+\sigma)=P\left(\frac{X-\mu}{\sigma}\le\frac{22+\sigma-20}{\sigma}\right)=P\left(Z<\frac{2+\sigma}{\sigma}\right)=0.9772$$

$$(2):S_2^{+\infty}=0.0228 \quad \rightarrow \quad P(Z>2)=0.0228 \quad \rightarrow \quad P(Z<2)=0.9772$$

$$(1),(2):\frac{2+\sigma}{\sigma}=2 \quad \rightarrow \quad \sigma=2$$

مثال ۴ اگر X دارای توزیع $N(\mu , 100)$ باشد و داشته باشیم $P(X>124)=0.05$ و $Z_{0.05}=1.65$، آن‌گاه مقدار μ برابر است با: (اقتصاد ـ ۸۵)

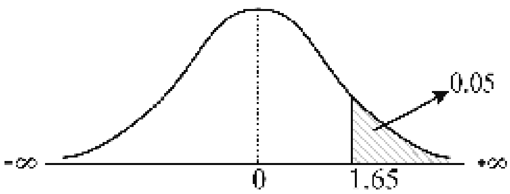
۱) 140.5 ۲) 121.9 ۳) 107.5 ۴) 104.5

حل: گزینه ۳ درست است.

با توجه به $X \sim N(\mu,100)$، مقدار $\sigma^2=100$ و درنتیجه $\sigma=10$ است، بنابراین:

$$(1)P(X>124)=P\left(\frac{X-\mu}{\sigma}>\frac{124-\mu}{\sqrt{100}}\right)=P\left(Z>\frac{124-\mu}{10}\right)=0.05$$

$$(2)Z_{0.05}=1.65 \rightarrow$$



$$\rightarrow P(Z>1.65)=0.05$$

$$(1),(2)\ \frac{124-\mu}{10}=1.65 \rightarrow \mu=107.5$$

مثال ۵ توزیع متغیر تصادفی X نرمال با میانگین 100 و انحراف معیار 10 است. اگر $P(X \ge x)=0.0495$ باشد، مقدار x چقدر است؟ (راهنمایی: $\int_{-4}^{-1.65} f(z)\,dz=0.0495$) (حسابداری ـ ۸۱)

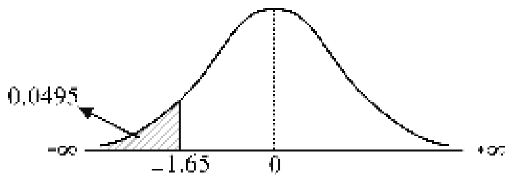
۱) 60 ۲) 83.5 ۳) 116.5 ۴) 140

حل: گزینه ۳ درست است.

$$(1)P(X \ge x)=P\left(\frac{X-\mu}{\sigma}\ge\frac{x-100}{10}\right)=P\left(Z\ge\frac{x-100}{10}\right)=0.0495$$

$$(2)\int_{\cancel{-4}\atop -\infty}^{-1.65} f(z)dz=P(Z<-1.65)=0.0495$$

$$(1),(2)\rightarrow -1.65=-\left(\frac{x-100}{10}\right)\rightarrow x=116.5$$



تقریب توزیع‌ها به وسیله توزیع نرمال

در شرایط مشخصی، توزیع‌های گسسته پواسون و دوجمله‌ای قابل تقریب به توزیع پیوسته نرمال هستند؛ یعنی در مواردی می‌توان از توزیع نرمال به‌جای این توزیع‌ها استفاده کرد.

۱- تقریب توزیع پواسون به نرمال

اگر میانگین توزیع پواسون (λ) به حدی بزرگ شود که $\lambda > 10$ باشد، آن‌گاه توزیع نرمال تقریب مناسبی برای توزیع پواسون خواهد بود (هرچه بزرگ‌تر شود تقریب بهتر است)، به طوری که:

توزیع پواسون		توزیع نرمال
$X \sim$ توزیع پواسون $\mu_X = \lambda$ $\sigma_X^2 = \lambda$	$\xrightarrow{\lambda > 10}$	$X \sim$ توزیع نرمال $\mu_X = \lambda$ $\sigma_X^2 = \lambda \rightarrow \sigma_X = \sqrt{\lambda}$

Z استاندارد در پواسون

برای تبدیل توزیع نرمال به نرمال استاندارد داریم:

$$\boxed{Z = \frac{X-\mu}{\sigma} \xrightarrow[\sigma=\sqrt{\lambda}]{\mu=\lambda} Z = \frac{X-\lambda}{\sqrt{\lambda}}}$$

مثال ۱ در یک توزیع پواسون با $\lambda = 64$ تقریب توزیع نرمال را در نظر می‌گیریم. عدد متناظر Z برای داده $x = 70$ کدام است؟

۱) 0.5 ۲) 0.25 ۳) 0.75 ۴) 1.25

حل: گزینه ۳ درست است.

$$Z = \frac{X-\lambda}{\sqrt{\lambda}} = \frac{70-64}{\sqrt{64}} = \frac{6}{8} = \frac{3}{4} = 0.75$$

مثال ۲ در یک توزیع پواسون با میانگین 16، مقدار $P(X<12)$ تقریباً چقدر است؟

۱) 0.16 ۲) 0.84 ۳) 0.68 ۴) 0.34

حل: گزینه ۱ درست است.

با توجه به آنکه در توزیع پواسون $\mu = \lambda = 16$ است، شرایط تقریب نرمال $(\lambda > 10)$ برقرار است و برای محاسبه احتمال از تقریب Z استفاده می‌کنیم.

$$P(X<12) = P\left(\frac{X-\mu}{\sigma} < \frac{12-\lambda}{\sqrt{\lambda}}\right) = P\left(Z < \frac{12-16}{4}\right) = P(Z<-1) = 0.16$$

۲- تقریب توزیع دوجمله‌ای به نرمال

اگر در یک توزیع دوجمله‌ای با پارامترهای n و p (n تعداد آزمایشات و p احتمال موفقیت) یکی از شرایط زیر برقرار باشد، توزیع نرمال با $\mu = np$ و $\sigma^2 = npq$ تقریب خوبی برای توزیع دوجمله‌ای است.

(a)

$X \sim Bin(n,p)$ $\mu = np$ $\sigma^2 = npq$	$\xrightarrow{np > 5 \text{ , } nq > 5}$ np و nq هرچه بیشتر از 5 باشند، تقریب بهتر است.	$X \sim N(\mu, \sigma^2)$ $\mu = np$ $\sigma^2 = npq$

(b)

$$\boxed{\begin{array}{l} X \sim Bin(n,p) \\ \mu = np \\ \sigma^2 = npq \end{array}} \xrightarrow{\text{برای } n \text{ های کوچک } p \simeq 0.5} \boxed{\begin{array}{l} X \sim N(\mu,\sigma^2) \\ \mu = np \\ \sigma^2 = npq \end{array}}$$

در شرایط مساوی تقریب (a) از تقریب (b) قوی‌تر و بهتر است.

Z استاندارد در دوجمله‌ای

برای تبدیل توزیع نرمال به نرمال استاندارد داریم:

$$\boxed{Z = \frac{X-\mu}{\sigma} \xrightarrow[\sigma=\sqrt{npq}]{\mu=np} Z = \frac{X-np}{\sqrt{npq}}}$$

مثال ۱ اگر n و p دو پارامتر توزیع دوجمله‌ای باشند، کدام‌یک از این موارد را می‌توان با توزیع نرمال تقریب زد؟

۱) $n=5$, $p=0.3$ ۲) $n=10$, $p=0.4$ ۳) $n=15$, $p=0.45$ ۴) $n=1000$, $p=0.5$

حل: گزینه ۴ درست است.

هرگاه در توزیع دوجمله‌ای $np>5$ و $nq>5$ باشد، آن‌گاه توزیع نرمال تقریب مناسبی برای بررسی توزیع دوجمله‌ای است:

گزینه۱) $n=5$, $p=0.3$, $q=0.7$ $\rightarrow \begin{cases} np=1.5<5 \\ nq=3.5<5 \end{cases} \rightarrow$ تقریب نرمال مناسب نیست.

گزینه۲) $n=10$, $p=0.4$, $q=0.6$ $\rightarrow \begin{cases} np=4<5 \\ nq=6>5 \end{cases} \rightarrow$ تقریب نرمال مناسب نیست.

گزینه۳) $n=15$, $p=0.45$, $q=0.55$ $\rightarrow \begin{cases} np=6.75>5 \\ nq=8.25>5 \end{cases} \rightarrow$ تقریب نرمال مناسب است.

گزینه۴) $n=1000$, $p=0.5$, $q=0.5$ $\rightarrow \begin{cases} np=500>5 \\ nq=500>5 \end{cases} \rightarrow$ تقریب نرمال مناسب است.

گزینه‌های ۳ و ۴ شرایط تقریب را دارند، اما ازآنجاکه در گزینه ۴، مقادیر np و nq به مراتب از عدد 5 بزرگ‌تر هستند، تقریب نرمال برای این گزینه بهتر از گزینه ۳ است.

مثال ۲ در یک توزیع دوجمله‌ای با پارامترهای $n=100$ و $p=\frac{1}{2}$ تقریب توزیع نرمال را در نظر بگیرید.

الف) پارامترهای تقریب نرمال را به دست آورید.

ب) عدد متناظر Z برای داده $x=55$ را به دست آورید.

حل:

الف) اگر در توزیع دوجمله‌ای $np>5$ و $nq>5$ شود، آن‌گاه از تقریب نرمال با پارامترهای $\mu=np$ و $\sigma^2=npq$ برای آن استفاده می‌شود:

$$\begin{cases} \mu = np = 100 \times \frac{1}{2} = 50 \\ \sigma^2 = npq = 100 \times \frac{1}{2} \times \frac{1}{2} = 25 \end{cases}$$

ب) رابطه $Z=\frac{X-\mu}{\sigma}$ برای متغیر X از نرمال و Z از نرمال استاندارد همواره برقرار است، در نتیجه بنا بر قسمت (الف) داریم:

$$\begin{cases} Z=\frac{X-\mu}{\sigma}=\frac{X-np}{\sqrt{npq}}=\frac{55-50}{5}=1 \\ x=55 \quad , \quad \mu=np=50 \quad , \quad \sigma^2=npq=25 \quad \to \quad \sigma=5 \end{cases}$$

مثال ۳ احتمال قبولی شرکت‌کنندگان در آزمونی 0.8 است، احتمال آنکه در یک نمونه 100 نفری حداقل 84 نفر قبول شوند، چقدر است؟

۱) 0.16 ۲) 0.84 ۳) 0.68 ۴) 0.34

حل: گزینه ۱ درست است.

با توجه به آنکه آزمون برای $n=100$ نفر تکرار شده و احتمال قبولی برای هر یک $p=0.8$ است، توزیع قبولی شرکت‌کنندگان در آزمون، دوجمله‌ای است با $n=100, p=0.8$.

حال از آنجاکه $nq=20>5, np=80>5$ است، استفاده از تقریب نرمال مناسب است:

$$P(X>84)=P\left(\frac{X-\mu}{\sigma}>\frac{84-np}{\sqrt{npq}}\right)=P\left(Z>\frac{84-80}{\sqrt{16}}\right)=P(Z>1)=0.16$$

تصحیح پیوستگی (دوجمله‌ای و پواسون)

هرگاه بخواهیم توزیع‌های گسسته مانند دوجمله‌ای و پواسون را تحت شرایط مطرح‌شده به نرمال که یک توزیع پیوسته است، تقریب بزنیم، برای محاسبه احتمال‌های مختلف، از تصحیح پیوستگی به شکل زیر استفاده می‌کنیم:

$$P(X=a)=P(a-0.5<X<a+0.5)$$

$$P(X\ge a)=P(X\ge a-0.5) \quad , \quad P(X>a)=P(X\ge a+0.5)$$

$$P(X\le a)=P(X\le a+0.5) \quad , \quad P(X<a)=P(X\le a-0.5)$$

$$P(a\le X\le b)=P(a-0.5\le X\le b+0.5)$$

مثال ۱ اگر بخواهیم $P(X\ge 10)$ را در توزیع دوجمله‌ای با توزیع نرمال تقریب بزنیم (در شرایطی که تقریب مجاز باشد)، کدام‌یک از موارد زیر را باید با تصحیح پیوستگی محاسبه کنیم؟

۱) $P(X\ge 9.5)$ ۲) $P(X\ge 10.5)$ ۳) $P(X\le 10.5)$ ۴) $P(X\le 9.5)$

حل: گزینه ۱ درست است.

$$P(X\ge 10)\equiv P(X\ge 10-0.5)=P(X\ge 9.5)$$

مثال ۲ اگر بخواهیم $P(X<8)$ را در توزیع پواسون به نرمال تقریب بزنیم (در شرایطی که $\lambda>10$ باشد)، کدام‌یک از موارد زیر با تصحیح پیوستگی حل شده است؟

۱) $P(X\le 8.5)$ ۲) $P(X\ge 7.5)$ ۳) $P(X\le 7.5)$ ۴) $P(X>8.5)$

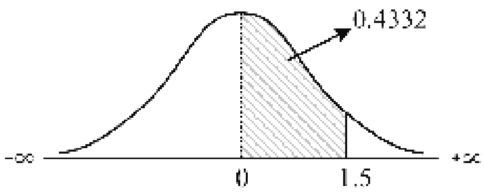
حل: گزینه ۳ درست است.

مثال ۳ متغیر X دارای توزیع پواسون با میانگین 25 است. $P(X \le 32)$ با تصحیح پیوستگی و استفاده از توزیع نرمال کدام است؟ $\left(S_0^{1.5} = 0.4332\right)$

۱) 0.1336 ۲) 0.4332 ۳) 0.9332 ۴) 0.5668

حل: گزینه ۳ درست است.

$$P(X \le 32) \equiv P(X \le 32 + 0.5) = P\left(\frac{X-\mu}{\sigma} \le \frac{32.5-\lambda}{\sqrt{\lambda}}\right) = P\left(Z \le \frac{32.5-25}{5}\right) = P(Z \le 1.5)$$



$$S_0^{1.5} = P(0 < Z < 1.5) = 0.4332$$

$$\rightarrow \; P(Z < 1.5) = 0.5 + 0.4332 = 0.9332$$

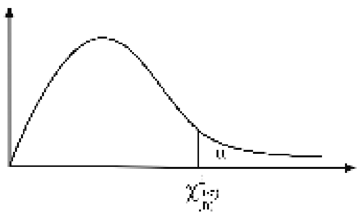
توجه: فقط هنگامی که در صورت سؤال ذکر شود با تصحیح پیوستگی، از تصحیح استفاده می‌کنیم در غیر این صورت مسئله را به شیوه معمول حل می‌کنیم.

توزیع‌های نتیجه‌شده از نرمال

توزیع کای اسکور (خی دو ، کای دو، مربع کای) (Chi – Square Distribution)

مقدمه: متغیر تصادفی χ^2 در خانواده‌ای از توزیع‌ها قرار دارد که هر یک با یک پارامتر n مشخص می‌شوند. مقدار n همان درجه آزادی یا تعداد مشاهدات مستقل در نمونه است.

مشخصات توزیع $\chi^2_{(n)}$ (کای اسکور با n درجه آزادی)



۱- یک توزیع تک نمایی است.

۲- مقادیر $\chi^2_{(n)}$ همگی مثبت هستند $\left(\chi^2_{(n)} > 0\right)$.

۳- توزیعی نامتقارن و دارای چولگی مثبت است.

۴- امید و واریانس توزیع کای دو

اگر متغیر تصادفی X دارای توزیع $\chi^2_{(n)}$ (کای‌دو با n درجه آزادی) باشد، امید و واریانس آن برابر است با:

امید ریاضی (میانگین)	$E(X) = n$ (درجه آزادی)
واریانس	$\sigma^2(X) = 2n$ (2 برابر درجه آزادی)
انحراف معیار	$\sigma_X = \sqrt{2n}$

یادآوری: $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\ , \Gamma(n) = (n-1)!$

مجموع و تفاضل دو متغیر χ^2

اگر کمیت‌های مستقل X و Y دارای توزیع χ^2 به ترتیب با m و n درجه آزادی باشند، آن‌گاه $X \pm Y$ دارای توزیع χ^2 با $m \pm n$ درجه آزادی است؛ به عبارت دیگر:

$$\begin{matrix} X \sim \chi^2_{(m)} \\ Y \sim \chi^2_{(n)} \end{matrix} \rightarrow \begin{cases} (X+Y) \sim \chi^2_{(m+n)} \\ (X-Y) \sim \chi^2_{(m-n)} \quad m>n \end{cases}$$

شکل توزیع χ^2 و درجه آزادی

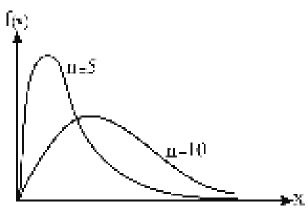
ازآنجاکه در توزیع $\chi^2_{(n)}$، درجه آزادی (n) تنها پارامتر توزیع است، شکل توزیع نیز باید توسط همین درجه آزادی مشخص شود.

شکل توزیع χ^2 بر اساس درجه آزادی (n) دو حالت دارد:

الف) اگر درجه آزادی کم باشد $(n \le 10)$، توزیع دارای چولگی مثبت است.

ب) اگر درجه آزادی بزرگ باشد $(n > 10)$، چولگی توزیع کم می‌شود و به سمت توزیع نرمال (قرینه) میل می‌کند.

$$\chi^2_{(n)} \xrightarrow{n>10} N\left(\mu = n\ , \sigma^2 = 2n\right)$$



رابطه χ^2 و نرمال استاندارد

بین توزیع χ^2 و توزیع نرمال استاندارد رابطه نزدیکی وجود دارد، به طوری که مجذور هر متغیر تصادفی نرمال استاندارد یک متغیر تصادفی χ^2 است. قضایای زیر، این رابطه را بررسی می‌کنند.

قضیه ۱: اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، آن‌گاه:

$$\chi^2_{(1)} = \left(\frac{X-\mu}{\sigma}\right)^2$$

نتیجه: اگر متغیرهای تصادفی مستقل $X_1, X_2, \dots, X_n$ دارای توزیع نرمال با میانگین μ_i و واریانس σ_i^2 باشند، آن‌گاه:

$$\chi^2_{(n)} = \sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

قضیه ۲: اگر متغیر تصادفی Z دارای توزیع نرمال استاندارد با میانگین 0 و واریانس 1 باشد، آن‌گاه:

$$\chi^2_{(1)} = Z^2$$

نتیجه: اگر متغیرهای تصادفی مستقل $Z_1, Z_2, \dots, Z_n$ دارای توزیع نرمال استاندارد با میانگین 0 و واریانس 1 باشند، آن‌گاه:

$$\chi^2_{(n)} = \sum_{i=1}^{n} z_i^2$$

نکته: طبق نتیجه بالا این توزیع، دیگر توزیع نرمال استاندارد نیست، چون واریانس برابر n است $(N(0,n))$.

مثال ۱ اگر Z_1، Z_2، Z_3 مستقل از هم دارای توزیع نرمال استاندارد باشند، توزیع $\sum_{i=1}^{3} z_i^2$ کدام است؟

۱) Z

۲) $t^2_{(3)}$

۳) $\chi^2_{(3)}$ (کای دو با 3 درجه آزادی)

۴) هیچ توزیعی را نشان نمی‌دهد.

حل: گزینه ۳ درست است.

با توجه به قضیه ۲ $\quad \rightarrow \quad \left(Z_1^2+Z_2^2+Z_3^2\right) \sim \chi^2_{(3)}$

مثال ۲ درصورتی‌که کمیت‌های مستقل X و Y بر طبق توزیع خی‌دو (کای دو) با واریانس‌های 16 و 10 باشند، کمیت $X + Y$ دارای چه توزیعی خواهد بود؟ (اقتصاد ـ ۸۴)

۱) نرمال با واریانس $4+\sqrt{5}$

۲) χ^2 با 26 درجه آزادی

۳) نرمال با واریانس 16

۴) χ^2 با 13 درجه آزادی

حل: گزینه ۴ درست است.

$$\left.\begin{array}{l}\sigma_X^2 = 2m = 16 \to m = 8 \text{ درجه آزادی} \\ \sigma_Y^2 = 2n = 10 \to n = 5 \text{ درجه آزادی}\end{array}\right. \to (X+Y) \sim \chi^2_{m+n} = \chi^2_{13}$$

دقت کنید اگر مقادیر 16 و 10 به عنوان امید ریاضی مطرح می‌شدند، همان درجه آزادی بودند و گزینه ۲ پاسخ درست بود، زیرا:

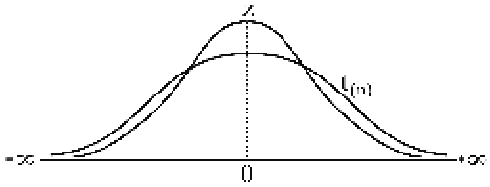
$$\begin{cases}\mu_X = m = 16 \quad \to \quad X \sim \chi^2_{(16)} \\ \mu_Y = n = 10 \quad \to \quad Y \sim \chi^2_{(10)}\end{cases} \to \quad (X+Y) \sim \chi^2_{(16+10)} = \chi^2_{26}$$

نکته:

۱- هرگاه X دارای توزیع $\chi^2_{(n)}$ با n درجه آزادی باشد، توزیع X معادل توزیع گاما با پارامترهای $r = \frac{n}{2}$ و $\lambda = \frac{1}{2}$ است.

۲- هرگاه X دارای توزیع $\chi^2_{(2)}$ با 2 درجه آزادی باشد، توزیع X معادل توزیع نمایی با پارامتر $\lambda = \frac{1}{2}$ است. (دقت کنید که این رابطه فقط برای $n = 2$ برقرار است.)

توزیع t–استیودنت (Student's t – Distribution)



توزیع t مانند توزیع نرمال متقارن است، اما نسبت به آن پراکندگی بیشتری دارد. هر متغیر t با یک درجه آزادی مشخص می‌شود که تعیین‌کننده ارتفاع و میزان پراکندگی توزیع است (درجه آزادی تعداد مشاهده مستقل در هر نمونه دلخواه از جامعه است).

همان‌طورکه در مبحث کشیدگی (فصل اول) مطرح شد، کشیدگی توزیع نرمال استاندارد برابر 3 است. حال ازآنجاکه پراکندگی توزیع t بیشتر از نرمال و درنتیجه، توزیع t کوتاه‌تر از توزیع نرمال است، کشیدگی آن حتماً کمتر از 3 است.

امید و واریانس توزیع t

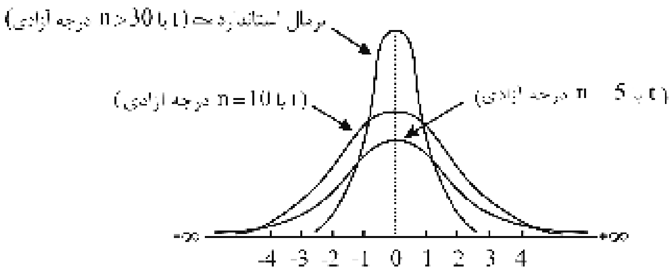
امید ریاضی (میانگین)	$E(X) = 0 \quad , \quad n > 1$
واریانس	$\sigma^2(X) = \dfrac{n}{n-2} \quad , \quad n > 2$
انحراف معیار	$\sigma_X = \sqrt{\dfrac{n}{n-2}} \quad , \quad n > 2$

رابطه توزیع t با توزیع نرمال استاندارد و کای‌دو

اگر متغیر تصادفی Z دارای توزیع **نرمال استاندارد** باشد $\left(Z \sim N(0,1)\right)$ و متغیر تصادفی $\chi^2_{(n)}$(کای اسکور با n درجه آزادی) مستقل از آن مفروض باشد، آن‌گاه کسر زیر دارای توزیع t **استیودنت با** n **درجه آزادی** است:

$$\boxed{t_{(n)} = \frac{Z}{\sqrt{\dfrac{\chi^2_{(n)}}{n}}}}$$

درواقع حاصل تقسیم یک متغیر نرمال استاندارد به جذر متغیر χ^2 به درجه آزادی‌اش، برابر توزیع t با درجه آزادی χ^2 است.



هر چه حجم نمونه مستقل یا همان n (درجه آزادی) بیشتر باشد، توزیع t متمرکزتر و پراکندگی آن کمتر می‌شود به طوری که برای نمونه‌های بزرگ $(n > 30)$، توزیع t تقریباً نرمال است.

رابطه مقادیر استاندارد و سطح زیر منحنی با کشیدگی

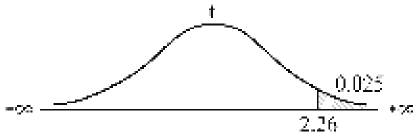
هرگاه یک توزیع متقارن، کوتاه‌تر یا بلندتر از توزیع نرمال استاندارد باشد، مقادیر استاندارد و سطح زیر منحنی برای آن توزیع نسبت به توزیع نرمال، به صورت زیر بررسی می‌شود:

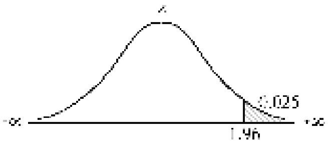
توزیع کوتاه‌تر از نرمال

هرگاه کشیدگی یک توزیع متقارن مانند t کمتر از کشیدگی نرمال استاندارد باشد (کوچک‌تر از 3) آنگاه:

۱- برای سطوح مشابه، مقدار استاندارد در توزیع مورد نظر **بزرگ‌تر** از مقدار استاندارد نرمال است.

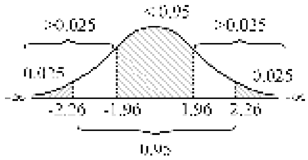
برای مثال، اگر سطح 0.025 را برای توزیع‌های نرمال و t درنظر بگیریم، داریم:

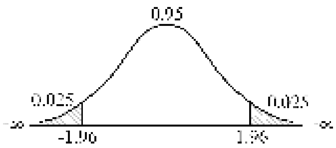




۲- برای هر بازه دلخواه از مقادیر استاندارد (بحرانی یا معنی‌دار) مانند $(-k,k)$، سطح بین دو مقدار، در توزیع نرمال و سطح خارج از دو مقدار، در توزیع مورد نظر بزرگ‌تر است.

برای مثال، اگر بازه $(-1.96,1.96)$ را برای توزیع‌های نرمال و t در نظر بگیریم، داریم:





توزیع بلندتر از نرمال

هرگاه کشیدگی یک توزیع متقارن بیشتر از کشیدگی نرمال استاندارد باشد (بزرگ‌تر از 3)، شرایط دقیقاً عکس حالتی است که توزیع مورد نظر کوتاه‌تر از نرمال باشد.

توزیع کوشی (Cauchy Distribution)

توزیع $t_{(1)}$ با یک درجه آزادی حاصل تقسیم دو متغیر تصادفی نرمال استاندارد است که در این وضعیت برابر با توزیعی به نام کوشی استاندارد با تابع چگالی زیر است که امید و واریانس ندارد.

$$t_{(1)}=\frac{Z}{Z_i}=\frac{1}{\pi\left(1+x^2\right)}$$

اثبات:

$$t_{(1)}=\frac{Z}{\sqrt{\frac{\chi^2_{(1)}}{1}}}=\frac{Z}{\sqrt{\frac{Z_i^2}{1}}}=\frac{Z}{Z_i}$$

مثال اگر Z_1 متغیری با توزیع نرمال استاندارد و Z_2 متغیری با توزیع χ^2 و با درجه آزادی k و مستقل از Z_1 باشد، متغیر $\frac{Z_1}{\sqrt{\frac{Z_2}{k}}}$ دارای توزیع:

۱) t است. ۲) F است.

۳) $\chi^2_{(3)}$ (کای دو با 3 درجه آزادی) است ۴) نرمال است.

حل: گزینه ۱ درست است.

$$t_{(k)}=\frac{Z}{\sqrt{\frac{\chi^2_{(k)}}{k}}}$$

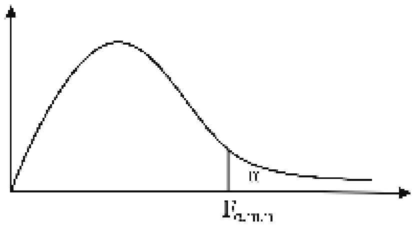
توزیع فیشر (F – Distribution)

توزیع F (فیشر) در خانواده‌ای از توزیع‌ها قرار دارد که با دو پارامتر m و n مشخص می‌شوند و از تقسیم دو متغیر تصادفی مستقل کای‌دو بر درجه آزادی آن‌ها به دست می‌آید.

$$F_{m,n}=\frac{\frac{\chi^2_{(m)}}{m}}{\frac{\chi^2_{(n)}}{n}}=\frac{n\chi^2_{(m)}}{m\chi^2_{(n)}}$$

مقدار n، درجه آزادی توزیع $\chi^2_{(n)}$ است که در صورت کسر قرار گرفته است.

مقدار m، درجه آزادی توزیع $\chi^2_{(m)}$ است که در مخرج کسر قرار گرفته است.



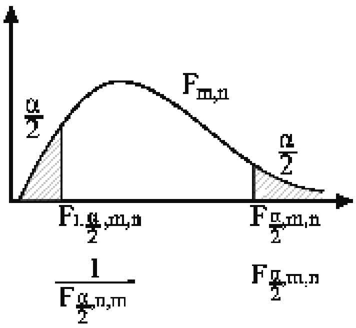
مشخصات توزیع $F_{m,n}$ (فیشر با m , n درجه آزادی)

۱- یک توزیع تک نمایی است.

۲- مقادیر $F_{m,n}$ همگی غیر منفی (مثبت) هستند $(F_{m,n}>0)$.

۳- توزیعی نامتقارن است که چولگی آن مثبت است.

۴- اگرچه توزیع F نامتقارن است اما نوعی تقارن معکوس در آن وجود دارد که بر اساس آن می‌توان حدود پایین را به صورت زیر به دست آورد:



$$F_{1-\frac{\alpha}{2},m,n}=\frac{1}{F_{\frac{\alpha}{2},n,m}}$$

۵- امید و واریانس توزیع F

اگر متغیر تصادفی X دارای توزیع $F_{m,n}$ با m و n درجه آزادی باشد، آن‌گاه امید و واریانس آن عبارت است از:

$$E(X)=\frac{n}{n-2}\qquad n>2$$

$$\sigma^2_X=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\ \overset{\text{در بعضی منابع}}{=\!=\!=}\ 2\left(\frac{1}{m}+\frac{1}{n}\right)\qquad n>4$$

۶- توزیع F با زیاد شدن درجات آزادی $(n,m>30)$ به سمت توزیع **نرمال** میل می‌کند.

$$F_{n,m}\sim N$$
$$n,m\to\infty$$

مثال ۱ اگر متغیرهای تصادفی U و V مستقل از یکدیگر بوده و هر یک به ترتیب دارای توزیع χ^2 با $m-1$ و $n-1$ درجه آزادی باشند، آن‌گاه متغیر تصادفی $\frac{(n-1)U}{(m-1)V}$ دارای چه توزیعی خواهد بود؟ (اقتصاد ـ ۷۳)

۱) $F_{(n-1,m-1)}$ ۲) $F_{(n,m)}$ ۳) $F_{(m-1,n-1)}$ ۴) $F_{(m,n)}$

حل: گزینه ۳ درست است.

$$\frac{(n-1)u}{(m-1)v}=\frac{\frac{u}{m-1}}{\frac{v}{n-1}}=\frac{\frac{\chi^2_{(m-1)}}{m-1}}{\frac{\chi^2_{(n-1)}}{n-1}}=F_{m-1,n-1}$$

مثال ۲ توزیع $F_{1,n}$ (فیشر با $n,1$ درجه آزادی) معادل کدام توزیع است؟

۱) $\chi^2_{(n)}$ ۲) Z^2 ۳) $t^2_{(n)}$ ۴) Z

حل: گزینه ۳ درست است.

$$F_{1,n}=\frac{\frac{\chi^2_{(1)}}{1}}{\frac{\chi^2_{(n)}}{n}}=\xrightarrow{Z^2=\chi^2_{(1)}}=\frac{\frac{Z^2}{1}}{\frac{\chi^2_{(n)}}{n}}=\left(\frac{Z}{\sqrt{\frac{\chi^2_{(n)}}{n}}}\right)^2=t^2_{(n)}$$

نتیجه: توزیع $F_{1,n}$ همان مجذور توزیع $t_{(n)}$ است.

$$\boxed{\left(t_{(n)}\right)^2=F_{1,n}}$$

مثال ۳ اگر $F_{0.05,2,10}=4.1$ باشد، $F_{0.95,10,2}$ کدام است؟

۱) 0.24 ۲) 4.1

۳) 3.1 ۴) باید جدول F در دسترس باشد.

حل: گزینه ۱ درست است.

$$F_{0.95,10,2}=\frac{1}{F_{0.05,2,10}}=\frac{1}{4.1}=0.24$$

مثال ۴ اگر X_1 و X_2 دارای توزیع نرمال استاندارد مستقل باشند، توزیع متغیر تصادفی $Y=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$ عبارت است از:

۱) مربع کای با یک درجه آزادی ۲) مربع کای با دو درجه آزادی

۳) توزیع F با یک و یک درجه آزادی ۴) توزیع t با یک درجه آزادی

حل: گزینه ۳ درست است.

$$X_1+X_2 \sim N(0,2)$$

$$X_1-X_2 \sim N(0,2)$$

$$Y=\frac{\left(\frac{X_1+X_2-0}{\sqrt{2}}\right)^2}{\left(\frac{X_1-X_2-0}{\sqrt{2}}\right)^2}=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}=\frac{Z_1^2}{Z_2^2}=\frac{\chi^2_{(1)}}{\chi^2_{(1)}}=F_{1,1}$$

مثال ۵ با افزایش حجم نمونه و میل کردن آن به سمت بی‌نهایت $(n\to\infty)$، توزیع‌های F, t, χ^2 به سمت ....

۱) توزیع پواسون میل می‌کنند.
۲) توزیع دوجمله‌ای میل می‌کنند.
۳) توزیع یکنواخت میل می‌کنند.
۴) توزیع نرمال میل می‌کنند.

حل: گزینه ۴ درست است.

در تمام توزیع‌های پیوسته $P(X=a)=0$ است.

تست‌های طبقه‌بندی شده

توزیع برنولی

۱. فرض کنید X_1 و X_2 متغیرهای تصادفی مستقل با تابع چگالی احتمال زیر باشند،

$$P(X=x)=\theta^x(1-\theta)^{1-x} \quad ; \quad x=0,1 \quad ; \; 0<\theta<1$$

در آن صورت $E\left(X_1^4 . X_2^4\right)$ برابر است با : (محیط زیست ــ ۸۸)

۱) θ^2 ۲) θ^4 ۳) $(1-\theta)^4$ ۴) $\theta^2(1-\theta)^2$

توزیع دوجمله‌ای

۲. به ازای کدام مقدار a، تابع $P(X=x)=\dfrac{\binom{4}{x}}{3a+1} \; ; \; x=0,1,2,3,4$، یک تابع احتمال است؟

(حسابداری و مدیریت ــ ۸۷)

۱) 6 ۲) 5 ۳) 4 ۴) 3

۳. شصت درصد افراد شرکت‌کننده در یک آزمون قبول شده‌اند. اگر X تعداد افراد قبول‌شده در هر انتخاب 96 نفری باشد، انحراف معیار X کدام است؟ (حسابداری و مدیریت ــ ۸۷)

۱) 4.8 ۲) 3.6 ۳) 5.4 ۴) 7.2

۴. ده درصد تراشه‌های تولیدی کارخانه‌ای معیوب است. اگر یک نمونه تصادفی 3 تایی از این تراشه‌ها انتخاب شود، احتمال مشاهده حداقل یک تراشه معیوب چند درصد است؟ (اقتصاد ــ۸۶)

۱) 23 ۲) 27 ۳) 73 ۴) 77

۵. ظرفیت هواپیمایی 360 نفر است ولی برای 400 نفر جا رزرو می‌شود. تعداد مسافرانی که جا رزرو کرده ولی برای پرواز حاضر نمی‌شوند به طور متوسط 40 نفر در هر پرواز است. احتمال اینکه همه 400 نفر برای پرواز حاضر شوند چقدر است؟ (اقتصاد ــ ۸۷)

۱) 0.1^{400} ۲) 0.9^{400} ۳) 0.1^{360} ۴) 0.9^{360}

۶. در یک توزیع دوجمله‌ای $E(X)=9$ و $Var(X)=6$ است. مقدار $P(X \geq 1)$ برابر است با: (اقتصاد ـ ۸۸)

۱) $\frac{2}{3}$ ۲) $1-\left(\frac{1}{3}\right)^{27}$ ۳) $\frac{1}{3}$ ۴) $1-\left(\frac{2}{3}\right)^{27}$

۷. احتمال جوانه زدن نوعی بذر $\frac{4}{5}$ است. اگر 5 عدد بذر از این نوع کاشته شود، با کدام احتمال فقط دو بذر جوانه می‌زند؟ (برنامه‌ریزی شهری ـ ۸۷)

۱) 0.128 ۲) 0.0924 ۳) 0.0512 ۴) 0.064

۸. احتمال موفقیت در یک آزمایش برنولی $\frac{3}{5}$ است. اگر X تعداد موفقیت‌ها در هر 24 مشاهده باشد، انحراف معیار X کدام است؟ (برنامه‌ریزی شهری ـ ۸۷)

۱) 1.2 ۲) 1.44 ۳) 2.4 ۴) 2.56

۹. در یک آزمایش برنولی احتمال موفقیت 98 درصد است. اگر X تعداد موفقیت‌ها در 25 بار تکرار این آزمایش باشد، انحراف معیار آن کدام است؟ (۸۸ ـ GIS)

۱) 0.49 ۲) 0.52 ۳) 0.63 ۴) 0.7

۱۰. معمولاً 80% کالای تولیدشده در یک کارخانه به طور استاندارد تولید می‌شود. از کالاهای تولید شده 3 کالا را انتخاب می‌کنیم. احتمال اینکه این 3 کالا معیوب باشند، چقدر است؟ (محیط زیست ـ ۸۶)

۱) 0.008 ۲) 0.488 ۳) 0.512 ۴) 0.982

۱۱. در یک توزیع دوجمله‌ای، میانگین برابر $\frac{8}{3}$ و حجم نمونه برابر 16 می‌باشد. واریانس چقدر است؟

(محیط زیست ـ ۸۶)

۱) $\frac{16\times 8}{3}$ ۲) $\frac{16}{3}$ ۳) $\frac{20}{9}$ ۴) $\sqrt{\frac{20}{9}}$

۱۲. متوسط تعداد دفعاتی که در 10 بار پرتاب یک سکه سالم خط خواهیم داشت عبارت است از: (محیط زیست ـ ۸۶)

۱) 2 ۲) 5 ۳) 7 ۴) 10

۱۳. متغیر تصادفی X از توزیع دوجمله‌ای با احتمال موفقیت 40% تبعیت می‌کند. در این صورت ضریب تغییرات جامعه برای یک نمونه 25 تایی برابر است با: (محیط زیست ـ ۸۶)

۱) $\frac{4}{10}$ ۲) $\frac{6}{10}$ ۳) $\frac{\sqrt{4}}{10}$ ۴) $\frac{\sqrt{6}}{10}$

۱۴. معمولاً 80% کالای تولیدشده در یک شرکت تولیدی به طور سالم تولید می‌شود. احتمال اینکه در یک نمونه 20 تایی حداکثر یک کالای معیوب وجود داشته باشد، چقدر است؟ (محیط زیست ـ ۸۷)

۱) $\binom{24}{5}\times(\%20)^{19}$ ۲) $\binom{5}{24}\times(\%80)^{19}$ ۳) $\binom{24}{5}\times(\%80)^{19}$ ۴) $\binom{5}{24}\times(\%20)^{19}$

۱۵. آزمونی چهار جوابی، دارای 20 سؤال است. شخصی به طور شانسی پاسخ‌ها را علامت می‌زند، اگر X نشان‌دهنده تعداد جواب‌های درست باشد، احتمال اینکه حداکثر به 9 سؤال جواب درست دهد از کدام رابطه باید محاسبه شود؟ **(محیط زیست ـ ۸۷)**

۱) $\binom{20}{9}(0.25)^9(0.75)^{11}$

۲) $\sum_{x=0}^{9}\binom{20}{x}(0.75)^x(0.25)^{20-x}$

۳) $\sum_{x=0}^{9}\binom{20}{x}(0.25)^x(0.75)^{20-x}$

۴) $\sum_{x=9}^{20}\binom{20}{x}(0.25)^x(0.75)^{20-x}$

۱۶. در سؤال قبل ، احتمال اینکه حداقل 3 جواب غلط داشته باشد، چیست؟ **(محیط زیست ـ ۸۷)**

۱) $\binom{20}{3}(0.25)^3(0.75)^{17}$

۲) $\sum_{x=0}^{17}\binom{20}{x}(0.75)^x(0.25)^{20-x}$

۳) $\sum_{x=3}^{20}\binom{20}{x}(0.75)^x(0.25)^{20-x}$

۴) $\sum_{x=0}^{17}\binom{20}{x}(0.25)^x(0.75)^{20-x}$

۱۷. در یک توزیع دوجمله‌ای، واریانس توزیع $\frac{2}{3}$ میانگین آن می‌باشد. حجم نمونه چقدر باشد تا میانگین جامعه برابر 20 باشد؟ **(محیط زیست ـ ۸۷)**

۱) 20 ۲) 40 ۳) 50 ۴) 60

۱۸. یک نوع بیماری با احتمال p افراد را مبتلا می‌کند. اگر n نفر را به طور تصادفی انتخاب کنیم توزیع تعداد افراد مبتلا .... است و امید ریاضی آن برابر است با .... **(محیط زیست ـ ۸۷)**

۱) فوق هندسی ، np ۲) فوق هندسی، $n(1-p)$ ۳) دوجمله‌ای ، np ۴) دوجمله ای، $n(1-p)$

توزیع چندجمله‌ای

۱۹.وزنه‌برداری در هر آزمون می‌تواند سه نوع امتیاز A، B و C را به ترتیب با احتمالات 0.5، 0.3 و 0.2 کسب نماید. احتمال اینکه در هفت بار آزمون امتیازات وی 2 بار A، 2 بار B و 3 بار C باشد، کدام است؟ **(حسابداری و مدیریت ـ ۸۶)**

۱) 0.0378 ۲) 0.0756 ۳) 0.0168 ۴) 0.0378

۲۰. تیراندازی باید 6 تیر به هدف رها کند، احتمال برخورد به داخل دایره وسط و بین دو دایره و خارج دو دایره به ترتیب 0.5، 0.4 و 0.1 است. احتمال اینکه از این 6 تیر، 3 تیر به وسط و 2 تیر بین دو دایره و 1 تیر به خارج اصابت کند کدام است؟ **(GIS ـ ۸۶)**

۱) 0.12 ۲) 0.18 ۳) 0.24 ۴) 0.36

۲۱.صفحه هدف متشکل از سه رنگ قرمز، سبز و زرد است. احتمال اصابت تیر به این رنگ‌ها به ترتیب 0.5، 0.3 و 0.2 است. از شش تیر رهاشده که به صفحه اصابت می‌کنند، با کدام احتمال 2 تیر به ناحیه قرمز، 1 تیر به ناحیه سبز و 3 تیر به ناحیه زرد برخورد می‌کند؟ **(GIS ـ ۸۸)**

۱) 0.018 ۲) 0.036 ۳) 0.045 ۴) 0.072

توزیع دوجمله‌ای منفی

۲۲.اگر کالایی معیوب باشد، فرد کنترل کیفیت با احتمال $\frac{3}{4}$ متوجه آن می‌شود. با کدام احتمال هفتمین کالای انتخابی، پنجمین کالای معیوبی می‌باشد که وی متوجه آن شده است؟ (GIS ـ ۸۷)

۱) $5\times 3^4\times 4^{-5}$ ۲) $5\times 3^5\times 4^{-6}$ ۳) $5\times 3^5\times 4^{-7}$ ۴) $5\times 3^6\times 4^{-7}$

توزیع هندسی

۲۳.دو تاس را با هم پرتاب می‌کنیم تا برای اولین بار مجموع دو عدد روشده 7 باشد، با کدام احتمال تعداد دفعات پرتاب‌شده فرد است؟ (حسابداری و مدیریت ـ ۸۶)

۱) $\frac{1}{2}$ ۲) $\frac{5}{9}$ ۳) $\frac{5}{11}$ ۴) $\frac{6}{11}$

۲۴.تاس سالمی را آن‌قدر پرتاب می‌کنیم تا عدد شش ظاهر شود. احتمال اینکه حداقل 2 پرتاب لازم باشد، چقدر است؟ (اقتصاد ـ ۸۸)

۱) $\frac{11}{36}$ ۲) $\frac{1}{6}$ ۳) $\frac{5}{36}$ ۴) $\frac{5}{6}$

توزیع فوق هندسی

۲۵.از جعبه‌ای که محتوی 12 عدد کالا است، 4 عدد آن معیوب است. به‌تصادف 2 تا را انتخاب می‌کنیم. اگر X تعداد کالای سالم انتخاب‌شده باشد، امید ریاضی X کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) $\frac{3}{2}$ ۲) $\frac{4}{3}$ ۳) $\frac{7}{6}$ ۴) $\frac{14}{11}$

۲۶.از 100 لامپ که 20 عدد آن غیراستاندارد است به طور تصادفی 5 لامپ انتخاب می‌شود. کمیت تصادفی X عبارت است از تعداد لامپ‌های غیراستاندارد بین 5 لامپ انتخاب‌شده، واریانس کمیت تصادفی X عبارت است از: (اقتصاد ـ ۸۶)

۱) 0.15 ۲) 0.77 ۳) 0.8 ۴) 0.95

۲۷.فرض کنید 2 شمع از 6 شمع یک اتومبیل 6 سیلندر معیوب هستند. اگر مکانیک به طور تصادفی 2 شمع را تعویض نماید، احتمال اینکه دقیقاً 2 شمع معیوب تعویض شده باشند برابر است با: (اقتصاد ـ ۸۷)

۱) $\frac{1}{2}$ ۲) $\frac{1}{3}$ ۳) $\frac{1}{15}$ ۴) $\frac{1}{30}$

۲۸.پنج درصد یک محموله 20 تایی از لامپ‌های روشنایی معیوب است. اگر یک مشتری 2 لامپ را خریداری کند، احتمال اینکه حداقل یکی از لامپ‌ها سالم باشد، چقدر است؟ (اقتصاد ـ ۸۸)

۱) 0.1 ۲) 0.5 ۳) 0.9 ۴) 1

۲۹. سه لامپ معیوب به طور غیر عمدی با 6 لامپ سالم مخلوط شده‌اند. از بین این لامپ‌ها، 2 عدد به‌تصادف انتخاب شده است. احتمال آنکه هر دو سالم باشند چقدر است؟ (محیط زیست ـ ۸۶)

۱) $\frac{2}{9}$ ۲) $\frac{6}{9}$ ۳) $\frac{5}{12}$ ۴) $\frac{7}{12}$

توزیع پواسون

۳۰. تعداد مشتری‌هایی که به یک فروشگاه مراجعه می‌کنند دارای توزیع پواسون با میانگین 3 مشتری در هر دقیقه است. با کدام احتمال در 80 ثانیه اول حداقل 2 مشتری مراجعه می‌کنند؟ $\left(e^{-4} = 0.018\right)$

(حسابداری و مدیریت ـ ۸۷)

۱) 0.915 ۲) 0.895 ۳) 0.870 ۴) 0.910

۳۱. به طور متوسط در هر دقیقه 10 نفر با یک مرکز مخابرات تماس می‌گیرند. با کدام احتمال در 30 ثانیه اول 4 نفر تماس می‌گیرند؟ $\left(e^{-5} = 0.007\right)$ (حسابداری و مدیریت ـ ۸۸)

۱) 0.168 ۲) 0.182 ۳) 0.196 ۴) 0.203

۳۲. به طور متوسط در هر شبانه‌روز 12 تصادف در یک شهر اتفاق می‌افتد. احتمال اینکه در 6 ساعت حداکثر یک تصادف اتفاق بیفتد، چند است؟ (اقتصاد ـ ۸۶)

۱) $1 - e^{-3}$ ۲) $1 - 3e^{-3}$ ۳) $4e^{-3}$ ۴) $3e^{-3}$

۳۳. به طور متوسط در هر 2 دقیقه یک نفر وارد کتابخانه مرکزی می‌شود. احتمال اینکه در 5 دقیقه بعد، حداقل یک نفر وارد کتابخانه شود برابر است با: (اقتصاد ـ ۸۷)

۱) $e^{-2.5}$ ۲) e^{-10} ۳) $1-e^{-10}$ ۴) $1-e^{-2.5}$

۳۴. به طور متوسط در هر 30 دقیقه یک مشتری وارد یک فروشگاه می‌شود. احتمال اینکه در 5 دقیقه آینده حداقل یک مشتری وارد فروشگاه شود چقدر است؟ (اقتصاد ـ ۸۸)

۱) $1-e^{-\frac{1}{6}}$ ۲) $1-e$ ۳) $1-e^{-1}$ ۴) $1-e^{\frac{1}{6}}$

۳۵. به طور متوسط با توزیع پواسون در هر دقیقه 8 اتومبیل از محلی در بزرگراه می‌گذرند. احتمال اینکه در 30 ثانیه لااقل 3 اتومبیل بگذرند کدام است؟ $\left(e^{-4} = 0.018\right)$ (برنامه‌ریزی شهری ـ ۸۸)

۱) 0.766 ۲) 0.792 ۳) 0.812 ۴) 0.823

۳۶. برای بررسی تعداد تخم مرغ‌های فاسد از محصولات یک مرغداری، هرگاه بدانیم به طور متوسط در هر شانه تخم مرغ 1 عدد تخم مرغ فاسد وجود دارد، احتمال اینکه در یک شانه دقیقاً 3 تخم مرغ فاسد وجود داشته باشد، چقدر است؟ (محیط زیست ـ ۸۷)

۱) $3e^{3}$ ۲) $\frac{e}{3!}$ ۳) $3e^{-3}$ ۴) $\frac{e^{-1}}{3!}$

۳۷. تجربه قبلی در مورد یک آزمون بیانگر این است که در مدت 60 ثانیه 2 سؤال پاسخ داده می‌شود. احتمال اینکه در مدت 30 ثانیه دو سؤال جواب داده شود، چقدر است؟ (محیط زیست ـ ۸۸)

۱) $\frac{1}{e}$ ۲) $\frac{2}{e}$ ۳) $\frac{1}{2e}$ ۴) $\frac{e}{3}$

تقریب توزیع دوجمله‌ای به پواسون

۳۸. در توزیع دوجمله‌ای $n = 4000$ و $p = 0.0015$ در تبدیل آن به توزیع پواسون، پارامتر توزیع کدام است؟

(برنامه‌ریزی شهری ـ ۸۷)

۱) 6 ۲) 3.6 ۳) 3 ۴) 2.4

۳۹.از هر 100 هزار واحد کالای موجود 125 واحد آن معیوب‌اند. اگر 1600 واحد از این کالا انتخاب شود، احتمال اینکه 4 عدد آن معیوب باشد کدام است؟ $\left(e^{-2}=0.135\right)$ (GIS ـ ۸۶)

۱) 0.09 ۲) 0.12 ۳) 0.18 ۴) 0.21

توزیع یکنواخت پیوسته

۴۰. تابع چگالی احتمال متغیر تصادفی X به صورت زیر است:

$$f(x)=\begin{cases}\frac{1}{4} & 1\le x\le 5\\ 0 & \text{x برای سایر مقادیر}\end{cases}$$

اگر واریانس X برابر $\frac{4}{3}$ باشد، ضریب تغییرات X کدام است؟ (اقتصاد ـ ۸۶)

۱) $\frac{2}{3\sqrt{3}}\times 100$ ۲) $\frac{4}{\sqrt{3}}\times 100$ ۳) $\frac{2\sqrt{3}}{4}\times 100$ ۴) $\frac{\sqrt{3}}{6}\times 100$

۴۱. توزیع یکنواخت $\alpha<x<\beta$; $f(x)=\frac{1}{\beta-\alpha}$ را در نظر بگیرید، $E\left(X^2\right)$ عبارت است از: (اقتصاد ـ ۸۷)

۱) $\frac{\beta^3-\alpha^3}{3(\beta-\alpha)}$ ۲) $\frac{\beta^3-\alpha^3}{(\beta-\alpha)}$ ۳) $\frac{(\alpha-\beta)^2}{12}$ ۴) $\frac{(\alpha+\beta)^2}{4}$

۴۲.اگر $f(x)=\begin{cases}\frac{1}{3} & 1<x<4\\ 0 & \text{سایر نقاط}\end{cases}$ تابع چگالی متغیر تصادفی X باشد، واریانس این متغیر تصادفی کدام است؟

(برنامه‌ریزی شهری ـ ۸۶)

۱) 0.25 ۲) 0.75 ۳) 1.25 ۴) 1.5

۴۳.سود شرکتی دارای توزیع یکنواخت بین -2 و 5 واحد پول است. واریانس سود این شرکت کدام است؟

(GIS ـ ۸۷)

۱) 4.08 ۲) 4.21 ۳) 3.85 ۴) 3.58

توزیع نمایی

۴۴.در تابع چگالی $f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}} & x>0\\ 0 & x\le 0\end{cases}$، میانگین X کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) $\frac{1}{2}$ ۲) 1 ۳) 2 ۴) 4

۴۵.تابع چگالی متغیر تصادفی X به صورت $x\ge 0$; $f(x)=e^{-x}$ است. واریانس X کدام است؟

(برنامه‌ریزی شهری ـ ۸۸)

۱) $\frac{1}{2}$ ۲) 1 ۳) 2 ۴) e

توزیع نرمال

۴۶. در 120 داده آماری دسته‌بندی شده، نمودار بافت‌نگار فراوانی مطلق متقارن است. مجموع این داده‌ها 840 و مجموع مربعات آن‌ها 6150 می‌باشد. تقریباً 95 درصد داده‌ها در کدام بازه است؟ (حسابداری و مدیریت ـ ۸۷)

۱) $(4 , 9)$ ۲) $(5 , 9)$ ۳) $(4 , 10)$ ۴) $(5 , 10)$

۴۷. در یک توزیع نرمال با میانگین 15.21 و واریانس 9، داده نظیر شصت و سومین صدک آن کدام است؟ $(P(Z<-0.33)=0.37)$ (حسابداری و مدیریت ـ ۸۷)

۱) 15.9 ۲) 16.3 ۳) 16.4 ۴) 16.2

۴۸. در یک امتحان ورودی کارشناسی ارشد، امتیاز شرکت‌کنندگان دارای توزیع نرمال با میانگین 75 و انحراف معیار 10 است. چه درصدی از شرکت‌کنندگان دارای امتیاز بیش از 85 می‌باشند؟ (اقتصاد ـ ۸۸)

۱) تقریباً 16% ۲) تقریباً 8% ۳) تقریباً 5% ۴) تقریباً 2.5%

۴۹. ضریب کشیدگی یا اوج تابع چگالی احتمال متغیر تصادفی X که متقارن است $\alpha_4 = 2$ است. احتمال اینکه متغیر تصادفی X کمیتی در فاصله دو انحراف معیار به طرفین میانگین را اختیار کند، چقدر است؟ (اقتصاد ـ ۸۸)

۱) حداقل 95% ۲) بیشتر از 95% ۳) تقریباً 95% ۴) کمتر از 95%

۵۰. در یک توزیع نرمال با میانگین 17.2 و واریانس 16، داده نظیر پنجاه و ششمین صدک آن کدام است؟ $(P(Z<-0.15)=0.44)$ (برنامه‌ریزی شهری ـ ۸۶)

۱) 18.4 ۲) 16.6 ۳) 18.2 ۴) 17.8

۵۱. در یک توزیع نرمال با میانگین 47 و واریانس 64، اگر به هر مقدار متغیر 5 واحد افزوده شود، آن‌گاه چند درصد داده‌های جدید بیشتر از 52 خواهد شد؟ (برنامه‌ریزی شهری ـ ۸۶)

۱) 48 ۲) 50 ۳) 52 ۴) 55

۵۲. نمرات ریاضی، توزیع نرمال با میانگین 14.5 و انحراف معیار 1.5 می‌باشد. اگر فردی از بین داوطلبان به‌تصادف انتخاب شود با کدام احتمال نمره وی بین 14.5 و 17 می‌باشد؟ $P(Z<-1.66)=0.048$ (برنامه‌ریزی شهری ـ ۸۷)

۱) 0.452 ۲) 0.476 ۳) 0.524 ۴) 0.548

۵۳. در یک توزیع نرمال با انحراف معیار 5 داریم $P(X\geq 9.8)=0.67$ و $P(Z<-0.44)=0.33$، میانگین این توزیع کدام است؟ (GIS ـ ۸۶)

۱) 8 ۲) 9 ۳) 11 ۴) 12

۵۴. اگر $P(Z\leq -1.5)=0.0668$ و X دارای توزیع نرمال با میانگین 25 و $P(X\geq 16)=0.9332$ باشد، انحراف معیار X کدام است؟ (GIS ـ ۸۷)

۱) 8 ۲) 6 ۳) 4 ۴) 3

۵۵. فرض کنید متغیر تصادفی دارای توزیع نرمال با میانگین 10 و واریانس 16 باشد، اگر احتمال کران پایین مقدار x برابر با 0.159 باشد، مقدار x چقدر است؟ (محیط زیست ـ ۸۶)

۱) 16 ۲) 14 ۳) 12 ۴) 8

۵۶.فرض کنید نمره درس آمار میان‌ترم کلاسی (X) دارای توزیع نرمال با میانگین 2 و واریانس 1 باشد و نمره درس آمار پایان‌ترم در همان کلاس (Y) دارای توزیع نرمال با میانگین 5 و واریانس 4 باشد (نمرات از 6 حساب شده است) و مدرس بخواهد نمره هر دانشجو را از رابطه $W=\frac{2Y+X}{3}$ حساب کند، توزیع W، امید ریاضی و واریانس کدام است؟ **(محیط زیست ــ ۸۷)**

۱) نرمال , 4 , $\frac{17}{9}$ ۲) نرمال , 4 , $\frac{17}{3}$ ۳) نرمال , 7 , 5 ۴) دوجمله‌ای , 4 , 5

تقریب توزیع دوجمله‌ای به نرمال

۵۷.تعداد 100 متقاضی به فروشگاهی مراجعه می‌کنند. احتمال آنکه هر یک خریدی را انجام دهند 0.8 است. احتمال آنکه حداقل 84 نفر خریدی را انجام دهند تقریباً چقدر است؟ **(اقتصاد ــ ۸۸)**

۱) %3 ۲) %16 ۳) %34 ۴) %47

توزیع کای ـ دو (خی‌دو، مربع کای، کای‌اسکور)

۵۸. اگر متغیر تصادفی X دارای توزیع کای دو $\left(\chi^2\right)$ با درجه آزادی 25 باشد، ضریب تغییرات این متغیر کدام است؟ **(اقتصاد ــ ۸۸)**

۱) %2 ۲) %25 ۳) %20 ۴) $\%20\sqrt{2}$

توزیع فیشر

۵۹.اگر $F_{0.1,4,8}=2.8$ باشد، مقدار $F_{0.9,8,4}$ کدام است؟ **(GIS ــ ۸۸)**

۱) 0.357 ۲) 0.753 ۳) 1.42 ۴) 2.41

پاسخ‌های تشریحی

توزیع برنولی

۱. گزینه ۱ درست است.

X_1 و X_2 دارای توزیع برنولی با پارامتر $p=\theta$ هستند، پس $E(X_1)=E(X_2)=\theta$ و داریم:

$$\begin{array}{c|cc} X_1 & 0 & 1 \\ \hline P(x) & 1-\theta & \theta \end{array} \quad \rightarrow \quad \begin{array}{c|cc} X_1^4 & 0^4 & 1^4 \\ \hline P(x_1^4) & 1-\theta & \theta \end{array} \quad \rightarrow \quad \begin{array}{c|cc} X_1^4 & 0 & 1 \\ \hline P(x_1^4) & 1-\theta & \theta \end{array}$$

بنابراین در توزیع برنولی هر توانی از X، دارای توزیع برنولی با همان پارامترهاست و ازآنجاکه X_1 و X_2 مستقل هستند، داریم:

$$E\left(X_1^4 . X_2^4\right)=E(X_1.X_2)=E(X_1)E(X_2)=\theta.\theta=\theta^2$$

توزیع دوجمله‌ای

۲. گزینه ۲ درست است.

توجه: یک حالت خاص از توزیع دوجمله‌ای که در آن $p=q=\frac{1}{2}$ باشد، دارای تابع احتمال زیر است:

$$P(x)=\binom{n}{x}p^x q^{n-x}=\binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x}=\binom{n}{x}\left(\frac{1}{2}\right)^n=\frac{\binom{n}{x}}{2^n} \quad ; \quad x=0,1,2,\dots,n$$

حال با دیدن تابع چگالی احتمال در صورت سؤال می‌توان متوجه شد که X دارای توزیع دوجمله‌ای با پارامترهای $p=q=\frac{1}{2}\ ,\ n=4$ است، بنابراین با توجه به تابع احتمال بالا داریم:

$$3a+1=2^n \xrightarrow{n=4} 3a+1=2^4 \rightarrow 3a=16-1=15 \rightarrow a=5$$

۳. گزینه ۱ درست است.

با توجه به اینکه احتمال موفقیت (قبول شدن) ثابت است $(p=0.6)$، توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n=96\ ,\ p=0.6\ ,\ q=0.4$.

می‌دانیم که در توزیع دوجمله‌ای، واریانس برابر است با:

$$\sigma^2 = npq = 96\times 0.6\times 0.4 = \frac{16\times 6\times 6\times 4}{100} \quad \rightarrow \quad \sigma = \frac{4\times 6\times 2}{10} = 4.8$$

۴ . گزینه ۲ درست است.

جامعه، نامحدود و احتمال، ثابت است، بنابراین توزیع تعداد موفقیت (معیوب بودن) در $n=3$ بار تکرار آزمایش (پیروزی، شکست) دوجمله‌ای است با $n=3\ ,\ p=0.1\ ,\ q=0.9$.

$$P(X\ge 1) = 1 - P(X=0) = 1-\binom{n}{x}p^x q^{n-x} = 1-\binom{3}{0}(0.1)^0(0.9)^3 = 1-(0.9)^3 = 0.271 \rightarrow 27 \text{ درصد}$$

۵ . گزینه ۲ درست است.

در هر پرواز احتمال موفقیت (مسافری جا رزرو کند و حاضر نشود) ثابت است $(p=?)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n=400$ و $p=?$ که میانگین آن برابر است با:

$$\mu = np = 40 \ \rightarrow\ 400p = 40 \ \rightarrow\ p = 0.1 \quad (\text{احتمال رزرو کردن و حاضر نشدن})$$

دقت کنید که وقتی همه 400 نفر برای پرواز حاضر شوند یعنی تعداد مسافرانی که جا رزرو کرده و حاضر نشده‌اند برابر صفر است.

X: تعداد مسافرانی که جا رزرو کرده و حاضر نشده‌اند

$$P(X=0) = \binom{n}{x}p^x q^{n-x} = \binom{400}{0}(0.1)^0(0.9)^{400} = (0.9)^{400}$$

۶ . گزینه ۴ درست است.

یادآوری: اگر متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای n و p باشد:

$$X\sim Bin(n,p) \ \rightarrow P(X=x) = \binom{n}{x}p^x q^{n-x} \quad ; \quad x=0,1,\dots,n \quad , \quad \begin{cases} E(X)=np \\ Var(X)=npq \end{cases}$$

بنابراین در این سؤال داریم:

$$\begin{cases} Var(X)=npq=6 \xrightarrow{np=9} 9q=6 \ \rightarrow\ q=\frac{6}{9}=\frac{2}{3} \ \rightarrow\ p=1-q=1-\frac{2}{3}=\frac{1}{3} \\ E(X)=np=9 \xrightarrow{p=\frac{1}{3}} n\times\frac{1}{3}=9 \ \rightarrow\ n=27 \end{cases}$$

$$P(X\ge 1) = 1-P(X<1) = 1-P(X=0) = 1-\binom{27}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{27-0} = 1-\left(\frac{2}{3}\right)^{27}$$

۷ . گزینه ۳ درست است.

احتمال موفقیت (جوانه زدن بذر) ثابت است $\left(p=\frac{4}{5}\right)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $p=\frac{4}{5}\ ,\ q=\frac{1}{5}\ ,\ n=5$.

X: تعداد بذرهای جوانه‌زده در نمونه

$$P(X=2) = \binom{n}{x}p^x q^{n-x} = \binom{5}{2}\left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right)^3 = \frac{10\times 16\times 1}{5^5} = 0.0512$$

۸ . گزینه ۳ درست است.

احتمال موفقیت در یک آزمایش برنولی $\frac{3}{5}$ است، بنابراین توزیع تعداد موفقیت در $n = 24$ آزمایش برنولی، دوجمله‌ای است با $n = 24 , q = \frac{2}{5} , p = \frac{3}{5}$.

می‌دانیم که در توزیع دوجمله‌ای، واریانس برابر است با:

$$Var(X) = npq = 24 \times \frac{3}{5} \times \frac{2}{5} = \frac{144}{25} \quad \to \quad \sigma = \frac{12}{5} = 2.4$$

۹ . گزینه ۴ درست است.

X، تعداد موفقیت‌ها در n بار تکرار آزمایش برنولی دوجمله‌ای است و $n = 25 , p = 0.98$.

در توزیع دوجمله‌ای با پارامترهای n و p داریم:

$$Var(X) = npq \quad \to \quad \sigma_X^2 = npq = 25 \times 0.98 \times 0.02 = 0.49 \quad \to \quad \sigma_X = \sqrt{0.49} = 0.7$$

۱۰ . گزینه ۱ درست است.

احتمال موفقیت (معیوب بودن) ثابت است $(p = 0.2)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n = 3 , p = 0.2 , q = 0.8$، بنابراین:

$$P(X = 3) = \binom{n}{x} p^x q^{n-x} = \binom{3}{3} (0.2)^3 (0.8)^{3-3} = (0.2)^3 = 0.008$$

۱۱ . گزینه ۳ درست است.

یادآوری: اگر متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای n و p باشد، آن‌گاه:

$$E(X) = np \quad , \quad Var(X) = npq$$

بنابراین داریم:

$$E(X) = np = \frac{8}{3} \xrightarrow{n=16} 16p = \frac{8}{3} \quad \to \quad p = \frac{1}{6} \quad , \quad q = \frac{5}{6}$$

$$Var(X) = npq = \frac{8}{3} \times \frac{5}{6} = \frac{20}{9}$$

۱۲. گزینه ۲ درست است.

احتمال خط آمدن در پرتاب یک سکه سالم (موفقیت) ثابت و برابر $p = \frac{1}{2}$ است، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n = 10 , p = q = \frac{1}{2}$ و داریم:

$$E(X) = np = 10 \times \frac{1}{2} = 5$$

۱۳ . گزینه ۴ درست است.

ضریب تغییرات : $CV = \frac{\sigma}{\mu} = \frac{\sqrt{6}}{10}$

می‌دانیم که امید و واریانس متغیر تصادفی X با توزیع $X \sim Bin(n = 25 , p = 0.4)$ برابر است با:

$$\mu = np = 25 \times 0.4 = 10$$

$$\sigma^2 = npq = 25 \times 0.4 \times 0.6 = 6$$

۱۴. گزینه ۳ درست است.

احتمال موفقیت (معیوب بودن) ثابت است $(p=0.2)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n=20\ ,\ p=0.2\ ,\ q=0.8$.

$$P(X\le 1)=P(X=0)+P(X=1)=\sum_{x=0}^{1}\binom{n}{x}p^x q^{n-x}=\binom{20}{0}(0.2)^0(0.8)^{20}+\binom{20}{1}(0.2)^1(0.8)^{19}$$

$$=(0.8)^{20}+4(0.8)^{19}=(0.8)^{19}(0.8+4)=\left(\frac{24}{5}\right)\times(0.8)^{19}$$

۱۵. گزینه ۳ درست است.

احتمال موفقیت (درست بودن جواب) ثابت است $\left(p=\frac{1}{4}\right)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n=20\ ,\ p=0.25\ ,\ q=0.75$.

X: تعداد جواب درست در 20 سؤال

$$P(X\le 9)=\sum_{x=0}^{9}\binom{20}{x}(0.25)^x(0.75)^{20-x}$$

توجه: حداکثر 9 جواب درست با حداقل 11 جواب غلط برابر است، بنابراین می‌توان احتمال را به صورت زیر نیز نوشت:

X: تعداد جواب غلط در 20 سؤال

$p=P(\text{غلط})=0.75\quad ,\quad q=0.25$

$$P(X\ge 11)=\sum_{x=11}^{20}\binom{20}{x}(0.75)^x(0.25)^{20-x}$$

۱۶. گزینه‌های ۳ و ۴ درست هستند.

احتمال موفقیت (غلط بودن جواب) ثابت است $\left(p=\frac{3}{4}\right)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n=20\ ,\ p=0.75\ ,\ q=0.25$.

X: تعداد جواب غلط در 20 سؤال

$$P(X\ge 3)=\sum_{x=3}^{20}\binom{20}{x}(0.75)^x(0.25)^{20-x}$$

توجه: حداقل 3 جواب غلط با حداکثر 17 جواب درست برابر است، بنابراین می‌توان احتمال را به صورت زیر نیز نوشت:

X: تعداد جواب درست در 20 سؤال

$p=P(\text{درست})=0.25\quad ,\quad q=0.75$

$$P(X\le 17)=\sum_{x=0}^{17}\binom{20}{x}(0.25)^x(0.75)^{20-x}$$

البته در کلید، گزینه ۴ درست بوده است درصورتی‌که گزینه ۳ هم درست است. با جایگذاری مقادیر مختلف x می‌توانید دو گزینه را بررسی کنید؛ حتماً به مقادیر احتمال یکسان خواهید رسید.

۱۷. گزینه ۴ درست است.

یادآوری:

$$X \sim Bin(n,p) \rightarrow \begin{cases} E(X) = np \\ Var(X) = npq \end{cases}$$

$$Var(X) = \frac{2}{3}E(X) \rightarrow npq = \frac{2}{3}np \rightarrow q = \frac{2}{3} \rightarrow p = \frac{1}{3}$$

$$E(X) = 20 \rightarrow np = 20 \xrightarrow{p=\frac{1}{3}} n \times \frac{1}{3} = 20 \rightarrow n = 60$$

۱۸. گزینه ۳ درست است.

احتمال موفقیت (ابتلا به بیماری) ثابت است (p)، بنابراین توزیع تعداد موفقیت در نمونه (n نفر) دوجمله‌ای است و امید ریاضی آن برابر است با:

X: تعداد افراد مبتلا به بیماری در n نفر

$$E(X) = np$$

توزیع چندجمله‌ای

۱۹. گزینه ۱ درست است.

نکته: هرگاه آزمایش دارای بیش از دو نتیجه ممکن باشد، به طوری که نتایج دوبه‌دو ناسازگار باشند و مجموع نتایج ممکن برابر یک شود، توزیع تعداد موفقیت‌ها در هر بار تکرار مستقل آزمایش، چندجمله‌ای است و تابع چگالی احتمال آن به صورت زیر است:

$$\binom{n}{x_1 \; x_2 \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \quad , \quad \sum_{i=1}^{k} p_i = 1$$

$$\begin{cases} \binom{7}{2 \;\; 2 \;\; 3} P(A)^2 P(B)^2 P(C)^3 = \frac{7!}{2!2!3!}(0.5)^2 (0.3)^2 (0.2)^3 = 210 \times 0.25 \times 0.09 \times 0.008 = 0.0378 \\ P(A) = 0.5 \, , \, P(B) = 0.3 \, , \, P(C) = 0.2 \end{cases}$$

۲۰. گزینه ۱ درست است.

با توجه به اینکه احتمال‌ها به دو حالت موفقیت و شکست محدود نیست، از توزیع چندجمله‌ای استفاده می‌کنیم:

$$p_1 = 0.5 \, , \, p_2 = 0.4 \, , \, p_3 = 0.1 \, , \sum p_i = 1$$

$$P(X_1 = 3 \, , X_2 = 2 \, , X_3 = 1) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} = \frac{6!}{3!2!1!}(0.5)^3 (0.4)^2 (0.1)^1 = 0.12$$

۲۱. گزینه ۲ درست است.

با توجه به اینکه بیش از دو حالت (پیروزی، شکست) برای مسئله وجود دارد، از توزیع چندجمله‌ای استفاده می‌کنیم:

$p_1 = P(\text{قرمز}) = 0.5 \quad , \quad p_2 = P(\text{سبز}) = 0.3 \quad , \quad p_3 = P(\text{زرد}) = 0.2 \quad , \quad n = 6$

$$P(X_1=2, X_2=1, X_3=3)=\binom{n}{x_1\ \ x_2\ \ x_3}p_1^{x_1}p_2^{x_2}p_3^{x_3}=\binom{6}{2\ \ 1\ \ 3}(0.5)^2(0.3)^1(0.2)^3$$

$$=\frac{6!}{2!\,1!\,3!}(0.5)^2(0.3)^1(0.2)^3=60\times0.25\times0.3\times0.008=0.036$$

توزیع دوجمله‌ای منفی

۲۲. گزینه ۴ درست است.

راه حل اول: احتمال موفقیت (متوجه شدن معیوب) ثابت است $\left(p=\frac{3}{4}\right)$، بنابراین توزیع تعداد تکرار برای رسیدن به rامین موفقیت، دوجمله‌ای منفی است با $p=\frac{3}{4}, q=\frac{1}{4}, r=5, x=7$.

$$P(X=7)=\binom{x-1}{r-1}p^r q^{x-r}=\binom{7-1}{5-1}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)^2=\frac{15\times3^5}{4^7}=5\times3^6\times4^{-7}$$

توجه: در توزیع دوجمله‌ای منفی به دنبال جای موفقیت هستیم، به همین دلیل از پسوند (اُمین) استفاده می‌کنیم.

راه حل دوم: این مسئله را بدون در نظر گرفتن جای موفقیت با توزیع دوجمله‌ای نیز می‌توان حل کرد و در آخر از اعداد قسمت ترکیب مسئله یک واحد کم کرد.

X، تعداد موفقیت‌ها، دارای توزیع دوجمله‌ای است (5 موفقیت در 7 بار تکرار آزمایش)، بنابراین:

$$\begin{cases} p=\frac{3}{4},\ q=\frac{1}{4},\ n=7 \\ P(X=5)=\binom{n}{x}p^x q^{n-x}=\binom{\cancel{7}^{\,6}}{\cancel{5}_{\,4}}p^5 q^{7-5}=\binom{6}{4}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)^2=\frac{15\times3^5}{4^7}=5\times3^6\times4^{-7} \end{cases}$$

توزیع هندسی

۲۳. گزینه ۴ درست است.

جامعه، نامحدود و احتمال، ثابت است، بنابراین توزیع تعداد تکرار آزمایش (پیروزی، شکست) برای رسیدن به اولین موفقیت، هندسی است.

X، تعداد دفعات پرتاب دو تاس برای رسیدن به اولین مجموع 7، دارای توزیع هندسی است.

$$\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\}$$

$$p=P(\text{مجموع دو تاس } 7)=\frac{6}{36}=\frac{1}{6},\quad q=\frac{5}{6}$$

تعداد دفعات پرتاب شدن تاس‌ها فرد باشد، یعنی:

$$P(\text{فرد } X)=P(X=1)+P(X=3)+P(X=5)+\ldots=\sum_{x:\ \text{فرد}}pq^{x-1}$$

$$\frac{1}{6}+\frac{1}{6}\times\left(\frac{5}{6}\right)^2+\frac{1}{6}\times\left(\frac{5}{6}\right)^4+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{\frac{1}{6}}{1-\left(\frac{5}{6}\right)^2}=\frac{\frac{1}{6}}{\frac{36-25}{36}}=\frac{6}{11}$$

یادآوری: تصاعد هندسی

$$1+q+q^2+q^3+\ldots=\frac{\text{جمله اول}}{1-\text{قدرنسبت}}=\frac{1}{1-q}$$

۲۴. گزینه ۴ درست است.

یادآوری: توزیع تعداد تکرار آزمایش برنولی (پیروزی، شکست) برای رسیدن به اولین موفقیت، هندسی است.

$$P(x)=pq^{x-1} \quad ; \quad x=1,2,\ldots,\infty$$

در این سؤال موفقیت، آمدن 6 در پرتاب تاس است که احتمال آن $p=\frac{1}{6}$ است.

تعداد پرتاب تاس تا رسیدن به 6 : $X\sim Ge\left(p=\frac{1}{6}\right)$

$$P(X\geq 2)=1-P(X<2)=1-P(X=1)=1-pq^{1-1}=1-\frac{1}{6}\left(\frac{5}{6}\right)^0=1-\frac{1}{6}=\frac{5}{6}$$

توزیع فوق هندسی

۲۵. گزینه ۲ درست است.

جامعه، محدود و پیش‌فرض، انتخاب بدون جایگذاری است. بنابراین توزیع تعداد موفقیت (سالم بودن) در n بار تکرار آزمایش (پیروزی، شکست)، فوق هندسی است.

$$\begin{cases} N=12 \quad , \quad n=2 \quad , \quad k=8 \quad (\text{سالم}) \\ E(X)=n\,.\frac{k}{N}=2\times\frac{8}{12}=\frac{4}{3} \end{cases}$$

توجه کنید که اگر به اشتباه توزیع X را دوجمله‌ای می‌دانستیم باز هم جواب همین بود، چون $n=2\,,\,p=\frac{k}{N}=\frac{8}{12}$ و $E(X)=np$ است. اما باید توزیع را درست تشخیص دهیم زیرا همیشه امید را نمی‌خواهند و ممکن است برای به دست آوردن سایر مقادیر دچار اشتباه شویم. وقتی جامعه محدود باشد، فقط درصورتی‌که انتخاب با جایگذاری در صورت سؤال ذکر شود، توزیع دوجمله‌ای است.

۲۶. گزینه ۲ درست است.

جامعه، محدود و پیش‌فرض، انتخاب بدون جایگذاری است، بنابراین توزیع تعداد موفقیت (غیراستاندارد بودن) در n بار تکرار آزمایش (پیروزی، شکست)، فوق هندسی است.

توزیع X فوق هندسی است با:

$$\begin{cases} N=100\,,n=5\,,k=20 \\ \sigma_X^2=n\left(\frac{k}{N}\right)\left(1-\frac{k}{N}\right)\frac{N-n}{N-1}=5\times\frac{20}{100}\times\frac{80}{100}\times\frac{100-5}{100-1}=5\times0.2\times0.8\times\frac{95}{99}=0.7676 \end{cases}$$

اگر به اشتباه توزیع X را دوجمله‌ای با $p=\frac{k}{N}$, $n=5$ تشخیص دهیم، تنها تفاوتشان در ضریب تصحیح $\frac{N-n}{N-1}$ است. توجه کنید که اگر انتخاب با جایگذاری در صورت سؤال ذکر می‌شد، توزیع، دوجمله‌ای و گزینه ۳ درست بود.

۲۷. گزینه ۳ درست است.

انتخاب به طور پیش‌فرض بدون جایگذاری در نظر گرفته می‌شود، بنابراین وقتی جامعه محدود است $(N=6)$ احتمال موفقیت (معیوب بودن) ثابت نیست و توزیع تعداد موفقیت در نمونه، فوق هندسی است با $k=2$, $n=2$, $N=6$.

$$P(X=2)=\frac{\binom{n}{x}\binom{N-n}{n-x}}{\binom{N}{n}}=\frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}}=\frac{1}{15}$$

۲۸. گزینه ۴ درست است.

جامعه، محدود و پیش‌فرض، انتخاب بدون جایگذاری است، بنابراین احتمال موفقیت (سالم بودن لامپ) ثابت نیست و توزیع تعداد موفقیت در نمونه فوق هندسی است با:

$N=20$, $n=2$, $k=0.95\times 20=19$ (تعداد لامپ‌های سالم)

$$P(X\geq 1)=1-P(X<1)=1-P(X=0)=1-\frac{\binom{19}{0}\binom{1}{\cancel{2}}^{0}}{\binom{20}{2}}=1-0=1$$

ازآنجاکه تعداد لامپ‌های معیوب، یکی است $(0.05\times 20=1)$ غیر ممکن است که در انتخاب 2 لامپ، هر دو معیوب باشند؛ به عبارت دیگر از بین 19 لامپ سالم و 1 لامپ معیوب، احتمال اینکه در انتخاب 2 لامپ حداقل یکی سالم باشد، حتماً 1 است.

۲۹. گزینه ۳ درست است.

جامعه، محدود و پیش‌فرض، انتخاب بدون جایگذاری است، بنابراین احتمال موفقیت در جامعه ثابت نیست و توزیع تعداد موفقیت در نمونه، فوق هندسی است با :

$N=3+6=9$, $n=2$, $k=6$ (سالم)

$$P(X=2)=\frac{\binom{k}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\frac{\binom{6}{2}\binom{3}{0}}{\binom{9}{2}}=\frac{\frac{6!}{2!4!}}{\frac{9!}{7!2!}}=\frac{15}{36}=\frac{5}{12}$$

توزیع پواسون

۳۰. گزینه ۴ درست است.

یادآوری: توزیع « X : تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda$, $Var(X)=\lambda$

متوسط تعداد مشتری (λ)	زمان
3	1 دقیقه = 60 ثانیه
?	80 ثانیه

$$\Rightarrow \lambda=\frac{3\times 80}{60}=4$$

یعنی به طور متوسط $\lambda = 4$ مشتری در 80 ثانیه به فروشگاه مراجعه می‌کند؛ در نتیجه:

$$\begin{cases} P(\text{حداقل 2 مشتری}) = P(X \ge 2) = 1 - P(X<2) = 1 - P(X=0) - P(X=1) = 1 - \sum_{x=0}^{1} \frac{e^{-\lambda}\lambda^{x}}{x!} \\ = 1 - e^{-4} - 4e^{-4} = 1 - 5e^{-4} \xlongequal{e^{-4}=0.018} 0.91 \\ X: \text{تعداد مشتری در 80 ثانیه} \quad , \quad \lambda = 4 \end{cases}$$

۳۱. گزینه ۲ درست است.

یادآوری: توزیع « X: تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

زمان	متوسط تعداد افراد (λ)
1 دقیقه = 60 ثانیه	10
30 ثانیه	?

$$\Rightarrow \quad \lambda = \frac{10 \times 30}{60} = 5$$

یعنی به طور متوسط $\lambda = 5$ نفر در 30 ثانیه با مرکز تماس می‌گیرند؛ درنتیجه:

$$\begin{cases} P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \xrightarrow[\lambda=5]{x=4} P(X=4) = \dfrac{e^{-5}5^{4}}{4!} = \dfrac{0.007 \times 625}{24} = 0.182 \\ X: \text{تعداد افرادی که در 30 ثانیه با مرکز تماس می‌گیرند} \quad , \quad \lambda = 5 \end{cases}$$

۳۲. گزینه ۳ درست است.

یادآوری: توزیع « X: تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

زمان	متوسط تعداد تصادفات (λ)
1 روز = 24 ساعت	12
6 ساعت	?

$$\Rightarrow \quad \lambda = \frac{12 \times 6}{24} = 3$$

یعنی به طور متوسط $\lambda = 3$ تصادف در 6 ساعت اتفاق می‌افتد؛ درنتیجه:

$$\begin{cases} P(X \le 1) = P(X=0) + P(X=1) = \sum_{x=0}^{x=1} \dfrac{e^{-\lambda}\lambda^{x}}{x!} = \dfrac{e^{-3}(3)^{0}}{0!} + \dfrac{e^{-3}(3)^{1}}{1!} = e^{-3} + 3e^{-3} = 4e^{-3} \\ X: \text{تعداد تصادفات در 6 ساعت} \quad , \quad \lambda = 3 \end{cases}$$

۳۳. گزینه ۴ درست است.

یادآوری: توزیع « X: تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

زمان	متوسط تعداد افراد (λ)
2 دقیقه	1
5 دقیقه	?

$$\Rightarrow \quad \lambda = \frac{1 \times 5}{2} = \frac{5}{2}$$

یعنی به طور متوسط $\lambda=\frac{5}{2}$ نفر در 5 دقیقه وارد کتابخانه می‌شود؛ درنتیجه:

$$\begin{cases} P(X\ge 1)=1-P(X=0)=1-\dfrac{e^{-\lambda}\lambda^x}{x!}=1-\dfrac{e^{-\frac{5}{2}}\left(\frac{5}{2}\right)^0}{0!}=1-e^{-\frac{5}{2}} \\ X:\text{تعداد افراد وارد شده در 5 دقیقه} \quad , \quad \lambda=\frac{5}{2} \end{cases}$$

۳۴ . گزینه ۱ درست است.

یادآوری: توزیع « X: تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

زمان	متوسط تعداد مشتری (λ)
30 دقیقه	1
5 دقیقه	?

$$\Rightarrow \lambda=\frac{1\times 5}{30}=\frac{1}{6}$$

یعنی به طور متوسط $\lambda=\frac{1}{6}$ مشتری در 5 دقیقه وارد فروشگاه می‌شود؛ درنتیجه:

$$\begin{cases} P(X\ge 1)=1-P(X<1)=1-P(X=0)=1-\dfrac{e^{-\frac{1}{6}}\left(\frac{1}{6}\right)^0}{0!}=1-e^{-\frac{1}{6}} \\ X:\text{تعداد مشتری در 5 دقیقه} \quad , \quad \lambda=\frac{1}{6} \end{cases}$$

۳۵ . گزینه ۱ درست است.

یادآوری: توزیع « X: تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

زمان	متوسط تعداد اتومبیل (λ)
1 دقیقه = 60 ثانیه	8
30 ثانیه	?

$$\Rightarrow \lambda=\frac{8\times 30}{60}=4$$

یعنی به طور متوسط $\lambda=4$ اتومبیل در 30 ثانیه عبور می‌کند؛ درنتیجه:

$$\begin{cases} P(X\ge 3)=1-P(X<3)=1-P(X=0)-P(X=1)-P(X=2)=1-\displaystyle\sum_{x=0}^{2}\frac{e^{-\lambda}\lambda^x}{x!} \\ =1-\dfrac{e^{-4}4^0}{0!}-\dfrac{e^{-4}4^1}{1!}-\dfrac{e^{-4}4^2}{2!}=1-e^{-4}(1+4+8)=1-13\times 0.018=0.766 \\ X:\text{تعداد اتومبیل در 30 ثانیه} \quad , \quad \lambda=4 \end{cases}$$

۳۶ . گزینه ۴ درست است.

یادآوری: توزیع « X: تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

$$\begin{cases} P(X=x)=\dfrac{e^{-\lambda}\lambda^x}{x!}\xrightarrow[\lambda=1]{x=3}P(X=3)=\dfrac{e^{-1}1^3}{3!}=\dfrac{e^{-1}}{3!} \\ X:\text{تعداد تخم مرغ‌های فاسد در یک شانه} \quad , \quad \lambda=1 \end{cases}$$

۳۷ . گزینه ۳ درست است.

یادآوری: توزیع «X : تعداد اتفاقات در واحد زمان یا مکان»، پواسون با پارامتر λ است. در توزیع پواسون، امید (متوسط) و واریانس توزیع با پارامتر توزیع برابر است، یعنی: $E(X)=\lambda \quad , \quad Var(X)=\lambda$

زمان	متوسط تعداد سؤال پاسخ داده‌شده (λ)
60 ثانیه	2
30 ثانیه	?

$$\Rightarrow \lambda=\frac{2\times 30}{60}=1$$

یعنی به طور متوسط $\lambda=1$ سؤال در 30 ثانیه پاسخ داده می‌شود؛ درنتیجه:

$$\begin{cases} P(X=x)=\dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=1]{x=2} P(X=2)=\dfrac{e^{-1}1^2}{2!}=\dfrac{e^{-1}}{2}=\dfrac{1}{2e} \\ X: \text{تعداد سؤال پاسخ داده‌شده در 30 ثانیه} \quad , \quad \lambda=1 \end{cases}$$

تقریب توزیع دوجمله‌ای به پواسون

۳۸ . گزینه ۱ درست است.

می‌دانیم که در تقریب توزیع دوجمله‌ای به پواسون، پارامتر توزیع پواسون $\lambda=np$ است.

$$\begin{cases} \lambda=np=4000\times 0.0015=6 \\ n=4000,\ p=0.0015 \end{cases}$$

دقت کنید که چون $n>100$ و $np<10$ است تقریب پواسون مناسب است.

۳۹ . گزینه ۱ درست است.

احتمال موفقیت (معیوب بودن) ثابت است $\left(p=\dfrac{125}{100000}\right)$، بنابراین توزیع مقدار موفقیت در نمونه دوجمله‌ای است با $p=0.00125\ ,\ n=1600$.

یادآوری: هرگاه در توزیع دوجمله‌ای $n\geq 100$ و $np<10$ باشد تقریب پواسون با پارامتر $\lambda=np$ مناسب است.

حال در این سؤال:

$$np=1600\times\frac{125}{100000}=2<10 \text{ و } n=1600\geq 100$$

بنابراین از تقریب پواسون با پارامتر $\lambda=np=2$ استفاده می‌کنیم:

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!} \rightarrow P(X=4)=\frac{e^{-2}2^4}{4!}=\frac{16}{24}e^{-2} \overset{e^{-2}=0.135}{=\!=} 0.09$$

توزیع یکنواخت پیوسته

۴۰. گزینه ۱ درست است.

راه حل اول:

هرگاه تابع چگالی پیوسته، یک عدد باشد، توزیع X یکنواخت پیوسته است، بنابراین، در این سؤال X دارای توزیع یکنواخت پیوسته است با $a=1\ ,\ b=5$.

$$f(x)=\frac{1}{b-a}=\frac{1}{5-1}=\frac{1}{4} \quad ; \quad 1<x<5$$

$$E(X)=\frac{a+b}{2}=\frac{1+5}{2}=3 \quad , \quad \sigma_X^2=\frac{(b-a)^2}{12}=\frac{(5-1)^2}{12}=\frac{16}{12}=\frac{4}{3}\to\sigma=\frac{2}{\sqrt{3}}$$

$$\text{ضریب تغییرات}: CV=\frac{\sigma}{\mu}\times 100=\frac{\frac{2}{\sqrt{3}}}{3}\times 100=\frac{2}{3\sqrt{3}}\times 100$$

اگرچه مقدار واریانس در صورت مسئله داده شده است اما به دست آوردن آن هم ساده بود.

راه حل دوم:

امید ریاضی را مستقیماً از رابطه زیر محاسبه می‌کنیم:

$$E(X)=\int_1^5 xf(x)dx=\int_1^5 x.\frac{1}{4}dx=\left[\frac{1}{8}x^2\right]_1^5=\frac{1}{8}(25-1)=\frac{24}{8}=3$$

۴۱. گزینه ۱ درست است.

یادآوری: هرگاه X دارای توزیع یکنواخت در فاصله (α,β) باشد:

$$f(x)=\frac{1}{\beta-\alpha} \qquad ; \qquad \alpha<x<\beta$$

$$E(X^2)=\int_\alpha^\beta x^2.f(x)dx=\int_\alpha^\beta x^2.\frac{1}{\beta-\alpha}dx=\frac{1}{\beta-\alpha}\left[\frac{1}{3}x^3\right]_\alpha^\beta=\frac{\beta^3-\alpha^3}{3(\beta-\alpha)}$$

البته می‌توانید با استفاده از مقادیر $E(X)=\frac{\alpha+\beta}{2}$, $Var(X)=\frac{(\beta-\alpha)^2}{12}$ نیز $E(X^2)$ را محاسبه کنید اما درگیر روابط ریاضی می‌شوید و دیرتر به جواب خواهید رسید.

$$E(X^2)=Var(X)+(E(X))^2$$

۴۲. گزینه ۲ درست است.

راه حل اول:

هرگاه تابع چگالی پیوسته یک عدد باشد، توزیع X یکنواخت پیوسته است. بنابراین، در این سؤال X دارای توزیع یکنواخت پیوسته با $a=1$, $b=4$.

$$f(x)=\frac{1}{b-a}=\frac{1}{4-1}=\frac{1}{3} \quad ; \quad 1<x<4$$

$$\sigma_X^2=\frac{(b-a)^2}{12}=\frac{(4-1)^2}{12}=\frac{9}{12}=\frac{3}{4}=0.75$$

راه حل دوم:

واریانس را مستقیماً از رابطه زیر محاسبه می‌کنیم:

$$\sigma_X^2=E(X^2)-(E(X))^2=\int_1^4 x^2.\frac{1}{3}dx-\left(\int_1^4 x.\frac{1}{3}dx\right)^2=\left[\frac{1}{9}x^3\right]_1^4-\left(\frac{1}{6}\left[x^2\right]_1^4\right)^2$$

$$=\left(\frac{1}{9}(64-1)\right)-\left(\frac{1}{6}(16-1)\right)^2=\frac{3}{4}$$

۴۳. گزینه ۱ درست است.

یادآوری: اگر X دارای توزیع یکنواخت در فاصله (a,b) باشد داریم:

$$Var(X)=\frac{(b-a)^2}{12}$$

در این سؤال X دارای توزیع یکنواخت در فاصله $(-2,5)$ است، بنابراین:

$$\mathrm{Var}(X)=\frac{(b-a)^2}{12}=\frac{(5-(-2))^2}{12}=\frac{49}{12}=4.08$$

توزیع نمایی

۴۴. گزینه ۳ درست است.

راه حل اول:

ازآنجاکه این تابع، تابع چگالی توزیع نمایی با $\lambda=\frac{1}{2}$ است:

$$\begin{cases} f(x)=\lambda e^{-\lambda x} \\ E(X)=\dfrac{1}{\lambda}=\dfrac{1}{\frac{1}{2}}=2 \end{cases}$$

راه حل دوم:

امید ریاضی را مستقیماً از رابطه زیر و به روش انتگرال‌گیری جزء به جزء محاسبه می‌کنیم:

$$E(X)=\int_0^{\infty} x f(x)dx=\int_0^{\infty} x.\frac{1}{2}e^{-\frac{x}{2}}dx=\left[x\left(-e^{-\frac{x}{2}}\right)-\left(2e^{-\frac{x}{2}}\right)\right]_0^{\infty}=2$$

یادآوری: انتگرال‌گیری به روش جزء به جزء: $\int u\cdot dv=u\cdot v-\int vdu$

در این سؤال:

$$u=x\ ,\ du=\frac{1}{2}e^{-\frac{x}{2}}\ \rightarrow\ du=dx\ ,\ v=-e^{-\frac{x}{2}}$$

توجه: می‌دانیم که راه دیگر محاسبه امید ریاضی برای $x>0$ برابر است با:

$$E(X)=\int_0^{\infty}\left[1-F(x)\right]dx$$

بنابراین در این سؤال داریم:

$$\begin{cases} E(X)=\displaystyle\int_0^{\infty}\left[1-\left(1-e^{-\frac{1}{2}x}\right)\right]dx=\int_0^{\infty} e^{-\frac{1}{2}x}dx=\left[-2e^{-\frac{1}{2}x}\right]_0^{\infty}=2 \\ F(x)=\displaystyle\int_0^{x} f(x)dx=\int_0^{x}\frac{1}{2}e^{-\frac{1}{2}x}dx=\left[-e^{-\frac{1}{2}x}\right]_0^{x}=1-e^{-\frac{1}{2}x} \end{cases}$$

این روش زمانی مورد استفاده قرار می‌گیرد که:

۱ـ تابع توزیع تجمعی $(F(x))$ داده شود و امید ریاضی را خواسته شود.

۲ـ نخواهیم از انتگرال‌گیری به روش جزء به جزء استفاده کنیم.

۴۵. گزینه ۲ درست است.

راه حل اول: تابع چگالی احتمال توزیع نمایی به صورت $f(x)=\lambda e^{-\lambda x}\ ;\ x>0$ است که به ازای $\lambda=1$ برابر با $f(x)=e^{-x}\ ;\ x>0$ می‌شود؛ در توزیع نمایی با پارامتر λ، داریم:

$$E(X)=\frac{1}{\lambda}\xrightarrow{\lambda=1}E(X)=1$$

$$Var(X)=\frac{1}{\lambda^2}\xrightarrow{\lambda=1}Var(X)=1$$

راه حل دوم: درصورتی‌که تشخیص ندهیم تابع چگالی داده‌شده، نمایی با پارامتر $\lambda=1$ است، باید واریانس را مستقیماً از رابطه زیر و به روش انتگرال‌گیری جزء به جزء محاسبه کنیم:

$$Var(X)=E\left(X^2\right)-\left(E(X)\right)^2=2-(1)^2=1$$

$$E\left(X^2\right)=\int x^2 f(x)dx=\int_0^{\infty}x^2e^{-x}dx=\left[x^2\left(-e^{-x}\right)-2x\left(e^{-x}\right)+2\left(-e^{-x}\right)\right]_0^{\infty}=2$$

$$E(X)=\int xf(x)dx=\int_0^{\infty}xe^{-x}dx=\left[x\left(-e^{-x}\right)-1\times\left(e^{-x}\right)\right]_0^{\infty}=1$$

توزیع نرمال

۴۶. گزینه ۳ درست است.

نمودار بافت‌نگار فراوانی مطلق، متقارن است، درنتیجه توزیع جامعه متقارن است و می‌توان آن را نرمال در نظر گرفت. همچنین می‌دانیم که در توزیع نرمال 0.95 داده‌ها در فاصله $\mu\pm2\sigma$ قرار دارند، بنابراین:

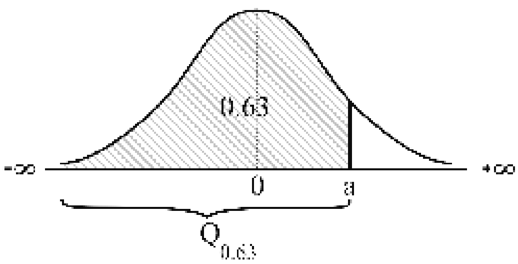
$$P(\mu\pm2\sigma)=0.95$$

$$\begin{cases}\mu\pm2\sigma=7\pm2\times1.5=7\pm3=(4,10)\\ \mu=\dfrac{\sum x_i}{N}=\dfrac{840}{120}=7\\ \sigma^2=\dfrac{\sum x_i^2}{N}-\left(\dfrac{\sum x_i}{N}\right)^2=\dfrac{6150}{120}-7^2=51.25-49=2.25\rightarrow\sigma=1.5\end{cases}$$

۴۷. گزینه ۴ درست است.

$$X\sim N\left(\mu=15.21\ ,\ \sigma^2=9\right)$$

$$Q_{0.63}=a\ \rightarrow\ P(X\le a)=0.63\ \rightarrow$$



$$(1)\ P(X\le a)=0.63\ \rightarrow\ P\left(\frac{X-\mu}{\sigma}<\frac{a-15.21}{3}\right)=0.63\rightarrow P\left(Z<\frac{a-15.21}{3}\right)=0.63$$

$$(2)\ P(Z<-0.33)=0.37\ \rightarrow\ P(Z>-0.33)=1-0.37=0.63$$

حال با توجه به آنکه $P(Z<A)=P(Z>B) \to A=-B$ است، داریم:

$$(1,2)\ \frac{a-15.21}{3}=-(-0.33) \quad \to \quad a=16.2$$

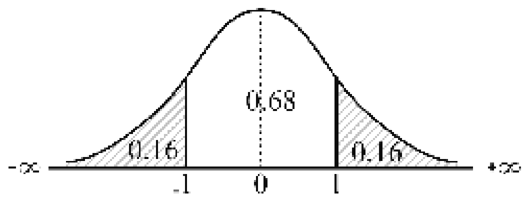
۴۸ . گزینه ۱ درست است.

امتیاز شرکت‌کنندگان : $X \sim N\left(\mu=75 \quad , \quad \sigma^2=10^2\right)$

$$P(X>85)=P\left(\frac{X-\mu}{\sigma}>\frac{85-75}{10}\right)=P\left(Z>\frac{10}{10}\right)=P(Z>1)=0.16=\%16$$

یادآوری:

$$P(Z<-1)=P(Z>1)=0.16$$

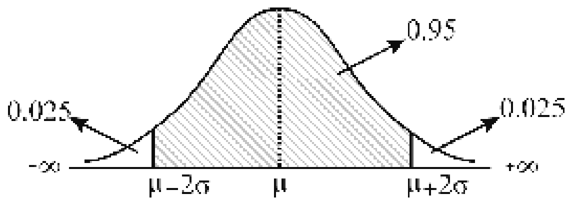


۴۹ . گزینه ۴ درست است.

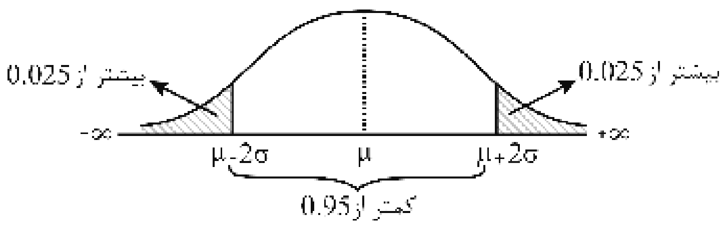
یادآوری:

۱- شاخص کشیدگی یا اوج در توزیع نرمال برابر $\alpha_4=3$ است.

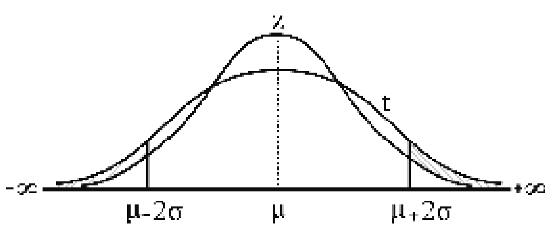
۲- در توزیع نرمال فاصله دو انحراف معیار از میانگین به صورت زیر است:



در این سؤال که شاخص کشیدگی توزیع جامعه $\alpha_4=2$ است، منحنی توزیع جامعه از توزیع نرمال کوتاه‌تر است (مانند توزیع t).



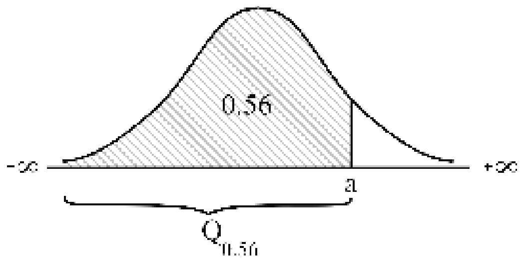
با توجه شکل مشخص می‌شود که در دو انتهای منحنی $>\mu+2\sigma$ و $<\mu-2\sigma$ مساحت (فضا) نسبت به توزیع نرمال بیشتر بوده و درنتیجه مساحت (فضای) بین دو انحراف، در این توزیع از توزیع نرمال کمتر است.



۵۰. گزینه ۴ درست است.

$X \sim N\left(\mu = 17.2\,,\ \sigma^2 = 16\right)$

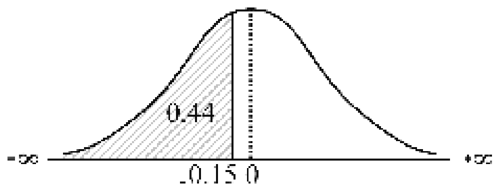
$P\left(X \le a\right) = 0.56 \ \rightarrow$



پنجاه و ششمین صدک: $Q_{0.56}$

$(1) P\left(X \le a\right) = 0.56 \ \rightarrow P\left(\frac{X-\mu}{\sigma} \le \frac{a-17.2}{4}\right) = 0.56$

$(2) P\left(Z < -0.15\right) = 0.44 \ \rightarrow \ P\left(Z > -0.15\right) = 1 - 0.44 = 0.56$



با توجه به آنکه $A = -B \ \rightarrow \ P\left(Z < A\right) = P\left(Z > B\right)$ است، داریم:

$(1),(2) \frac{a-17.2}{4} = -\left(-0.15\right) \rightarrow a = 17.8$

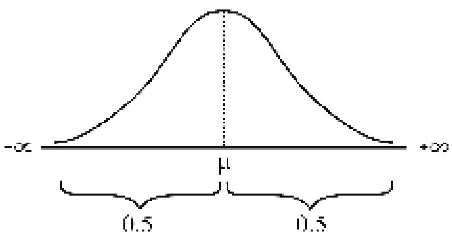
۵۱. گزینه ۲ درست است.

$X \sim N\left(\mu = 47, \sigma^2 = 64\right)$

متغیر جدید : $Y = X + 5$

$P\left(Y > 52\right) = P\left(X + 5 > 52\right) = P\left(X > 47\right) = P\left(X > \mu_X\right) = 0.5 = 50$ درصد

یادآوری:

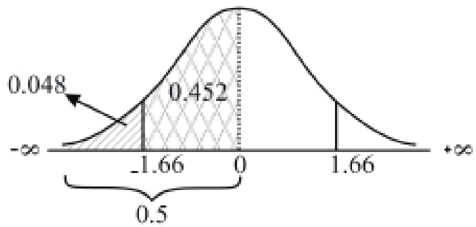


۵۲. گزینه ۱ درست است.

$X \sim N\left(\mu = 14.5\,,\ \sigma^2 = 1.5^2\right)$

$P\left(14.5 < X < 17\right) = P\left(\frac{14.5-14.5}{1.5} < \frac{X-\mu}{\sigma} < \frac{17-14.5}{1.5}\right) = P\left(0 < Z < 1.66\right) = 0.452$

$$P(Z<-1.66)=0.048$$
$$P(-1.66<Z<0)=P(0<Z<1.66)=0.5-0.048=0.452$$



۵۳. گزینه ۴ درست است.

$$X\sim N\left(\mu=?\ ,\ \sigma^2=5^2\right)$$
$$1)\ P(X\ge 9.8)=P\left(\frac{X-\mu}{\sigma}\ge\frac{9.8-\mu}{5}\right)=P\left(Z\ge\frac{9.8-\mu}{5}\right)=0.67$$
$$2)\ P(Z<-0.44)=0.33\ \rightarrow\ P(Z>-0.44)=0.67$$

حال با توجه به آنکه $P(Z>A)=P(Z>B)\rightarrow\ A=B$ است، داریم:

$$(1)\ ,\ (2)=\frac{9.8-\mu}{5}=-0.44\ \rightarrow\ \mu=9.8+2.2=12$$

۵۴. گزینه ۲ درست است.

$$X\sim N\left(\mu=25,\sigma^2=?\right)$$
$$1)\ P(X\ge 16)=P\left(\frac{X-\mu}{\sigma}\ge\frac{16-25}{\sigma}\right)=P\left(Z\ge\frac{-9}{\sigma}\right)=0.9332$$
$$2)\ P(Z\le -1.5)=0.0668\rightarrow P(Z\ge -1.5)=1-0.0668=0.9332$$

حال با توجه به آنکه $P(Z>A)=P(Z>B)\rightarrow A=B$ است، داریم:

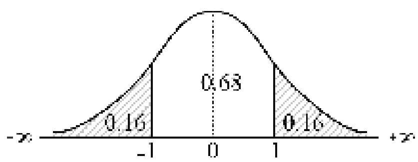
$$(1)\ ,\ (2)\ \ \frac{-9}{\sigma}=-1.5\ \rightarrow\ \sigma=6$$

۵۵. گزینه ۲ درست است.

$$X\sim N\left(\mu=10\ \ ,\ \ \sigma^2=16\right)$$
$$(1)\ P(X>a)=0.159\simeq 0.16\ \rightarrow\ P\left(\frac{X-\mu}{\sigma}>\frac{a-10}{\sqrt{16}}\right)=0.16\ \rightarrow\ P\left(Z>\frac{a-10}{4}\right)\simeq 0.16$$

یادآوری:

$$(2)\ P(Z>1)=P(Z<-1)=0.16$$



حال با توجه به نکته $P(Z>a)=P(Z>b)\ \rightarrow\ a=b$ داریم.

$$(1),(2)\rightarrow\ \frac{a-10}{4}=1\ \rightarrow\ a=10+4=14$$

۵۶ . گزینه ۱ درست است.

$$X \sim N\left(\mu = 2 \;,\; \sigma^2 = 1\right) \quad , \quad Y \sim N\left(\mu = 5 \;,\; \sigma^2 = 4\right)$$

$$W = \frac{2Y + X}{3}$$

یادآوری: هر ترکیب خطی از توزیع نرمال، باز هم نرمال است.

بنابراین توزیع W نرمال بوده و امید و واریانس آن برابر است با:

$$E(W) = E\left(\frac{2Y+X}{3}\right) = \frac{2E(Y)+E(X)}{3} = \frac{2\times 5+2}{3} = \frac{12}{3} = 4$$

$$Var(W) = Var\left(\frac{2Y+X}{3}\right) = \frac{2^2 Var(Y) + Var(X)}{3^2} = \frac{4\times 4+1}{9} = \frac{17}{9}$$

$$W \sim N\left(\mu = 4 \;,\; \sigma^2 = \frac{17}{9}\right)$$

توجه: چون از همبستگی X و Y در صورت سؤال صحبتی نشده است، آن‌ها را مستقل در نظر می‌گیریم. اگرچه در این سؤال نیز واضح است که نمره میان‌ترم از پایان‌ترم مستقل است، بنابراین در محاسبه واریانس W جزء کواریانس X و Y صفر شده و در محاسبه نوشته نشده است.

یادآوری:

$$\begin{cases} E(aX+bY) = aE(X) + bE(Y) & X \text{ و } Y \text{ مستقل یا وابسته} \\ Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + \underbrace{\overset{0}{\cancel{2abCov(X,Y)}}}_{\text{در شرایط استقلال}} & \end{cases}$$

تقریب توزیع دوجمله‌ای به نرمال

۵۷ . گزینه ۲ درست است.

احتمال موفقیت (خرید کردن) ثابت است $(p = 0.8)$، بنابراین توزیع تعداد موفقیت در نمونه دوجمله‌ای است با $n = 100 \;,\; p = 0.8 \;,\; q = 0.2$.

یادآوری: هرگاه در توزیع دوجمله‌ای $(nq, np) > 5$ باشد، تقریب آن به نرمال با $\mu = np \;,\; \sigma^2 = npq$ مناسب است.

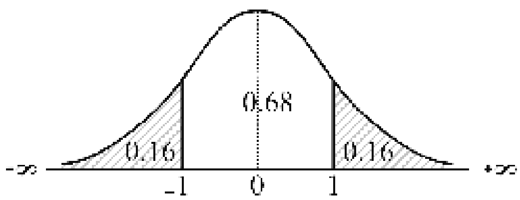
در این سؤال نیز $nq = 100\times 0.2 = 20 > 5 \quad , \quad np = 100\times 0.8 = 80 > 5$ است بنابراین تقریب آن به نرمال مناسب است با:

$$\begin{cases} \mu = np = 100\times 0.8 = 80 \\ \sigma^2 = npq = 100\times 0.8\times 0.2 = 16 \end{cases}$$

$$P(X \geq 84) = P\left(\frac{X-np}{\sqrt{npq}} \geq \frac{84-80}{\sqrt{16}}\right) = P(Z \geq 1) = 0.16 = \%16$$

یادآوری:

$$P(Z \leq -1) = P(Z \geq 1) = 0.16$$



توزیع کای ـ دو (خی‌دو، مربع کای، کای‌اسکور)

۵۸ . گزینه ۴ درست است.

یادآوری: اگر متغیر تصادفی X دارای توزیع χ^2 با درجه آزادی (n) باشد:

$$X \sim \chi^2_{(n)} \rightarrow \begin{cases} E(X) = n \\ Var(X) = 2n \end{cases}$$

بنابراین برای محاسبه ضریب تغییرات X که دارای توزیع χ^2 با درجه آزادی $n = 25$ است، داریم:

$$CV_X = \frac{\sigma_X}{\mu_X} = \frac{\sqrt{2n}}{n} = \frac{\sqrt{2\times 25}}{25} = \frac{5\sqrt{2}}{25} = \frac{\sqrt{2}}{5} = 0.2\sqrt{2} = \%20\sqrt{2}$$

توزیع فیشر

۵۹ . گزینه ۱ درست است.

یادآوری: در توزیع فیشر (F) با n_1 و n_2 درجه آزادی، رابطه زیر برقرار است:

$$F_{\alpha, n_1, n_2} = \frac{1}{F_{1-\alpha, n_2, n_1}}$$

بنابراین در این سؤال داریم:

$$F_{0.9,8,4} = \frac{1}{F_{0.1,4,8}} = \frac{1}{2.8} = 0.357$$

خودآزمایی

توزیع دوجمله‌ای

۱. **تاسی را 144 بار می‌ریزیم. واریانس دفعاتی که عدد 6 بالا قرار گیرد، کدام است؟**

۱) $\sqrt{20}$ ۲) 20 ۳) 40 ۴) 400

۲. **یک جفت تاس سالم را 18 بار پرتاب می‌کنیم. انتظار دارید چند بار هر دو عدد روشده مضرب 3 باشند؟**

۱) 2 ۲) 3 ۳) 5 ۴) 4

۳. **سکه سالمی 20 بار پرتاب می‌شود. اگر متغیر تصادفی X تعداد یک روی ظاهرشده از سکه باشد، مطلوب است امید ریاضی و واریانس X.**

۱) $15 , \sqrt{15}$ ۲) $20 , 5$ ۳) $10 , \sqrt{5}$ ۴) $10 , 5$

۴. **اگر $P(X=x)=\dfrac{\binom{4}{x}}{16}$، تابع احتمال بر روی مجموعه $\{0,1,2,3,4\}$ باشد، واریانس X کدام است؟**

۱) 1.75 ۲) 1.5 ۳) 1.25 ۴) 1

۵. **در یک توزیع دوجمله‌ای اگر میانگین برابر 30 و واریانس برابر 10 باشد، مقدار p (احتمال موفقیت) کدام است؟**

۱) $\frac{1}{3}$ ۲) $\frac{2}{3}$ ۳) $\sqrt{\frac{1}{3}}$ ۴) $\sqrt{\frac{1}{3}}$

۶. **میزان برگشت از خریدهای شرکتی 10% خریدها بوده است. احتمال اینکه با انتخاب تصادفی 4 قلم از خریدهای شرکت هیچ یک برگشت نشده باشند، چقدر است؟**

۱) 0.6561 ۲) 0.36 ۳) 0.0001 ۴) 0.09

۷. **احتمال فارغ‌التحصیل شدن پذیرفته‌شدگان یک دانشکده برابر 0.8 بوده است. چقدر احتمال دارد که از بین 4 دانشجوی این دانشکده، 3 نفر فارغ‌التحصیل شوند؟**

۱) 0.4096 ۲) 0.384 ۳) 0.48 ۴) 0.6

۸. ده سکه همتراز را با هم پرتاب می‌کنیم. اگر X را تعداد دفعات ظاهر شدن شیر در نظر بگیریم، به طور متوسط انتظار داریم چند شیر ظاهر شود؟

۱) 5 ۲) 10 ۳) 100 ۴) 1024

۹. اگر X تعداد پرتاب‌های منجر به گل یک بازیکن۴ بسکتبال باشد و میانگین این پرتاب‌ها 20 و واریانس آن‌ها $\frac{20}{3}$ باشد، آن‌گاه احتمال اینکه در 5 پرتاب حداقل 1 پرتاب گل شود، چقدر است؟

۱) $\frac{1}{243}$ ۲) $\frac{80}{243}$ ۳) $\frac{239}{243}$ ۴) $\frac{242}{243}$

توزیع چندجمله‌ای

۱۰. در یک کارگاه تک‌تولیدی، 50 درصد کالاها مرغوب، 40 درصد متوسط و 10 درصد نامرغوب‌اند. اگر 5 عدد از این کالاها به‌تصادف برداشته شود، با کدام احتمال 2 عدد مرغوب، 2 عدد متوسط و یک عدد نامرغوب‌اند؟

۱) 0.12 ۲) 0.15 ۳) 0.16 ۴) 0.24

۱۱. از انباری با تعداد زیادی کالا که 50% آن‌ها از نوع الف، 30% از نوع ب و 20% از نوع ج است، با انتخاب تصادفی 3 کالا، چقدر احتمال دارد که از هر نوع یک عدد برداشته شده باشد؟

۱) 0.03 ۲) 0.18 ۳) 0.06 ۴) 0.09

توزیع دوجمله‌ای منفی

۱۲. اگر تا انهدام کامل یک هدف، به سوی آن شلیک شود و فرض کنیم که احتمال اصابت هر راکت به هدف 0.3 است، برای انهدام کامل هدف، اصابت دو راکت لازم است. احتمال اینکه با شلیک پنجمین راکت، هدف کاملاً نابود شود، چقدر است؟

۱) 0.6225 ۲) 0.1235 ۳) 0.2425 ۴) 0.4245

۱۳. اگر فوتبالیستی 80% پنالتی‌هایش را وارد دروازه کند، چقدر احتمال دارد که پنجمین پنالتی او سومین موفقیتش باشد؟

۱) 0.2048 ۲) 0.1536 ۳) 0.1229 ۴) 0.0819

توزیع هندسی

۱۴. تاسی را آن‌قدر می‌ریزیم تا بالاخره عدد یک بالا قرار گیرد. احتمال آنکه در سومین نوبت ریختن تاس، برای اولین دفعه عدد یک بالا قرار گیرد، کدام است؟

۱) $\frac{1}{216}$ ۲) $\frac{1}{36}$ ۳) $\frac{1}{6}$ ۴) $\frac{25}{216}$

۱۵. تاسی را آن‌قدر می‌ریزیم تا سرانجام عدد شش ظاهر شود. واریانس تعداد دفعاتی که باید منتظر باشیم تا عدد شش ظاهر شود، چقدر است؟

۱) $\frac{6}{5}$ ۲) $\frac{25}{6}$ ۳) 6 ۴) 30

توزیع فوق هندسی

۱۶. از بین 10 نفر از کارمندان شرکتی که 7 نفر آن‌ها تحصیلات دانشگاهی دارند، 5 نفر به‌تصادف انتخاب می‌شوند. اگر X تعداد افرادی باشد که تحصیلات دانشگاهی دارند، واریانس X چقدر است؟

۱) 0.81 ۲) 0.94 ۳) 0.58 ۴) 1.05

۱۷. از دوازده تخم‌مرغ که سه عدد آن شکسته و بقیه سالم‌اند، سه تخم مرغ را به‌تصادف انتخاب می‌کنیم. احتمال آنکه هر سه سالم باشند، چقدر است؟

۱) $\frac{3}{4}$ ۲) $\frac{3}{12}$ ۳) $\frac{23}{45}$ ۴) $\frac{21}{55}$

توزیع پواسون

۱۸. تعداد سرکشی‌های اضطراری به یک خط تولید دارای توزیع پواسون با متوسط 3 سرکشی در روز است. احتمال اینکه در یک روز هیچ سرکشی‌ای صورت نگرفته باشد، چقدر است؟

۱) e^{-3} ۲) $3e^{-3}$ ۳) $2e^{-3}$ ۴) $\frac{1}{3}e^{-3}$

۱۹. به‌طور متوسط با توزیع پواسون در هر دقیقه 2 اتومبیل برای زدن بنزین بدون سرب به پمپ بنزینی مراجعه می‌کنند. احتمال اینکه در پنج دقیقه، 2 اتومبیل مراجعه کند، چقدر است؟

۱) $50\,e^{-10}$ ۲) $2e^{-5}$ ۳) $4e^{-4}$ ۴) $10e^{-10}$

۲۰. میانگین و واریانس توزیع پواسون به ترتیب (از راست به چپ) کدام است؟

۱) λ , λ ۲) $\sqrt{\lambda} , \lambda$ ۳) $\frac{1}{\lambda^2} , \lambda$ ۴) $\frac{1}{\lambda^2} , \frac{1}{\lambda}$

۲۱. تعداد مشتریانی که به بانک مراجعه می‌کنند دارای توزیع پواسون با میانگین 2 مشتری در هر دقیقه است. با کدام احتمال در 1.5 دقیقه اول کمتر از 4 مشتری مراجعه می‌کنند؟ $\left(e^{-3} = 0.05\right)$

۱) 0.325 ۲) 0.425 ۳) 0.55 ۴) 0.65

۲۲. در کدام‌یک از توزیع‌های زیر مقدار انحراف معیار همواره با جذر مقدار میانگین برابر است؟

۱) توزیع نرمال ۲) توزیع دوجمله‌ای ۳) توزیع نمایی ۴) توزیع پواسون

۲۳. یک تایپیست به‌طور متوسط در هر 5 صفحه، 1 لغت را غلط تایپ می‌کند. احتمال اینکه در 10 صفحه 2 غلط تایپی داشته باشد، چقدر است؟

۱) e^{-2} ۲) $\frac{1}{2}e^{-2}$ ۳) $2e^{-2}$ ۴) $4e^{-2}$

تقریب توزیع دوجمله‌ای به پواسون

۲۴. احتمال بروز عوارض جانبی در مقابل مصرف نوعی دارو در یک بیمار 0.002 است. اگر 500 بیمار این دارو را مصرف کرده باشند، احتمال آنکه فقط 2 نفرشان دچار عوارض جانبی شوند، چقدر است؟

۱) $\frac{1}{2e}$ ۲) $\frac{2}{e^2}$ ۳) $\frac{2}{e}$ ۴) $\frac{1}{e^2}$

۲۵. معمولاً $\frac{1}{100}$ مسافران هواپیماها به موقع به پرواز نمی‌رسند. احتمال آنکه از 200 مسافر یک پرواز 3 نفر به موقع نرسند، کدام است؟ $(e = 2.7)$

۱) 0.000001　۲) 0.01　۳) 0.1　۴) 0.18

توزیع یکنواخت پیوسته

۲۶. میانگین متغیر تصادفی X با تابع چگالی زیر چقدر است؟

$$f(x)=\begin{cases}\frac{2}{3} & ; \quad -1<x<\frac{1}{2}\\ 0 & ; \quad \text{در غیر این صورت}\end{cases}$$

۱) $\frac{3}{4}$　۲) $\frac{1}{4}$　۳) $-\frac{3}{4}$　۴) $-\frac{1}{4}$

۲۷. واریانس متغیر تصادفی X در سؤال قبل چقدر است؟

۱) $\frac{1}{16}$　۲) $\frac{3}{16}$　۳) $\frac{9}{4}$　۴) $\frac{4}{9}$

۲۸. اگر محموله‌ای متشکل از 10 قلم کالا باشد که وزن هر قلم کالا دارای توزیع یکنواخت با پارامترهای $\alpha = 13$ و $\beta = 25$ تن است، کدام‌یک از عبارت‌های زیر در مورد محموله درست است؟

۱) $\sigma^2 = 190$، $\mu = 120$ و توزیع یکنواخت　۲) $\sigma^2 = 190$، $\mu = 120$ و توزیع نرمال

۳) $\sigma^2 = 120$، $\mu = 190$ و توزیع نرمال　۴) $\sigma^2 = 120$، $\mu = 190$ و توزیع نامشخص

توزیع نمایی

۲۹. متغیر تصادفی X با توزیع نمایی با پارامتر $\lambda = 3$ مفروض است. واریانس متغیر تصادفی X چقدراست؟

۱) 3　۲) 9　۳) $\frac{1}{3}$　۴) $\frac{1}{9}$

۳۰. به‌طور متوسط با توزیع نمایی هر 5 دقیقه یک مشتری به بانکی مراجعه می‌کند. امید ریاضی تعداد افرادی که در هر دقیقه مراجعه می‌کنند چقدر است؟

۱) 5　۲) $\frac{1}{5}$　۳) 12　۴) $\frac{1}{12}$

۳۱. در سؤال قبل، واریانس تعداد افرادی که در هر 5 دقیقه مراجعه می‌کنند، چقدر است؟

۱) $\frac{1}{5}$　۲) 1　۳) $\frac{1}{25}$　۴) 5

۳۲. فرض کنید T یک متغیر تصادفی از نوع نمایی باشد. کدام‌یک از روابط زیر برقرار است؟

۱) $P(T > a + b \mid T < a) = P(T > b)$　۲) $P(T > a + b \mid T > b) = P(T > a)$

۳) $P(T > a + b \mid T > b) = P(T > b)$　۴) $P(T > a + b \mid T > a) = P(T > a)$

توزیع نرمال

۳۳. زمان لازم برای انجام کارهای بانکی یک مشتری به طور متوسط 2 دقیقه با انحراف معیار 40 ثانیه است. در توزیع نرمال، 5 درصد از مشتریان بیشترین زمان را گرفته‌اند. حداقل این زمان چند ثانیه است؟ $\left(S_0^{1.65} = 0.45\right)$

۱) 186 ۲) 176 ۳) 172 ۴) 152

۳۴. نمرات آزمون داوطلبان یک توزیع نرمال با میانگین 66 و انحراف معیار 4 است، چند درصد از این داوطلبان نمراتی بین $(60 , 72)$ دارند؟ $\left(\int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 0.0668\right)$

۱) 43.34 ۲) 46.66 ۳) 86.64 ۴) 93.32

۳۵. با فرض اینکه نمرات درس ریاضی دانشجویان یک دانشگاه دارای توزیع نرمال با میانگین 14 و واریانس 16 باشد، احتمال اینکه یک دانشجو نمره‌ای کمتر از 10 کسب کند، چقدر است؟

۱) 0.32 ۲) 0.68 ۳) 0.50 ۴) 0.16

۳۶. در آزمون سراسری میانگین نمرات 60 و انحراف معیار نمرات 20 است. اگر %10 شرکت‌کنندگان بتوانند قبول شوند، حداقل نمره قبولی چقدر است؟ $\left(\int_{-\infty}^{1.28} = 0.9\right)$

۱) 75.6 ۲) 85 ۳) 75 ۴) 85.6

۳۷. در یک بررسی از یک توزیع نرمال $\mu = 100$ و $\sigma = 15$ مشاهده شده است. چه نسبتی از این افراد 130 یا بالاتر از 130 هستند؟ $\left(\int_{-2}^{+2} f_Z(z) d_Z = 0.9544\right)$

۱) 0.0228 ۲) 0.0456 ۳) 0.0912 ۴) 0.4772

۳۸. برای رسم نمودار احتمال در صفحه احتمال نرمال برای داده‌های طبقه‌بندی شده، از کدام‌یک از موارد زیر استفاده می‌شود؟

۱) محور افقی حدود طبقات، محور عمودی فراوانی نسبی
۲) محور افقی فراوانی نسبی، محور عمودی حدود طبقات
۳) محور افقی حدود بالای طبقات، محور عمودی فراوانی تجمعی یا نسبی
۴) محور افقی حدود بالای طبقات، محور عمودی فراوانی نسبی تجمعی

توزیع نرمال استاندارد

۳۹. نمرات احمد و محمود در درس ریاضی 18 و 12 و نمرات استاندارد شده آن‌ها به ترتیب 1 و -1 است، میانگین و انحراف معیار نمرات ریاضی کل دانشجویان به ترتیب کدام است؟

۱) 2 , 15 ۲) 3 , 15 ۳) 2 , 16 ۴) 3 , 16

تقریب توزیع دوجمله‌ای به نرمال

۴۰. اگر n و p پارامترهای توزیع دوجمله‌ای باشند، در کدام حالت دقت تخمین احتمال با تبدیل به توزیع نرمال بیشتر است؟

۱) $n = 20 , p = 0.4$ ۲) $n = 20 , p = 0.45$ ۳) $n = 30 , p = 0.25$ ۴) $n = 5 , p = 0.6$

تقریب توزیع پواسون به نرمال

۴۱. توزیع پواسونی با $\lambda = 36$ را در نظر بگیرید. پارامترهای تقریب نرمال آن کدام است؟

۱) $\sigma=36, \mu=36$ ۲) $\sigma=36, \mu=6$ ۳) $\sigma=6, \mu=36$ ۴) $\sigma=6, \mu=6$

توزیع کای ـ دو، (خی‌دو، مربع کای، کای‌اسکور)

۴۲. اگر X دارای توزیع نرمال با میانگین صفر و واریانس 1 باشد، آن‌گاه $Y=X^2$ دارای توزیع:

۱) نرمال است.
۲) χ^2 با یک درجه آزادی است.
۳) t با یک درجه آزادی است.
۴) χ^2 با $\frac{1}{2}$ درجه آزادی است.

۴۳. درجه آزادی یک توزیع کای ـ مربع 5 است. میانگین و واریانس آن از چپ به راست کدام است؟

۱) $5 , 5$ ۲) $10 , 5$ ۳) $10 , 10$ ۴) $5 , 10$

توزیع فیشر

۴۴. اگر X_1 و X_2 دارای توزیع نرمال استاندارد مستقل باشند، توزیع متغیر تصادفی $Y=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$ کدام است؟

۱) مربع کای با یک درجه آزادی
۲) مربع کای با دو درجه آزادی
۳) توزیع F با یک و یک درجه آزادی
۴) توزیع t با یک درجه آزادی

پاسخنامه

۱	۱	●	۳	۴
۲	●	۲	۳	۴
۳	۱	۲	۳	●
۴	۱	۲	۳	●
۵	۱	●	۳	۴
۶	●	۲	۳	۴
۷	●	۲	۳	۴
۸	●	۲	۳	۴
۹	۱	۲	۳	●
۱۰	●	۲	۳	۴
۱۱	۱	●	۳	۴
۱۲	۱	●	۳	۴
۱۳	۱	۲	●	۴
۱۴	۱	۲	۳	●
۱۵	۱	۲	۳	●
۱۶	۱	۲	●	۴
۱۷	۱	۲	۳	●
۱۸	●	۲	۳	۴
۱۹	●	۲	۳	۴
۲۰	●	۲	۳	۴
۲۱	۱	۲	۳	●
۲۲	۱	۲	۳	●
۲۳	۱	۲	●	۴
۲۴	●	۲	۳	۴
۲۵	۱	۲	۳	●
۲۶	۱	۲	۳	●
۲۷	۱	●	۳	۴
۲۸	۱	۲	●	۴
۲۹	۱	۲	۳	●
۳۰	۱	●	۳	۴
۳۱	۱	●	۳	۴
۳۲	۱	●	۳	۴
۳۳	●	۲	۳	۴
۳۴	۱	۲	●	۴
۳۵	۱	۲	۳	●
۳۶	۱	۲	۳	●
۳۷	●	۲	۳	۴
۳۸	۱	۲	۳	●
۳۹	۱	●	۳	۴
۴۰	۱	●	۳	۴
۴۱	۱	۲	●	۴
۴۲	۱	●	۳	۴
۴۳	۱	۲	۳	●
۴۴	۱	۲	●	۴
۴۵	۱	۲	۳	۴
۴۶	۱	۲	۳	۴
۴۷	۱	۲	۳	۴
۴۸	۱	۲	۳	۴
۴۹	۱۱	۲	۳	۴
۵۰	۱	۲	۳	۴
۵۱	۱	۲	۳	۴
۵۲	۱	۲	۳	۴
۵۳	۱	۲	۳	۴
۵۴	۱	۲	۳	۴
۵۵	۱	۲	۳	۴
۵۶	۱	۲	۳	۴
۵۷	۱	۲	۳	۴
۵۸	۱	۲	۳	۴
۵۹	۱	۲	۳	۴
۶۰	۱	۲	۳	۴

برای دریافت پاسخ تشریحی سؤالات به سایت **www.Tourani.ir** مراجعه کنید.

فصل ۵ توزیع‌های نمونه‌ای و برآورد

مقدمه

در بسیاری از موارد پژوهشگران به دنبال تعیین شاخص‌های جامعه هستند اما دسترسی به این شاخص‌ها به دلایلی به صورت مستقیم از طریق سرشماری امکان‌پذیر نیست، ازاین‌رو از سرشماری جامعه صرف‌نظر کرده، به نمونه‌ای از آن اکتفا می‌کنند؛ نمونه (Sample) تعداد محدودی از عناصر جامعه است که بیان‌کننده تمام ویژگی‌های آن باشد. دلایل نمونه‌گیری عبارت‌اند از:

۱- هزینه: جمع‌آوری اطلاعات از نمونه، هزینه بسیار کمتری نسبت به سرشماری تمام عناصر جامعه دارد.

۲- به‌روز بودن: ازآنجاکه در نمونه، داده‌های کمتری جمع‌آوری و تحلیل می‌شود، زمانی که اطلاعات برای تصمیم‌گیری سریع مورد نیاز باشد، نمونه اغلب اطلاعاتی بهنگام‌تر از سرشماری به دست می‌دهد.

۳- درستی و صحت: ازآنجاکه در یک کار تحقیقی کوچک بهتر می‌توان خطاهای جمع‌آوری داده‌ها را کنترل کرد، نمونه اغلب اطلاعاتی به درستی سرشماری و یا حتی درست‌تر از آن فراهم می‌کند.

۴- زمان: سرشماری کل جامعه به زمانی طولانی نیاز دارد به طوری که گاهی زمان تحقیق و دسترسی به تمام عناصر جامعه به اندازه‌ای طولانی است که کاربرد آن را منتفی می‌کند.

۵- آزمون تخریب‌کننده: وقتی آزمونی موجب خراب شدن یک کالا می‌شود، باید نمونه‌گیری انجام داد. برای مثال، برای کنترل لامپ‌های یک کارخانه باید نمونه‌ای از لامپ‌ها تهیه شود.

روش‌های نمونه‌گیری

برای آنکه نمونه انتخاب‌شده از جامعه بتواند پیش‌بینی درست و دقیقی از شاخص‌های جامعه داشته باشد، باید نماینده واقعی جامعه باشد زیرا یک نمونه اریب به هیچ وجه نمی‌تواند تخمین درستی از شاخص جامعه را نشان دهد.
اریبی نمونه را می‌توان با به کار بردن روش‌های نمونه‌گیری مناسب و درست و با در نظر گرفتن شرایط جامعه، کاهش داد تا به نااریبی برسد. در این وضعیت، استنباط شاخص‌های جامعه از چنین نمونه‌هایی دارای پایایی خواهد بود. این امر به این علت است که اصول نمونه‌گیری تصادفی، پایه آمار استنباطی (برآورد و آزمون) را تشکیل می‌دهد و بر اساس شانس و احتمال است.

انواع روش‌های نمونه‌گیری

صرف‌نظر از اینکه استفاده از کدام روش آمار استنباطی مورد نظر است، قدرت آن روش بستگی به روشی دارد که برای انتخاب نمونه به کار می‌رود. در عمل، میزان دقت مورد نظر و هزینه تخصیص داده شده برای نمونه‌گیری در انتخاب نوع نمونه‌گیری نقش تعیین‌کننده‌ای دارد. انواع روش‌های نمونه‌گیری عبارت‌اند از:

۱- نمونه‌گیری تصادفی ساده

۲- نمونه‌گیری تصادفی منظم (سیستماتیک)

۳- نمونه‌گیری تصادفی گروهی (طبقه‌ای)

۴- نمونه‌گیری تصادفی خوشه‌ای

۵- نمونه‌گیری تصادفی مرحله‌ای

در ادامه به بررسی این روش‌ها می‌پردازیم.

۱- نمونه‌گیری تصادفی ساده (Simple Random Sampling)

در نمونه‌گیری تصادفی ساده، هر یک از عناصر جامعه برای انتخاب شدن شانس مساوی دارند (هم تراز هستند). در این حالت افراد یا اشیا به صورت تصادفی از لیست تهیه‌شده از جامعه انتخاب می‌شوند و باید ویژگی‌هایی همانند ویژگی‌های همان جامعه‌ای که از آن انتخاب می‌شوند، داشته باشند. نمونه حاصل از این روش، **نمونه تصادفی ساده** (Simple Random Sample) نامیده می‌شود.

✓ **دقت کنید!**

به صورت پیش‌فرض، منظور از «نمونه» و «نمونه تصادفی» همان «نمونه تصادفی ساده» است.

مثال یک نمونه تصادفی ساده چگونه نمونه‌ای است؟ (اقتصاد ـ ۷۳)

۱) همه عناصر جامعه شانس مساوی در انتخاب شدن داشته باشند و همه نمونه‌های ممکن هم‌شانس باشند.

۲) همه عناصر جامعه شانس مساوی داشته باشند که در نمونه انتخاب شوند.

۳) همه نمونه‌های ممکن هم‌شانس باشند.

۴) نماینده خوبی از کل جامعه آماری باشد.

حل: گزینه ۲ درست است.

نمونه‌گیری تصادفی ساده به دو روش انجام می‌شود:

۱) قرعه‌کشی: قرعه‌کشی با هر یک از روش‌های معمول آن نوعی نمونه‌برداری است اما فراهم آوردن وسایل قرعه‌کشی بی‌نقص، مخصوصاً در گروه‌های بزرگ، کار آسانی نیست.

۲) جدول اعداد تصادفی: در جدول اعداد تصادفی، ارقام 0 تا 9 در تعدادی سطر و ستون گردآوری شده‌اند. ترتیب استخراج و تنظیم این اعداد به صورت کاملاً تصادفی، با روش‌ها و وسایلی مانند قرعه‌کشی و رایانه انجام می‌شود. یکی از مشکلات این روش، تهیه و تدوین فهرست افراد جامعه آماری است، زیرا در بسیاری از موارد چنین کاری قبلاً انجام نشده است.

۲- نمونه‌گیری تصادفی منظم (سیستماتیک) (Systematic Random Sampling)

در این روش، شکل تغییریافته نمونه‌گیری تصادفی ساده به کار گرفته می‌شود. این روش برای حالتی که واحدهای جامعه از 1 تا N شماره‌گذاری شده‌اند به صورت زیر انجام می‌شود:

فرض کنید بخواهیم یک نمونه 20 نفره از یک لیست 500 نفره انتخاب کنیم. یعنی از هر 25 نفر که در جامعه است یک نمونه انتخاب شود. با استفاده از فهرست اعداد تصادفی عددی را به‌تصادف بین 1 تا 25 انتخاب می‌کنیم، سپس با جمع کردن متوالی 25 با عدد انتخاب‌شده نمونه‌های بعدی به دست می‌آیند. این روش برای آن دسته از جوامع آماری که کد از پیش تعیین

شده و مرتبی دارند، کاربرد فراوان دارد. با مشخص شدن اولین عضو، بقیه اعضا نیز مشخص می‌شوند. این خاصیت از یک سو مزیت است زیرا اولاً انجام آن آسان است و ثانیاً پرهزینه نیست، اما از سوی دیگر چون شانس را از بقیه اعضا می‌گیرد، عیب محسوب می‌شود (مانند شماره کارمندی و شماره دانشجویی). یکی از ویژگی‌های این نمونه‌گیری انتخاب نمونه به طور یکنواخت در گروه‌های مختلف است.

مثال مناسب‌ترین روش نمونه‌گیری از اتومبیل‌های سواری که وارد یک بزرگراه می‌شوند، کدام است؟ (اقتصاد ـ ۸۶)

۱) سیستماتیک ۲) ساده ۳) خوشه‌ای ۴) طبقه‌بندی شده

حل: گزینه ۱ درست است.

۳- نمونه‌گیری تصادفی گروهی (طبقه‌ای) (Stratified Random Sampling)

در این روش جامعه را به گروه‌های متجانس (همگن) تقسیم می‌کنیم به طوری که عناصر هر گروه دارای ویژگی مشابهی باشند و اختلاف بین گروه‌ها نیز قابل توجه باشد. پس از تقسیم جامعه به گروه‌های متجانس (همگن)، از هر گروه، نمونه مورد نظر به روش تصادفی ساده یا منظم گرفته می‌شود؛ با این کار، تخمین‌های بهتری از پارامترهای جامعه به دست می‌آید. نکته مهم این است که این روش در جوامعی مورد استفاده قرار می‌گیرد که از نظر صفت مورد نظر ناهمگون هستند، مانند بررسی میزان درآمد کارمندان یک کارخانه؛ در این حالت، ابتدا کارمندان را به طبقات کم‌درآمد، متوسط و پردرآمد تقسیم می‌کنیم به طوری که افراد درون هر طبقه درآمدهای نزدیک به هم داشته باشند و درعین‌حال تفاوت درآمد بین دو طبقه زیاد باشد و سپس از هر طبقه به صورت جداگانه نمونه‌گیری می‌کنیم.

مثال برای نمونه‌گیری از یک جامعه آماری که دارای خصوصیات (ویژگی‌های) ناهمگن بوده ولی قابل تقسیم به طبقات یا گروه‌های همگن باشد، از کدام روش نمونه‌گیری استفاده می‌شود؟ (برنامه‌ریزی شهری ـ ۸۳)

۱) تصادفی طبقه‌بندی شده ۲) خوشه‌ای ۳) تصادفی بدون جایگزینی ۴) تصادفی با جایگزینی

حل: گزینه ۱ درست است.

۴- نمونه‌گیری تصادفی خوشه‌ای (Cluster Random sampling)

هرگاه جامعه مورد نظر خیلی وسیع و گسترده باشد مانند وضعیت معاش یا تحصیل کارمندان یک شهر بزرگ یا یک کشور، برای نمونه‌گیری ابتدا سازمان‌ها یا اداراتی را به روش تصادفی ساده یا سیستماتیک (منظم) انتخاب می‌کنیم، سپس کارمندان مورد نیاز را با استفاده از همین روش به دست می‌آوریم؛ در اینجا واحد نمونه‌گیری خوشه‌ای، سازمان بوده است. علت استفاده از این روش، بیشتر به دلیل کم‌هزینه بودن آن است.

۵- نمونه‌گیری تصادفی مرحله‌ای (Stage Random Sampling)

این روش، شکل گسترش‌یافته نمونه‌گیری خوشه‌ای است. در این حالت، نمونه‌گیری از جامعه طی چند مرحله انجام می‌شود. یعنی انتخاب نمونه از نمونه دیگر، بدین صورت که ابتدا بین خوشه‌ها یک نمونه را انتخاب می‌کنیم سپس به‌جای اینکه از تمام واحدهای درون خوشه استفاده کنیم از آن واحدها نمونه دیگری استخراج می‌کنیم. برای مثال چند سازمان از شهر انتخاب می‌کنیم سپس از بین هر سازمان چند واحد را معین می‌کنیم سپس عناصر نمونه را با فرض همگن بودن جامعه به صورت تصادفی به دست می‌آوریم.

در نمونه‌گیری مرحله‌ای فقط تعدادی از خوشه‌ها مورد بررسی قرار می‌گیرند و مابقی آن‌ها کنار گذاشته می‌شوند، درنتیجه شرط همگن بودن جامعه در این روش نقش مهمی دارد. در مواردی که جامعه تحت بررسی، چندان همگن نیست می‌توان ابتدا آن را طبقه‌بندی کرد (زیرا روش طبقه‌بندی، طبقاتی ایجاد می‌کند که اعضای آن تا حد زیادی همگن هستند)، سپس در داخل هر طبقه نمونه‌گیری خوشه‌ای انجام داد.

پارامتر و آماره

هر **پارامتر** یک ویژگی از جامعه است درحالی‌که همان ویژگی در نمونه یک **آماره** نامیده می‌شود.
برای مثال، میانگین محاسبه‌شده از سرشماری جامعه، یک پارامتر است. حال اگر از جامعه نمونه‌گیری کنیم، میانگین حاصل از نمونه یک آماره خواهد بود.
فرض کنید جامعه‌ای با 6 عنصر به صورت $2,0,3,1,5,7$ باشد. در این صورت، میانگین جامعه (μ) برابر است با:

$$\mu=\frac{2+0+3+1+5+7}{6}=3$$

حال فرض کنید نمونه‌ای تصادفی به حجم $n=4$ از جامعه انتخاب کنیم:
ـ اگر نمونه به صورت $2,0,3,1$ باشد، میانگین نمونه $(\bar{X})$ برابر است با:

$$\bar{x}=\frac{\sum x}{n}=\frac{2+0+3+1}{4}=1.5$$

ـ اگر نمونه به صورت $3,1,7,5$ باشد، میانگین نمونه $(\bar{X})$ برابر است با:

$$\bar{x}=\frac{\sum x}{n}=\frac{3+1+7+5}{4}=4$$

همان‌طورکه ملاحظه می‌شود پارامتر جامعه $(\mu=3)$ مقداری ثابت و منحصربه‌فرد دارد درحالی‌که آماره $\bar{X}$ حاصل از نمونه، یک متغیر تصادفی است زیرا از یک نمونه به نمونه دیگر در حال تغییر است.
پارامتر (Parameter): اصطلاحی که در مورد جامعه استفاده می‌شود و خصوصیتی از آن را بیان می‌کند، مانند میانگین جامعه (μ)، واریانس جامعه (σ^2)، نسبت جامعه (p).
آماره (Statistic): اصطلاحی که در مورد نمونه استفاده می‌شود و خصوصیتی از آن را بیان می‌کند، مانند میانگین نمونه $(\bar{X})$، واریانس نمونه (S^2)، نسبت نمونه $(\bar{p})$.

مقایسه پارامتر و آماره

۱) هر **پارامتر** عددی **ثابت** است درحالی‌که هر **آماره** یک **متغیر تصادفی** است چون از یک نمونه به نمونه دیگر تغییر می‌کند.
۲) هر **پارامتر** در جامعه **مجهول** است در حالی‌که هر **آماره** فاقد پارامتر مجهول جامعه و **معلوم** است.
نتیجه: هر پارامتر جامعه یک عدد مجهول است که مقدار آن تنها از طریق سرشماری قابل محاسبه است. ازآنجاکه سرشماری همواره امکان‌پذیر نیست، مقدار پارامتر از طریق آماره‌ها در نمونه برآورد و آزمون می‌شود.
پارامترها و آماره‌های مهم

گروه	شاخص	نوع عمل	مشخصات	میانگین	واریانس	نسبت	ضریب همبستگی
جامعه	پارامتر θ	سرشماری	مجهول و ثابت	μ	σ^2	p	ρ
نمونه	آماره $\hat{\theta}$	نمونه‌گیری	معلوم و متغیر	$\bar{X}$	S^2	$\bar{p}$	r

مثال $\bar{X}$ به عنوان برآوردکننده‌ای از μ: (اقتصاد ـ ۷۰)
۱) یک متغیر تصادفی است اگر که μ متغیر تصادفی باشد.
۲) یک متغیر تصادفی است درحالی‌که μ ثابت است.
۳) یک کمیت ثابت است در حالی‌که μ متغیر باشد.
۴) یک کمیت ثابت است و μ معین ثابت است.

حل: گزینه ۲ درست است.

توزیع‌های نمونه‌ای (Sample Distributions)

همان‌طور‌که دیدیم برای استنباط (برآورد، آزمون) یک پارامتر جامعه باید از یک آماره استفاده کرد که تابعی از نمونه و درنتیجه یک متغیر تصادفی است.

توزیع آماره: هر آماره دارای یک تابع احتمال است که بر اساس نمونه‌های تصادفی n تایی که به طور مکرر از جامعه انتخاب شده، به دست می‌آید. به این تابع در اصطلاح «توزیع نمونه‌ای آماره» می‌گویند.

توزیع نمونه‌ای یک آماره به جامعه‌ای بستگی دارد که نمونه از آن به دست آمده است. اگر اندازه نمونه n تایی بزرگ باشد، آماره به‌دست‌آمده از نمونه‌های مکرر که به صورت نمودار فراوانی نسبی ترسیم می‌شود تقریب بسیار خوبی برای توزیع نمونه‌ای واقعی است.

هرگونه استنباط از آماره به توزیع آن بستگی دارد؛ توزیع آماره مشخص می‌کند که آیا آماره به کار رفته، با پارامتر جامعه متناسب است یا خیر؟

در ادامه به بررسی توزیع‌های نمونه‌ای برای آماره‌های مختلف به شرح زیر می‌پردازیم:

۱ـ توزیع میانگین نمونه $(\bar{X})$

۲ـ توزیع تفاضل یا مجموع میانگین دو نمونه $(\bar{X}_1 \pm \bar{X}_2)$

۳ـ توزیع نسبت (نرخ) موفقیت نمونه $(\bar{p})$

۴ـ توزیع تفاضل یا مجموع نسبت دو نمونه $(\bar{p}_1 \pm \bar{p}_2)$

۵ـ توزیع واریانس نمونه (S^2)

۶ـ توزیع نسبت واریانس دو نمونه $\left(S_1^2 / S_2^2\right)$

میانگین نمونه

آماره‌های مختلفی مانند میانه (Md)، میانگین نمونه $(\bar{X})$، ... برای تخمین (برآورد) میانگین جامعه (μ) وجود دارد که در مقایسه آن‌ها با هم، $\bar{X}$ (میانگین نمونه) دارای خواص مطلوب‌تری است؛ به همین دلیل برای تخمین میانگین جامعه (μ) از میانگین نمونه $(\bar{X})$ و توزیع آن به نحو وسیعی استفاده می‌شود.

تعریف: اگر $X_1, X_2, \ldots, X_n$ یک نمونه مستقل n تایی از جامعه‌ای با میانگین μ_X و واریانس σ_X^2 باشد، به طوری که $E(X_1)=\mu_X, \ldots, E(X_n)=\mu_X$ و $\sigma_{X_1}^2=\sigma_X^2, \ldots, \sigma_{X_n}^2=\sigma_X^2$، آن‌گاه میانگین نمونه به صورت زیر تعریف می‌شود:

$$\bar{X}=\frac{\sum_{i=1}^{n} x_i}{n}$$

امید و واریانس میانگین نمونه

$$\mu_{\bar{X}} = E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{\overset{\mu_X}{\cancel{E(X_1)}} + \ldots + \overset{\mu_X}{\cancel{E(X_n)}}}{n} = \frac{n\mu_X}{n} = \mu_X$$

$$\sigma^2_{\bar{X}} = \sigma^2\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{\overset{\sigma^2_X}{\cancel{\sigma^2_{X_1}}} + \ldots + \overset{\sigma^2_X}{\cancel{\sigma^2_{X_n}}}}{n^2} = \frac{n\sigma^2_X}{n^2} = \frac{\sigma^2_X}{n}$$

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

بنابراین:

امید میانگین نمونه برابر با میانگین جامعه است.	$\mu_{\bar{X}} = E(\bar{X}) = \mu$
واریانس میانگین نمونه برابر با حاصل تقسیم واریانس جامعه بر n است.	$\sigma^2_{\bar{X}} = \frac{\sigma^2_x}{n}$
انحراف معیار میانگین نمونه برابر با حاصل تقسیم انحراف معیار جامعه بر $\sqrt{n}$ است.	$\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{n}}$

مثال ۱ توزیع نمونه‌گیری $\bar{X}$ دارای میانگین 120 است. میانگین واقعی جامعه آماری چقدر است؟

۱) 120 ۲) 130 ۳) 110 ۴) اطلاعات کافی نیست.

حل: گزینه ۱ درست است.

$$\begin{cases} E(\bar{X}) = \mu_{\bar{X}} = 120 = \mu_X \\ \mu_X : \text{میانگین جامعه} \\ \mu_{\bar{X}} = \bar{X} \ (\text{میانگین نمونه}) \end{cases}$$

مثال ۲ کدام رابطه درست است؟

۱) $E(\bar{X}) = \mu_X$ ۲) $E(\bar{X}) = \mu_{\bar{X}}$ ۳) $\mu_X = \bar{X}$ ۴) ۱ و ۲

حل: گزینه ۴ درست است.

توجه: میانگین نمونه با میانگین جامعه برابر نیست $(\mu_X \neq \bar{X})$ بلکه امید میانگین نمونه با میانگین جامعه برابر است.

مثال ۳ توزیع نمونه‌گیری $\bar{X}$ دارای انحراف معیار 2 است. اگر انحراف معیار جامعه آماری 12 باشد، مقدار n چقدر است؟

۱) 6 ۲) 36 ۳) 144 ۴) 72

حل: گزینه ۲ درست است.

$$\begin{cases} \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \rightarrow 2 = \frac{12}{\sqrt{n}} \rightarrow \sqrt{n} = 6 \rightarrow n = 36 \\ \sigma_{\bar{X}} = 2 \quad , \quad \sigma_X = 12 \end{cases}$$

مثال ۴ توزیع میانگین‌های نمونه یک جامعه نامحدود با میانگین 10 و انحراف معیار 2 دارای واریانس 1 خواهد بود اگر تعداد نمونه عبارت باشد از: (مدیریت ـ ۷۴)

۱) 4 ۲) 10 ۳) 16 ۴) 33

حل: گزینه ۱ درست است.

$$\begin{cases} \sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} \rightarrow 1 = \dfrac{2}{\sqrt{n}} \rightarrow \sqrt{n} = 2 \rightarrow n = 4 \\ \sigma_X = 2 , \sigma_{\bar{X}}^2 = 1 \rightarrow \sigma_{\bar{X}} = 1 \end{cases}$$

واریانس جامعه نامعلوم

در این شرایط از $S_X^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ واریانس نمونه استفاده می‌کنیم و داریم:

$$S_{\bar{X}}^2 = \frac{S_X^2}{n} \quad , \quad S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$$

رابطه بین تعداد نمونه و $\sigma_{\bar{X}}^2$ یا $\sigma_{\bar{X}}$

با توجه به روابط $\sigma_{\bar{X}}^2 = \dfrac{\sigma_X^2}{n}$ و $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}}$ مشخص است که رابطه بین n و $\left(\sigma_{\bar{X}}, \sigma_{\bar{X}}^2\right)$ معکوس است، یعنی با افزایش n مقدار $\left(\sigma_{\bar{X}}, \sigma_{\bar{X}}^2\right)$ کاهش و با کاهش n مقدار آن‌ها افزایش می‌یابد.

پارامتر / تغییرات نمونه	واریانس	انحراف معیار	پراکندگی	کشیدگی
کاهش تعداد نمونه (n)	افزایش می‌یابد	افزایش می‌یابد	افزایش می‌یابد	کاهش می‌یابد
افزایش تعداد نمونه (n)	کاهش می‌یابد	کاهش می‌یابد	کاهش می‌یابد	افزایش می‌یابد

مثال ۱ اگر بخواهیم انحراف معیار میانگین نمونه‌ای $\left(\sigma_{\bar{X}}\right)$ بر اساس حجم نمونه $n = 64$ تایی از جامعه‌ای که دارای انحراف معیار 6 است به نصف کاهش یابد، حجم نمونه باید چند تا شود؟ (اقتصاد ـ ۸۲)

۱) 128 ۲) 182 ۳) 256 ۴) 320

حل: گزینه ۳ درست است.

$$\begin{cases} \sigma_{\bar{X}} \rightarrow \dfrac{1}{2}\sigma_{\bar{X}} \rightarrow \dfrac{\sigma_X}{\sqrt{n}} \rightarrow \dfrac{1}{2}\dfrac{\sigma_X}{\sqrt{n}} = \dfrac{\sigma_X}{\sqrt{4n}} \rightarrow \text{n باید 4 برابر شود.} \rightarrow n = 64 \times 4 = 256 \\ n = 64 \ , \ \sigma_X = 6 \end{cases}$$

در صورتی انحراف معیار میانگین نمونه‌ای نصف خواهد شد که حجم نمونه چهار برابر شود.

مثال ۲ در رابطه $\sigma_{\bar{X}}^2=\frac{\sigma_X^2}{n}$ با افزایش حجم نمونه چه اتفاقی می‌افتد؟

۱) کشیدگی توزیع $\bar{X}$ کمتر از توزیع X می‌شود.
۲) کشیدگی توزیع $\bar{X}$ بیشتر از توزیع X می‌شود.
۳) چولگی توزیع $\bar{X}$ بیشتر از توزیع X می‌شود.
۴) دنباله‌های توزیع $\bar{X}$ طولانی‌تر می‌شود.

حل: گزینه ۲ درست است.

جامعه محدود N تایی (فوق‌هندسی)

درصورتی‌که نمونه n تایی از جامعه محدود N تایی به صورت بدون جایگذاری انتخاب شود، در شرایطی که $\frac{n}{N}>0.05$ باشد، برای محاسبه $\sigma_{\bar{X}}^2$ باید از ضریب تصحیح (کاهش) $\frac{N-n}{N-1}$ استفاده کرد. درواقع این ضریب ناشی از همبستگی منفی بین مشاهدات نمونه است. در این شرایط $E(\bar{X})=\mu_X$ است و از این نظر تفاوتی با جامعه نامحدود ندارد؛ به عبارت دیگر:

$$\begin{cases}\mu_{\bar{X}}=E(\bar{X})=\mu_X\\ \sigma_{\bar{X}}^2=\frac{N-n}{N-1}\cdot\frac{\sigma_X^2}{n}\\ \sigma_{\bar{X}}=\sqrt{\frac{N-n}{N-1}}\cdot\frac{\sigma_X}{\sqrt{n}}\end{cases}$$

توجه:

۱- جامعه به طور پیش‌فرض نامحدود در نظر گرفته می‌شود مگر آنکه حجم آن (N) ذکر شود.

۲- درصورتی‌که $\frac{n}{N}\le 0.05$ باشد، دیگر نیازی به ضریب تصحیح $\frac{N-n}{N-1}$ نخواهد بود.

نتیجه:

وضعیت $\bar{X}$ (میانگین نمونه) با توجه به جامعه محدود یا نامحدود به شکل زیر خلاصه می‌شود:

جامعه محدود N تایی	جامعه نامحدود
✓ انتخاب بدون جایگذاری از جامعه محدود	✓ انتخاب با جایگذاری از جامعه محدود
✓ $\left(\frac{n}{N}>0.05\right)$	✓ جامعه محدود، بدون جایگذاری، $\frac{n}{N}\le 0.05$ ✓ انتخاب از جامعه نامحدود (بدون‌جایگذاری، با جایگذاری)
$E(\bar{X})=\mu_X$	$E(\bar{X})=\mu_X$
$\sigma_{\bar{X}}^2=\frac{N-n}{N-1}\cdot\frac{\sigma_X^2}{n}$	$\sigma_{\bar{X}}^2=\frac{\sigma_X^2}{n}$
$\sigma_{\bar{X}}=\sqrt{\frac{N-n}{N-1}}\cdot\frac{\sigma_X}{\sqrt{n}}$	$\sigma_{\bar{X}}=\frac{\sigma_X}{\sqrt{n}}$

مثال ۳ ضریب عامل تصحیح واریانس یعنی $\frac{N-n}{N-1}$ وقتی در محاسبه واریانس میانگین نمونه‌ای مورد استفاده قرار می‌گیرد که نسبت $\frac{n}{N}$: (اقتصاد ـ ۷۳)

۱) بزرگ‌تر از %10 باشد.
۲) کوچک‌تر از %5 باشد.
۳) بزرگ‌تر از %5 باشد.
۴) کوچک‌تر از %10 باشد.

حل: گزینه ۳ درست است.

توزیع میانگین نمونه

فرض کنید $X_1, X_2, \ldots, X_n$ نمونه مستقل n تایی از جامعه‌ای با میانگین μ و واریانس σ^2 باشد.

۱- جامعه نرمال، واریانس جامعه (σ^2) معلوم

توزیع $\bar{X}$ در این شرایط **مستقل از حجم نمونه** (n) و برای هر تعداد از آن $(n \geq 1)$ همیشه **نرمال** است.

۲- جامعه نرمال، واریانس جامعه (σ^2) نامعلوم

توزیع $\bar{X}$ در این وضعیت وابسته به حجم نمونه (n) است. در شرایطی که $n > 30$ باشد، توزیع $\bar{X}$ نرمال و در شرایطی که $n \leq 30$ باشد، توزیع آن $t_{(n-1)}$ با $(n-1)$ درجه آزادی خواهد بود.

	اندازه نمونه	توزیع $\bar{X}$	میانگین $\bar{X}$	واریانس $\bar{X}$	انحراف معیار $\bar{X}$	متغیر استاندارد
الف)	$n > 30$ (نمونه بزرگ)	نرمال	$E(\bar{X}) = \mu$	$S_{\bar{X}}^2 = \frac{S_X^2}{n}$	$S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{\frac{S_X}{\sqrt{n}}}$
ب)	$n \leq 30$ (نمونه کوچک)	$t_{(n-1)}$	$E(\bar{X}) = \mu$	$S_{\bar{X}}^2 = \frac{S_X^2}{n}$	$S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$	$t_{n-1} = \frac{\bar{X} - \mu}{\frac{S_X}{\sqrt{n}}}$

۳- جامعه دلخواه (نامعلوم)، واریانس جامعه (σ^2) معلوم

توزیع $\bar{X}$ در این شرایط **وابسته** به **حجم نمونه** (n) است و به دو صورت زیر بررسی می‌شود:

الف) $n > 30$ **(حجم نمونه زیاد)**

توزیع $\bar{X}$ در این وضعیت با توجه به **قضیه حد مرکزی نرمال** خواهد بود.

اندازه نمونه	توزیع $\bar{X}$	میانگین $\bar{X}$	واریانس $\bar{X}$	انحراف معیار $\bar{X}$	متغیر استاندارد
$n > 30$ (نمونه بزرگ)	نرمال (قضیه حد مرکزی)	$E(\bar{X}) = \mu$	$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$	$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$

قضیه حد مرکزی (Central Limit Theorem)

اگر $X_1, X_2, \ldots, X_n$ دنباله‌ای از متغیرهای تصادفی مستقل با توزیع یکسان و با میانگین μ و واریانس σ^2 باشد، آن‌گاه توزیع مجموع متغیرها به صورت زیر است:

$$\frac{\sum_{i=1}^{n} x_i - n\mu}{\sigma\sqrt{n}} \approx \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

وقتی $n \to \infty$ میل کند توزیع مجموع متغیرها به سمت توزیع نرمال استاندارد میل می‌کند.

اگر یک نمونه تصادفی مستقل n تایی از جامعه‌ای دلخواه (غیرنرمال) با میانگین μ و واریانس σ^2 انتخاب شود، در شرایطی که n (تعداد نمونه) بزرگ باشد $(n > 30)$، توزیع $\bar{X}$ (میانگین نمونه) تقریباً نرمال خواهد بود و داریم:

$$\begin{cases} E(\bar{X}) = \mu_{\bar{X}} = \mu_X \\ \sigma_{\bar{X}}^2 = \dfrac{\sigma_X^2}{n} \\ \sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} \end{cases} \longrightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ب) $n \le 30$ **(حجم نمونه کوچک)**

توزیع $\bar{X}$ در این وضعیت **نامعلوم** است و به کمک **قضیه چی‌بی‌شف** میانگین جامعه (μ) را تخمین می‌زنیم.

قضیه دوم چی‌بی‌شف

بنا بر قضیه چی‌بی‌شف (که در فصل اول مطرح شد)، درصورتی‌که مشاهدات جامعه (X) دارای توزیع غیرنرمال (نامعلوم) باشند، آن‌گاه:

$$(\mathrm{I})\ P\left(\left| X - \mu_X \right| \le \varepsilon\right) \ge 1 - \frac{\sigma_X^2}{\varepsilon^2} \quad , \quad (\mathrm{II})\ P\left(\left| X - \mu_X \right| \ge \varepsilon\right) \le \frac{\sigma_X^2}{\varepsilon^2}$$

حال فرض کنید، به‌جای مشاهدات (X) از میانگین نمونه $(\bar{X})$ استفاده کنیم؛ در این صورت $\mu_{\bar{X}} = \mu_X$ و $\sigma_{\bar{X}}^2 = \dfrac{\sigma_X^2}{n}$ است و قضیه چی بی‌شف به صورت زیر مطرح می‌شود:

قضیه: اگر $X_1, X_2, \ldots, X_n$ یک نمونه مستقل n تایی و $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$، میانگین نمونه باشد، آن‌گاه:

$$(\mathrm{I})\ P\left(\left| \bar{X} - \mu_X \right| \le \varepsilon\right) \ge 1 - \frac{\sigma_X^2}{n\varepsilon^2}$$

$$(\mathrm{II})\ P\left(\left| \bar{X} - \mu_X \right| \ge \varepsilon\right) \le \frac{\sigma_X^2}{n\varepsilon^2}$$

قانون اعداد بزرگ و چی‌بی‌شف

درصورتی‌که بخواهیم روابط به‌دست‌آمده از قضیه دوم چی‌بی‌شف را به صورت حدی بیان کنیم روابط زیر صادق است:

قانون ضعیف: $(\mathrm{I})\ \lim\limits_{n \to \infty} P\left(\left| \bar{X} - \mu \right| \le \varepsilon\right) = 1$

زمانی که نمونه به سمت ∞ میل می‌کند، احتمال آنکه میانگین نمونه $(\bar{X})$ و میانگین جامعه (μ) برابر شوند، یک است.

قانون قوی: $(\mathrm{II})\ \lim\limits_{n \to \infty} P\left(\left| \bar{X} - \mu \right| \ge \varepsilon\right) = 0$

زمانی که نمونه به سمت ∞ میل می‌کند، احتمال آنکه میانگین نمونه $(\bar{X})$ و میانگین جامعه (μ) متفاوت باشند، صفر است.

مثال ۱ دستگاه بسته‌بندی مواد غذایی روی 50 گرم تنظیم شده است، انحراف معیار وزن بسته‌ها 2 گرم است. تعداد $n = 10$ بسته به طور تصادفی انتخاب شده، احتمال آنکه میانگین این 10 بسته بین 46 تا 54 گرم باشد کدام است؟ (مدیریت ـ ۷۱)

۱) 0.9 ۲) 0.99 ۳) 0.975 ۴) 0.999

حل: گزینه ۳ درست است.

با توجه به آنکه توزیع جامعه نامعلوم، واریانس جامعه معلوم و $n < 30$ است از قضیه چی‌بی‌شف استفاده می‌کنیم؛ درعین‌حال چون احتمال حدود میانگین نمونه $(\bar{X})$ خواسته شده است، از شکل (II) قضیه دوم چی‌بی‌شف استفاده می‌کنیم.

$$\begin{cases} P(46 < \bar{X} < 54) = P(46-\mu < \bar{X}-\mu < 54-\mu) = P(-4 < \bar{X}-\mu < 4) \\ = P\left(|\bar{X}-\mu| < \underbrace{4}_{\varepsilon}\right) \ge 1 - \dfrac{\sigma^2}{n\varepsilon^2} = 1 - \dfrac{4}{10 \times 16} = 0.975 \text{ حداقل} \\ \mu = 50 \ , \ \sigma = 2 \ , \ n = 10 \end{cases}$$

مثال ۲ اگر یک توزیع آماری پیوسته دارای چولگی شدیدی باشد، مقدار نمونه دست‌کم چقدر باشد تا آماره $\bar{X}$ از تقریب نرمال برخوردار شود؟

۱) 30 ۲) 50 ۳) 60 ۴) 100

حل: گزینه ۱ درست است.

با توجه به حالت (۳ـ الف) درصورتی‌که تعداد نمونه بزرگ‌تر از 30 باشد توزیع جامعه به توزیع نرمال تقریب می‌خورد.

مثال ۳ حجم نمونه چقدر باشد تا توزیع $\bar{X}$ همان توزیع X شود؟

۱) $n = N$ ۲) $n = 1$ ۳) $n = 30$ ۴) $n > 30$

حل: گزینه ۲ درست است.

همان‌طورکه می‌دانیم وجه تمایز $\bar{X}$ و توزیع X در مقدار واریانس آن‌ها است چون مقدار میانگین هر دو برابر است بنابراین داریم:

$$\begin{cases} \mu_{\bar{X}} = \mu_X = \mu \\ \sigma^2_{\bar{X}} = \dfrac{\sigma^2_X}{n} \xrightarrow{n=1} \sigma^2_{\bar{X}} = \sigma^2_X \end{cases}$$

تنها درصورتی‌که تعداد **نمونه برابر یک** باشد $(n = 1)$ مقدار واریانس توزیع $\bar{X}$ و توزیع X برابر خواهد شد و در نهایت هر دو توزیع $\bar{X}$ و X یکسان هستند.

مثال ۴ میانگین توزیع نمره‌های دانشجویان یک دانشکده 52 و انحراف معیار آن 15 است. احتمال اینکه میانگین یک نمونه تصادفی 100 نفره کمتر از 55 باشد، چقدر است؟

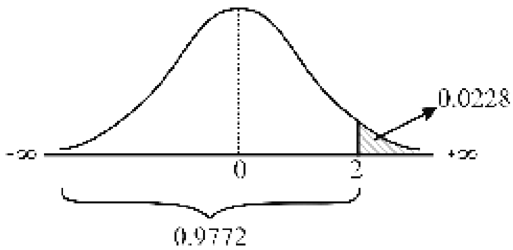
۱) 0.5 ۲) 0.228 ۳) 0.9772 ۴) 0.5793

حل: گزینه ۳ درست است.

با توجه به حالت (۳ـ الف) توزیع $\bar{X}$ به صورت $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2_X}{n}\right)$ است، بنابراین:

$$\begin{cases} P(\bar{X} < 55) = P\left(\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \dfrac{55-52}{\frac{15}{\sqrt{100}}}\right) = P(Z < 2) = 0.9772 \\ Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \\ \mu = 52 \ , \ \sigma = 15 \ , \ n = 100 > 30 \end{cases}$$

یادآوری:



مثال ۵ میانگین نمرات دانشجویان یک دانشگاه 14.5 و انحراف معیار آن 4 است، اگر میانگین نمرات یک نمونه تصادفی 100 نفره از آنان بیشتر از 15 باشد، آن‌گاه عدد نرمال (Z) کدام است؟ (برنامه‌ریزی شهری ـ ۸۴)

۱) 0.25 ۲) 1.25 ۳) 1.5 ۴) 2.5

حل: گزینه ۲ درست است.

با توجه به حالت (۳ ـ الف) توزیع $\bar{X}$ به صورت $\bar{X} \sim N\left(\mu, \frac{\sigma_X^2}{n}\right)$ است، بنابراین:

$$\begin{cases} P(\bar{X} > 15) = P\left(\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \dfrac{15-14.5}{\frac{4}{\sqrt{100}}}\right) = P(Z > 1.25) \\ \mu = 14.5 \ , \ \sigma = 4 \ , \ n = 100 > 30 \end{cases}$$

مثال ۶ توزیع نمرات ارزشیابی یک سازمان نرمال با میانگین 14.5 و انحراف معیار 6 است. اگر یک نمونه 25 نفری از بین آنان انتخاب شود، با کدام احتمال میانگین نمرات ارزشیابی آنان بین 16 و 13 است؟ $\left(S_0^{1.25} = 0.3944\right)$ (حسابداری ـ ۸۴)

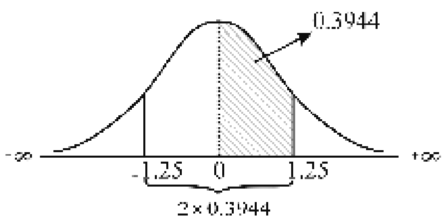
۱) 0.3944 ۲) 0.6056 ۳) 0.7888 ۴) 0.8944

حل: گزینه ۳ درست است.

با توجه به حالت (۱) توزیع $\bar{X}$ به صورت $\bar{X} \sim N\left(\mu, \frac{\sigma_X^2}{n}\right)$ است، بنابراین:

$$\begin{cases} P(13 < \bar{X} < 16) = P\left(\dfrac{13-14.5}{\frac{6}{\sqrt{25}}} < \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \dfrac{16-14.5}{\frac{6}{\sqrt{25}}}\right) = P(-1.25 < Z < 1.25) = 0.7888 \\ X \sim N(\mu = 14.5, \sigma^2 = 6^2) \ \rightarrow \ \mu = 14.5 \ , \ \sigma = 6 \ , \ n = 25 \end{cases}$$

$$S_0^{1.25} = P(0 < Z < 1.25) = 0.3944$$
$$\rightarrow P(-1.25 < Z < 1.25) = 2 \times 0.3944 = 0.7888$$



مثال ۷ در یک نمونه‌گیری تصادفی به حجم $n>1$ از جامعه‌ای نرمال با میانگین μ و واریانس σ_X^2، توزیع نمونه‌ای $(\bar{X}-\mu)$ کدام‌یک از موارد زیر است؟ (اقتصاد ـ ۸۲)

۱) $(\bar{X}-\mu)\sim\chi^2_{\alpha,(n-1)}$

۲) $(\bar{X}-\mu)\sim N\left(\mu,\sigma^2_{(X-\mu)}\right)$

۳) $(\bar{X}-\mu)\sim N\left(0,\frac{\sigma_X^2}{n}\right)$

۴) $(\bar{X}-\mu)\sim N\left(0,\frac{\sigma_X^2}{n}\right)$

حل: گزینه ۳ درست است.

با توجه به حالت (۱) توزیع $\bar{X}$ به صورت $\bar{X}\sim N\left(\mu,\frac{\sigma_X^2}{n}\right)$ است. از طرفی با توجه به بحث مربوط به توزیع نرمال هر ترکیب خطی از این توزیع مانند $\bar{X}\pm\mu$ دارای توزیع نرمال است و کافی است میانگین و واریانس آن را محاسبه کنیم.

$$\bar{X}\sim N\left(\mu,\frac{\sigma_X^2}{n}\right)\to\begin{cases}E(\bar{X})=\mu\\ Var(\bar{X})=\frac{\sigma_X^2}{n}\end{cases}\to\begin{cases}E(\bar{X}-\mu)=E(\bar{X})-\mu=\mu-\mu=0\\ Var(\bar{X}-\underbrace{\cancel{\mu}}_{0})=Var(\bar{X})=\sigma_{\bar{X}}^2=\frac{\sigma_X^2}{n}\end{cases}\to(\bar{X}-\mu)\sim N\left(0,\frac{\sigma_X^2}{n}\right)$$

توجه: در این سؤال اگر توزیع $\bar{X}+\mu$ خواسته شده بود، $(\bar{X}+\mu)\sim N\left(2\mu,\frac{\sigma_X^2}{n}\right)$ برقرار بود.

مثال ۸ بین گزاره‌های زیر کدام گزاره نادرست است؟ (اقتصاد ـ ۸۱)

۱) در نمونه‌گیری از جامعه دلخواه با میانگین μ و انحراف معیار σ توزیع $\bar{X}$ به n بستگی ندارد.

۲) در نمونه‌گیری از جامعه دلخواه با میانگین μ و انحراف معیار σ وقتی n بزرگ است توزیع $\bar{X}$ نرمال است.

۳) در نمونه‌گیری از جامعه نرمال با میانگین μ و انحراف معیار σ توزیع $\bar{X}$ نرمال است.

۴) در نمونه‌گیری از جامعه نرمال با میانگین μ و انحراف معیار σ توزیع $\bar{X}$ به حجم n بستگی ندارد.

حل: گزینه ۱ درست است.

گزینه ۱: جامعه غیرنرمال و واریانس معلوم است پس توزیع $\bar{X}$ بنا بر **حالت ۳** به حجم n وابسته است. درصورتی‌که $n>30$ باشد بنا بر **حالت (۳ ـ الف) نرمال** است و درصورتی‌که $n\le 30$ باشد بنا بر **حالت (۳ ـ ب) توزیع** $\bar{X}$ **نامعلوم** است.

گزینه ۲: بنا بر حالت (۳ ـ الف) توزیع $\bar{X}$ نرمال است.

گزینه ۳: بنا بر حالت (۱) توزیع $\bar{X}$ نرمال است.

گزینه ۴: بنا بر حالت (۱) توزیع $\bar{X}$ مستقل از حجم نمونه (n) است.

مثال ۹ از قاعده چی‌بی‌شف در کدام مورد زیر برای تخمین μ_X استفاده می‌شود؟ (مدیریت ـ ۸۳)

۱) توزیع $\bar{X}$ نرمال باشد.

۲) توزیع $\bar{X}$ نامعلوم باشد.

۳) توزیع جامعه آماری نرمال باشد.

۴) توزیع $\bar{X}$ از نوع t استودنت باشد.

حل: گزینه ۲ درست است.

بنا بر حالت (۳ـ ب) هرگاه توزیع $\bar{X}$ نامعلوم باشد از قاعده چی‌بی‌شف استفاده می‌شود.

توزیع تفاضل یا مجموع میانگین دو نمونه

اگر n_1 و n_2 دو نمونه مستقل با میانگین‌های $\bar{X}_1$ و $\bar{X}_2$ و واریانس‌های S_1^2 و S_2^2 از دو جامعه با میانگین‌های μ_1 و μ_2 و واریانس‌های σ_1^2 و σ_2^2 باشند، برای بررسی توزیع مربوط به $\bar{X}_1 \pm \bar{X}_2$ به صورت زیر عمل می‌کنیم:

۱- دو جامعه نرمال با واریانس‌های معلوم σ_1^2 و σ_2^2 ($n_1 \geq 1$ و $n_2 \geq 1$ دلخواه) یا

دو جامعه غیرنرمال با واریانس‌های معلوم σ_1^2, σ_2^2 و $(n_2 > 30, n_1 > 30)$ (قضیه حد مرکزی)

در این شرایط توزیع $\bar{X}_1 \pm \bar{X}_2$ نرمال است و داریم:

$$\begin{cases} \bar{X}_1 \pm \bar{X}_2 \sim N\left(\mu_1 \pm \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right) \\ E\left(\bar{X}_1 \pm \bar{X}_2\right) = \mu_1 \pm \mu_2 \\ \sigma^2\left(\bar{X}_1 \pm \bar{X}_2\right) = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2} \\ \sigma\left(\bar{X}_1 \pm \bar{X}_2\right) = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}} \end{cases} \longrightarrow \quad Z = \frac{\left(\bar{X}_1 \pm \bar{X}_2\right) - \left(\mu_1 \pm \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

۲- دو جامعه نرمال با واریانس‌های نامعلوم ($n_1 > 30$, $n_2 > 30$ یا نمونه‌ها بزرگ)

در این شرایط توزیع $\bar{X}_1 \pm \bar{X}_2$ با توجه به بزرگ بودن نمونه‌ها باز هم نرمال است و به‌جای σ_1^2 و σ_2^2 از S_1^2 و S_2^2 استفاده می‌کنیم:

$$\begin{cases} \bar{X}_1 \pm \bar{X}_2 \sim N\left(\mu_1 \pm \mu_2, \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right) \\ E\left(\bar{X}_1 \pm \bar{X}_2\right) = \mu_1 \pm \mu_2 \\ \sigma^2\left(\bar{X}_1 \pm \bar{X}_2\right) = \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2} \\ \sigma\left(\bar{X}_1 \pm \bar{X}_2\right) = \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}} \end{cases} \longrightarrow \quad Z = \frac{\left(\bar{X}_1 \pm \bar{X}_2\right) - \left(\mu_1 \pm \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

۳- دو جامعه نرمال با واریانس‌های نامعلوم ($n_1 \leq 30, n_2 \leq 30$ یا $n_1 + n_2 \leq 30$ یا نمونه‌ها کوچک)

در این شرایط توزیع $\bar{X}_1 \pm \bar{X}_2$ با توجه به کوچک بودن نمونه‌ها دارای توزیع t–استیودنت است که با توجه به فرض $\sigma_1^2 = \sigma_2^2$ (برابری واریانس دو جامعه) یا $\sigma_1^2 \neq \sigma_2^2$ (عدم برابری واریانس دو جامعه) دو وضعیت متفاوت به شرح زیر به وجود می‌آید:

الف) فرض برابری واریانس دو جامعه $\left(\sigma_1^2 = \sigma_2^2\right)$

در شرایطی که واریانس دو جامعه نامعلوم، نمونه‌ها کوچک و فرض برابری واریانس دو جامعه مطرح باشد، می‌توان هر دو جامعه را یک جامعه فرض کرده و از واریانس ترکیبی (آمیخته) S_p^2 برای هر دو استفاده کرد. در این وضعیت $\bar{X}_1 \pm \bar{X}_2$ با توجه به شرایط مطرح‌شده دارای توزیع t با $n_1 + n_2 - 2$ درجه آزادی خواهد بود.

$$\begin{cases} \bar{X}_1 \pm \bar{X}_2 \sim t_{n_1+n_2-2}\left(\mu_1 \pm \mu_2, \dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}\right) \\ E\left(\bar{X}_1 \pm \bar{X}_2\right) = \mu_1 \pm \mu_2 \\ \sigma^2\left(\bar{X}_1 \pm \bar{X}_2\right) = \dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2} \\ S_p^2 = \dfrac{\left(n_1 - 1\right)S_1^2 + \left(n_2 - 1\right)S_2^2}{n_1 + n_2 - 2} \\ \sigma\left(\bar{X}_1 \pm \bar{X}_2\right) = S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} \end{cases} \longrightarrow t_{n_1+n_2-2} = \frac{\left(\bar{X}_1 \pm \bar{X}_2\right) - \left(\mu_1 \pm \mu_2\right)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

اگر $n_1 = n_2 = n$ باشد داریم:

$$S_p^2 = \frac{(n-1)S_1^2 + (n-1)S_2^2}{\underset{2(n-1)}{\cancel{n+n-2}}} = \frac{S_1^2 + S_2^2}{2}$$

ب) فرض عدم برابری واریانس دو جامعه $\left(\sigma_1^2 \neq \sigma_2^2\right)$

در شرایطی که واریانس دو جامعه نامعلوم، نمونه‌ها کوچک و فرض عدم برابری واریانس دو جامعه مطرح باشد از S_1^2 و S_2^2 به‌جای σ_1^2 و σ_2^2 استفاده می‌کنیم. در این وضعیت $\bar{X}_1 \pm \bar{X}_2$ با توجه به شرایط مطرح‌شده دارای توزیع t با r درجه آزادی به صورت زیر است:

$$\begin{cases} \bar{X}_1 \pm \bar{X}_2 \sim t_r\left(\mu_1 \pm \mu_2, \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right) \\ E\left(\bar{X}_1 \pm \bar{X}_2\right) = \mu_1 \pm \mu_2 \\ \sigma^2\left(\bar{X}_1 \pm \bar{X}_2\right) = \dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2} \\ r = \dfrac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\dfrac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(S_2^2/n_2\right)^2}{n_2 - 1}} \end{cases} \longrightarrow t_r = \frac{\left(\bar{X}_1 \pm \bar{X}_2\right) - \left(\mu_1 \pm \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

توجه:

۱- در شرایطی که در صورت مسئله واریانس‌های σ_1^2 و σ_2^2 مطرح نشوند آن‌ها را به طور پیش‌فرض نامعلوم در نظر می‌گیریم.

۲- در شرایطی که در صورت مسئله در مورد واریانس‌ها صحبتی نشود به طور پیش‌فرض واریانس دو جامعه را نامعلوم و نابرابر در نظر می‌گیریم $\left(\sigma_1^2 \neq \sigma_2^2\right)$.

۳- در شرایطی که در صورت مسئله n_1 و n_2 ذکر نشود آن‌ها را به طور پیش‌فرض کوچک و $n_1 \leq 30$ و $n_2 \leq 30$ در نظر می‌گیریم.

۴- در شرایطی که توزیع جامعه در صورت مسئله قید نشود، جامعه را نرمال در نظر می‌گیریم.

مثال ۱ توزیع مجموع دو متغیر نرمال مستقل کدام است؟

۱) نرمال ۲) کای ـ مربع ۳) فیشر ۴) t−استیودنت

حل: گزینه ۱ درست است.

مثال ۲ دو متغیر تصادفی مستقل X و Y دارای توزیع نرمال با میانگین‌های یکسان و واریانس‌های به ترتیب برابر با 8 و 4 می‌باشند. بر اساس دو نمونه تصادفی به اندازه‌های 16 از جامعه (متغیر) X و 8 از جامعه (متغیر) Y برآوردکننده‌های میانگین دو جامعه به ترتیب $\bar{X}$ و $\bar{Y}$ به دست آمده، توزیع $\left(\bar{X}-\bar{Y}\right)$ کدام است؟ (اقتصاد ـ ۸۳)

۱) $\left(\bar{X}-\bar{Y}\right)\sim N(0,12)$ ۲) $\left(\bar{X}-\bar{Y}\right)\sim N(0,1)$

۳) $\left(\bar{X}-\bar{Y}\right)\sim N(0,4)$ ۴) $\left(\bar{X}-\bar{Y}\right)\sim N(0,8)$

حل: گزینه ۲ درست است.

با توجه به حالت (۱) توزیع $\bar{X}-\bar{Y}$ نرمال است با:

$$\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right) \rightarrow \begin{cases} E\left(\bar{X}-\bar{Y}\right)=\mu_1-\mu_2 \xrightarrow{\mu_1=\mu_2} 0 \\ Var\left(\bar{X}-\bar{Y}\right)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}=\frac{8}{16}+\frac{4}{8}=1 \end{cases} \rightarrow \left(\bar{X}-\bar{Y}\right)\sim N(0,1)$$

مثال ۳ واریانس توزیع نمونه‌ای تفاوت میانگین‌ها در نمونه‌های کوچک کدام‌یک از موارد زیر است؟ (واریانس جامعه‌های اصلی معلوم نیست اما می دانیم برابرند). (اقتصاد ـ ۷۴)

۱) $\sigma^2_{\bar{X}_1-\bar{X}_2}=\frac{S_1^2+S_2^2}{n_1+n_2}$

۲) $\sigma^2_{\bar{X}_1-\bar{X}_2}=\frac{nS_1^2+nS_2^2}{n_1+n_2-2}$

۳) $\sigma^2_{\bar{X}_1-\bar{X}_2}=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$

۴) $\sigma^2_{\bar{X}_1-\bar{X}_2}=\frac{nS_1^2}{n_1}+\frac{nS_2^2}{n_2}$

حل: گزینه ۳ درست است.

با توجه به حالت (۳ ـ الف) توزیع $\bar{X}_1-\bar{X}_2$ عبارت است از $t_{(n_1+n_2-2)}$ با:

$$E\left(\bar{X}_1-\bar{X}_2\right)=\mu_1-\mu_2$$

$$\begin{cases} \sigma^2\left(\bar{X}_1-\bar{X}_2\right)=\dfrac{S_p^2}{n_1}+\dfrac{S_p^2}{n_2} \\ S_p^2=\dfrac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{n_1+n_2-2} \end{cases} \rightarrow \quad \sigma^2\left(\bar{X}_1-\bar{X}_2\right)=\frac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

مثال ۴ اگر یک نمونه 100تایی از جامعه اول با واریانس 9 و یک نمونه 25 تایی از جامعه دوم با واریانس 4 انتخاب شوند و این دو نمونه مستقل از یکدیگر باشند، انحراف معیار تفاضل میانگین دو نمونه کدام است؟ (حسابداری و مدیریت ـ ۸۶)

۱) 0.25 ۲) 0.5 ۳) 1.25 ۴) 1.5

حل: گزینه ۲ درست است.

با توجه به حالت (۱) (قضیه حد مرکزی) توزیع $\bar{X}_1 \pm \bar{X}_2$ نرمال است:

$$\bar{X}_1 \pm \bar{X}_2 \sim N\left(\mu_1 \pm \mu_2\,,\ \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

$$\sigma^2\left(\bar{X}_1-\bar{X}_2\right)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}=\frac{9}{100}+\frac{4}{25}=0.25 \rightarrow$$ انحراف معیار تفاضل میانگین دو نمونه : $\sigma\left(\bar{X}_1-\bar{X}_2\right)=0.5$

نسبت (نرخ) موفقیت نمونه

در مواردی مانند محاسبه درصد کالاهای معیوب، درصد افراد بی‌سواد، درصد افراد بیکار، نسبت مدیران وظیفه مدار و ... در یک جامعه، به دنبال آن هستیم که نسبت عناصری از جامعه را تخمین بزنیم که دارای ویژگی مشترکی (معیوب بودن، بی‌سواد بودن،...) هستند.

درعین‌حال چون صفت مورد بررسی یک صفت کیفی است، بسیاری از این تحقیق‌ها از مقیاس اسمی یا رتبه‌ای برخوردار هستند که پارامتر توصیف آن‌ها نسبت موفقیت است.

تعریف: در یک آزمایش دوجمله‌ای با نسبت موفقیت p درصورتی‌که با n بار تکرار آزمایش به x موفقیت رسیده باشیم، آن‌گاه:

نسبت (نرخ) موفقیت در نمونه : $\bar{p}=\dfrac{x}{n}$

در این صورت $\bar{p}$ (نرخ نمونه) آماره‌ای خواهد بود که از طریق بررسی توزیع مربوط به آن می‌توانیم به تخمین پارامتر $\left(p=\dfrac{X}{N}\right)$ نسبت واقعی در جامعه بپردازیم:

میانگین نسبت نمونه : $E\left(\bar{p}\right)=E\left(\dfrac{X}{n}\right)=\dfrac{E\left(X\right)}{n}=\dfrac{np}{n}=p$

واریانس نسبت نمونه : $\sigma_{\bar{p}}^2=\sigma^2\left(\dfrac{X}{n}\right)=\dfrac{\sigma_X^2}{n^2}=\dfrac{npq}{n^2}=\dfrac{pq}{n}$

توجه: توزیع X تعداد موفقیت در نمونه دوجمله‌ای است با $E\left(X\right)=np\ ,\ \sigma_X^2=npq$.

میانگین نسبت نمونه $\left(\mu_{\bar{p}}\right)$ برابر با نسبت جامعه (p) است.	$\mu_{\bar{p}} = E(\bar{p}) = p$
واریانس نسبت نمونه $\left(\sigma_{\bar{p}}^2\right)$ برابر با $\frac{pq}{n}$ است.	$\sigma_{\bar{p}}^2 = \frac{pq}{n}$
انحراف معیار نسبت نمونه $\left(\sigma_{\bar{p}}\right)$ برابر با $\sqrt{\frac{pq}{n}}$ است.	$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$

مثال ۱ اگر $\bar{p}$ برآوردگر p باشد، آن‌گاه $\sigma_{\bar{p}}$ برابر است با:

۱) $\sqrt{n\bar{p}(1-\bar{p})}$ ۲) $\sqrt{\frac{p(1-p)}{n}}$ ۳) $\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ ۴) $\sqrt{np(1-p)}$

حل: گزینه ۲ درست است.

با توجه به آنکه $p+q=1$ برقرار است:

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

مثال ۲ 0.50 افراد یک شهر بی‌سواد هستند. درصورتی‌که در یک نمونه 100 عضوی از این شهر 40 نفر بی‌سواد باشند، نرخ بی‌سوادی و انحراف معیار نرخ بی‌سوادی کدام است؟

۱) 0.05 , 0.4 ۲) 0.04 , 0.5 ۳) 0.03 , 0.4 ۴) هیچ‌کدام

حل: گزینه ۱ درست است.

$$\begin{cases} \text{نرخ بی‌سوادی} = \bar{p} = \frac{x}{n} = \frac{40}{100} = 0.4 \\ \text{انحراف معیار نرخ بی‌سوادی} = \sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.5 \times 0.5}{100}} = 0.05 \\ n = 100 \;,\; x = 40 \;,\; p = 0.5 \end{cases}$$

نسبت جامعه مجهول

هنگام محاسبه $\sigma_{\bar{p}}^2 = \frac{pq}{n}$ یا $\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$ هرگاه مقدار p (نسبت در جامعه) موجود نباشد، از $\bar{p} = \frac{x}{n}$(نرخ نمونه) استفاده می‌کنیم و داریم:

$$\begin{cases} \sigma_{\bar{p}}^2 = \frac{pq}{n} \\ \sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} \end{cases} \xrightarrow{\text{p (نسبت جامعه) مجهول}} \begin{cases} S_{\bar{p}}^2 = \frac{\bar{p}\bar{q}}{n} \\ S_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}} \end{cases}$$

مثال ۳ در بررسی درصد بیکاری در یک نمونه 100 نفری از جامعه‌ای، مشخص شد که 20 نفر بیکار هستند. نرخ بیکاری و انحراف معیار نرخ بیکاری کدام است؟

۱) 0.04 , 0.2 ۲) 0.02 , 0.8 ۳) 0.16 , 0.2 ۴) 0.08 , 0.2

حل: گزینه ۱ درست است.

$$\begin{cases} \text{نرخ بیکاری} = \bar{p} = \frac{x}{n} = \frac{20}{100} = 0.2 \\ \text{انحراف معیار نرخ بیکاری } S_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{0.2 \times 0.8}{100}} = 0.04 \\ n = 100 \;,\; x = 20 \end{cases}$$

همان‌طورکه دیده می‌شود به علت مجهول بودن p (نرخ بیکاری در جامعه) از $\bar{p}$ (نرخ بیکاری در نمونه) برای محاسبه $\sigma_{\bar{p}}$ استفاده شده است.

مثال ۴ برای تخمین نسبت موفقیت‌ها در جامعه‌ای، دو نمونه تصادفی مستقل به حجم‌های n_1 و n_2 از جامعه گرفته و برآوردکننده $\hat{p}=\frac{x_1+x_2}{n_1+n_2}$ پیشنهاد شده است که در آن x_1 و x_2 تعداد موفقیت‌ها در نمونه اول و دوم است. کمیت انتظاری (امید ریاضی) این برآوردکننده کدام است؟ (اقتصاد ـ ۸۵)

۱) p ۲) $\frac{2\mu_X}{n_1+n_2}$ ۳) $\frac{p_1+p_2}{n_1+n_2}$ ۴) $\frac{2p}{n_1+n_2}$

حل: گزینه ۱ درست است.

هرگاه در جامعه‌ای دوجمله‌ای به دنبال نسبت یا درصد موفقیت باشیم، روابط زیر همیشه برقرار است:

$$\bar{p}_i=\frac{x_i}{n_i}\rightarrow\begin{cases}E(\bar{p}_i)=p\\E(X_i)=n_ip\end{cases}$$

$$\hat{p}=\frac{x_1+x_2}{n_1+n_2}\rightarrow E(\hat{p})=E\left(\frac{X_1+X_2}{n_1+n_2}\right)=\frac{E(X_1)+E(X_2)}{n_1+n_2}=\frac{n_1p+n_2p}{n_1+n_2}=\frac{(n_1+n_2)p}{n_1+n_2}=p$$

مثال ۵ برای تخمین نسبت موفقیت‌ها در جامعه‌ای، دو نمونه تصادفی مستقل به حجم‌های n_1 و n_2 از جامعه گرفته و برآوردکننده $\hat{p}=\frac{\bar{p}_1+\bar{p}_2}{n_1+n_2}$ پیشنهاد شده است که در آن x_1 و x_2 تعداد موفقیت‌ها در نمونه اول و دوم است. کمیت انتظاری (امید ریاضی) این برآوردکننده کدام است؟

۱) p ۲) $\frac{2\mu_x}{n_1+n_2}$ ۳) $\frac{p_1+p_2}{n_1+n_2}$ ۴) $\frac{2p}{n_1+n_2}$

حل: گزینه ۴ درست است.

هرگاه در جامعه‌ای دوجمله‌ای به دنبال نسبت یا درصد موفقیت هستیم روابط زیر همیشه برقرار است:

$$\bar{p}_i=\frac{x_i}{n_i}\rightarrow\begin{cases}E(\bar{p}_i)=p\\E(X_i)=n_ip\end{cases}$$

$$\hat{p}=\frac{\bar{p}_1+\bar{p}_2}{n_1+n_2}\rightarrow E(\hat{p})=E\left(\frac{\bar{p}_1+\bar{p}_2}{n_1+n_2}\right)=\frac{E(\bar{p}_1)+E(\bar{p}_2)}{n_1+n_2}=\frac{p+p}{n_1+n_2}=\frac{2p}{n_1+n_2}$$

جامعه محدود و بدون جایگذاری

درصورتی‌که همه شرایط زیر برقرار باشد:

۱- جامعه محدود باشد (N)،

۲- نمونه‌گیری بدون جایگذاری انجام شود،

۳- نسبت تعداد نمونه (n) به تعداد جامعه (N) بیشتر از 0.05 باشد $\left(\frac{n}{N}>0.05\right)$، برای محاسبه $\sigma_{\bar{p}}^2$ (واریانس نسبت نمونه) باید از ضریب تصحیح (کاهش) $\frac{N-n}{N-1}$ استفاده کنیم. در این شرایط $E(\bar{p})=p$ است و از این نظر تفاوتی با جامعه نامحدود ندارد.

$$\begin{cases} \sigma_{\bar{p}}^2 = \dfrac{N-n}{N-1} \cdot \dfrac{pq}{n} \\ \sigma_{\bar{p}} = \sqrt{\dfrac{N-n}{N-1}} \cdot \sqrt{\dfrac{pq}{n}} \\ E(\bar{p}) = p \end{cases}$$

یادآوری: $\dfrac{n}{N} > 0.05 \rightarrow n > \%5N$ یا $N < 20n$

✓ **دقت کنید!**

کمیت $\dfrac{N-n}{N-1}$ یک عامل اصلاح‌کننده برای واریانس نسبت نمونه $\left(\sigma_{\bar{p}}^2\right)$ در جامعه محدود (N) با شرایط بالا است و در اکثر اوقات یعنی بروز یکی از حالات زیر به آن توجهی نمی‌شود و آن را در محاسبات به کار نمی‌برند.

شرایط چشم‌پوشی از $\dfrac{N-n}{N-1}$

در صورت بروز یکی از حالات زیر از $\dfrac{N-n}{N-1}$ برای تصحیح $\sigma_{\bar{p}}^2$ استفاده نکرده و از آن چشم‌پوشی می‌کنند:

۱- جامعه نامحدود باشد.

۲- نمونه‌گیری با جایگذاری باشد.

۳- نسبت نمونه (n) به جامعه (N) حداکثر 0.05 باشد $\left(\dfrac{n}{N} \le 0.05\right)$.

درواقع در شرایط بالا، کمیت $\dfrac{N-n}{N-1}$ برابر با 1 در نظر گرفته می‌شود.

توزیع نسبت (نرخ) موفقیت نمونه

با فرض آنکه حجم نمونه به اندازه کافی بزرگ باشد $(n > 30)$، بنا بر قضیه حد مرکزی آماره $\bar{p}$ (نرخ نمونه) توزیع نرمال دارد و به صورت زیر بررسی می‌شود:

تعداد نمونه	توزیع $\bar{p}$	میانگین $\bar{p}$	واریانس $\bar{p}$	انحراف معیار $\bar{p}$	متغیر استاندارد
$n > 30$ (نمونه بزرگ)	نرمال	$\mu_{\bar{p}} = E(\bar{p}) = p$	$\sigma_{\bar{p}}^2 = \dfrac{pq}{n}$	$\sigma_{\bar{p}} = \sqrt{\dfrac{pq}{n}}$	$Z = \dfrac{\bar{p}-p}{\sqrt{\dfrac{pq}{n}}}$
$n \le 30$ (نمونه کوچک)	دوجمله‌ای	$\mu_{\bar{p}} = E(\bar{p}) = p$	$\sigma_{\bar{p}}^2 = \dfrac{pq}{n}$	$\sigma_{\bar{p}} = \sqrt{\dfrac{pq}{n}}$	ندارد

مثال ۱ رفتار توزیع نمونه‌گیری $\bar{p}$ در یک نمونه تصادفی 200 تایی دارای کدام توزیع پیوسته‌ای است؟

۱) یکنواخت ۲) نمایی ۳) نرمال ۴) هیچ‌کدام

حل: گزینه ۳ درست است.

در شرایطی که تعداد نمونه بزرگ $(n > 30)$ باشد رفتار توزیع $\bar{p}$ نرمال است.

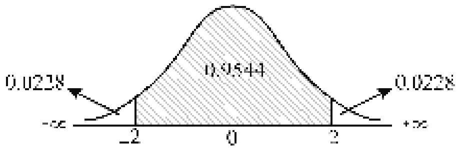
مثال ۲ 50 درصد کارمندان سازمان مرد هستند. احتمال آنکه حداقل 60 درصد افراد یک نمونه تصادفی 100 نفره مرد باشد چقدر است؟

۱) 0.9772 ۲) 0.0228 ۳) 0.4772 ۴) 0.683

حل: گزینه ۲ درست است.

$$\begin{cases} P(\bar{p} \geq 0.6) = P\left(\dfrac{\bar{p}-p}{\sqrt{\dfrac{pq}{n}}} \geq \dfrac{0.6-0.5}{\sqrt{\dfrac{0.5\times 0.5}{100}}}\right) = P(Z \geq 2) = 0.0228 \\ Z = \dfrac{\bar{p}-p}{\sqrt{\dfrac{pq}{n}}} \\ \bar{p} = 0.6 \ , \ p = 0.5 \ , \ n = 100 \end{cases}$$

یادآوری: در توزیع نرمال استاندارد همان‌طورکه می‌دانیم:



$$P(Z < -2) = 0.0228 \ , \ P(-2 < Z < 2) = 0.9544 \ , \ P(Z > 2) = 0.0228$$

توزیع تفاضل یا مجموع نسبت دو نمونه

اگر برای مقایسه دو جامعه، معیاری کمّی مورد نظر باشد از تخمین $\bar{X}_1 \pm \bar{X}_2$ استفاده می‌شود ولی درصورتی‌که معیار مقایسه دو جامعه یک صفت کیفی باشد از مقایسه نسبت موفقیت دو جامعه آماری استفاده می‌شود.

پیش‌فرض دو جامعه را دوجمله‌ای به صورت $\text{Bin}(n_1,p_1), \text{Bin}(n_2,p_2)$ در نظر می‌گیریم.

اگر x_1 تعداد افراد (اشیا) دارای صفت کیفی مورد نظر در نمونه گرفته‌شده از جامعه اول باشد نسبت موفقیت در جامعه اول برابر است با: $\bar{p}_1 = \dfrac{x_1}{n_1}$

$$\bar{\mathbf{p}}_1 = \frac{\mathbf{x_1}}{\mathbf{n_1}} \xrightarrow{E(X_1)=n_1p_1} \begin{cases} \mathbf{E}(\bar{\mathbf{p}}_1) = E\left(\dfrac{X_1}{n_1}\right) = \dfrac{E(X_1)}{n_1} = \dfrac{n_1p_1}{n_1} = \mathbf{p_1} \\ \mathbf{Var}(\bar{\mathbf{p}}_1) = \text{Var}\left(\dfrac{X_1}{n_1}\right) = \dfrac{\text{Var}(X_1)}{n_1^2} = \dfrac{n_1p_1q_1}{n_1^2} = \dfrac{\mathbf{p_1q_1}}{\mathbf{n_1}} \end{cases}$$

و اگر x_2 تعداد افراد (اشیا) دارای صفت کیفی مورد نظر در نمونه گرفته‌شده از جامعه دوم باشد نسبت موفقیت در جامعه دوم برابر است با: $\bar{p}_2 = \dfrac{x_2}{n_2}$

$$\bar{\mathbf{p}}_2 = \frac{\mathbf{x_2}}{\mathbf{n_2}} \xrightarrow{E(X_2)=n_2p_2} \begin{cases} \mathbf{E}(\bar{\mathbf{p}}_2) = E\left(\dfrac{X_2}{n_2}\right) = \dfrac{E(X_2)}{n_2} = \dfrac{n_2p_2}{n_2} = \mathbf{p_2} \\ \mathbf{Var}(\bar{\mathbf{p}}_2) = \text{Var}\left(\dfrac{X_2}{n_2}\right) = \dfrac{\text{Var}(X_2)}{n_2^2} = \dfrac{n_2p_2q_2}{n_2^2} = \dfrac{\mathbf{p_2q_2}}{\mathbf{n_2}} \end{cases}$$

در این شرایط توزیع $\bar{p}_1 \pm \bar{p}_2$ عبارت است از:

توزیع تفاضل یا مجموع نسبت دو نمونه کوچک	دوجمله‌ای
توزیع تفاضل یا مجموع نسبت دو نمونه بزرگ	نرمال
میانگین مجموع یا تفاضل نسبت دو نمونه	$E\left(\bar{p}_1 \pm \bar{p}_2\right) = p_1 \pm p_2$
واریانس مجموع یا تفاضل نسبت دو نمونه	$Var\left(\bar{p}_1 \pm \bar{p}_2\right) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$
انحراف معیار مجموع یا تفاضل نسبت دو نمونه	$\sqrt{Var\left(\bar{p}_1 \pm \bar{p}_2\right)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$
متغیر استاندارد مجموع یا تفاضل نسبت دو نمونه $Z = \frac{X-\mu}{\sigma}$	$Z = \frac{\left(\bar{p}_1 \pm \bar{p}_2\right) - \left(p_1 \pm p_2\right)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

توجه کنید که در این مقایسه فرض بر بزرگ بودن نمونه‌ها و مستقل بودن دو نمونه از یکدیگر است.

توزیع واریانس نمونه

اگر $X_1, X_2, \ldots, X_n$ نمونه‌ای مستقل از یک جامعه نرمال با میانگین μ و واریانس σ^2 باشد، آن‌گاه واریانس نمونه S^2 با یکی از دو روش زیر قابل محاسبه است:

$$S^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}$$: میانگین جامعه (μ) نامعلوم

$$S^2 = \frac{\sum\left(x_i - \mu\right)^2}{n}$$: میانگین جامعه (μ) معلوم

✓ دقت کنید!

با آنکه به طور پیش‌فرض میانگین جامعه (μ) نامعلوم در نظر گرفته می‌شود، اما در صورت معلوم بودن میانگین جامعه (μ)، مقدار دقیق‌تری برای S^2 به دست خواهد آمد.

با آنکه به دست آوردن توزیع S^2 بسیار مشکل است اما تحت شرایط خاصی می‌توان توزیع S^2 را به صورت زیر تعیین کرد:

تعریف ۱: (پیش‌فرض) درصورتی‌که S^2 واریانس یک نمونه n تایی از جامعه‌ای نرمال با میانگین نامعلوم (مجهول) (μ) باشد، آن‌گاه:

$$S^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}$$

برای بررسی توزیع مربوط به آماره S^2 از آماره $\frac{(n-1)S^2}{\sigma^2}$ **با توزیع** $\chi^2_{(n-1)}$ **(کای مربع) با $n-1$ درجه آزادی** به صورت زیر استفاده می‌شود:

$$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma^2} \xrightarrow{S^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}} \chi^2_{(n-1)} = \frac{\sum\left(x_i - \bar{x}\right)^2}{\sigma^2}$$

مثال ۱ در یک جامعه نرمال احتمال اینکه واریانس یک نمونه 81 تایی کمتر از واریانس جامعه باشد، کدام است؟ (اقتصاد ـ ۸۲)

۱) $P\left(\chi^2_{(80)} \le 81\right)$ ۲) $P\left(F_{1,80} \le 5\right)$ ۳) $P\left(F_{1,80} \le 6\right)$ ۴) $P\left(\chi^2_{(80)} \le 80\right)$

حل: گزینه ۴ درست است.

$$\begin{cases} P\left(S^2 < \sigma^2\right) = P\left(\dfrac{(n-1)S^2}{\sigma^2} < \dfrac{(n-1)\sigma^2}{\sigma^2}\right) = P\left(\chi^2_{n-1} < n-1\right) = P\left(\chi^2_{80} < 80\right) \\ n = 81 \end{cases}$$

مثال ۲ اگر $n = 10$ و $S^2_X = 80$ و $\sigma^2_X = 65$ باشد، مقدار استاندارد کای مربع کدام است؟

۱) 11.08 ۲) 1.23 ۳) 15.32 ۴) 8.125

حل: گزینه ۱ درست است.

$$\begin{cases} \chi^2_{n-1} = \dfrac{(n-1)S^2}{\sigma^2} \rightarrow \chi^2_9 = \dfrac{9 \times 80}{65} = 11.08 \\ n = 10 \end{cases}$$

مثال ۳ وقتی نمونه‌گیری از یک جامعه نرمال انجام شود، متغیر تصادفی $\dfrac{(n-1)S^2}{\sigma^2}$ دارای توزیع χ^2 با چند درجه آزادی است؟

(اقتصاد ـ ۷۰)

۱) $n-1$ ۲) n ۳) $n+1$ ۴) $n-2$

حل: گزینه ۱ درست است.

تعریف ۲: درصورتی‌که میانگین جامعه (μ) معلوم باشد، آن‌گاه:

$$S^2 = \frac{\sum (x_i - \mu)^2}{n}$$

در این‌صورت برای بررسی توزیع مربوط به آماره S^2 از آماره $\dfrac{nS^2}{\sigma^2}$ **با توزیع** $\chi^2_{(n)}$ **(کای مربع) با** n **درجه آزادی** به شرح زیر استفاده می‌شود:

$$\chi^2_{(n)} = \frac{nS^2}{\sigma^2} \xrightarrow{S^2 = \frac{\sum(x_i-\mu)^2}{n}} \chi^2_{(n)} = \frac{\sum(x_i-\mu)^2}{\sigma^2}$$

توجه: پیش‌فرض میانگین جامعه (μ) نامعلوم است مگر آنکه در صورت سؤال مطرح شود.

امید و واریانس آماره S^2

یادآوری: در هر توزیع $\chi^2_{(v)}$ با v درجه آزادی همواره داریم:

$$E(\chi^2_{(v)}) = v$$

$$\sigma^2(\chi^2_{(v)}) = 2v$$

با توجه به تعاریف ۱ و ۲ برای توزیع آماره S^2 محاسبه امید و واریانس آماره در دو حالت قابل محاسبه است.

۱- در صورتی‌که به طور پیش‌فرض میانگین جامعه (μ) نامعلوم در نظر گرفته شود:

$$\boxed{S^2=\frac{\sum(x_i-\bar{x})^2}{n-1} \longrightarrow \begin{cases} E\left(S^2\right)=\sigma^2 \\ Var\left(S^2\right)=\dfrac{2\sigma^4}{n-1}\end{cases}}$$

اثبات:

$$\begin{cases} \mathbf{E}\left(\chi^2_{(n-1)}\right)=\mathbf{n-1} \;\to\; E\left(\dfrac{(n-1)S^2}{\sigma^2}\right)=n-1 \;\to\; \dfrac{n-1}{\sigma^2}E\left(S^2\right)=n-1 \to \mathbf{E}\left(\mathbf{S^2}\right)=\boldsymbol{\sigma^2} \\ \chi^2_{(n-1)}=\dfrac{(n-1)S^2}{\sigma^2}\end{cases}$$

$$\begin{cases} \mathbf{Var}\left(\chi^2_{(n-1)}\right)=\mathbf{2(n-1)} \;\to\; Var\left(\dfrac{(n-1)S^2}{\sigma^2}\right)=2(n-1) \;\to\; \dfrac{(n-1)^2}{\sigma^4}Var(S^2)=2(n-1) \;\to\; \mathbf{Var(S^2)}=\dfrac{\mathbf{2\sigma^4}}{\mathbf{n-1}} \\ \chi^2_{(n-1)}=\dfrac{(n-1)S^2}{\sigma^2}\end{cases}$$

۲- درصورتی‌که میانگین جامعه (μ) معلوم باشد:

$$\boxed{S^2=\frac{\sum(x_i-\mu)^2}{n} \longrightarrow \begin{cases} E\left(S^2\right)=\sigma^2 \\ Var\left(S^2\right)=\dfrac{2\sigma^4}{n}\end{cases}}$$

اثبات:

$$\begin{cases} \mathbf{E}\left(\chi^2_{(n)}\right)=\mathbf{n} \;\to\; E\left(\dfrac{nS^2}{\sigma^2}\right)=n \;\to\; \dfrac{n}{\sigma^2}E\left(S^2\right)=n \to \mathbf{E}\left(\mathbf{S^2}\right)=\boldsymbol{\sigma^2} \\ \chi^2_{(n)}=\dfrac{nS^2}{\sigma^2}\end{cases}$$

$$\begin{cases} \mathbf{Var}\left(\chi^2_{(n)}\right)=\mathbf{2n} \;\to\; Var\left(\dfrac{nS^2}{\sigma^2}\right)=2n \;\to\; \dfrac{n^2}{\sigma^4}Var(S^2)=2n \;\to\; \mathbf{Var(S^2)}=\dfrac{\mathbf{2\sigma^4}}{\mathbf{n}} \\ \chi^2_{(n)}=\dfrac{nS^2}{\sigma^2}\end{cases}$$

مثال ۴ متغیر تصادفی X بر طبق قانون نرمال با امید ریاضی μ و واریانس σ^2 توزیع شده است. در نمونه‌ای به حجم n تابع نمونه‌ای **(Statistic)** به صورت $\theta=f\left(x_1,x_2,...,x_n\right)=S^2=\dfrac{\sum(x_i-\bar{x})^2}{n-1}$ تعریف می‌شود. واریانس S^2 کدام است؟ (اقتصاد ـ ۷۵)

۱) $Var\left(S^2\right)=\dfrac{2\sigma^4}{n-1}$ ۲) $Var\left(S^2\right)=\dfrac{2n\sigma^4}{(n-1)^2}$ ۳) $Var\left(S^2\right)=\dfrac{2\sigma^4}{n^2}$ ۴) $Var\left(S^2\right)=\dfrac{2\sigma^4}{(n-1)^2}$

حل: گزینه ۱ درست است.

مثال ۵ متغیر تصادفی X بر طبق قانون نرمال با امید ریاضی μ و واریانس σ^2 توزیع شده است. در نمونه‌ای به حجم n تابع نمونه‌ای **(Statistic)** به صورت $\theta=f(x_1,x_2,\dots,x_n)=S^2=\frac{\sum(x_i-\bar{x})^2}{n}$ تعریف می‌شود. امید و واریانس S^2 به ترتیب کدام است؟ (اقتصاد ـ ۷۵)

۱) $\sigma^2\ ,\ \frac{2\sigma^4}{n-1}$

۲) $\frac{n-1}{n}\sigma^2\ ,\ \frac{2(n-1)\sigma^4}{n^2}$

۳) $\frac{n}{n-1}\sigma^2\ ,\ \frac{2\sigma^4}{n^2}$

۴) $\frac{n-1}{n}\sigma^2\ ,\ \frac{2n\sigma^4}{(n-1)^2}$

حل: گزینه ۲ درست است.

$$\mathrm{Var}(S^2)=\mathrm{Var}\left(\frac{\sum(x_i-\bar{x})^2}{n}\right)=\mathrm{Var}\left(\frac{n-1}{n-1}\frac{\sum(x_i-\bar{x})^2}{n}\right)$$
$$=\frac{(n-1)^2}{n^2}\mathrm{Var}\left(\frac{\sum(x_i-\bar{x})^2}{n-1}\right)=\frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1}=\frac{2(n-1)\sigma^4}{n^2}$$
$$E(S^2)=E\left(\frac{\sum(x_i-\bar{x})^2}{n}\right)=E\left(\frac{n-1}{n-1}\frac{\sum(x_i-\bar{x})^2}{n}\right)=\frac{n-1}{n}E\left(\frac{\sum(x_i-\bar{x})^2}{n-1}\right)=\frac{n-1}{n}\sigma^2$$

S^2	آماره	توضیحات	توزیع	امید	واریانس
$S^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$	$\frac{(n-1)S^2}{\sigma^2}$	میانگین نامعلوم	$\chi^2_{(n-1)}$	$E(S^2)=\sigma^2$	$\mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1}$
$S^2=\frac{\sum(x_i-\mu)^2}{n}$	$\frac{nS^2}{\sigma^2}$	میانگین معلوم	$\chi^2_{(n)}$	$E(S^2)=\sigma^2$	$\mathrm{Var}(S^2)=\frac{2\sigma^4}{n}$

توزیع نسبت واریانس‌های دو نمونه

فرض کنید S_1^2 و S_2^2 واریانس‌های دو نمونه مستقل با حجم n_1 و n_2 از دو جامعه نرمال با واریانس‌های نامعلوم σ_1^2 و σ_2^2 و میانگین‌های نامعلوم μ_1 و μ_2 باشند. ازآنجاکه آماره S_1^2/S_2^2 توزیع مشخصی را نشان نمی‌دهد، با در نظر گرفتن روابط زیر توزیع مربوط به آماره $\frac{(S_1^2/\sigma_1^2)}{(S_2^2/\sigma_2^2)}$ در شرایطی که واریانس دو جامعه برابر باشند $(\sigma_1^2=\sigma_2^2)$ همان توزیع آماره S_1^2/S_2^2 است.

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{\frac{1}{(n_1-1)}\times\frac{(n_1-1)S_1^2}{\sigma_1^2}}{\frac{1}{(n_2-1)}\times\frac{(n_2-1)S_2^2}{\sigma_2^2}}\xrightarrow[\chi^2_{(n_2-1)}=\frac{(n_2-1)S_2^2}{\sigma_2^2}]{\chi^2_{(n_1-1)}=\frac{(n_1-1)S_1^2}{\sigma_1^2}}\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{\frac{\chi^2_{(n_1-1)}}{(n_1-1)}}{\frac{\chi^2_{(n_2-1)}}{(n_2-1)}}=F_{n_1-1,n_2-1}$$

به عبارت دیگر:

$$\frac{\left(S_1^2/\sigma_1^2\right)}{\left(S_2^2/\sigma_2^2\right)} \sim F_{n_1-1,n_2-1} \xrightarrow{\sigma_1^2=\sigma_2^2} \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$$

تعریف: آماره $\frac{\left(S_1^2/\sigma_1^2\right)}{\left(S_2^2/\sigma_2^2\right)}$ دارای توزیع F (فیشر) با $\left(n_1-1\right)$ درجه آزادی برای صورت و $\left(n_2-1\right)$ درجه آزادی برای مخرج کسر است.

میانگین جوامع معلوم

درصورتی‌که میانگین دو جامعه $\left(\mu_1,\mu_2\right)$ معلوم باشد برای محاسبه واریانس نمونه $\left(S^2\right)$ به‌جای رابطه $S^2=\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}$ باید از رابطه $S^2=\frac{\sum\left(x_i-\mu\right)^2}{n}$ که دقیق‌تر است، استفاده کنیم. در این صورت توزیع آماره به صورت زیر است:

$$\frac{\left(S_1^2/\sigma_1^2\right)}{\left(S_2^2/\sigma_2^2\right)} \sim F_{n_1,n_2} \xrightarrow{\sigma_1^2=\sigma_2^2} \frac{S_1^2}{S_2^2} \sim F_{n_1,n_2}$$

پیش‌فرض، میانگین جامعه (μ) نامعلوم و واریانس دو جامعه برابر است.

مثال برای توزیع F کدام تعریف را می‌پذیرید؟ (اقتصاد ـ ۷۴)

۱) توزیع نسبت واریانس دو جامعه اصلی است.

۲) توزیع نسبت واریانس نمونه به واریانس جامعه اصلی است.

۳) توزیع نسبت واریانس جامعه اصلی به واریانس نمونه با درجه آزادی معلوم است.

۴) توزیع نسبت واریانس دو نمونه است.

حل: گزینه ۴ درست است.

استنباط آماری

به فرآیندی که در آن از نتایج حاصل از نمونه برای «**برآورد پارامتری از جامعه**» یا «**آزمودن فرضی درباره یکی از پارامترهای جامعه**» استفاده می‌شود، **استنباط آماری** گویند.
امروزه مسئله برآورد و آزمون، کاربردهای فراوانی در زمینه‌های اقتصادی، بازرگانی و ... دارد. برای مثال ممکن است در یک مطالعه اقتصادی تمایل داشته باشیم متوسط درآمد خانوارهای ساکن در یک شهر را بدانیم. برای این کار می‌توان نمونه‌ای از خانوارهای شهر را انتخاب و متوسط درآمد آن‌ها را محاسبه کرد، سپس این مقدار را برآوردی از درآمد متوسط تمام خانوارهای شهر در نظر گرفت؛ همچنین اگر حدس یا ادعایی درباره این پارامتر از جامعه وجود داشته باشد می‌توانیم با استفاده از این برآورد، آن را آزمون کنیم.
در ادامه، مبحث «برآورد» و در فصل بعد «آزمون فرض»، برای استنباط آماری درباره پارامترهای جامعه مورد تحلیل و بررسی قرار می‌گیرند.

برآورد (تخمین) (Estimation)

هرگاه برای یافتن پارامتر مجهول جامعه به کمک نتایج حاصل از نمونه به مقدار یا محدوده‌ای برسیم، می‌گوییم برآوردی از پارامتر را به دست آورده‌ایم.
در این فصل توضیح خواهیم داد که چگونه می‌توان پارامترهای مورد نظر جامعه را با نتایج حاصل از نمونه‌ای که از جامعه انتخاب شده است، برآورد کرد.

انواع برآورد

هنگام برآورد پارامتر جامعه بسته به اینکه بخواهیم مقدار یا محدوده‌ای برای پارامتر به دست آوریم، به دو نوع برآورد خواهیم رسید:
۱- برآورد نقطه‌ای
۲- برآورد فاصله‌ای
در برآورد نقطه‌ای، از مقدار آماره $\hat{\theta}$ حاصل از نمونه برای تخمین پارامتر θ استفاده می‌کنیم، در حالی‌که در برآورد فاصله‌ای با تعیین حدودی اطراف آماره $\hat{\theta}$، به فاصله‌ای می‌رسیم که با احتمال مشخصی مطمئن هستیم پارامتر θ در آن قرار دارد.
در ادامه با تشریح هر یک از دو نوع برآورد خواهیم دید که برآورد فاصله‌ای نسبت به برآورد نقطه‌ای دقت بیشتری دارد چراکه برای یافتن یک عدد مجهول، همواره داشتن یک فاصله، درجه صحت بیشتری نسبت به یک مقدار دارد که معلوم نیست همان مقدار مجهول است یا خیر.

مزیت برآورد فاصله‌ای به برآورد نقطه‌ای

برآورد فاصله‌ای، به این علت که با تعیین یک فاصله به نام فاصله اطمینان درجه صحت و حدود خطا مشخص می‌شود، دقت بیشتری نسبت به برآورد نقطه‌ای دارد.

مثال علت اینکه برآورد فاصله‌ای به برآورد نقطه‌ای ترجیح داده می‌شود این است که: (اقتصاد ـ ۷۵)

۱) برآوردهای بدون تورش و سازگارتری به دست می‌دهد.
۲) برآوردهای سازگار و کاراتری به دست می‌دهد.
۳) برآوردهای بدون تورش، سازگار و کاراتری به دست می‌دهد.
۴) درجه صحت برآورد را مشخص می‌کند.

حل: گزینه ۴ درست است.

برآورد نقطه‌ای (Point Estimation)

هرگاه برای برآورد پارامتر مجهول θ از آماره $\hat{\theta}$ در نمونه استفاده کنیم آن‌گاه $\hat{\theta}$ یک برآوردکننده نقطه‌ای و مقدار آن یک برآورد نقطه‌ای برای پارامتر θ است.

برای مثال، فرض کنید نمونه تصادفی $1,2,4,7,4$ با حجم $n=5$ از جامعه‌ای که پارامترهای آن نامشخص است، انتخاب شده باشد. اکنون چنانچه بخواهیم میانگین جامعه (μ) را برآورد کنیم ساده‌ترین راه این است که میانگین نمونه $(\bar{x})$ را به عنوان تخمینی از آن ارائه دهیم:

$$\bar{x}=\frac{\sum x}{n}=3.6$$

در رابطه بالا، $\bar{X}=\frac{\sum x}{n}$ یک **برآوردکننده نقطه‌ای** و مقدار آن (3.6) یک **برآورد نقطه‌ای** برای میانگین جامعه است.

تعریف: آماره $\hat{\theta}$ را یک **برآوردکننده نقطه‌ای** برای پارامتر θ گوییم هرگاه رابطه‌ای داشته باشد که با استفاده از نتایج نمونه تنها یک مقدار را به عنوان برآورد نقطه‌ای θ ارائه دهد.

برآوردکننده‌های نقطه‌ای مهم

ـ آماره $\bar{X}=\frac{\sum x}{n}$ یک برآوردکننده نقطه‌ای برای پارامتر μ (میانگین جامعه) است.

ـ آماره $\bar{p}=\frac{x}{n}$ یک برآوردکننده نقطه‌ای برای پارامتر p (نسبت جامعه) است.

ـ آماره‌های $S^2=\frac{\sum(x-\mu)^2}{n}$ و $S^2=\frac{\sum(x-\bar{x})^2}{n-1}$ برآوردکننده‌های نقطه‌ای برای σ^2 (واریانس جامعه) هستند.

مثال ۱ در یک نمونه 100تایی از جامعه‌ای برای برآورد میانگین جامعه (μ) درصورتی‌که واریانس نمونه $S^2=16$ باشد، برآورد نقطه‌ای انحراف معیار میانگین نمونه کدام است؟

۱) 4 ۲) 0.16 ۳) 0.4 ۴) 16

حل: گزینه ۳ درست است.

$$\begin{cases} S_{\bar{X}}=\frac{S_X}{\sqrt{n}}=\frac{4}{\sqrt{100}}=0.4 \\ n=100\ ,\ S_X^2=16 \rightarrow S_X=4 \end{cases}$$

مثال ۲ در یک نمونه 100تایی برای برآورد نسبت بیکاری در جامعه (p)، 90 نفر بیکار بودند. برآورد نقطه‌ای نرخ بیکاری و انحراف معیار نرخ بیکاری کدام است؟

۱) 0.3 , 0.9 ۲) 0.03 , 0.09 ۳) 0.3 , 0.09 ۴) 0.03 , 0.9

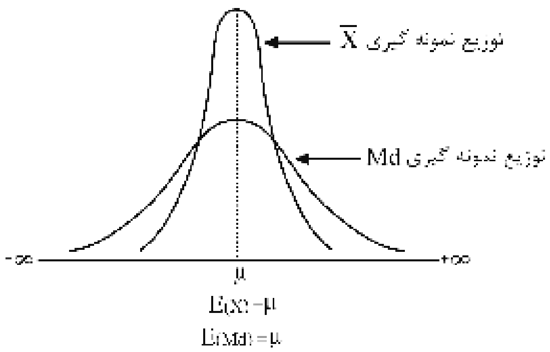
حل: گزینه ۴ درست است.

$$\begin{cases} (\text{برآورد نرخ بیکاری})\ \bar{p}=\frac{x}{n}=\frac{90}{100}=0.9 \\ (\text{انحراف معیار نرخ بیکاری})\ \sigma_{\bar{p}}=\sqrt{\frac{\bar{p}\bar{q}}{n}}=\sqrt{\frac{0.9\times 0.1}{100}}=0.03 \end{cases}$$

چون برآوردهای نقطه‌ای تابعی از نمونه تصادفی هستند، بنابراین برای هر پارامتر مجهول برآوردکننده‌های مختلفی وجود دارد، برای همین باید در بین برآوردکننده‌های مختلف برای یک پارامتر مجهول بهترین آن‌ها را انتخاب کنیم.

خواص مطلوب برآوردکننده‌های نقطه‌ای

برای هر پارامتر از جامعه برآوردکننده‌های متفاوتی وجود دارد که باید بهترین آن‌ها برای استنباط درباره پارامتر انتخاب شود.
برای مثال $\bar{X}$ (میانگین) و Md (میانه) دو برآوردکننده برای تخمین μ (میانگین جامعه) هستند که امید هر دوی آن‌ها با μ برابر است اما همان‌طورکه در ادامه خواهیم دید، $\bar{X}$ پراکندگی کمتری نسبت به Md دارد و برآوردکننده مطلوب‌تری است.



خواصی که از روی آن‌ها می‌توان بهترین برآوردکننده را برای استنباط درباره پارامتر جامعه انتخاب کرد به‌شرح زیر است:
۱- ناریبی (بدون تورش بودن، ناتور بودن)
۲- کارایی (حداقل واریانس)
۳- حداقل میانگین مجذور خطا (MSE)
۴- سازگاری (پایداری)

ناریبی (Unbias)

هرگاه میانگین یک آماره دقیقاً بر پارامتر منطبق شود، آن‌گاه آن آماره برآوردکننده‌ای ناریب (بدون تورش، ناتور) برای پارامتر جامعه است.

پارامتر = (آماره) E $\longleftrightarrow$ آماره برآوردکننده ناریب برای پارامتر

تعریف: اگر آماره $\hat{\theta}$ یک برآوردکننده دلخواه برای پارامتر θ باشد، آن‌گاه:

$\hat{\theta}$ ناریب برای θ $\longleftrightarrow$ $E(\hat{\theta})=\theta$ یا $E(\hat{\theta}-\theta)=0$

یادآوری:
ازآنجاکه پارامتر هر جامعه (θ) یک عدد ثابت و امید هر عدد برابر با همان عدد است پس $E(\theta)=\theta$ و خواهیم داشت:
$\hat{\theta}$ ناریب $\longleftrightarrow E(\hat{\theta})=\theta \longleftrightarrow E(\hat{\theta})=E(\theta) \longleftrightarrow E(\hat{\theta}-\theta)=0$

مثال ۱ برآوردکننده $\hat{\theta}$ برای پارامتر θ یک برآوردکننده ناریب است اگر: (اقتصاد ـ ۷۰)

۱) $\sigma^2(\hat{\theta})=\theta^2$ ۲) $P(|\hat{\theta}-\theta|\geq\varepsilon)=0$ ۳) $E(\theta)=\hat{\theta}$ ۴) $E(\hat{\theta})=\theta$

حل: گزینه ۴ درست است.

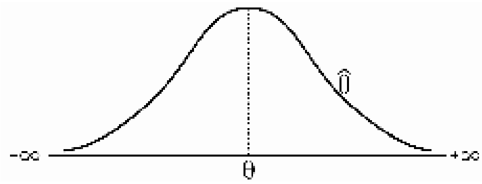
اریبی یا تورش (Bias)

اگر (آماره $E(\hat{\theta})$ با پارامتر θ متفاوت باشد ($E(\hat{\theta}) > \theta$ یا $E(\hat{\theta}) < \theta$) آن‌گاه آماره $\hat{\theta}$ اریب است و این تفاوت به عنوان میزان اریبی (تورش) شناخته می‌شود:

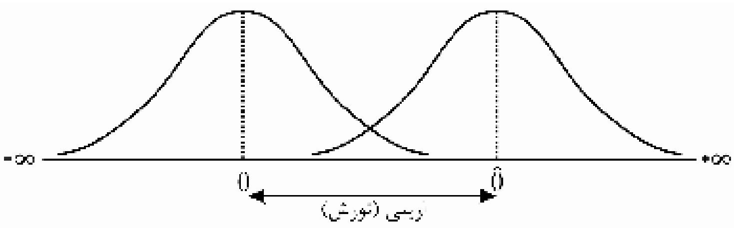
$$\boxed{\text{اریبی (تورش)} = E\left(\hat{\theta}-\theta\right) = E\left(\hat{\theta}\right)-\theta}$$

توجه: همان‌طورکه می‌دانیم θ پارامتر جامعه یک عدد ثابت است و امید ریاضی هر عدد ثابت، خود عدد ثابت است $E(\theta)=\theta$، بنابراین:

$$E\left(\hat{\theta}-\theta\right) = E\left(\hat{\theta}\right) - E(\theta) = E\left(\hat{\theta}\right)-\theta$$



الف) $\hat{\theta}$ نااریب



ب) $\hat{\theta}$ اریب

نتیجه:

الف) اگر $\hat{\theta}$ **نااریب** باشد $\left(E\left(\hat{\theta}\right)=\theta\right)$، آن‌گاه میزان اریبی (تورش) برابر صفر خواهد بود و برعکس.

$$\text{اریبی} = E\left(\hat{\theta}\right)-\theta = E\left(\hat{\theta}-\theta\right) = 0 \xleftrightarrow{E\left(\hat{\theta}\right)=\theta} \text{آماره } \hat{\theta} \text{ نااریب}$$

ب) اگر $\hat{\theta}$ **اریب** باشد، درصورتی‌که $E\left(\hat{\theta}\right) > \theta$ برقرار باشد آن‌گاه $\hat{\theta}$ گرایش به این دارد که پارامتر θ را زیادتر برآورد کند و اگر $E\left(\hat{\theta}\right) < \theta$ باشد، آن‌گاه $\hat{\theta}$ گرایش به این دارد که پارامتر θ را کمتر برآورد کند.

$$\text{اریبی} = E\left(\hat{\theta}\right)-\theta = E\left(\hat{\theta}-\theta\right) \neq 0 \xleftrightarrow{E\left(\hat{\theta}\right)\neq\theta} \text{آماره } \hat{\theta} \text{ اریب}$$

مثال ۲ اگر (آماره) E بزرگ‌تر از پارامتر باشد، کدام‌یک از این گزینه‌ها درست است؟ (حسابداری ـ ۸۳)

۱) آماره دارای اریبی است.

۲) آماره دارای کمترین واریانس است.

۳) آماره سازگار است.

۴) باید نوع آماره مشخص باشد، تا بتوان اظهار نظر کرد.

حل: گزینه ۱ درست است.

مثال ۳ تابع چگالی احتمال X به صورت $\begin{cases} f(x) = \frac{1}{\theta} \\ 0 < x < \theta \end{cases}$ است. به ازای چه مقداری از k تخمین‌زن $\hat{\theta} = \frac{X}{k}$ می‌تواند تخمین‌زنی ناتور (ناریب) برای پارامتر θ باشد؟

۱) $\frac{1}{2}$ ۲) 1 ۳) 2 ۴) 3

حل: گزینه ۱ درست است.

برآوردکننده $\hat{\theta}$ را برای تخمین پارامتر θ ناریب گوییم هرگاه شرط $E(\hat{\theta}) = \theta$ برقرار باشد. ازآنجاکه برای کنترل این شرط به $E(X)$ نیاز داریم، ابتدا آن را محاسبه می‌کنیم:

راه حل اول:

$$E(X) = \int_0^\theta x f(x) dx = \int_0^\theta \frac{x}{\theta} dx = \left[\frac{x^2}{2\theta}\right]_0^\theta = \frac{\theta}{2}$$

راه حل دوم: ازآنجاکه $f(x)$ تابع توزیع یکنواخت پیوسته است، داریم:

$$E(X) = \frac{\theta + 0}{2} = \frac{\theta}{2}$$

حال می‌توانیم شرط ناریبی را کنترل کنیم تا مقدار k به دست آید:

ناریبی: $E(\hat{\theta}) = \theta \to E\left(\frac{X}{k}\right) = \theta \to k = \frac{E(X)}{\theta} = \frac{\frac{\theta}{2}}{\theta} = \frac{1}{2}$

مثال ۴ کدام‌یک از گزینه‌های زیر یک برآوردکننده ناریب برای پارامتر θ از جامعه‌ای با تابع چگالی زیر است؟

$$\begin{cases} f(x) = \frac{1}{\theta} \\ 0 < x < \theta \end{cases}$$

۱) $\hat{\theta} = 2X$ ۲) $\hat{\theta} = \frac{X}{2}$ ۳) $\hat{\theta} = X$ ۴) $\hat{\theta} = 4X$

حل: گزینه ۱ درست است.

برآوردکننده $\hat{\theta}$ را برای تخمین پارامتر θ ناریب گوییم هرگاه شرط $E(\hat{\theta}) = \theta$ برقرار باشد. ازآنجاکه برای کنترل این شرط در تمام گزینه‌ها به $E(X)$ نیاز داریم، ابتدا آن را محاسبه می‌کنیم:

$$E(X) = \int_0^\theta x f(x) dx = \int_0^\theta \frac{x}{\theta} dx = \left[\frac{x^2}{2\theta}\right]_0^\theta = \frac{\theta}{2}$$

حال کافی است شرط ناریبی را برای گزینه‌ها کنترل کنیم:

گزینه ۱ : $E(\hat{\theta}) = E(2X) = 2E(X) = 2 \times \frac{\theta}{2} = \theta$ ناریب✓

گزینه ۲ : $E(\hat{\theta}) = E\left(\frac{X}{2}\right) = \frac{1}{2}E(X) = \frac{1}{2} \times \frac{\theta}{2} = \frac{\theta}{4}$ اریب

گزینه ۳ : $E(\hat{\theta}) = E(X) = \frac{\theta}{2}$ اریب

گزینه ۴ : $E(\hat{\theta}) = E(4X) = 4E(X) = 4 \times \frac{\theta}{2} = 2\theta$ اریب

مثال ۵ اگر $\hat{\theta}$ برآوردکننده پارامتر θ با اریب (تورش) $K\theta-2$ باشد، آن‌گاه کدام برآوردکننده زیر نااریب (بدون تورش) است؟

۱) $\frac{\hat{\theta}}{K+1}$ ۲) $\frac{\hat{\theta}+2}{K+1}$ ۳) $\frac{2\hat{\theta}}{2K+1}$ ۴) $\frac{1}{(K+1)\hat{\theta}}$

حل: گزینه ۲ درست است.

ابتدا رابطه مربوط به اریبی را در نظر می‌گیریم:

$$\text{اریبی} = E\left(\hat{\theta}\right)-\theta = K\theta-2$$

حال ازآنجاکه شرط نااریبی برآوردکننده دلخواه T برای پارامتر θ برابر است با $E(T)=\theta$، کافی است در رابطه بالا θ را به سمت راست تساوی منتقل کنیم و بقیه مقادیر را در سمت چپ، داخل E قرار دهیم تا برآوردکننده نااریب T مشخص شود:

$$E\left(\hat{\theta}\right)-\theta = K\theta-2 \rightarrow E\left(\hat{\theta}\right)+2 = K\theta+\theta \rightarrow E\left(\frac{\hat{\theta}+2}{K+1}\right)=\theta$$

مثال ۶ فرض کنید u_1 و u_2 برآوردکننده‌های نااریب و مستقل برای پارامتر θ باشند. ضریب a چقدر باشد تا آماره $T=\frac{u_1}{2}+\frac{u_2}{a}$ یک برآوردکننده نااریب برای θ باشد؟

۱) 2 ۲) 4 ۳) 1 ۴) 5

حل: گزینه ۱ درست است.

ازآنجاکه u_1 و u_2 برآوردکننده‌های نااریب برای پارامتر θ هستند، داریم:

$$E(u_1)=\theta \quad , \quad E(u_2)=\theta$$

حال برای آنکه مقدار a را برای نااریبی آماره T به دست آوریم کافی است شرط نااریبی را برای آن برقرار کنیم:

$$\text{شرط نااریبی} : E(T)=\theta \rightarrow \frac{\overset{\theta}{\cancel{E(u_1)}}}{2}+\frac{\overset{\theta}{\cancel{E(u_2)}}}{a}=\theta \rightarrow \frac{1}{2}+\frac{1}{a}=1 \rightarrow a=2$$

مثال ۷ به ازای چه مقداری از k برآوردکننده $\hat{\theta}=\frac{X}{k}$ برآوردکننده بدون تورشی (نااریبی) از پارامتر θ جامعه‌ای است که دارای تابع احتمال گسسته مقابل است؟ (اقتصاد ـ ۸۵)

$$\begin{cases} f(x)=\theta^{x}(1-\theta)^{1-x} & ; \quad x=0,1 \\ f(x)=0 & ; \quad \text{برای سایر مقادیر } x \end{cases}$$

۱) $\frac{1}{2}$ ۲) 1 ۳) 2 ۴) 5

حل: گزینه ۲ درست است.

با توجه به اینکه این تابع، تابع چگالی توزیع برنولی با پارامتر θ است داریم: $X \sim \text{برنولی} \rightarrow E(X)=\theta$

حال اگر بخواهیم $\hat{\theta}$ برای θ نااریب باشد باید $E\left(\hat{\theta}\right)=\theta$ شود، بنابراین:

$$E\left(\hat{\theta}\right)=\theta \rightarrow E\left(\frac{X}{k}\right)=\frac{E(X)}{k}=\frac{\theta}{k}=\theta \quad \rightarrow k=1$$

اگر تشخیص ندهیم که تابع چگالی، تابع چگالی توزیع برنولی است باید امید X را محاسبه کنیم:

$$E(X)=\sum_{x=0}^{1} xf(x)=\sum_{x=0}^{1} x\theta^{x}(1-\theta)^{1-x}=0\times\theta^{0}\times(1-\theta)^{1-0}+1\times\theta^{1}\times(1-\theta)^{1-1}=\theta$$

آماره‌های ناریب برای پارامترهای جامعه

۱- آماره‌های ناریب برای میانگین جامعه

اگر $X_1, X_2, \dots, X_n$ نمونه‌ای مستقل از جامعه‌ای با میانگین μ و $E(X_1)=\mu$ و ... و $E(X_n)=\mu$ باشد، آن‌گاه:

الف) میانگین نمونه : $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\xrightarrow{\bar{X} \text{ تخمین‌زننده ناریب برای } \mu}$ $E(\bar{X})=\mu$

ب) میانه : Md $\xrightarrow{Md \text{ تخمین‌زننده ناریب برای } \mu}$ $E(Md)=\mu$

ج) $\bar{X} = \frac{\sum_{i=a}^{b} x_i}{b-a+1}$ $\xrightarrow{\bar{X} \text{ تخمین‌زننده ناریب برای } \mu}$ $E(\bar{X})=\mu$

✓ **دقت کنید!**

قسمت (الف) حالت خاصی از (ج) است. زمانی که $a=1$ و $b=n$ باشد.

د) $\begin{cases} T = \frac{\sum_{i=1}^{n} a_i x_i}{b} = \frac{a_1x_1+\dots+a_nx_n}{b} \\ a_1+a_2+\dots+a_n=b \end{cases}$ $\xrightarrow{T \text{ تخمین‌زننده ناریب برای } \mu}$ $E(T)=\mu$

(مجموع ضرایب x_i برابر است با b)

ه) $\begin{cases} T = \sum_{i=1}^{n} a_i x_i = a_1x_1+\dots+a_nx_n \\ a_1+a_2+\dots+a_n=1 \end{cases}$ $\xrightarrow{T \text{ تخمین‌زننده ناریب برای } \mu}$ $E(T)=\mu$

(مجموع ضرایب x_i برابر است با 1)

تبصره: در روابط (ج)، (د) و (ه) می‌توان به‌جای x_i از $\bar{x}_i$ استفاده کرد زیرا $E(\bar{X}_i)=E(X_i)=\mu$

برای اثبات قسمت‌های (ج) و (د) و (ه) می‌توانیم به صورت زیر عمل کنیم:

$$\text{ج) } E(T) = \frac{\overset{\mu}{\cancel{E(X_a)}}+\dots+\overset{\mu}{\cancel{E(X_b)}}}{b-a+1} = \frac{(b-a+1)\mu}{(b-a+1)} = \mu$$

$$\text{د) } E(T) = \frac{a_1\overset{\mu}{\cancel{E(X_1)}}+\dots+a_n\overset{\mu}{\cancel{E(X_n)}}}{b} = \frac{\overset{b}{\cancel{(a_1+\dots+a_n)}}\mu}{b} = \mu$$

$$\text{ه) } E(T) = a_1\overset{\mu}{\cancel{E(X_1)}}+\dots+a_n\overset{\mu}{\cancel{E(X_n)}} = \overset{1}{\cancel{(a_1+\dots+a_n)}}\mu = \mu$$

مثال ۱ درصورتی‌که X_1, X_2, X_3 دارای توزیعی با میانگین μ باشند آن‌گاه کدام‌یک از آماره‌های T_1, T_2, T_3 برای میانگین جامعه ناریب هستند؟

$$T_1 = \frac{3X_1 + X_2}{3} \quad , \quad T_2 = \frac{3X_1 + 2X_2 + X_3}{6} \quad , \quad T_3 = \frac{X_1 + 3X_2 - X_3}{3}$$

۱) T_1 ۲) T_2 , T_3 ۳) T_1 , T_3 ۴) T_1 , T_2

حل: گزینه ۲ درست است.

برای آنکه آماره‌های T_1 , T_2 , T_3 برای μ جامعه ناریب باشند باید امید (میانگین) هر کدام از آن‌ها برابر μ شود.

با توجه به حالت «د» داریم:

$$T_1 = \frac{3X_1 + X_2}{3} \xrightarrow{3+1\neq 3} E(T_1) \neq \mu \to \text{اریب } T_1$$

$$T_2 = \frac{3X_1 + 2X_2 + X_3}{6} \xrightarrow{3+2+1=6} E(T_2) = \mu \to \text{ناریب } T_2$$

$$T_3 = \frac{X_1 + 3X_2 - X_3}{3} \xrightarrow{1+3-1=3} E(T_3) = \mu \to \text{ناریب } T_3$$

درعین‌حال میزان اریبی T_1 به صورت زیر محاسبه می‌شود:

$$\text{اریبی} = E(\text{آماره}) - \text{پارامتر} = E(T_1) - \mu = \frac{4\mu}{3} - \mu = \frac{\mu}{3}$$

مثال ۲ درصورتی‌که $T = 2X + (1-\alpha)X$ برآوردکننده (آماره) میانگین جامعه باشد و X نیز دارای میانگین μ باشد به ازای کدام مقدار از α، T یک برآوردکننده ناریب خواهد بود؟

۱) $\frac{1}{2}$ ۲) 4 ۳) 2 ۴) $\frac{1}{4}$

حل: گزینه ۳ درست است.

راه حل اول: برای ناریب بودن کافی است $E(T) = \mu$ شود، درنتیجه:

$$E(T) = E(2X + (1-\alpha)X) = 2\mu + (1-\alpha)\mu = \mu \to \alpha = 2$$

راه حل دوم: با توجه به حالت «ه» داریم:

$$\begin{cases} T = 2x + (1-\alpha)x \xrightarrow{2+(1-\alpha)=1} E(T) = \mu \to \text{ناریب } T \\ 2 + (1-\alpha) = 1 \to \alpha = 2 \end{cases}$$

مثال ۳ اگر $X_1 , X_2 , \dots , X_n$ مشاهدات حاصل از نمونه تصادفی به حجم n از جامعه‌ای با میانگین μ_X و واریانس σ^2 باشند، برآورد اریب (باتورش) میانگین جامعه μ_X را از بین آماره‌های زیر انتخاب کنید؟ (اقتصاد ـ ۷۳)

۱) $\bar{X} = \frac{\sum_{i=1}^{n-1} x_i}{n-1}$ ۲) $\bar{X} = \frac{\sum_{i=3}^{n} x_i}{n-2}$ ۳) $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n-1}$ ۴) $\bar{X} = \frac{\sum_{i=2}^{n-1} x_i}{n-2}$

حل: گزینه ۳ درست است.

بنابر حالت (ج) داریم:

$$\text{گزینه ۱}: E(\bar{X})=\frac{E\left(\sum_{i=1}^{n-1}x_i\right)}{n-1}=\frac{(n-1-1+1)\,\mu}{n-1}=\mu \qquad \text{گزینه ۲}: E(\bar{X})=\frac{E\left(\sum_{i=3}^{n}x_i\right)}{n-2}=\frac{(n-3+1)\,\mu}{n-2}=\mu$$

$$\text{گزینه ۳}: E(\bar{X})=\frac{E\left(\sum_{i=1}^{n}x_i\right)}{n-1}=\frac{(n-1+1)\,\mu}{n-1}=\frac{n}{n-1}\mu \qquad \text{گزینه ۴}: E(\bar{X})=\frac{E\left(\sum_{i=2}^{n-1}x_i\right)}{n-2}=\frac{(n-1-2+1)\,\mu}{n-2}=\mu$$

مثال ۴ اگر $x_1, x_2, \dots, x_n$ نمونه‌ای تصادفی از یک جامعه دلخواه باشد، کدام‌یک از برآوردکننده‌های زیر برای μ ناریب (ناتور) است؟ (اقتصاد ـ ۷۹)

۱) $\hat{\mu}_1=\frac{2\sum_{i=1}^{n}(x_i+1)}{n}$

۲) $\hat{\mu}_2=\frac{1}{2}\left(\frac{\sum_{i=1}^{k}x_i}{k}-\frac{\sum_{i=1+k}^{n}x_i}{n-k}\right)$

۳) $\hat{\mu}_3=\frac{2\sum_{i=1}^{n}(x_i-1)}{n}$

۴) $\hat{\mu}_4=\frac{1}{2}\left(\frac{\sum_{i=1}^{k}x_i}{k}+\frac{\sum_{i=1+k}^{n}x_i}{n-k}\right)$

حل: گزینه ۴ درست است.

اگر ابتدا بر اساس حالت (ج) گزینه‌های ۲ و ۴ را بررسی کنیم، گزینه ۴ ناریب است.

$$\bar{X}_1=\frac{\sum_{i=a}^{b}x_i}{b-a+1}\xrightarrow{a=1,b=k}\frac{\sum_{i=1}^{k}x_i}{k}$$

$$\bar{X}_2=\frac{\sum_{i=a}^{b}x_i}{b-a+1}\xrightarrow{a=k+1,b=n}\frac{\sum_{i=k+1}^{n}x_i}{n-k}$$

$$E(\hat{\mu}_2)=\frac{1}{2}E\left(\frac{\sum_{i=1}^{k}x_i}{k}-\frac{\sum_{i=1+k}^{n}x_i}{n-k}\right)=\frac{1}{2}E(\bar{X}_1-\bar{X}_2)=\frac{1}{2}(\mu-\mu)=0$$

$$E(\hat{\mu}_4)=\frac{1}{2}E\left(\frac{\sum_{i=1}^{k}x_i}{k}+\frac{\sum_{i=1+k}^{n}x_i}{n-k}\right)=\frac{1}{2}E(\bar{X}_1+\bar{X}_2)=\frac{1}{2}(\mu+\mu)=\mu$$

برای درک بهتر، گزینه‌های ۱ و ۳ را نیز بررسی می‌کنیم:

$$E(\hat{\mu}_1)=E\left(\frac{2\sum_{i=1}^{n}(x_i+1)}{n}\right)=2\times\left(E\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)+E\left(\frac{\sum_{i=1}^{n}1}{n}\right)\right)=2\left(E(\bar{X})+E(1)\right)=2(\mu+1)$$

$$E(\hat{\mu}_3)=E\left(\frac{2\sum_{i=1}^{n}(x_i-1)}{n}\right)=2\times\left(E\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)-E\left(\frac{\sum_{i=1}^{n}1}{n}\right)\right)=2\left(E(\bar{X})-E(1)\right)=2(\mu-1)$$

مثال۵ دو نمونه تصادفی مستقل به حجم‌های n_1 و n_2 از جامعه نرمالی گرفته شده است و میانگین‌های $\bar{X}_1$ و $\bar{X}_2$ محاسبه شده است. برآوردکننده (تخمین‌زن) بدون تورش θ برای برآورد میانگین جامعه کدام است؟ (اقتصاد ـ ۷۷)

۱) $\theta=\frac{\bar{X}_1}{n_1}+\frac{\bar{X}_2}{n_2}$

۲) $\theta=\frac{n_1\bar{X}_1+n_2\bar{X}_2}{n_1+n_2}$

۳) $\theta=\frac{n_1\bar{X}_1+n_2\bar{X}_2}{n_1n_2}$

۴) $\theta=n_1\bar{X}_1+n_2\bar{X}_2$

حل: گزینه ۲ درست است.

با توجه به آنکه همواره $E(\bar{X}_i)=\mu$، کافی است شرط نااریبی را برای همه گزینه‌ها بررسی کنیم:

گزینه۱ : $E(\theta)=\frac{\mu}{n_1}+\frac{\mu}{n_2}\neq\mu$

گزینه ۲ : $E(\theta)=\frac{n_1\mu+n_2\mu}{n_1+n_2}=\mu$

گزینه۳ : $E(\theta)=\frac{n_1\mu+n_2\mu}{n_1n_2}\neq\mu$

گزینه ۴ : $E(\theta)=n_1\mu+n_2\mu\neq\mu$

۲ـ آماره‌های نااریب برای نسبت جامعه

درصورتی‌که به دنبال نسبت (درصد) یک صفت کیفی در جامعه باشیم (نسبت بیکاری ـ نسبت بیسوادی) با در نظر گرفتن جامعه به صورت دوجمله‌ای با نسبت p برای صفت مورد نظر $\left(E(X)=np,\sigma_X^2=npq\right)$:

الف) در صورت انتخاب یک نمونه nتایی که xتای آن صفت مورد نظر را داشته باشند آماره $\bar{p}=\frac{x}{n}$ (نرخ نمونه) یک برآوردکننده نااریب برای نسبت p در جامعه خواهد بود:

$$\bar{p}=\frac{x}{n}\xrightarrow{\bar{p}\text{ تخمین‌زننده نااریب برای } p}E(\bar{p})=p$$

یادآوری:

$$E(\bar{p})=E\left(\frac{X}{n}\right)=\frac{E(X)}{n}=\frac{np}{n}=p$$

ب) درصورتی‌که دو نمونه مستقل به حجم n_1 و n_2 از جامعه انتخاب کنیم که به ترتیب x_1 و x_2 تای هر یک از نمونه‌ها صفت مورد نظر را داشته باشند، آن‌گاه $\bar{p}_1=\frac{x_1}{n_1}$ و $\bar{p}_2=\frac{x_2}{n_2}$ به ترتیب نسبت موفقیت در نمونه‌ها است و:

$$E(X_1)=n_1p\ \rightarrow\ E(\bar{p}_1)=p$$

$$E(X_2)=n_2p\ \rightarrow\ E(\bar{p}_2)=p$$

در این وضعیت آماره‌های زیر برآوردکننده‌های نااریب برای نسبت p در جامعه خواهند بود:

۱ـ $$T=\frac{x_1+x_2}{n_1+n_2}\xrightarrow{T\text{ برآوردکننده نااریب برای } p}E(T)=p$$

$$E(T)=E\left(\frac{X_1+X_2}{n_1+n_2}\right)=\frac{E(X_1)+E(X_2)}{n_1+n_2}=\frac{n_1p+n_2p}{n_1+n_2}=p$$

۲- $\boxed{T = \dfrac{n_1x_1 + n_2x_2}{n_1^2 + n_2^2} \xrightarrow{\text{برآوردکننده ناریب برای } p \ T} E(T) = p}$

$$E(T) = E\left(\frac{n_1X_1 + n_2X_2}{n_1^2 + n_2^2}\right) = \frac{n_1\,\overbrace{\cancel{E(X_1)}}^{n_1p} + n_2\,\overbrace{\cancel{E(X_2)}}^{n_2p}}{n_1^2 + n_2^2} = \frac{n_1^2p + n_2^2p}{n_1^2 + n_2^2} = p$$

۳- $\boxed{T = \dfrac{\bar{p}_1 + \bar{p}_2}{2} \xrightarrow{\text{برآوردکننده ناریب برای } p \ T} E(T) = p}$

$$E(T) = E\left(\frac{\bar{p}_1 + \bar{p}_2}{2}\right) = \frac{E(\bar{p}_1) + E(\bar{p}_2)}{2} = \frac{p + p}{2} = p$$

مثال ۱ دو نمونه تصادفی مستقل از هم به حجم‌های n_1 و n_2 از جامعه گرفته شده است و نسبت‌های $\bar{p}_1 = \dfrac{x_1}{n_1}$ و $\bar{p}_2 = \dfrac{x_2}{n_2}$ محاسبه شده است. برآوردکننده (تخمین‌زن) بدون تورش $\hat{\theta}$ برای برآورد $p = \dfrac{X}{N}$ در جامعه کدام است؟ (اقتصاد ـ ۷۸)

۱) $\hat{\theta} = \dfrac{x_1}{n_1} + \dfrac{x_2}{n_2}$ ۲) $\hat{\theta} = \dfrac{x_1 + x_2}{n_1 + n_2}$ ۳) $\hat{\theta} = \dfrac{n_1x_1 + n_2x_2}{n_1 + n_2}$ ۴) $\hat{\theta} = \dfrac{x_1 + x_2}{n_1n_2}$

حل: گزینه ۲ درست است.

بنا بر حالت (۱) ، فقط آماره $\hat{\theta}$ در گزینه ۲ برآوردکننده ناریب برای p است، زیرا:

$$E(\hat{\theta}) = E\left(\frac{X_1 + X_2}{n_1 + n_2}\right) = \frac{E(X_1) + E(X_2)}{n_1 + n_2} = \frac{n_1p + n_2p}{n_1 + n_2} = p$$

درعین‌حال آماره‌های سایر گزینه ها اریب هستند زیرا:

گزینه۱ : $E(\hat{\theta}) = E\left(\dfrac{X_1}{n_1} + \dfrac{X_2}{n_2}\right) = \dfrac{E(X_1)}{n_1} + \dfrac{E(X_2)}{n_2} = \dfrac{n_1p}{n_1} + \dfrac{n_2p}{n_2} = 2p \neq p$

گزینه۳ : $E(\hat{\theta}) = E\left(\dfrac{n_1X_1 + n_2X_2}{n_1 + n_2}\right) = \dfrac{n_1E(X_1) + n_2E(X_2)}{n_1 + n_2} = \dfrac{n_1^2p + n_2^2p}{n_1 + n_2} \neq p$

گزینه۴ : $E(\hat{\theta}) = E\left(\dfrac{X_1 + X_2}{n_1n_2}\right) = \dfrac{E(X_1) + E(X_2)}{n_1n_2} = \dfrac{n_1p + n_2p}{n_1n_2} \neq p$

مثال ۲ برای تخمین نسبت موفقیت‌ها در جامعه‌ای دو نمونه تصادفی مستقل به حجم‌های n_1 و n_2 از جامعه گرفته شده و برآوردکننده زیر پیشنهاد شده است که در آن x_1 و x_2 تعداد موفقیت‌ها در نمونه اول و دوم است. کمیت انتظاری این برآوردکننده کدام است؟ (اقتصاد ـ ۸۷)

$$\hat{p} = \frac{1}{2}\left(\frac{x_1}{n_1} + \frac{x_2}{n_2}\right)$$

۱) p ۲) μ_x ۳) $\dfrac{p}{2n_1n_2}$ ۴) $\dfrac{(n_1 + n_2)\mu_X}{n_1 + n_2}$

حل: گزینه ۱ درست است.

همان‌طورکه دیده می‌شود آماره $\hat{p}$ در این سؤال همان گزینه ۱ در مثال قبلی است که در $\frac{1}{2}$ ضرب شده و به صورت یک آماره ناریب درآمده است:

$$E(\hat{p}) = \frac{1}{2}\left(\frac{n_1 p}{n_1} + \frac{n_2 p}{n_2}\right) = p$$

۳- آماره‌های ناریب برای واریانس جامعه

اگر $X_1, X_2, \ldots, X_n$ نمونه تصادفی مستقل به حجم n از جامعه نرمالی با میانگین μ و واریانس σ^2 باشد آن‌گاه:

الف) در شرایطی که μ (میانگین جامعه) نامعلوم باشد از $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ (میانگین نمونه) استفاده کرده و برآوردکننده ناریب S^2 به شکل زیر بررسی می‌شود:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \xrightarrow{\text{برآوردکننده ناریب برای } \sigma^2 \; S^2} E(S^2) = \sigma^2$$

البته از بسط رابطه بالا به $S^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$ می‌رسیم که $E(S^2) = \sigma^2$ است.

ب) در شرایطی که μ (میانگین جامعه) معلوم باشد، دیگر نیازی به $\bar{X}$ نیست و به برآوردکننده بهتری نسبت به حالت (الف) می‌رسیم:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} \xrightarrow{\text{برآوردکننده ناریب برای } \sigma^2 \; S^2} E(S^2) = \sigma^2$$

البته از بسط رابطه بالا به $\frac{\sum x_i^2}{n} - \mu^2$ می‌رسیم که $E(S^2) = \sigma^2$ است.

نکته:

۱- در شرایط برابر آماره S^2 در وضعیت (ب) مطلوب‌تر از (الف) است.

۲- وضعیت‌های (الف) و (ب) تنها آماره‌های ناریب برای σ^2 (واریانس جامعه) هستند، بنابراین هر رابطه دیگری برای S^2 یک برآوردکننده اریب برای σ^2 است.

مثال ۱ اگر کمیت‌های $X_1, X_2, \ldots, X_n$ نمونه تصادفی با حجم n از جامعه نرمال باشد، کدام‌یک از توابع نمونه‌ای (آماره یا Statistic) زیر تخمین با تورش (اریب) از واریانس σ^2 است؟ (اقتصاد ـ ۷۱)

۱) $f(x_1, \ldots, x_n) = \frac{\sum(x_i - \bar{x})^2}{n-1}$

۲) $f(x_1, \ldots, x_n) = \frac{\sum(x_i - \bar{x})^2}{n}$

۳) $f(x_1, \ldots, x_n) = \frac{\sum(x_i - \mu)^2}{n}$

۴) $f(x_1, \ldots, x_n) = \frac{\sum x_i^2}{n} - \mu^2$

حل: گزینه ۲ درست است.

مثال ۲ اگر $S^2=\frac{\sum(x_i-\bar{x})^2}{n}$ باشد، آن‌گاه $E(S^2)$ کدام است؟

۱) σ^2 ۲) $\frac{n-1}{n}\sigma^2$ ۳) $\frac{n}{n-1}\sigma^2$ ۴) $(n-1)\sigma^2$

حل: گزینه ۲ درست است.

آماره $S^2=\frac{\sum(x_i-\bar{x})^2}{n}$ یک برآوردکننده اریب است زیرا در صورت استفاده از $\bar{x}$ باید مخرج کسر $n-1$ باشد تا S^2 با توجه به وضعیت (الف) یک آماره نااریب برای σ^2 باشد بنابراین، به کمک وضعیت (الف) می‌توانیم $E(S^2)$ را در این وضعیت به دست آوریم:

$$E(S^2)=E\left(\frac{n-1}{n-1}\frac{\sum(x_i-\bar{x})^2}{n}\right)=\frac{n-1}{n}E\left(\frac{\sum(x_i-\bar{x})^2}{n-1}\right)=\frac{n-1}{n}\sigma^2$$

مثال ۳ درصورتی‌که $S^2=\frac{\sum(x_i-\mu)^2}{n-1}$ باشد آن‌گاه $E(S^2)$ کدام است؟

۱) σ^2 ۲) $\frac{n-1}{n}\sigma^2$ ۳) $\frac{n}{n-1}\sigma^2$ ۴) $n\sigma^2$

حل: گزینه ۳ درست است.

در صورت استفاده از μ باید مخرج کسر n باشد تا S^2 با توجه به وضعیت (ب) یک برآوردکننده نااریب برای σ^2 باشد، بنابراین S^2 در این حالت اریب است و $E(S^2)$ به صورت زیر محاسبه می‌شود:

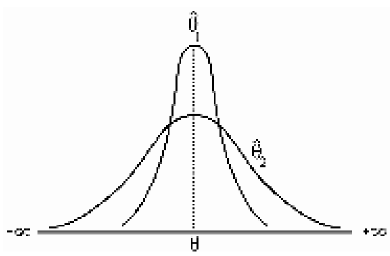
$$E(S^2)=E\left(\frac{n}{n}\frac{\sum(x_i-\mu)^2}{n-1}\right)=\frac{n}{n-1}E\left(\frac{\sum(x_i-\mu)^2}{n}\right)=\frac{n}{n-1}\sigma^2$$

کارایی (حداقل واریانس) (Minimal Variance)

بعد از بررسی شرط نااریبی برای آماره‌های مربوط به پارامتر جامعه از بین آماره‌هایی که شرط نااریبی را دارند، آماره‌ای مطلوب‌تر است که واریانس کمتری داشته و در اطراف پارامتر جامعه به شدت متمرکز باشد.

تعریف: از بین دو برآوردکننده نااریب $\hat{\theta}_1$ و $\hat{\theta}_2$ برای پارامتر θ، برآوردکننده‌ای کاراتر است که واریانس کمتری داشته و متمرکزتر باشد.

$$Var(\hat{\theta}_1)<Var(\hat{\theta}_2)\xrightarrow{\hat{\theta}_1 \text{ و } \hat{\theta}_2 \text{ نااریب}}\hat{\theta}_1 \text{ کاراتر از } \hat{\theta}_2$$



$\hat{\theta}_1$, $\hat{\theta}_2$ هر دو برآوردکننده‌های ناریب برای θ هستند اما $\hat{\theta}_1$ از $\hat{\theta}_2$ کاراتر است چون واریانس کمتری دارد یا به عبارت دیگر، حول θ متمرکزتر است.

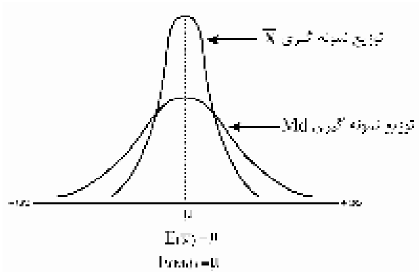
مثال ۱ اگر برای برآورد پارامتر جامعه (μ) از $\bar{X}$ و میانه (Me) استفاده کنیم: (اقتصاد ـ ۷۴)

۱) $\bar{X}$ کاراتر از Me خواهد بود. ۲) هر دو با تورش خواهند بود.

۳) هر دو ناسازگار خواهند بود. ۴) واریانس هر دو برابر خواهند بود.

حل: گزینه ۱ درست است.

همان‌طورکه قبلاً مطرح شد $\bar{X}$ (میانگین نمونه) و Md (میانه) برآوردکننده‌های (بدون تورش) برای μ (میانگین جامعه) هستند، اما واریانس آن‌ها با هم متفاوت است و البته $Var(\bar{X}) < Var(Md)$ و درنتیجه $\bar{X}$ کاراتر از Md است.



$$\begin{cases} \sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n} \\ \sigma_{Md}^2 = \dfrac{\pi}{2}\dfrac{\sigma^2}{n} \end{cases} \rightarrow \sigma_{\bar{X}}^2 < \sigma_{Md}^2 \rightarrow (\text{کاراتر از } Md \ \bar{X})$$

کارایی نسبی (Efficiency)

درصورتی‌که $\hat{\theta}_1$ و $\hat{\theta}_2$ برآوردکننده‌های ناریب برای پارامتر θ باشند:

$$\boxed{\text{کارایی نسبی } \hat{\theta}_1 \text{ به } \hat{\theta}_2 = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}}$$

بعد از محاسبه کارایی نسبی یکی از دو وضعیت زیر ممکن است بروز کند:

$$\begin{cases} \text{الف) } \hat{\theta}_1 \text{ کاراتر از } \hat{\theta}_2 \longrightarrow Var(\hat{\theta}_1) < Var(\hat{\theta}_2) \longrightarrow \dfrac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)} > 1 \\ \text{ب) } \hat{\theta}_2 \text{ کاراتر از } \hat{\theta}_1 \longrightarrow Var(\hat{\theta}_2) < Var(\hat{\theta}_1) \longrightarrow \dfrac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)} < 1 \end{cases}$$

$\bar{X}$ کاراترین برآوردکننده μ

آماره ناریب $\bar{X} = \dfrac{\sum x_i}{n}$ (میانگین نمونه) از هر آماره ناریب دیگر به صورت $T = \dfrac{a_1X_1 + \ldots + a_kX_k}{b}$ برای میانگین جامعه (μ) کاراتر است و واریانس کمتری دارد زیرا طبق خواص میانگین کمترین مجذور انحرافات را دارد.

مثال ۲ از بین برآوردکننده‌های ناریب μ به صورت زیر کدام‌یک کارایی بیشتری دارد؟ (x_i ها نمونه‌های مستقل‌اند.)

۱) $T_1 = \dfrac{X_1 + X_2 + X_3}{3}$ ۲) $T_2 = \dfrac{2X_1 + X_2 + X_3}{4}$

۳) $T_3 = \dfrac{3X_1 + 4X_2 - X_3}{6}$ ۴) $T_4 = \dfrac{X_1 + 3X_2 + X_3}{5}$

حل: گزینه ۱ درست است.

راه حل اول:

با توجه به نکته مطرح‌شده آماره $\bar{X} = \frac{X_1+X_2+X_3}{3} = \frac{\sum_{i=1}^{3} X_i}{3} = T_1$ از بقیه آماره‌ها کاراتر است.

راه حل دوم: در بین آماره‌های ناریب برآوردکننده‌ای کاراتر است که واریانس کمتری داشته باشد، بنابراین:

$$\sigma^2(T_1) = \frac{\sigma^2+\sigma^2+\sigma^2}{9} = \frac{\sigma^2}{3}$$

$$\sigma^2(T_2) = \frac{4\sigma^2+\sigma^2+\sigma^2}{16} = \frac{6}{16}\sigma^2$$

$$\sigma^2(T_3) = \frac{9\sigma^2+16\sigma^2+\sigma^2}{36} = \frac{26}{36}\sigma^2$$

$$\sigma^2(T_4) = \frac{\sigma^2+9\sigma^2+\sigma^2}{25} = \frac{11}{25}\sigma^2$$

با توجه به واریانس آماره‌های T_1, T_2, T_3, T_4 مشاهده می‌شود که T_1 واریانس کمتری دارد و کاراتر است.

یادآوری:

$$\sigma^2(aX_1+bX_2) = a^2\sigma_{X_1}^2 + b^2\sigma_{X_2}^2 + 2ab\,Cov(X_1,X_2) \xrightarrow{X_2,X_1 \text{ مستقل}} a^2\sigma_{X_1}^2 + b^2\sigma_{X_2}^2$$

مثال ۳ متغیر X توزیع نرمال با میانگین μ و واریانس σ^2 دارد. از این جامعه نمونه‌ای به حجم 3 انتخاب می‌شود. کارایی تخمین‌زننده $\hat{X} = \frac{X_1+2X_2+3X_3}{6}$ نسبت به $\bar{X} = \frac{X_1+X_2+X_3}{3}$ چیست؟

۱) $\frac{6}{126}$ ۲) $\frac{36}{130}$ ۳) $\frac{36}{126}$ ۴) $\frac{36}{42}$

حل: گزینه ۴ درست است.

کارایی $\hat{X}$ به $\bar{X}$ برابر با واریانس $\bar{X}$ به $\hat{X}$ است.

$$\begin{cases} \dfrac{Var(\bar{X})}{Var(\hat{X})} = \dfrac{\frac{\sigma^2}{3}}{\frac{14}{36}\sigma^2} = \dfrac{36}{42} \\ Var(\hat{X}) = Var\left(\dfrac{X_1+2X_2+3X_3}{6}\right) = \dfrac{1}{36}\left(\sigma^2+4\sigma^2+9\sigma^2\right) = \dfrac{14}{36}\sigma^2 \\ Var(\bar{X}) = Var\left(\dfrac{X_1+X_2+X_3}{3}\right) = \dfrac{\sigma^2+\sigma^2+\sigma^2}{9} = \dfrac{\sigma^2}{3} \end{cases}$$

مثال ۴ سه تخمین‌زننده $T_1 = \frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3$، $T_2 = \frac{1}{4}X_1+\frac{1}{4}X_2+\frac{2}{4}X_3$ و $T_3 = \frac{1}{8}X_1+\frac{2}{8}X_2+\frac{5}{8}X_3$ وجود دارند. کدام رابطه بین واریانس آن‌ها برقرار است؟ (محیط زیست ـ ۸۶)

۱) $Var(T_1) < Var(T_2) < Var(T_3)$ ۲) $Var(T_2) < Var(T_3) < Var(T_1)$

۳) $Var(T_2) < Var(T_1) < Var(T_3)$ ۴) $Var(T_3) < Var(T_2) < Var(T_1)$

حل: گزینه ۱ درست است.

راه حل اول:

آماره $\bar{X} = \frac{X_1+X_2+X_3}{3} = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 = T_1$ همان میانگین نمونه است و بنا بر خواص میانگین کمترین مقدار واریانس را دارد و این موضوع تنها در گزینه ۱ دیده می‌شود.

راه حل دوم :

ابتدا واریانس هریک از سه تخمین‌زن را محاسبه کرده سپس آن‌ها را با هم مقایسه می‌کنیم:

$$\left.\begin{aligned} Var(T_1) &= \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 = \frac{1}{3}\sigma^2 \\ Var(T_2) &= \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{4}{16}\sigma^2 = \frac{6}{16}\sigma^2 \\ Var(T_3) &= \frac{1}{64}\sigma^2 + \frac{4}{64}\sigma^2 + \frac{25}{64}\sigma^2 = \frac{30}{64}\sigma^2 \end{aligned}\right\} \longrightarrow Var(T_1) < Var(T_2) < Var(T_3)$$

مثال ۵ اگر توزیع X نرمال بوده و دو تخمین‌زننده $S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$ و $\hat{\sigma}^2 = \frac{\sum(x_i-\mu)^2}{n}$ برای تخمین σ_X^2 مورد نظر باشد، به ازای $n = 10$ ضریب کارایی (نسبت واریانس $\hat{\sigma}^2$ به S^2) چیست؟ (اقتصاد ـ ۸۷)

۱) 0.19 ۲) 0.81 ۳) 0.9 ۴) 1.1

حل: گزینه ۳ درست است.

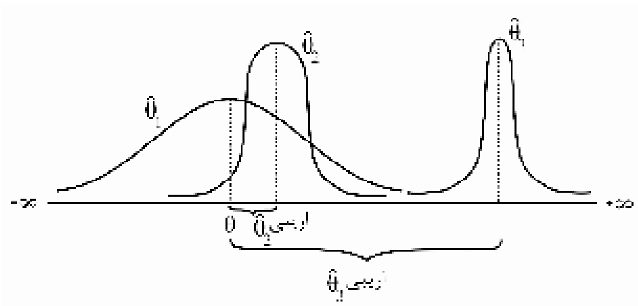
$$\frac{Var(\hat{\sigma}^2)}{Var(S^2)} = \frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}} = \frac{n-1}{n} = 0.9$$

دقت کنید، درصورتی‌که نسبت واریانس S^2 به $\hat{\sigma}^2$ خواسته می‌شد، آن‌گاه جواب برابر $\frac{n}{n-1} = 1.1$ و گزینه ۴ درست بود.

حداقل میانگین مجذور خطا (Mean Square Error)

همان‌طورکه در شرط **حداقل واریانس** بیان شد: «هنگام مقایسه بین دو برآوردکننده **ناریب**، برآوردکننده‌ای کاراتر است که **کمترین واریانس** «پراکندگی» را داشته باشد».

حال اگر بخواهیم دو برآوردکننده دلخواه (اریب یا ناریب) را با هم مقایسه کنیم، معیار انتخاب دیگر کمترین واریانس نخواهد بود، زیرا با توجه به شکل زیر در مورد برآوردکننده‌های $\hat{\theta}_1$ (ناریب) و $\hat{\theta}_2$ و $\hat{\theta}_3$ (اریب) می‌بینیم که:



$\hat{\theta}_3$: اگرچه کمترین واریانس را دارد ولی مقدار اریبی آن زیاد است، بنابراین مناسب نیست.

$\hat{\theta}_1$: اگرچه کمترین اریبی را دارد (نااریب) اما مقدار واریانس آن زیاد است، بنابراین قابل انتخاب نیست.

$\hat{\theta}_2$: دارای بهترین ترکیب از کمترین اریبی و کمترین واریانس است، بنابراین بهترین عملکرد را دارد.

با توجه به دلایل شهودی که در بالا مطرح شد، برای مقایسه برآوردکننده‌های دلخواه (اریب، نااریب) باید از ملاکی استفاده کنیم که هم واریانس و هم اریبی را در نظر بگیرد، یعنی هم پراکندگی برآوردکننده حول میانگین آن (واریانس) و هم پراکندگی برآوردکننده در اطراف پارامتر واقعی θ (اریبی) باید بررسی شود.

تعریف: بهترین ملاک برای مقایسه دو برآوردکننده دلخواه (اریب، نااریب) که همزمان اریبی و واریانس برآوردکننده‌ها را در نظر می‌گیرد، ملاک MSE یا میانگین مجذور خطا است و به صورت زیر محاسبه می‌شود:

$$\text{میانگین مجذور خطا}: \text{MSE}(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right] \qquad (۱)$$

که از بسط رابطه (۱) به رابطه محاسباتی‌تری به صورت زیر می‌رسیم:

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(E(\hat{\theta})-\theta\right)^2 = \text{واریانس} + (\text{اریبی،تورش})^2 = \text{مجذور اریبی} + \text{واریانس} \qquad (۲)$$

ملاک مقایسه

در بین دو برآوردکننده دلخواه **(اریب یا نااریب)** برآوردکننده‌ای **بهتر (کاراتر)** است که:

«حداقل MSE» یا **« حداقل میانگین مجذور خطا»** یا «**حداقل (واریانس** + ^{2}(**تورش**))» را داشته باشد.

نکته: با توجه به رابطه (2) برای برآوردکننده $\hat{\theta}$ یعنی:

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{اریبی})^2$$

الف) اگر $\hat{\theta}$ نااریب باشد، میانگین مجذور خطا (MSE) برابر با واریانس برآوردکننده نااریب $\hat{\theta}$ است.

$$\hat{\theta}\ \text{نااریب} \xrightarrow{\text{اریبی} = 0} \boxed{\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})}$$

ب) اگر $\hat{\theta}$ اریب باشد، میانگین مجذور خطا (MSE) بزرگ‌تر از واریانس برآوردکننده اریب $\hat{\theta}$ است.

$$\hat{\theta}\ \text{اریب} \xrightarrow{\text{اریبی} \neq 0} \boxed{\text{MSE}(\hat{\theta}) > \text{Var}(\hat{\theta})}$$

بنابراین همواره $\text{MSE}(\hat{\theta}) \geq \text{Var}(\hat{\theta})$ است.

نکته: با توجه به روابط مربوط به محاسبه میانگین مجذور خطا (MSE) همواره $\text{MSE}(\hat{\theta}) \geq 0$ است و در شرایطی که واریانس و اریبی صفر شود حداقل مقدار MSE حاصل شده و $\text{MSE}(\hat{\theta}) = 0$ خواهد شد. (رجوع شود به شرط سازگاری $\hat{\theta}$)

مثال ۱ کدام‌یک از عبارت‌های زیر بیان خطای متوسط مجذوری (MSE) تخمین‌زن (برآوردکننده) T برای پارامتر θ از قانون توزیع احتمال است؟ (اقتصاد ـ ۷۳)

۱) $[E(T)-\theta]^2$ ۲) $E(T-\theta)^2$ ۳) $E(T)-\theta$ ۴) $E(T^2)-\theta^2$

حل: گزینه ۲ درست است.

در بعضی مسایل، $\hat{\theta}$ را با T نمایش می‌دهند؛ بنابراین داریم:

$$MSE(T) = E(T-\theta)^2 = Var(T) + \left[\underbrace{E(T)-\theta}_{\text{اریبی}}\right]^2$$

دقت کنید، گزینه ۳ تعریف اریبی (تورش) است.

مثال ۲ برای برآوردکننده $Y = \dfrac{X_1 + 2X_2 + X_3}{4}$: (اقتصاد ـ ۷۳)

۱) $V(Y) = MSE(Y)$
۲) $MSE(Y) > V(Y)$
۳) $MSE(Y) < V(Y)$
۴) $E(Y) \neq \mu$

حل: گزینه ۱ درست است.

ازآنجاکه آماره Y نااریب است:

$$E(Y) = \frac{\mu + 2\mu + \mu}{4} = \mu$$

با توجه به حالت (الف) از نکته ۱ داریم:

$$MSE(Y) = Var(Y)$$

دقت کنید، اگر آماره Y اریب بود، آن‌گاه بنا بر حالت (ب) از نکته۱ گزینه ۲ درست بود.

مثال ۳ کدام‌یک از تعاریف زیردرمورد میانگین مجذور خطا (MSE) صحیح است؟ (اقتصاد ـ ۷۵)

۱) مجذور تورش یک برآوردکننده
۲) مجذور انحرافات از میانگین یک برآوردکننده اریب
۳) واریانس به اضافه اریب برآوردکننده
۴) واریانس یک برآوردکننده نااریب

حل: گزینه ۴ درست است.

برای آماره دلخواه: (MSE) میانگین مجذور خطا = واریانس + $(\text{اریبی})^2$

برای آماره نااریب: (MSE) میانگین مجذور خطا = واریانس + $(\underbrace{\cancel{\text{اریبی}}}_{0})^2$ = واریانس برآوردکننده نااریب

مثال ۴ چنانچه $\hat{\theta} = \bar{X} + \dfrac{1}{n}$ برآوردکننده‌ای از میانگین جامعه باشد، MSE این برآوردکننده برابر است با: (اقتصاد ـ ۷۵)

۱) $\dfrac{n\sigma_X^2 + 1}{n^2}$
۲) $\dfrac{\sigma_X^2 + 1}{n}$
۳) $\dfrac{\sigma_X^2 - 1}{n}$
۴) $\dfrac{n\sigma_X^2 + 1}{n}$

حل: گزینه ۱ درست است.

$$\begin{cases} MSE = Var(\hat{\theta}) + (\text{اریبی})^2 = Var(\hat{\theta}) + \left[\underbrace{E(\hat{\theta})-\theta}_{\text{اریبی}}\right]^2 = \dfrac{\sigma^2}{n} + \left(\dfrac{1}{n}\right)^2 = \dfrac{n\sigma^2+1}{n^2} \\ Var(\hat{\theta}) = Var(\bar{X} + \underbrace{\cancel{\tfrac{1}{n}}}_{0}) = Var(\bar{X}) = \dfrac{\sigma^2}{n} \\ \text{اریبی} = E(\hat{\theta}) - \theta = E\left(\bar{X} + \dfrac{1}{n}\right) - \mu = E(\bar{X}) + \dfrac{1}{n} - \mu = \mu + \dfrac{1}{n} - \mu = \dfrac{1}{n} \end{cases}$$

مثال ۵ اگر بخواهیم بهترین برآوردکننده را از بین برآوردکننده‌های اریب و نااریب انتخاب کنیم، معیار گزینش عبارت خواهد بود از: (اقتصاد ـ ۷۶)

۱) کمترین واریانس ۲) کمترین میانگین مجذور خطا (MSE)

۳) نااریب بودن ۴) نااریب بودن به علاوه نااریب بودن مجانبی

حل: گزینه ۲ درست است.

در بین دو برآوردکننده دلخواه (اریب و نااریب) برآوردکننده‌ای بهتر (کاراتر) است که MSE (میانگین مجذور خطا) کمتری داشته باشد یعنی دارای حداقل واریانس و اریبی باشد.

مثال ۶ به منظور برآورد میانگین جامعه بر اساس یک نمونه تصادفی سه‌تایی، دو برآوردکننده A و B زیر پیشنهاد شده است. برای تشخیص آنکه کدام‌یک برآوردکننده بهتری است، چه ملاکی کفایت می‌کند؟ (اقتصاد ـ ۸۷)

$$B=\frac{X_1+X_2+X_3}{3}+2 \qquad A=\frac{X_1+X_2+X_3}{2}$$

۱) تورش ۲) واریانس ۳) واریانس + $(\text{تورش})^2$ ۴) واریانس + تورش

حل: گزینه ۳ درست است.

ازآنجاکه برآورد کننده های A و B اریب هستند:

$$E(B)=\frac{\mu+\mu+\mu}{3}+2=\mu+2\neq\mu$$

$$E(A)=\frac{\mu+\mu+\mu}{2}=\frac{3\mu}{2}\neq\mu$$

بهترین ملاک مقایسه «واریانس + $(\text{تورش})^2$ = MSE» است.

مثال ۷ به منظور برآورد میانگین جامعه بر اساس یک نمونه تصادفی 2 تایی، برآوردکننده‌های A و B زیر پیشنهاد شده‌اند. برای تشخیص آنکه کدام‌یک مناسب‌تر است چه ملاکی کفایت می‌کند؟ (اقتصاد ـ ۸۵)

$$A=\frac{2X_1+3X_2}{5} \quad , \quad B=\frac{X_1+X_2}{2}+2$$

۱) تورش ۲) واریانس ۳) واریانس+$(\text{تورش})^2$ ۴) واریانس + تورش

حل: گزینه ۳ درست است.

ازآنجاکه برآوردکننده A نااریب و B اریب است:

$$E(A)=\frac{2\mu+3\mu}{5}=\mu$$

$$E(B)=\frac{\mu+\mu}{2}+2=\mu+2\neq\mu$$

بهترین ملاک مقایسه «واریانس + $(\text{تورش})^2$ = MSE» است.

مثال ۸ دو برآوردکننده $\hat\theta_1$ و $\hat\theta_2$ با ویژگی‌های زیر برای برآورد پارامتر θ پیشنهاد شده است: $Var(\hat\theta_2)=50$ و $E(\hat\theta_2-\theta)=6$ و $Var(\hat\theta_1)=90$ و $E(\hat\theta_1-\theta)=0$ آن‌گاه: (اقتصاد ـ ۸۶)

۱) $\hat\theta_2$ مناسب‌تر است چون کاراتر است.

۲) $\hat\theta_1$ مناسب‌تر است زیرا یک برآوردکننده نااریب (بدون تورش) است.

۳) $\hat\theta_1$ مناسب‌تر است چون میانگین مجذور خطای آن (MSE) کمتر است.

۴) $\hat\theta_2$ مناسب‌تر است چون میانگین مجذور خطای آن (MSE) کمتر است.

حل: گزینه ۴ درست است.

ازآنجاکه:

$\hat{\theta}_1$ اریبی آماره $: E(\hat{\theta}_1 - \theta) = 0 \quad \rightarrow$ نااریب

$\hat{\theta}_2$ اریبی آماره $: E(\hat{\theta}_2 - \theta) = 6 \quad \rightarrow$ اریب

آماره $\hat{\theta}_1$ نااریب و $\hat{\theta}_2$ اریب است بنابراین بهترین ملاک مقایسه استفاده از میانگین مجذور خطای (MSE) دو آماره به شرح زیر است:

$$\begin{cases} MSE(\hat{\theta}_1) = Var(\hat{\theta}_1) + \left[E(\hat{\theta}_1 - \theta)\right]^2 = 90 + 0^2 = 90 \\ MSE(\hat{\theta}_2) = Var(\hat{\theta}_2) + \left[E(\hat{\theta}_2 - \theta)\right]^2 = 50 + 6^2 = 86 \end{cases} \longrightarrow MSE(\hat{\theta}_2) < MSE(\hat{\theta}_1)$$

مثال ۹ وقتی می‌گوییم برآوردکننده $\hat{\theta}_1$ به طور مجانبی کاراتر از $\hat{\theta}_2$ است، مفهوم آن است که: (اقتصاد ـ ۷۰)

۱) $\lim\limits_{n \to \infty} E(\hat{\theta}_1) < \lim\limits_{n\to\infty} E(\hat{\theta}_2)$

۲) $E(\hat{\theta}_1) < E(\hat{\theta}_2)$

۳) $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$

۴) $\lim\limits_{n \to \infty} V(\hat{\theta}_1) < \lim\limits_{n \to \infty} V(\hat{\theta}_2)$

حل: گزینه ۳ درست است.

به طور مجانبی یعنی در شرایطی که $n \to \infty$ برود.

می‌دانیم که هرگاه $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$ باشد $\hat{\theta}_1$ به طور نسبی کاراتر از $\hat{\theta}_2$ است. در شرایطی هم که $n \to \infty$ می‌رود همین اصل صادق است یعنی هرگاه $\lim\limits_{n \to \infty} MSE(\hat{\theta}_1) < \lim\limits_{n \to \infty} MSE(\hat{\theta}_2)$ باشد $\hat{\theta}_1$ به طور مجانبی کاراتر از $\hat{\theta}_2$ است. چنین عبارتی در گزینه‌ها نیست اما به این نکته توجه کنیم که اگر در صورت سؤال گفته می‌شد $\hat{\theta}_1$ و $\hat{\theta}_2$ نااریب‌اند و یا به طور مجانبی نااریب‌اند، گزینه ۴ می‌توانست جواب درست باشد چون می‌دانیم در شرایطی که $\hat{\theta}$ نااریب باشد $MSE(\hat{\theta})$ با $Var(\hat{\theta})$ برابر است.

$$MSE(\hat{\theta}) = (\text{اریبی})^2 + Var(\hat{\theta})$$

سازگاری (پایداری) (Consistency)

برآوردکننده‌ای سازگار است که وقتی حجم نمونه (n) خیلی بزرگ شود احتمال اینکه تخمین به‌دست‌آمده به پارامتر جامعه نزدیک باشد به یک خیلی نزدیک شود. به عبارت دیگر با افزایش حجم نمونه، توزیع نمونه‌گیری حول پارامتر جامعه متمرکز و متمرکزتر شود، یعنی وقتی n به سمت بی‌نهایت میل می‌کند، واریانس توزیع نمونه‌گیری نیز به سمت صفر میل کند و نهایتاً توزیع بر روی پارامتر جامعه فرو ریزد. به عبارت دیگر، آماره $\hat{\theta}$ خصوصیت سازگاری را خواهد داشت، هرگاه یکی از تعاریف ۱ یا ۲ یا ۳ برای آن برقرار باشد:

تعریف ۱: برآوردکننده $\hat{\theta}$ برای تخمین پارامتر θ سازگار است هرگاه:

اگر حجم نمونه به طور نامتناهی افزایش یابد $(n \to \infty)$ آن‌گاه آماره $\hat{\theta}$ به سمت پارامتر θ میل کند ($\hat{\theta}$ روی θ متمرکز شود):

آماره $\hat{\theta}$ سازگار (پایدار) $\quad \rightleftarrows \quad \lim\limits_{n\to\infty} \hat{\theta} = \theta$

نتیجه:

با توجه به این تعریف می‌توانیم تمرکز $\hat{\theta}$ روی θ را به صورت حدی و به کمک احتمال (قضیه چی‌بی‌شف) به صورت زیر بیان کنیم:

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1 \rightleftarrows \text{آماره } \hat{\theta} \text{ سازگار (پایدار)}$$

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right|>\varepsilon\right)=0 \rightleftarrows \text{آماره } \hat{\theta} \text{ سازگار (پایدار)}$$

مثال ۱ آماره $\bar{X}$ یک آماره سازگار است چون وقتی n به سمت بی‌نهایت میل می‌کند $\bar{x}$ به سمت .... میل می‌کند؟ (مدیریت ـ ۸۳)

۱) ∞ ۲) $N\mu_X$ ۳) μ_X ۴) 0

حل: گزینه ۳ درست است.

ازآنجاکه آماره $\bar{X}$ (میانگین نمونه) برای پارامتر μ (میانگین جامعه) سازگار است، داریم:

$$\lim_{n\to\infty} \bar{x}=\mu$$

مثال ۲ آماره $\bar{p}$ یک آماره سازگار است چون وقتی n به سمت بی‌نهایت میل می‌کند، $\bar{p}$ به سمت .... میل می‌کند.

۱) ∞ ۲) Np ۳) p ۴) 0

حل: گزینه ۳ درست است.

ازآنجاکه آماره $\bar{p}$ (نسبت نمونه) برای پارامتر p (نسبت جامعه) سازگار است، داریم:

$$\lim_{n\to\infty} \bar{p}=p$$

مثال ۳ آماره S^2 (واریانس نمونه) یک آماره سازگار است چون وقتی n به سمت بی‌نهایت میل می‌کند، S^2 به سمت .... میل می‌کند.

۱) ∞ ۲) σ^2 ۳) $\frac{\sigma^2}{N}$ ۴) 0

حل: گزینه ۲ درست است.

ازآنجاکه آماره S^2 (واریانس نمونه) برای پارامتر σ^2 (واریانس جامعه) سازگار است، داریم:

$$\lim_{n\to\infty} S^2=\sigma^2$$

توجه: منظور از آماره S^2 در اینجا $\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}$ یا $\frac{\sum\left(x_i-\mu\right)^2}{n}$ است.

مثال ۴ رابطه $\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1$ بیان‌کننده کدام‌یک از موارد زیر است؟ (اقتصاد ـ ۷۱)

۱) سازگاری (پایداری) برآوردنقطه‌ای ۲) ناتور بودن (ناریب بودن) برآورد نقطه‌ای

۳) کافی بودن برآورد نقطه‌ای ۴) کافی بودن

حل: گزینه ۱ درست است.

تعریف ۲: برآوردکننده $\hat{\theta}$ برای تخمین پارامتر θ سازگار است، هرگاه، اگر حجم نمونه به طور نامتناهی افزایش یابد $(n \to \infty)$ آن‌گاه واریانس و اریبی آماره $\hat{\theta}$ به سمت صفر میل کند.

$$\text{آماره } \hat{\theta} \text{ سازگار (پایدار)} \iff \begin{cases} \lim\limits_{n\to\infty} Var\left(\hat{\theta}\right)=0 \\ \lim\limits_{n\to\infty} \text{اریبی}=0 \qquad \left(\text{نااریب } E\left(\hat{\theta}\right)=\theta\right) \end{cases}$$

مثال ۱ آماره $\bar{X}$ یک آماره سازگار است چون وقتی n به سمت بی‌نهایت میل می‌کند واریانس $\bar{X}$ به سمت ……… میل می‌کند.

۱) ∞ ۲) $N\mu_X$ ۳) μ_X ۴) 0

حل: گزینه ۴ درست است.

ازآنجاکه آماره $\bar{X}$ به عنوان برآوردکننده μ دارای واریانس $\sigma_{\bar{X}}^2=\frac{\sigma_X^2}{n}$ است، داریم:

$$\lim_{n\to\infty}\sigma_{\bar{X}}^2=\lim_{n\to\infty}\frac{\sigma_X^2}{n}=0$$

مثال ۲ آماره $\bar{p}$ یک آماره سازگار است، چون وقتی n به سمت بی‌نهایت میل می‌کند واریانس $\bar{p}$ به سمت ……… میل می‌کند.

۱) ∞ ۲) Np ۳) p ۴) 0

حل: گزینه ۴ درست است.

ازآنجاکه آماره $\bar{p}$ به عنوان برآوردکننده p دارای واریانس $\sigma_{\bar{p}}^2=\frac{pq}{n}$ است، داریم:

$$\begin{cases}\lim\limits_{n\to\infty}\sigma_{\bar{p}}^2=\lim\limits_{n\to\infty}\frac{pq}{n}=0\\ (0\le p\,,\,q\le 1)\end{cases}$$

مثال ۳ آماره $S^2=\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}$ (واریانس نمونه) یک آماره سازگار است چون وقتی n به سمت بی‌نهایت میل می‌کند واریانس S^2 به سمت ……… میل می‌کند.

۱) ∞ ۲) σ^2 ۳) $\frac{\sigma^2}{n}$ ۴) 0

حل: گزینه ۴ درست است.

ازآنجاکه آماره S^2 به عنوان برآوردکننده σ^2 دارای واریانس $Var\left(S^2\right)=\frac{2\sigma^4}{n-1}$ است، داریم:

$$\lim_{n\to\infty}Var\left(S^2\right)=\lim_{n\to\infty}\frac{2\sigma^4}{n-1}=0$$

یادآوری: آماره $S^2=\frac{\sum\left(x_i-\mu\right)^2}{n}$ نیز سازگار است زیرا:

$$\lim_{n\to\infty}Var\left(S^2\right)=\lim_{n\to\infty}\frac{2\sigma^4}{n}=0$$

مثال۴ اگر $\hat{\theta}$ برآوردکننده ناریب و سازگار θ باشد، کدام‌یک از موارد زیر درست است؟

۱) $E\left(\frac{1}{\hat{\theta}}\right)=\frac{1}{\theta}$ ۲) $V\left(\frac{1}{\hat{\theta}}\right)=\theta$ ۳) $\lim\limits_{n\to\infty} P\left(\frac{1}{\hat{\theta}}\right)=\frac{1}{\theta}$ ۴) $MSE\left(\frac{1}{\hat{\theta}}\right)=V\left(\frac{1}{\hat{\theta}}\right)$

حل: گزینه ۳ درست است.

✓ **دقت کنید!**

اولاً، برای سازگار بودن یک آماره حتماً شرط نمونه بزرگ $(n\to\infty)$ لازم است.

ثانیاً، اگر $\hat{\theta}$ به طور دقیق بر θ منطبق شود، آن‌گاه حتماً $\frac{1}{\hat{\theta}}$ نیز بر $\frac{1}{\theta}$ منطبق خواهد شد.

تعریف۳: برآوردکننده $\hat{\theta}$ برای تخمین پارامتر θ سازگار است، هرگاه، اگر حجم نمونه به طور نامتناهی افزایش یابد $(n\to\infty)$، میانگین مجذور خطا (MSE) حداقل مقدار خود را اختیار کرده و برابر صفر شود $(MSE=0)$.

$$MSE\left(\hat{\theta}\right)=\cancel{Var\left(\hat{\theta}\right)}^{\,0}+\cancel{(\text{اریبی})^2}^{\,0}=0 \iff \begin{cases}\lim\limits_{n\to\infty} Var\left(\hat{\theta}\right)=0\\ \lim\limits_{n\to\infty} \text{اریبی}=0\end{cases} \iff \text{آماره } \hat{\theta} \text{ سازگار(پایدار)}$$

یادآوری: میانگین مجذور خطا (MSE) همواره $MSE\geq 0$ است و زمانی که حداقل مقدار خود را برای یک آماره اختیار کند بهترین عملکرد را خواهد داشت.

مثال وقتی می‌گوییم برآوردکننده (تخمین‌زن)، $\hat{\theta}_1$ سازگار یا پایدار است، به مفهوم آن است که: (اقتصاد ـ ۷۰)

۱) اریب و واریانس $\hat{\theta}_1$ با بزرگ شدن n به سمت صفر میل می‌کند.

۲) امید ریاضی $\hat{\theta}_1$ و واریانس آن برابر صفر است.

۳) امید ریاضی $\hat{\theta}_1$ برابر θ و واریانس آن برابر صفر است.

۴) میانگین مجذور خطای این برآوردکننده برابر صفر است.

حل: گزینه ۱ درست است.

با توجه به تعریف ۲، سازگاری گزینه ۱ تمام شرایط را داشته و اما گزینه ۳ با توجه به تعریف ۲ سازگاری شرط مربوط به $n\to\infty$ (n بزرگ) را مطرح نکرده است و گزینه ۴ نیز با توجه به تعریف ۳ شرط مربوط به $n\to\infty$ (n بزرگ) را مطرح نکرده است، بنابراین گزینه‌های ۳ و ۴ درست نیستند.

روش‌های برآورد نقطه‌ای

برای تخمین پارامتر مجهول θ به صورت نقطه‌ای می‌توان به یکی از روش‌های زیر عمل کرد:

۱- روش گشتاوری

۲- روش حداکثر درستنمایی

۱- روش گشتاوری (Method of Moment Estimation)

درصورتی‌که به دنبال تخمین پارامترهای مجهول جامعه از روش گشتاوری باشیم به ترتیب زیر عمل می‌کنیم:

الف) به تعداد پارامترهای مجهول خواسته‌شده، **امید ریاضی** را **برحسب پارامتر مجهول** با استفاده ازتابع داده‌شده به شرح زیر به دست می‌آوریم:

اگر 1 پارامتر مجهول باشد، $E(X)$ را محاسبه می‌کنیم:

$$E(X)=\sum\!\!\!\!\int xf(x)dx$$

اگر 2 پارامتر مجهول باشد، $E(X)$ و $E(X^2)$ را محاسبه می‌کنیم:

$$E(X)=\sum\!\!\!\!\int xf(x)dx \quad , \quad E(X^2)=\sum\!\!\!\!\int x^2f(x)dx$$

و به همین ترتیب اگر n پارامتر مجهول باشد، $E(X)$ تا $E(X^n)$ را محاسبه می‌کنیم.

ب) به تعداد پارامترهای مجهول خواسته‌شده، **میانگین نمونه** را که مقداری **معلوم** است به شرح زیر محاسبه می‌کنیم:

اگر 1 پارامتر مجهول باشد، $\bar{X}$ را محاسبه می‌کنیم:

$$\bar{X}=\frac{\sum X_i}{n}$$

اگر 2 پارامتر مجهول باشد، $\bar{X}$ و $\overline{X^2}$ را محاسبه می‌کنیم:

$$\bar{X}=\frac{\sum X_i}{n} \quad , \quad \overline{X^2}=\frac{\sum X_i^2}{n}$$

و به همین ترتیب اگر n پارامتر مجهول باشد، $\bar{X}$ تا $\overline{X^n}$ را محاسبه می‌کنیم.

ج) با توجه به روابط به‌دست‌آمده از مراحل (الف) و (ب)، دستگاه معادلات زیر را تشکیل می‌دهیم و پارامترهای مجهول را برحسب میانگین نمونه محاسبه می‌کنیم:

اگر 1 پارامتر مجهول باشد، مقدار پارامتر را از معادله زیر برآورد می‌کنیم:

$$E(X)=\bar{X}$$

اگر 2 پارامتر مجهول باشد، مقدار پارامترها را از دو معادله زیر برآورد می‌کنیم:

$$E(X)=\bar{X}\ ,\ E(X^2)=\overline{X^2}$$

و به همین ترتیب اگر n پارامتر مجهول باشد، مقدار پارامترها را از n معادله زیر برآورد می‌کنیم:

$$E(X)=\bar{X}\ ,\ \dots\ ,\ E(X^n)=\overline{X^n}$$

✓ **دقت کنید!**

اگر داده‌های نمونه در مسئله وجود نداشته باشد، از همان $\bar{X}$ و $\overline{X^2}$ و ... استفاده می‌کنیم.

مثال ۱ نمونه تصادفی $0.8, 0.3, 0.4, 0.5$ را از توزیعی با چگالی زیردر نظر می‌گیریم. یک برآورد گشتاوری پارامتر θ چیست؟

$$f(x;\theta)=\begin{cases}\frac{2}{\theta^2}(\theta-x) & 0<x<\theta\\ 0 & \text{سایر نقاط} ,\end{cases}$$

۱) $\frac{1}{3}$ ۲) 3 ۳) $\frac{3}{2}$ ۴) $\frac{1}{2}$

حل: گزینه ۳ درست است.

تنها یک پارامتر مجهول (θ) خواسته شده است، بنابراین:

الف) $$E(X)=\int_0^\theta x\cdot f(x)dx=\int_0^\theta \frac{2x}{\theta^2}(\theta-x)dx=\frac{1}{\theta^2}\left[\theta x^2-\frac{2}{3}x^3\right]_0^\theta=\frac{\theta}{3}$$

ب) $$\bar{X}=\frac{\sum X_i}{n}=\frac{0.3+0.4+0.5+0.8}{4}=0.5=\frac{1}{2}$$

ج) با استفاده از روابط به دست آمده از روابط (الف) و (ب) دستگاه زیر را تشکیل می‌دهیم:

$$E(X)=\bar{X}\ \rightarrow\ \frac{\theta}{3}=\frac{1}{2}\ \rightarrow \theta=\frac{3}{2}$$

مثال ۲ نمونه تصادفی $X_1,\dots,X_n$ را از توزیعی با چگالی زیر مشاهده کرده‌ایم، برآورد گشتاوری m کدام است؟

$$f(x;m)=\begin{cases} mx^{m-1} & , \quad 0<x<1 \\ 0 & , \quad \text{سایر نقاط} \end{cases}$$

۱) $\frac{\bar{x}}{1+\bar{x}}$ ۲) $\frac{\bar{x}}{1-\bar{x}}$ ۳) $\frac{1+\bar{x}}{\bar{x}}$ ۴) $\frac{1-\bar{x}}{\bar{x}}$

حل: گزینه ۲ درست است.

تنها یک پارامتر مجهول خواسته شده است، بنابراین:

الف) $$E(X)=\int_0^1 x\cdot f(x)dx=\int_0^1 x\cdot mx^{m-1}dx=\int_0^1 mx^m dx=\frac{m}{m+1}\left[x^{m+1}\right]_0^1=\frac{m}{m+1}$$

ب) $$\bar{X}=\frac{\sum X_i}{n}$$

ج) با استفاده از روابط به‌دست‌آمده از مراحل (الف) و (ب) دستگاه زیر را تشکیل می‌دهیم:

$$E(X)=\bar{X}\ \rightarrow\ \frac{m}{m+1}=\bar{x}\ \rightarrow\ m\bar{x}+\bar{x}=m\ \rightarrow\ m-m\bar{x}=\bar{x}\ \rightarrow\ m=\frac{\bar{x}}{1-\bar{x}}$$

با توجه به آنکه داده‌های نمونه موجود نیست از $\bar{x}$ در رابطه استفاده می‌کنیم و θ را با استفاده از آن به دست می‌آوریم.

مثال ۳ نمونه تصادفی $0.5, 0.3, 0.4, 0.2, 0.6$ را از تابع توزیع $F(x;\theta)=x^\theta\ ,\ 0<x<1\ ,\ \theta>0$ مشاهده کرده‌ایم. برآورد گشتاوری θ برابر است با:

۱) $\frac{1}{3}$ ۲) $\frac{3}{4}$ ۳) $\frac{2}{3}$ ۴) $\frac{1}{5}$

حل: گزینه ۳ درست است.

یادآوری: اگر X متغیر تصادفی پیوسته، $f_X(x)$ تابع چگالی و $F_X(x)$ تابع توزیع تجمعی باشد، داریم:

$$f_X(x)=F'_X(x)=\left(x^\theta\right)'=\theta x^{\theta-1}$$

تنها یک پارامتر مجهول خواسته شده است، بنابراین:

الف) $$E(X)=\int_0^1 x\cdot f(x)dx=\int_0^1 x\cdot\theta x^{\theta-1}dx=\int_0^1 \theta x^\theta dx=\frac{\theta}{\theta+1}\left[x^{\theta+1}\right]_0^1=\frac{\theta}{\theta+1}$$

ب) $$\bar{X}=\frac{\sum X_i}{n}=\frac{0.2+0.6+0.4+0.3+0.5}{5}=0.4$$

ج) با استفاده از روابط به دست آمده از مراحل (الف) و (ب) دستگاه زیر را تشکیل می‌دهیم:

$$E(X)=\bar{X} \rightarrow \frac{\theta}{\theta+1}=0.4 \rightarrow 0.4\theta+0.4=\theta \rightarrow 0.6\theta=0.4 \rightarrow \theta=\frac{2}{3}$$

مثال ۴ اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی n تایی از توزیع دوجمله‌ای با پارامترهای n و p (p معلوم) باشد، آن‌گاه برآوردگر n با روش گشتاوری کدام است؟

۱) $\frac{\bar{X}}{p}$ ۲) $\frac{p}{\bar{X}}$ ۳) $p\bar{X}$ ۴) $\frac{\bar{X}}{1-p}$

حل: گزینه ۱ درست است.

یادآوری: در توزیع دوجمله‌ای با پارامترهای n و p داریم:

$$f(x)=\binom{n}{x}p^x q^{n-x} \; ; x=0,1,...,$$
$$E(X)=np$$

تنها یک پارامتر مجهول وجود دارد (p معلوم است و فقط n مجهول است)، بنابراین:

الف)
$$E(X)=np$$

ب)
$$\bar{X}=\frac{\sum X_i}{n}$$

ج) با استفاده از روابط به‌دست‌آمده از مراحل (الف) و (ب) دستگاه زیر را تشکیل می‌دهیم:

$$E(X)=\bar{X} \rightarrow np=\bar{x} \rightarrow n=\frac{\bar{X}}{p}$$

با توجه به آنکه داده‌های نمونه موجود نیست از $\bar{X}$ در رابطه بالا استفاده می‌کنیم و n را با استفاده از آن به دست می‌آوریم.

✓ **دقت کنید!**

اگر n معلوم و p مجهول باشد، برآورد گشتاوری p در توزیع دوجمله‌ای عبارت است از:
$$\hat{p}=\frac{\bar{X}}{n}$$

قضیه: اگر X دارای توزیع نرمال با میانگین مجهول μ و واریانس مجهول σ^2 باشد، در این صورت برآورد گشتاوری این پارامترها عبارت است از:

$$\hat{\mu}=\bar{X}$$
$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}=\overline{X^2}-\bar{X}^2$$

اثبات:

طبق روش گشتاوری برای برآورد دو پارامتر داریم:

الف)

$$\begin{cases} E(X)=\mu \\ E(X^2)=\sigma^2+\mu^2 \end{cases}$$

برای محاسبه $E(X^2)$، از رابطه واریانس به صورت زیر استفاده می‌کنیم:

$$\sigma^2=E(X^2)-E(X)^2 \rightarrow E(X^2)=\sigma^2+\mu^2$$

ب)

$$\begin{cases} \bar{X} = \dfrac{\sum X_i}{n} \\ \overline{X^2} = \dfrac{\sum X_i^2}{n} \end{cases}$$

ج)

$$\begin{cases} E(X) = \bar{X} \\ E\left(X^2\right) = \overline{X^2} \end{cases} \to \begin{cases} \mu = \bar{X} \\ \sigma^2 + \underbrace{\cancel{\mu^2}}_{\bar{X}^2} = \overline{X^2} \end{cases} \to \sigma^2 = \overline{X^2} - \bar{X}^2$$

پس:

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \overline{X^2} - \bar{X}^2 = \dfrac{\sum X_i^2}{n} - \left(\dfrac{\sum X_i}{n}\right)^2 = \dfrac{\sum\left(X_i - \bar{X}\right)^2}{n} \end{cases}$$

تبصره: در شرایطی که میانگین جامعه (μ) معلوم باشد در برآورد گشتاوری واریانس و انحراف معیار، در رابطه بالا به‌جای $\bar{X}$ از مقدار μ استفاده می‌شود.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n} = \overline{X^2} - \mu^2$$

مثال۵ فرض کنید $1,2,3,4,5$ یک نمونه تصادفی از چگالی نرمال با میانگین μ و واریانس σ^2 باشد، برآورد گشتاوری برای μ و σ^2 برابر است با:

۱) $\hat{\mu} = 3\,,\ \sigma^2 = 2$ ۲) $\hat{\mu} = 3\,,\ \hat{\sigma}^2 = 2.5$ ۳) $\hat{\mu} = 4\,,\hat{\sigma}^2 = 2$ ۴) $\hat{\mu} = 4\,,\hat{\sigma}^2 = 2.5$

حل: گزینه ۱ درست است.

$$\begin{cases} \hat{\mu} = \bar{X} = 3 \\ \hat{\sigma}^2 = \overline{X^2} - \bar{X}^2 = 11 - 3^2 = 2 \\ \bar{X} = \dfrac{\sum X_i}{n} = \dfrac{1+2+3+4+5}{5} = 3 \\ \overline{X^2} = \dfrac{\sum X_i^2}{n} = \dfrac{1^2+2^2+3^2+4^2+5^2}{5} = \dfrac{55}{5} = 11 \end{cases}$$

درعین‌حال برای برآورد واریانس از رابطه زیر نیز می‌توان استفاده کرد:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n} = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5} = \frac{10}{5} = 2$$

۲- روش حداکثر درستنمایی (Maximum Likelihood Estimation)

اگر $f_X(x;\theta)$ تابع احتمال متغیر تصادفی X با پارامتر مجهول θ از جامعه‌ای باشد، برای تخمین پارامتر θ به روش «حداکثر درستنمایی (MLE)» به صورت زیر عمل می‌کنیم:

الف) تابع احتمال درستنمایی $L(\theta)$ را بر اساس یک نمونه مستقل n تایی $(X_n,\dots,X_2,X_1)$ به صورت زیر تشکیل می‌دهیم:

$$L(\theta)=f(x_1,x_2,\dots,x_n;\theta)=f(x_1;\theta)\times\dots\times f(x_n;\theta)=\prod_{i=1}^{n}(x_i;\theta)$$

تبصره ۱: در شرایطی که تابع درستنمایی به‌دست‌آمده در قسمت (الف) به صورت توانی باشد، برای راحتی محاسبات در مرحله بعد می‌توان از آن لگاریتم (Ln) گرفت.

ب) حال باید به دنبال نقطه‌ای باشیم که به ازای آن تابع احتمال درستنمایی $L(\theta)$ ماکزیمم شود؛ این نقطه همان «برآورد حداکثر درستنمایی» برای پارامتر مجهول θ است.

تبصره ۲: یک روش معمول برای به دست آوردن نقطه ماکزیمم آن است که از تابع مشتق گرفته و آن را برابر صفر قرار دهیم؛ در این وضعیت نقطه به‌دست‌آمده (بحرانی) می‌تواند همان مقدار ماکزیمم تابع درستنمایی باشد؛ به عبارت دیگر:

$$L'(\theta)=0 \quad \text{یا} \quad (\ln L(\theta))'=0 \rightarrow \theta=\text{برآورد حداکثر درستنمایی}$$

یادآوری: در توابع لگاریتمی داریم:

$$\begin{cases} 1)\ (\ln u)'=\dfrac{u'}{u} \\ 2)\ \ln\left(\prod_{i=1}^{n}X_i\right)=\ln(x_1\times x_2\times\dots\times x_n)=\ln x_1+\ln x_2+\dots+\ln x_n=\sum_{i=1}^{n}\ln x_i \\ 3)\ \ln\left(\dfrac{x}{y}\right)=\ln x-\ln y \end{cases}$$

مثال ۱ متغیر تصادفی X دارای جدول توزیع احتمال زیر است:

$X=x$	-1	0	1
$P_\theta(X=x)$	$\frac{\theta}{2}$	$1-\theta$	$\frac{\theta}{2}$

$(0\le\theta\le1)$

نمونه تصادفی $-1,1,-1,0$ را از X مشاهده کرده‌ایم. برآورد درستنمایی ماکزیمم (MLE) برای θ برابر است با:

۱) $\frac{1}{4}$ ۲) $\frac{1}{2}$ ۳) $\frac{1}{3}$ ۴) $\frac{3}{4}$

حل: گزینه ۴ درست است.

الف) ابتدا تابع درستنمایی $L(\theta)$ را بر اساس نمونه مستقل $(-1,1,-1,0)$ به صورت زیر به دست می‌آوریم:

$$L(\theta)=f(0;\theta)\times f(-1;\theta)\times f(1;\theta)\times f(-1;\theta)=(1-\theta)\times\frac{\theta}{2}\times\frac{\theta}{2}\times\frac{\theta}{2}=(1-\theta)\frac{\theta^3}{8}$$

ب) حال برای به دست آوردن حداکثر درستنمایی، $L'(\theta)=0$ را تشکیل می‌دهیم:

$$L'(\theta)=0\rightarrow\left((1-\theta)\frac{\theta^3}{8}\right)'=0 \rightarrow -\frac{\theta^3}{8}+(1-\theta)\frac{3\theta^2}{8}=0\rightarrow\frac{\theta^2}{8}(-\theta+3(1-\theta))=0$$

$$\rightarrow\frac{\theta^2}{8}(-\theta+3-3\theta)=0\rightarrow\frac{\theta^2}{8}(-4\theta+3)=0\rightarrow\begin{cases}\theta=0\\ \theta=\dfrac{3}{4}\end{cases}$$

مثال ۲ اگر $X_1,X_2,\dots,X_n$ یک نمونه تصادفی به اندازه n از توزیعی با تابع چگالی $f(x)=\theta(1-\theta)^x\ ,\ x=0,1,2,3,\dots$ باشد، برآورد θ^{-1} به روش درستنمایی ماکزیمم (MLE) برابر است با:

۱) $\frac{1}{\bar{X}+1}$ ۲) $\frac{1}{\bar{X}}$ ۳) $\bar{X}+1$ ۴) $\bar{X}$

حل: گزینه ۳ درست است.

الف) ابتدا تابع درستنمایی $L(\theta)$ را به صورت زیر به دست می‌آوریم:

$$L(\theta)=f(X_1;\theta)\times\ldots\times f(X_n;\theta)=\theta(1-\theta)^{X_1}\times\ldots\times\theta(1-\theta)^{X_n}=\theta^n(1-\theta)^{\sum_{i=1}^{n}X_i}$$

با توجه به توانی بودن تابع برای راحتی کار در مرحله بعد از طرفین رابطه بالا Ln می‌گیریم:

$$\operatorname{Ln}L(\theta)=n\ln\theta+\sum_{i=1}^{n}X_i\ln(1-\theta)$$

ب) حال برای به دست آوردن حداکثر درستنمایی $(\ln L(\theta))'=0$ را تشکیل می‌دهیم:

$$(\ln L(\theta))'=0\ \rightarrow\ \left(\frac{n}{\theta}+-\frac{\sum X_i}{1-\theta}\right)=0\ \rightarrow\ \frac{n}{\theta}=\frac{\sum X_i}{1-\theta}\ \rightarrow\ \frac{\sum X_i}{n}=\frac{1-\theta}{\theta}\ \rightarrow\ \bar{X}=\frac{1-\theta}{\theta}$$

$$\rightarrow\ \theta=\frac{1}{1+\bar{X}}\ \rightarrow\ \boxed{\frac{1}{\theta}=\bar{X}+1}$$

مثال ۳ فرض کنید $X_1,\ldots,X_n$ یک نمونه تصادفی از توزیعی با تابع چگالی احتمال $f(x;\theta)=\theta x^{-\theta-1}\ ,\ x>0$ باشد. MLE برای θ کدام است؟

۱) $\frac{1}{\bar{X}}$
۲) $\frac{n}{\prod_{i=1}^{n}X_i}$
۳) $\frac{n}{\sum_{i=1}^{n}\ln X_i}$
۴) $\frac{\prod_{i=1}^{n}X_i}{n}$

حل: گزینه ۳ درست است.

الف) ابتدا تابع درستنمایی را به صورت زیر به دست می‌آوریم:

$$L(\theta)=f(x_1,\theta)\times f(x_2,\theta)\times\ldots\times f(x_n,\theta)=\theta x_1^{-\theta-1}\times\theta x_2^{-\theta-1}\times\ldots\times\theta x_n^{-\theta-1}=\theta^n\prod_{i=1}^{n}x_i^{-\theta-1}$$

با توجه به توانی بودن تابع، برای راحتی کار در مرحله بعد از آن Ln می‌گیریم:

$$\ln L(\theta)=n\ln\theta+\ln\left(\prod_{i=1}^{n}x_i^{-\theta-1}\right)=n\ln\theta+(-\theta-1)\sum_{i=1}^{n}\ln x_i$$

ب) حال برای به دست آوردن حداکثر درستنمایی $(\ln L(\theta))'=0$ را تشکیل می‌دهیم:

$$(\ln L(\theta))'=0\ \rightarrow\ \frac{n}{\theta}-\sum\ln x_i=0\ \rightarrow\ \sum\ln x_i=\frac{n}{\theta}\rightarrow\theta=\frac{n}{\sum\ln x_i}$$

مثال ۴ فرض کنید X نمایانگر تعداد تلفن‌های رسیده در فواصل 5 دقیقه به یک دانشکده باشد (X دارای توزیع پواسون با پارامتر λ است). بر اساس 12 مشاهده مستقل زیر، برآورد درستنمایی ماکزیمم پارامتر λ کدام است؟

1,2,1,1,2,4,0,1,0,1,1,0

۱) $\frac{6}{7}$
۲) $\frac{13}{12}$
۳) $\frac{7}{6}$
۴) $\frac{12}{13}$

حل: گزینه ۳ درست است.

یادآوری: تابع چگالی توزیع پواسون با پارامتر λ عبارت است از:

$$f(x)=\frac{e^{-\lambda}\lambda^x}{x!}\quad;\quad x=0,1,\ldots$$

الف) تابع درستنمایی را به صورت زیر تشکیل می‌دهیم:

$$L(\lambda)=f(X_1,\lambda)\times f(X_2,\lambda)\times\ldots\times f(X_n,\lambda)=\frac{e^{-\lambda}\lambda^{X_1}}{X_1!}\times\ldots\times\frac{e^{-\lambda}\lambda^{X_n}}{X_n!}=\frac{e^{-n\lambda}\ \lambda^{\sum X_i}}{\prod_{i=1}^{n}X_i!}$$

با توجه به اینکه تابع درستنمایی، توانی است برای راحتی کار در مرحله بعد از تابع درستنمایی Ln می‌گیریم:

$$\ln L(\lambda)=-n\lambda+\sum X_i\ln\lambda-\ln\left(\prod_{i=1}^{n}X_i!\right)$$

ب) حال برای به دست آوردن حداکثر درستنمایی $(\ln L(\lambda))'=0$ را تشکیل می‌دهیم:

$$(\ln L(\lambda))'=0\ \rightarrow\ -n+\frac{\sum X_i}{\lambda}=0\ \rightarrow\ \lambda=\frac{\sum X_i}{n}=\bar{X}$$

با توجه به اینکه داده‌های نمونه در دست است، میانگین آن‌ها برآورد حداکثر درستنمایی λ خواهد بود:

$$\hat{\lambda}=\bar{X}=\frac{\sum X_i}{n}=\frac{1+2+1+1+2+4+0+1+0+1+1+0}{12}=\frac{14}{12}=\frac{7}{6}$$

قضیه: اگر X دارای توزیع نرمال با میانگین مجهول μ و واریانس مجهول σ^2 باشد، در این صورت برآورد MLE این پارامترها مانند روش گشتاوری عبارت است از:

$$\hat{\mu}=\bar{X}$$

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}=\overline{X^2}-\bar{X}^2$$

تبصره: اگر میانگین جامعه (μ) معلوم باشد، در برآورد درستنمایی واریانس و انحراف معیار، در رابطه بالا به‌جای $\bar{X}$ از مقدار μ استفاده می‌شود.

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}=\overline{X^2}-\mu^2$$

مثال ۵ یافته‌های یک نمونه تصادفی 3تایی از $N(\mu,\sigma^2)$ عبارت است از $X_1=1\ ,\ X_2=3\ ,\ X_3=5$. برآورد درستنمایی ماکزیمم (MLE) پارامترهای μ و σ^2 به ترتیب کدام است؟

۱) $4\ ,\ 3$
۲) $5\ ,\ \frac{8}{3}$
۳) $2\ ,\ 5$
۴) $3\ ,\ \frac{8}{3}$

حل: گزینه ۴ درست است.

$$\hat{\mu}=\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{1+3+5}{3}=3$$

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}=\frac{(1-3)^2+(3-3)^2+(5-3)^2}{3}=\frac{8}{3}$$

حالت خاص تابع درستنمایی

اگر حدود x به پارامتر θ در تابع چگالی وابسته باشد، مشتق گرفتن در مرحله (ب) بی‌فایده است؛ در چنین شرایطی از قواعد زیر برای برآورد پارامتر استفاده می‌کنیم:

الف) اگر $x \ge \theta$ باشد آن‌گاه: $\boxed{\theta \text{ حداکثر درستنمایی} = \min(X_1, X_2, ..., X_n)}$

ب) اگر $x \le \theta$ باشد آن‌گاه: $\boxed{\theta \text{ حداکثر درستنمایی} = \max(X_1, X_2, ..., X_n)}$

مثال ۶ فرض کنید $X_1, ..., X_n$ یک نمونه تصادفی n تایی از توزیعی با چگالی احتمال زیر باشد:

$$f(x) = \frac{2x}{1-\theta^2} \quad ; \theta \le x < 1$$

برآوردگر حداکثر درستنمایی (MLE) پارامتر θ کدام است؟

۱) $\hat{\theta} = \max(X_1, ..., X_n)$ ۲) $\hat{\theta} = \min(X_1, ..., X_n)$

۳) $\hat{\theta} = \dfrac{\min(X_1, ..., X_n) + \max(X_1, ..., X_n)}{2}$ ۴) $\hat{\theta} = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}$

حل: گزینه ۲ درست است.

با توجه به نکته گفته شده در قسمت«الف» داریم: $x \ge \theta \rightarrow \theta \text{ حداکثر درستنمایی} = \min(X_1, X_2, \dots, X_n)$

اثبات:

الف) ابتدا تابع درستنمایی $L(\theta)$ را بر اساس نمونه مستقل $(X_n, ..., X_2, X_1)$ به صورت زیر به دست می‌آوریم:

$$L(\theta) = f(x_1;\theta) \times f(x_2;\theta) \times ... \times f(x_n;\theta) = \frac{2x_1}{1-\theta^2} \times \frac{2x_2}{1-\theta^2} \times ... \times \frac{2x_n}{1-\theta^2} = \frac{2^n \prod_{i=1}^{n} x_i}{\left(1-\theta^2\right)^n}$$

در این قسمت، مشتق‌گیری از تابع درستنمایی ما را به نتیجه نمی‌رساند، بنابراین برای آنکه $L(\theta)$ نسبت به θ ماکزیمم شود، باید $\left(1-\theta^2\right)$ در مخرج کسر مینیمم شود، درنتیجه باید θ را ماکزیمم کنیم تا $\left(1-\theta^2\right)$ مینیمم شود، از طرفی همواره $\theta \le x < 1$ است، بنابراین ماکزیمم مقدار θ برابر با مینیمم مقدار x خواهد بود؛ به عبارت دیگر $\theta = \min(X_1, X_2, ..., X_n)$.

برآورد فاصله‌ای (فاصله اطمینان) (Interval Estimation)

در برآورد نقطه‌ای به دنبال یک آماره $\left(\hat{\theta}\right)$ از روی داده‌های نمونه برای تخمین پارامتر مجهول جامعه (θ) بودیم؛ اما می‌دانیم که این تخمین‌زن ممکن است دقیقاً برابر مقدار پارامتر نباشد بلکه در عین نزدیکی به مقدار واقعی پارامتر بیشتر یا کمتر از آن باشد. برای حل این مشکل می‌توان احتمال قرار گرفتن θ پارامتر جامعه را در یک فاصله به دست آورد. می‌دانیم که برای به دست آوردن هر فاصله دو عدد باید به دست آید، حد بالای فاصله و حد پایین فاصله؛ این دو مقدار با اضافه کردن و کم کردن مقدار ثابتی از برآورد نقطه‌ای پارامتر جامعه $\left(\hat{\theta}\right)$ به دست می‌آیند.

$$\begin{cases} \text{حد پایین فاصله} = \hat{\theta}-e \\ \text{حد بالای فاصله} = \hat{\theta}+e \end{cases} \rightarrow \quad \hat{\theta}-e<\theta<\hat{\theta}+e$$

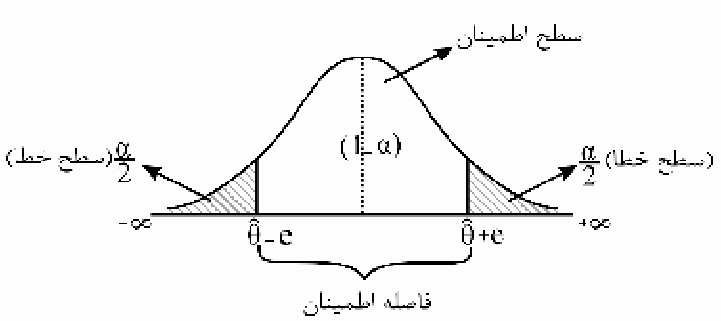
e همان مقدار ثابت است که حد بالا و پایین فاصله به کمک آن به دست می‌آید. e را در اصطلاح آماری خطای برآورد و یا دقت می‌نامند و فاصله به‌دست‌آمده به این روش را فاصله اطمینان می‌گویند.

تعریف: در برآورد فاصله‌ای با تعیین فاصله‌ای به نام فاصله اطمینان، حدودی را مشخص می‌کنیم که با احتمال مشخصی به نام سطح اطمینان $(1-\alpha)$ یقین داریم پارامتر مجهول جامعه (θ) در آن فاصله قرار دارد.

به عبارت دیگر هر فاصله اطمینان برای پارامتر θ، فاصله‌ای است به صورت $\hat{\theta}\pm e$ که با احتمال $(1-\alpha)$ اطمینان داریم، پارامتر θ در آن قرار دارد، به عبارت ریاضی داریم:

$$P\left(\hat{\theta}-e<\theta<\hat{\theta}+e\right)=(1-\alpha)$$

در این وضعیت:



$$\begin{cases} \textbf{فاصله اطمینان}: \hat{\theta}-e<\theta<\hat{\theta}+e \\ \textbf{سطح اطمینان}: (1-\alpha) \\ \textbf{سطح خطا (بحرانی یا معنی‌دار)}: \frac{\alpha}{2}+\frac{\alpha}{2}=\alpha \end{cases}$$

$$\boxed{\text{سطح خطا} = 1 - \text{سطح اطمینان}}$$

سطح اطمینان و سطح خطا

سطوح اطمینان $(1-\alpha)$ معمولاً به صورت 90%، 95% و 99% در نظر گرفته می‌شود که در این صورت سطح خطای آن‌ها α به صورت 10% , 5% و 1% خواهد بود.

✓ **دقت کنید!**

مقدار پیش‌فرض برای سطح اطمینان، 95% است.

مفهوم سطح اطمینان $(1-\alpha)$

هرگاه با توجه به مشاهدات نمونه یک فاصله اطمینان $\left(\hat{\theta}\pm e\right)$ در سطح اطمینان $(1-\alpha)$ و یا سطح خطای α برای پارامتر جامعه (θ) به دست می‌آوریم:

$$P\left(\theta<\hat{\theta}-e\right)+P\left(\theta>\hat{\theta}+e\right)=\alpha$$

یا

$$P\left(\hat{\theta}-e<\theta<\hat{\theta}+e\right)=(1-\alpha)$$

این فاصله به مفهوم آن است که:

الف) اگر 100 بار نمونه‌گیری را تکرار کنیم $100\times(1-\alpha)$ بار فاصله‌ای را به دست می‌آوریم که θ را در برخواهد داشت.

ب) اگر 100 بار نمونه‌گیری را تکرار کنیم $100\times\alpha$ بار فاصله‌ای را به دست می‌آوریم که θ را در بر نخواهد داشت.

ج) با احتمال $(1-\alpha)$ اطمینان داریم پارامتر جامعه در آن فاصله قرار خواهد داشت.

د) با احتمال α اطمینان داریم پارامتر جامعه در آن فاصله قرار نخواهد داشت.

مثال وقتی از روی مشاهدات نمونه فاصله اعتمادی فی‌المثل در سطح $\alpha=\%5$ برای میانگین جامعه‌ای می‌سازیم، به مفهوم آن است که: (اقتصاد ـ ۷۰)

۱) با احتمال 5% میانگین جامعه در آن فاصله قرار دارد.

۲) اگر نمونه‌گیری را 100 بار تکرار کنیم 95 بار فاصله‌ای به دست می‌آوریم که μ را در بر دارد.

۳) با اعتماد 95% میانگین جامعه در آن فاصله قرار دارد.

۴) با اعتماد 95% میانگین نمونه در آن فاصله قرار دارد.

حل: گزینه ۲ درست است.

فاصله اطمینان در سطح $\alpha=\%5$ یا $1-\alpha=\%95$ هر یک از موارد زیر را نشان می‌دهد.

- از 100 نمونه‌گیری 95 بار میانگین جامعه در آن فاصله قرار دارد و 5 بار قرار ندارد.
- با احتمال 95% میانگین جامعه در آن فاصله قرار دارد و با احتمال 5% قرار ندارد.

نحوه ساختن فاصله اطمینان

برای ایجاد فاصله اطمینان برای پارامتر مجهول جامعه، ابتدا بهترین برآوردکننده نقطه‌ای $\left(\hat{\theta}\right)$ را برای آن انتخاب می‌کنیم، سپس با استفاده از تابع توزیع نمونه‌ای مرتبط که از این به بعد به آن «تابع محوری» می‌گوییم، به نحوی که خواهیم دید، حدودی را مشخص می‌کنیم که با احتمال $(1-\alpha)$ اطمینان داریم پارامتر جامعه در آن قرار دارد.

برای مثال، فرض کنید می‌خواهیم یک فاصله اطمینان $(1-\alpha)$ برای میانگین جامعه (μ) ایجاد کنیم. در این وضعیت همان‌طورکه دیدیم آماره $\bar{X}$ بهترین برآورد نقطه‌ای برای آن است. حال با فرض آنکه جامعه اصلی نرمال و واریانس جامعه معلوم باشد، تابع توزیع نمونه‌ای $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ خواهد بود که در اینجا به آن «تابع محوری» می‌گوییم.

$$\text{تابع محوری}: Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad\rightarrow\quad \text{فاصله اطمینان}: \bar{X}\pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

که در آن $Z_{\frac{\alpha}{2}}$ ضریب اطمینان است.

✓ **دقت کنید!**

هر فاصله اطمینان، بر اساس نتایج حاصل از نمونه و با استفاده از آماره معلوم $(\hat{\theta})$ ساخته می‌شود. درنتیجه فاقد پارامتر مجهول (θ) است.

برای مثال، هنگام تخمین پارامتر مجهول میانگین جامعه (μ) توسط آماره معلوم $(\bar{X})$:

فاصله اطمینان است $\bar{X} \pm e$

فاصله اطمینان نیست $\mu \pm e$

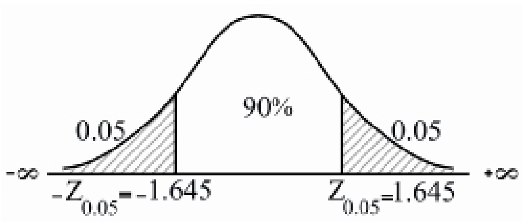
ضریب اطمینان

هنگام تعیین فاصله اطمینان برای یک پارامتر دلخواه از جامعه، با استفاده از توزیع نمونه‌ای مربوط به آن باید از یکی از توزیع‌های Z، t، χ^2 یا F استفاده کنیم. در این وضعیت به مقادیر این توزیع‌ها با توجه به سطوح اطمینان 90%، 95% یا 99% در جدول‌های پیوست کتاب نیاز است؛ این مقادیر همان ضرایب اطمینان در فواصل اطمینان هستند.

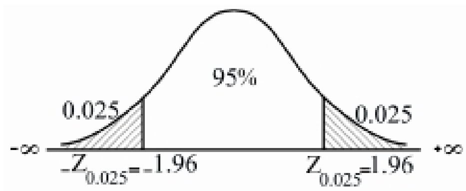
ضریب اطمینان Z (نرمال استاندارد) برای سطوح 90%، 95% و 99%

ازآنجاکه توزیع بیشتر برآوردکننده‌های آماری نرمال استاندارد (Z) است، مقادیر جدول Z که از این به بعد به آن ضریب اطمینان می‌گوییم، برای سطوح اطمینان 90%، 95% و 99% به صورت زیر است:

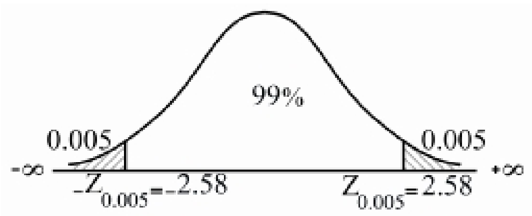
$$P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = (1-\alpha)$$



$P(-1.645 < Z < 1.645) = \%90$



$P(-1.96 < Z < 1.96) = \%95$



$P(-2.58 < Z < 2.58) = \%99$

روابط بین سطح خطا (α)، سطح اطمینان $(1-\alpha)$ و ضریب اطمینان (Z_α)

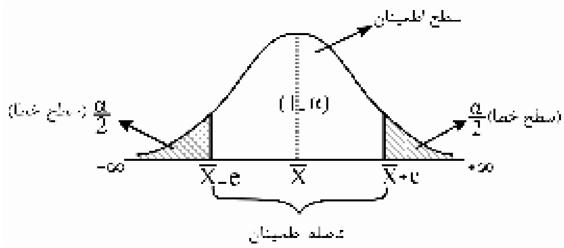
هرچه سطح اطمینان $(1-\alpha)$ بیشتر شود، مقدار α یا $\frac{\alpha}{2}$ (سطح معنی‌داری، سطح خطا) کمتر شده و درنتیجه ضریب اطمینان افزایش می‌یابد.

ضریب اطمینان	سطح اطمینان	سطح خطا
$Z_{\alpha}=1.28\ ,\ Z_{\frac{\alpha}{2}}=1.645$	$(1-\alpha)=0.90$	$\alpha=0.1$
$Z_{\alpha}=1.645\ ,\ Z_{\frac{\alpha}{2}}=1.96$	$(1-\alpha)=0.95$	$\alpha=0.05$
$Z_{\alpha}=2.32\ ,\ Z_{\frac{\alpha}{2}}=2.58$	$(1-\alpha)=0.99$	$\alpha=0.01$
افزایش $\downarrow$	افزایش $\downarrow$	کاهش $\downarrow$

نتیجه: ضرایب اطمینان $|Z_{\alpha}|$ یا $|Z_{\frac{\alpha}{2}}|$ با سطح اطمینان $(1-\alpha)$ رابطه مستقیم و با سطح خطا یا سطح معنی‌داری (α) رابطه عکس دارد.

برآورد فاصله‌ای میانگین جامعه

با فرض آنکه در جامعه‌ای با میانگین μ و واریانس σ^2 یک نمونه تصادفی به حجم n انتخاب کرده باشیم، $\bar{X}$ (میانگین نمونه) به عنوان بهترین برآوردکننده نقطه‌ای برای μ (میانگین جامعه) در تعیین فاصله اطمینان برای میانگین جامعه به صورت زیر استفاده می‌شود:



$Z=\dfrac{\bar{X}-\mu}{\frac{\sigma_X}{\sqrt{n}}}$ یا t : تابع محوری

در این حالت فاصله اطمینان برای میانگین جامعه به صورت $\bar{X}\pm e$ خواهد بود که e میزان دقت یا خطای برآورد است. حالات مختلف فاصله اطمینان بر اساس شرایطی که در توزیع نمونه‌ای $\bar{X}$ دیده شد، به صورت زیر است:

شرایط	فاصله اطمینان
۱) جامعه اصلی نرمال و σ^2 معلوم و $n\geq 1$	$\left(\bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma_X}{\sqrt{n}}\ ,\ \bar{X}+Z_{\frac{\alpha}{2}}\frac{\sigma_X}{\sqrt{n}}\right)$
۲) جامعه اصلی نرمال و σ^2 نامعلوم	
الف) $n>30$	$\left(\bar{X}-Z_{\frac{\alpha}{2}}\frac{S_X}{\sqrt{n}}\ ,\ \bar{X}+Z_{\frac{\alpha}{2}}\frac{S_X}{\sqrt{n}}\right)$
ب) $n\leq 30$	$\left(\bar{X}-t_{(n-1),\frac{\alpha}{2}}\frac{S_X}{\sqrt{n}}\ ,\ \bar{X}+t_{(n-1),\frac{\alpha}{2}}\frac{S_X}{\sqrt{n}}\right)$
۳) جامعه اصلی غیرنرمال و σ^2 معلوم	
الف) $(n>30)$ (قضیه حد مرکزی)	$\left(\bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma_X}{\sqrt{n}}\ ,\ \bar{X}+Z_{\frac{\alpha}{2}}\frac{\sigma_X}{\sqrt{n}}\right)$
ب) $n\leq 30$ (قضیه چی‌بی‌شف)	$\left(\bar{X}-\sqrt{\frac{1}{\alpha}}\times\frac{\sigma_X}{\sqrt{n}}\ ,\ \bar{X}+\sqrt{\frac{1}{\alpha}}\times\frac{\sigma_X}{\sqrt{n}}\right)$

مثال ۱ کمیت تصادفی X برطبق قانون نرمال با واریانس $\sigma^2 = 25$ توزیع شده است. از این جامعه نمونه‌ای به حجم $n = 100$ به طور تصادفی انتخاب می‌کنیم و تخمین امید ریاضی $\bar{x} = 180$ به دست آمده است. فاصله اعتماد 95% برای میانگین (μ) جامعه کدام است؟ (اقتصاد ـ ۷۱)

۱) $(179.02\,,180.98)$ ۲) $(179.18\,,180.82)$

۳) $(178.835\,,811.65)$ ۴) $(178.712\,,181.288)$

حل: گزینه ۱ درست است.

با توجه به حالت (۱) داریم:

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \mu \in \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\begin{cases} \mu \in \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 180 \pm 1.96 \times \frac{5}{\sqrt{100}} = (179.02\,,180.98) \\ \bar{x} = 180\,,\ n = 100\,,\ \sigma^2 = 25 \rightarrow \sigma = 5\,,\alpha = 0.05 \rightarrow Z_{0.025} = 1.96 \end{cases}$$

مثال ۲ تقاضای کالا (X) با توزیع نرمال برای 36 روز، دارای میانگین $\bar{x} = 40$ و $S = 5$ به دست آمده است. تخمین فاصله‌ای برای میانگین جامعه (μ) برابر است با: $(\alpha = 0.05)$ (مدیریت ـ ۷۱)

۱) $(38.37\,,41.63)$ ۲) $(36.7\,,43.3)$ ۳) $(37.85\,,42.15)$ ۴) $(38.64\,,41.36)$

حل: گزینه ۱ درست است.

با توجه به حالت (۲ ـ الف) داریم:

$$Z = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \rightarrow \mu \in \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

$$\begin{cases} \mu \in \bar{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} = 40 \pm 1.96 \frac{5}{\sqrt{36}} = (38.37, 41.63) \\ \bar{x} = 40\,,\ S = 5\,,\ n = 36 > 30, \alpha = 0.05 \rightarrow Z_{0.025} = 1.96 \end{cases}$$

مثال ۳ نمرات یک نمونه تصادفی 3 تایی از دانشجویان کلاسی که دارای توزیع نرمال است، 15،16 و 17 بوده است. فاصله اطمینان 90% میانگین نمرات دانشجویان کدام است؟ $(t \approx 3)$ (اقتصاد ـ ۸۶)

۱) $15.3-16.7$ ۲) $14.3-17.7$ ۳) $13.9-18.1$ ۴) $13.7-18.3$

حل: گزینه ۲ درست است.

با توجه به حالت (۲ـ ب) داریم:

$$t_{(n-1)} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \rightarrow \mu \in \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

مقدار t داده شده است، اما $\bar{X}$ و S باید با توجه به داده‌های نمونه برآورد شوند.

نکته: هرگاه تعداد نمونه $n = 3$ باشد و اعداد نمونه متوالی باشند، واریانس نمونه یک است و میانگین اعداد، برابر با عدد وسط است.

در این سؤال نیز $n = 3$ و اعداد نمونه متوالی‌اند، پس داریم:

$$\bar{X} = \text{داده وسط} = 16 \,,\, S^2 = 1 \;\to\; S = 1$$

اگر از طریق رابطه S^2 نیز محاسبه کنیم به همین اعداد می‌رسیم:

$$\begin{cases} \bar{X} = \dfrac{\sum x_i}{n} = \dfrac{15+16+17}{3} = \dfrac{48}{3} = 16 \\ S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1} = \dfrac{(15-16)^2 + (16-16)^2 + (17-16)^2}{3-1} = \dfrac{1+0+1}{2} = 1 \;\to\; S = 1 \end{cases}$$

$$\begin{cases} \mu \in \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \dfrac{S}{\sqrt{n}} = 16 \pm 3 \times \dfrac{1}{\sqrt{3}} = 16 \pm \sqrt{3} = (14.3\,, 17.7) \\ \bar{x} = 16 \;,\; S^2 = 1 \;,\; n = 3 \le 30 \;,\; t_{0.05,2} \simeq 3 \end{cases}$$

مثال ۴ یک نمونه تصادفی از 64 لامپ نشان می‌دهد که عمر متوسط نمونه 350 ساعت است. یک فاصله اطمینان 95 درصد برای متوسط طول عمر واقعی لامپ‌ها با فرض $\sigma_X = 100$ عبارت است از: (اقتصاد ـ ۸۵)

۱) 150 تا 550 ۲) 154 تا 546 ۳) 250.5 تا 449.5 ۴) 325.5 تا 374.5

حل: گزینه ۴ درست است.

با توجه به حالت (۳ـ الف) داریم:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to \mu \in \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\begin{cases} \mu \in \bar{X} \pm Z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}} = 350 \pm 1.96 \times \dfrac{100}{\sqrt{64}} = 350 \pm 24.5 = (325.5\,, 374.5) \\ \bar{x} = 350 \;,\; \sigma = 100 \;,\; n = 64 > 30 \,, \alpha = 0.05 \;\to\; Z_{0.025} = 1.96 \end{cases}$$

مثال ۵ از جامعه‌ای با توزیع غیرنرمال، نمونه‌ای متشکل از $n = 25$ مشاهده با میانگین $\bar{x} = 100$ و انحراف معیار $\sigma = 25$ تهیه شده است. فاصله اطمینان 0.95 برای میانگین جامعه (μ) برابر است با: (اقتصاد ـ ۷۳)

۱) 100 ± 4.47 ۲) 100 ± 20 ۳) $100 \pm 4.47 \dfrac{25}{25}$ ۴) $100 \pm 4.47(5)$

حل: گزینه ۴ درست است.

با توجه به حالت (۳ ـ ب) داریم:

$$\begin{cases} \mu \in \bar{X} \pm \sqrt{\dfrac{1}{\alpha}} \cdot \dfrac{\sigma}{\sqrt{n}} = 100 \pm \sqrt{\dfrac{1}{0.05}} \times \dfrac{25}{\sqrt{25}} = 100 \pm 4.47\,(5) \\ \bar{x} = 100 \;,\; \sigma = 25 \;,\; n = 25 \le 30 \;,\; \alpha = 0.05 \end{cases}$$

محاسبه خطا یا دقت (e)

در برآورد فاصله برای میانگین جامعه دقت یا خطا عبارت است از:

$$\boxed{\text{خطا} : e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

در رابطه بالا، در حالات مختلف به‌جای $Z_{\frac{\alpha}{2}}$ از $t_{\frac{\alpha}{2}}$ و $\sqrt{\frac{1}{\alpha}}$ و همین‌طور در صورت مجهول بودن σ از S استفاده می‌شود.

مثال ۱ کمیت تصادفی X برطبق قانون نرمال با واریانس $\sigma^2 = 16$ توزیع شده است. برای ارزیابی پارامتر μ نمونه‌ای به حجم $n = 64$ انتخاب و تخمین امید ریاضی $\bar{x} = 120$ به دست آمده است. حداکثر خطای حدی (ε) با احتمال 0.95 چقدر است؟ (اقتصاد ـ ۷۱)

۱) 2.58 ۲) 1.9 ۳) 0.98 ۴) 1.98

حل: گزینه ۳ درست است.

با توجه به حالت (۱) داریم:

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \mu \in \bar{X} \pm \underbrace{Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}_{\text{خطا}}$$

$$\begin{cases} \text{خطا}: e = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = Z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \times \frac{4}{\sqrt{64}} = 0.98 \\ n = 64, \sigma^2 = 16 \rightarrow \sigma = 4 \, , \, \alpha = 0.05 \rightarrow Z_{0.025} = 1.96 \end{cases}$$

مثال ۲ اگر تعداد نمونه $n = 5$، میانگین و انحراف معیار نمونه به ترتیب 50 و 1.581 باشد و جامعه آماری نرمال باشد، مقدار خطای نمونه‌گیری (ε) در برآورد میانگین جامعه کدام است؟ (راهنمایی: $t_{0.025} = 2.776$ و $Z_{0.025} = 1.96$) (حسابداری ـ ۸۲)

۱) 1.39 ۲) 1.96 ۳) 4.38 ۴) 6.21

حل: گزینه ۲ درست است.

با توجه به حالت (۲ـ ب) داریم:

$$t_{(n-1)} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \rightarrow \mu \in \bar{X} \pm \underbrace{t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}}_{\text{خطا}}$$

$$\begin{cases} \text{خطا}: e = t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} = 2.776 \times \frac{1.581}{\sqrt{5}} = 1.96 \\ n = 5 \, , \, S = 1.581 \, , \, t_{0.025} = 2.776 \end{cases}$$

محاسبه طول فاصله (2e)

برای محاسبه طول فاصله در برآورد فاصله میانگین جامعه به صورت زیر عمل می‌کنیم:

$$\text{فاصله اطمینان}: \left(\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \, , \, \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

$$\text{طول فاصله} = \text{حد بالا} - \text{حد پایین} = \left(\bar{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 2Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2e$$

$$\boxed{\text{طول فاصله} = 2e = 2\,Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}$$

در رابطه بالا در حالات مختلف به‌جای $Z_{\frac{\alpha}{2}}$ از $t_{\frac{\alpha}{2}}$ و $\sqrt{\frac{1}{\alpha}}$ و همین‌طور در مواردی که σ مجهول باشد از S استفاده می‌شود.

محاسبه تعداد نمونه (n)

در برآورد فاصله برای میانگین جامعه، تعداد نمونه با توجه به مقدار e برابر است با:

$$e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \rightarrow e^2 = Z_{\frac{\alpha}{2}}^2 \frac{\sigma^2}{n} \rightarrow \boxed{n = \frac{Z_{\frac{\alpha}{2}}^2 \sigma^2}{e^2}}$$

در رابطه بالا در حالات مختلف به‌جای $Z_{\frac{\alpha}{2}}$ از $t_{\frac{\alpha}{2}}$ و $\sqrt{\frac{1}{\alpha}}$ و همین‌طور در صورت مجهول بودن σ از S استفاده می‌کنیم.

مثال ۱ اگر انحراف معیار جامعه 20 و میزان خطای برآورد 5 باشد، حداقل تعداد نمونه لازم برای به دست آوردن فاصله اطمینان 95 درصد میانگین کدام است؟ $(Z | \alpha = 0.025) = 1.96$ (مدیریت ـ ۷۳)

۱) 62 ۲) 8 ۳) 157 ۴) 347

حل: گزینه ۱ درست است.

$$\begin{cases} e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \rightarrow 5 = 1.96 \times \frac{20}{\sqrt{n}} \rightarrow \sqrt{n} = 7.84 \rightarrow n \simeq 62 \\ e = 5 , \sigma = 20 , \alpha = 0.05 \rightarrow Z_{0.025} = 1.96 \end{cases}$$

مثال ۲ به منظور برآورد میانگین یک جامعه نرمال با واریانس 4، در نظر است یک نمونه تصادفی انتخاب گردد. اگر دقت برآورد 0.4 باشد، حجم نمونه تحقیق در سطح خطای 5%، کدام است؟ (حسابداری ـ ۸۱)

۱) 94 ۲) 97 ۳) 100 ۴) 103

حل: گزینه ۲ درست است.

$$\begin{cases} e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \rightarrow 0.4 = 1.96 \times \frac{2}{\sqrt{n}} \rightarrow \sqrt{n} = 9.8 \rightarrow n \simeq 97 \\ e = 0.4 , \sigma^2 = 4 \rightarrow \sigma = 2 , \alpha = 0.05 \rightarrow Z_{0.025} = 1.96 \end{cases}$$

تأثیر عوامل مختلف بر خطا یا دقت (e) و طول فاصله (2e)

درصورتی‌که به طور مثال برای محاسبه خطای برآورد فاصله اطمینان میانگین از روابط زیر استفاده کنیم:

$$e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{یا} \quad Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \quad \text{یا} \quad t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

با توجه به روابط بالا، ارتباط عوامل σ و S (انحراف معیار جامعه یا نمونه)، $t_{\frac{\alpha}{2}}$ یا $Z_{\frac{\alpha}{2}}$ (ضریب اطمینان)، $(1-\alpha)$ (سطح اطمینان)، α (سطح خطا) و n (تعداد نمونه) با e (خطا) و 2e (طول فاصله) به شرح زیر است:

عامل	ارتباط با e (خطا) و 2e (طول فاصله)	ارتباط با دقت برآورد
(۱) σ (انحراف معیار جامعه)	مستقیم	معکوس
(۲) S (انحراف معیار نمونه)	مستقیم	معکوس
(۳) $t_{\frac{\alpha}{2}}$ یا $Z_{\frac{\alpha}{2}}$ (ضریب اطمینان)	مستقیم	معکوس
(۴) $(1-\alpha)$ (سطح اطمینان)	مستقیم	معکوس
(۵) α (سطح خطا)	معکوس	مستقیم
(۶) n تعداد نمونه	معکوس	مستقیم

درباره موارد (۴) و (۵) به مثال زیر دقت کنید:

سطح اطمینان	خطا	ضریب اطمینان
$(1-\alpha)=0.95$	$\longrightarrow \alpha=0.05$	$\longrightarrow Z_{\frac{\alpha}{2}}=1.96$
$(1-\alpha)=0.99$	$\longrightarrow \alpha=0.01$	$\longrightarrow Z_{\frac{\alpha}{2}}=2.58$

همان‌طورکه مشاهده می‌شود با افزایش سطح اطمینان $(1-\alpha)$ یا کاهش سطح خطا (α)، ضریب اطمینان بیشتر شده و درنتیجه خطا (e) افزایش می‌یابد.

مثال ۱ در برآورد فاصله‌ای برای میانگین جامعه نرمال، درصورتی‌که انحراف معیار نمونه را دو برابر و حجم نمونه را 4 برابر کنیم، خطای برآورد چه تغییری می‌کند؟ (نمونه بزرگ)

۱) 2 برابر می‌شود. ۲) نصف می‌شود. ۳) 4 برابر می‌شود. ۴) تغییری نمی‌کند.

حل: گزینه ۴ درست است.

با توجه به شرط نمونه بزرگ $(n>30)$ برای محاسبه خطا از رابطه $e=Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$ استفاده می‌کنیم:

$$e_{old}=Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \xrightarrow{\text{اعمال تغییرات}} e_{new}=Z_{\frac{\alpha}{2}}\frac{2S}{\sqrt{4n}}=Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}=e_{old}$$

همان‌طورکه مشاهده می‌شود خطای برآورد هیچ تغییری نمی‌کند.

مثال ۲ در برآورد فاصله‌ای برای میانگین جامعه‌ای نرمال، حجم نمونه را 4 برابر می‌کنیم. در صورت ثابت ماندن سایر عوامل طول فاصله چه تغییری می‌کند؟ (نمونه بزرگ)

۱) 2 برابر می‌شود. ۲) نصف می‌شود. ۳) 4 برابر می‌شود. ۴) تغییری نمی‌کند.

حل: گزینه ۲ درست است.

با توجه به شرط نمونه بزرگ $(n>30)$ برای محاسبه خطا از رابطه $e=Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$ استفاده می‌کنیم:

$$e_{old}=Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \xrightarrow{\text{اعمال تغییرات}} e_{new}=Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{4n}}=Z_{\frac{\alpha}{2}}\frac{S}{2\sqrt{n}}=\frac{1}{2}e_{old}$$

با توجه به آنکه خطا نصف (e) شده است، طول فاصله $(2e)$ نیز نصف می‌شود.

مثال ۳ دربرآورد فاصله‌ای پارامتر، هرچه سطح اطمینان (Level of Confidence) بیشتر شود، با فرض ثابت ماندن سایر عوامل: (اقتصاد ـ ۷۵)

۱) دقت برآورد بیشتر می‌شود. ۲) دقت برآورد تغییرنمی‌کند.

۳) دقت برآورد کمتر می‌شود. ۴) سطح معنا α زیاد می‌شود.

حل: گزینه ۳ درست است.

با توجه به رابطه (۴) از جدول، با افزایش سطح اطمینان خطای برآورد افزایش می‌یابد و درنتیجه دقت برآورد کاهش می‌یابد.

مثال ۴ کدام‌یک از موارد زیر در مورد فاصله اطمینان یک پارامتر آماری مصداق ندارد؟ هرقدر ....، طول فاصله اطمینان کمتر می‌شود. (اقتصاد ـ ۸۰)

۱) حجم نمونه بیشتر باشد ۲) واریانس تخمین‌زننده نقطه‌ای کمتر شود

۳) واریانس جامعه آماری کمتر شود ۴) ضریب اطمینان بالاتر رود

حل: گزینه ۴ درست است.

گزینه۱، با توجه به رابطه (۶) از جدول درست است.

گزینه۲، با توجه به رابطه (۲) از جدول درست است.

گزینه۳، با توجه به رابطه (۱) از جدول درست است.

گزینه۴، با توجه به رابطه (۳) از جدول نادرست است.

بنابراین تنها گزینه ۴ در مورد کاهش طول فاصله اطمینان مصداق ندارد.

مثال ۵ در ساختن فاصله اطمینان برای میانگین، اگر حجم نمونه افزایش پیدا کند، کدام عبارت درست است؟ (اقتصاد ـ ۸۱)

۱) طول فاصله اطمینان کاهش می‌یابد.
۲) طول فاصله اطمینان افزایش می‌یابد.
۳) طول فاصله اطمینان بدون تغییر می‌ماند.
۴) واریانس نمونه‌ای، S^2، افزایش می‌یابد.

حل: گزینه ۱ درست است.

با توجه به رابطه (۶) از جدول، تعداد نمونه با خطا (e) و طول فاصله $(2e)$ رابطه معکوس دارد.

مثال ۶ اگر حجم نمونه به $\frac{1}{4}$ تقلیل یابد، طول فاصله اطمینان $100(1-\alpha)$ درصد میانگین جامعه: (اقتصاد ـ ۷۴)

۱) در صورت عدم تغییر دیگر شرایط 4 برابر می‌شود.
۲) $4\bar{X}$ افزایش می‌یابد.
۳) نصف می‌شود.
۴) در صورت عدم تغییر انحراف معیار نمونه، S_X، دوبرابر می‌شود.

حل: گزینه ۴ درست است.

$$e_{old} = Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \xrightarrow{\text{اعمال تغییرات}} e_{new} = Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{\frac{n}{4}}} = Z_{\frac{\alpha}{2}} \frac{2S}{\sqrt{n}} = 2e_{old}$$

همان‌طورکه دیده می‌شود، در صورت ثابت بودن عوامل S (انحراف معیار نمونه) و $Z_{\frac{\alpha}{2}}$ (ضریب اطمینان)، مقدار خطا (e) دو برابر شده و درنتیجه طول فاصله $(2e)$ نیز دو برابر می‌شود.

برآورد فاصله‌ای تفاضل یا مجموع میانگین دو جامعه

فرض کنید دو جامعه نرمال با میانگین‌های μ_1 و μ_2 و واریانس‌های σ_1^2 و σ_2^2 وجود داشته باشند و نمونه‌هایی مستقل به تعداد n_1 از جامعه اول و n_2 از جامعه دوم انتخاب کنیم، برای تعیین فاصله اطمینان مجموع یا اختلاف میانگین دو جامعه، آماره ناریب $\bar{X}_1 \pm \bar{X}_2$ با کمترین واریانس بهترین انتخاب است، بنابراین:

$$\left.\begin{aligned} E\left(\bar{X}_1 \pm \bar{X}_2\right) &= \mu_1 \pm \mu_2 \\ \sigma^2\left(\bar{X}_1 \pm \bar{X}_2\right) &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{aligned}\right\} \longrightarrow \text{تابع محوری} : Z \text{ یا } t = \frac{\left(\bar{X}_1 \pm \bar{X}_2\right) - \left(\mu_1 \pm \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

در این حالت فاصله اطمینان به صورت $(\bar{X}_1 \pm \bar{X}_2) \pm e$ است و با توجه به حالات مختلف به صورت زیر بررسی می‌شود:

شرایط	فاصله اطمینان
۱) دو جامعه نرمال ، واریانس ها معلوم یا دو جامعه غیرنرمال، واریانس‌ها معلوم و $(n_1, n_2) > 30$ (قضیه حد مرکزی)	$(\bar{X}_1 \pm \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
۲) دو جامعه نرمال، واریانس‌ها نامعلوم و $n_2 > 30 , n_1 > 30$	$(\bar{X}_1 \pm \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
۳) دو جامعه نرمال، واریانس‌ها نامعلوم و $n_2 \le 30, n_1 \le 30$ یا $(n_1 + n_2 \le 30)$ الف) با فرض $\sigma_1^2 = \sigma_2^2$ (برابری واریانس دو جامعه)	$(\bar{X}_1 \pm \bar{X}_2) \pm t_{(n_1+n_2-2), \frac{\alpha}{2}} \times S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ واریانس ترکیبی $= S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ درجه آزادی $= n_1 + n_2 - 2$
ب) با فرض $\sigma_1^2 \neq \sigma_2^2$ (عدم برابری واریانس دو جامعه)	$(\bar{X}_1 \pm \bar{X}_2) \pm t_{r, \frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ درجه آزادی : $r = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1} + \frac{\left(S_2^2/n_2\right)^2}{n_2-1}}$

محاسبه دقت یا خطا (e)

مقدار خطا یا دقت در فاصله اطمینان مربوط به $\mu_1 \pm \mu_2$ با توجه به حالات مختلف به شرح زیر است:

حالات (۱) و (۲)
$$e = Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad , \quad e = Z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

حالت (۳-الف)
$$e = t_{(n_1+n_2-2), \frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

حالت (۳-ب)
$$e = t_{(r), \frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

و طول فاصله برابر 2e است.

مثال ۱ وزن مسافران و بار همراه آنان در یک پرواز دارای توزیع نرمال است. بر اساس اطلاعات در مورد میانگین و واریانس از یک نمونه تصادفی n_1 تایی از وزن مسافران و یک نمونه تصادفی n_2 تایی مستقل از بار مسافران، فاصله اطمینان $(1-\alpha)$ درصد برای مجموع میانگین وزن مسافر و بار همراه وی $(\mu_1+\mu_2)$ کدام است؟ (اقتصاد ـ ۸۳)

۱) $(\bar{X}_1+\bar{X}_2)\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

۲) $(\mu_1+\mu_2)\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

۳) $(\bar{X}_1+\bar{X}_2)\pm t_{\frac{\alpha}{2},(n_1+n_2-2)}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

۴) $(\bar{X}_1+\bar{X}_2)\pm t_{\frac{\alpha}{2},r}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

حل: گزینه ۴ درست است.

با توجه به آنکه به طور پیش‌فرض «واریانس جوامع نامعلوم و نابرابر» و «نمونه‌ها کوچک $n_1, n_2 < 30$» در نظر گرفته می‌شود، حالت (۳-ب) رخ می‌دهد و فاصله اطمینان برای مجموع میانگین‌ها $(\mu_1+\mu_2)$ عبارت است از:

$$\mu_1+\mu_2 \in (\bar{X}_1+\bar{X}_2)\pm t_{r,\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

مثال ۲ برآورد فاصله‌ای مجموع متوسط وزن مسافر (μ_1) و بار همراه وی (μ_2) در یک پرواز بر اساس یک نمونه تصادفی n_1 تایی از وزن مسافران و یک نمونه تصادفی n_2 تایی مستقل از بار همراه مسافران کدام است؟ (اقتصاد ـ ۸۶)

۱) $(\bar{X}_1+\bar{X}_2)\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

۲) $(\bar{X}_1-\bar{X}_2)\pm t_{\frac{\alpha}{2},r}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

۳) $(\bar{X}_1-\bar{X}_2)\pm Z_{\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

۴) $(\mu_1+\mu_2)\pm t_{\frac{\alpha}{2},r}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

حل: گزینه ۱ درست است.

با پیش‌فرض نرمال بودن جوامع، شرایط این مسئله دقیقاً مانند مثال قبل است و برآورد فاصله برای مجموع میانگین دو جامعه عبارت است از:

$$\mu_1+\mu_2 \in (\bar{X}_1+\bar{X}_2)\pm t_{r,\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

که در گزینه‌ها وجود ندارد.

دقت کنید، گزینه ۴ نادرست است، زیرا در یک فاصله اطمینان همه چیز با توجه به نتایج حاصل از نمونه‌گیری بوده و معلوم است و به هیچ وجه نباید از پارامتر مجهول (در این مثال $\mu_1+\mu_2$) در آن استفاده کرد.

انتخاب گزینه درست:

با کمی دقت در گزینه‌ها متوجه می‌شویم که:

ـ گزینه‌های ۲ و ۳ نادرست هستند، زیرا برآورد تفاضل میانگین را نشان می‌دهند، درحالی‌که سؤال برآورد مجموع میانگین را خواسته است.

ـ گزینه ۴ نیز با توجه به دلایلی که مطرح شد اصلاً فاصله اطمینان نیست.

بنابراین تنها گزینه قابل انتخاب گزینه ۱ است که متأسفانه کلید سازمان سنجش نیز بوده است!!

برآورد فاصله‌ای نسبت جامعه

برای تخمین نسبت یا نرخ موفقیت در جامعه‌ای با انتخاب نمونه‌ای n تایی درصورتی‌که x موفقیت مشاهده شود، رابطه $\bar{p} = \frac{x}{n}$ را نسبت نمونه در نظر می‌گیریم. **توزیع** $\bar{p}$ همان‌طورکه در بحث نمونه‌گیری مطرح شد، در نمونه‌های بزرگ **نرمال** است بنابراین:

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}\,\bar{q}}{n}}}$$: تابع محوری

بنابراین فاصله اطمینان نسبت جامعه عبارت است از:

$$\boxed{\left(\bar{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\ ,\ \bar{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\right)}$$

توجه: در شرایطی که p جامعه مجهول باشد از $\bar{p}$ نسبت آن در نمونه استفاده می‌شود و اگر نسبت نمونه نیز داده نشده باشد، به طور پیش‌فرض p و q را برابر $\frac{1}{2}$ می‌دانیم.

مثال ۱ معاون اداری مالی دانشگاهی بر اساس یک نمونه تصادفی 100تایی از دانشجویان مشاهده کرده است که 80 نفر از آن‌ها از کمک‌هزینه تحصیلی استفاده می‌کنند. فاصله اطمینان 90 درصدی نسبت واقعی دانشجویانی که از کمک‌هزینه تحصیلی استفاده می‌کنند، کدام است؟ (اقتصاد ـ ۸۳)

۱) 0.8 ± 0.0656 ۲) 0.8 ± 0.0784 ۳) 0.8 ± 0.0520 ۴) 0.8 ± 0.0822

حل: گزینه ۱ درست است.

$$\begin{cases} \bar{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}} = 0.8 \pm 1.64\sqrt{\frac{0.8 \times 0.2}{100}} = 0.8 \pm 0.0656 \\ n = 100,\ x = 80 \\ \bar{p} = \frac{x}{n} = \frac{80}{100} = 0.8\ ,\ \alpha = 0.1 \to Z_{0.05} = 1.64 \end{cases}$$

محاسبه دقت یا خطا (e)

در برآورد فاصله برای نسبت جامعه دقت یا خطا عبارت است از:

$$\boxed{e = Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}}$$

مثال ۲ به منظور ارزیابی کیفیت محصولات تولید شده، تعداد 200 واحد محصول را به طور تصادفی انتخاب کرده‌ایم که بین آن‌ها 40 محصول نقص‌دار مشاهده شده است. دقت تخمین (یا خطای حدی) در سطح احتمال 0.95 کدام است؟(اقتصاد ـ ۷۲)

۱) 0.0554 ۲) 0.1025 ۳) 0.0046 ۴) 0.55

حل: گزینه ۱ درست است.

$$\begin{cases} e = Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}} = 1.96 \times \sqrt{\frac{0.2 \times 0.8}{200}} = 0.0554 \\ \bar{p} = \frac{x}{n} = \frac{40}{200} = 0.2\ ,\ \alpha = 0.05 \to\ Z_{0.025} \simeq 1.96 \\ n = 200\ ,\ x = 40\ \text{(تعداد افراد دارای صفت مورد نظر)} \end{cases}$$

محاسبه طول فاصله (2e)

برای محاسبه طول فاصله در برآورد فاصله‌ای نسبت جامعه، به صورت زیر عمل می‌کنیم:

$$\text{فاصله اطمینان}: \left(\bar{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\ ,\ \bar{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\right)$$

$$\text{طول فاصله} = \left(\bar{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\right) - \left(\bar{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\right) = 2Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}} = 2e$$

$$\boxed{\text{طول فاصله} = 2e = 2Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}}}$$

محاسبه تعداد نمونه (n)

در برآورد فاصله‌ای برای نسبت جامعه، تعداد نمونه با توجه به مقدار e برابر است با:

$$e = Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\,\bar{q}}{n}} \rightarrow e^2 = Z^2_{\frac{\alpha}{2}}\frac{\bar{p}\,\bar{q}}{n} \longrightarrow \boxed{n = \frac{Z^2_{\frac{\alpha}{2}}\,\bar{p}\,\bar{q}}{e^2}}$$

مثال۳ مأمور کنترل کیفیت در صدد تعیین حجم نمونه برای برآورد درصد کالاهای معیوب است. وی معتقد است نسبت مذکور بیش از 0.20 نیست. اگر دقت برآورد 0.04 باشد، حداقل حجم نمونه در سطح خطای %5 چقدر است؟ $(Z \mid \alpha = 0.025) = 1.96$ (مدیریت ـ ۷۹)

۱) 8 ۲) 10 ۳) 385 ۴) 481

حل: گزینه ۳ درست است.

$$\begin{cases} e = Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \rightarrow 0.04 = 1.96\sqrt{\frac{0.2\times 0.8}{n}} \rightarrow \sqrt{n} = 19.6 \rightarrow n \simeq 385 \\ e = 0.04\ , p = 0.2\ , \alpha = 0.05 \rightarrow Z_{0.025} = 1.96 \end{cases}$$

مثال۴ اگر بخواهیم نرخ بیکاری در جامعه را در سطح معنی‌دار %5 با حداکثر خطای تخمین %1 برآورد کنیم، حجم نمونه انتخابی چقدر باشد؟ $(Z_{\%2.5} = 2)$ (اقتصاد ـ ۷۳)

۱) 1000 ۲) 10000 ۳) 20000 ۴) 100000

حل: گزینه ۲ درست است.

چون هیچ اطلاعاتی از نسبت (نرخ بیکاری) نداریم p و q را پیش‌فرض برابر $\frac{1}{2}$ در نظر می‌گیریم.

$$\begin{cases} e = Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \rightarrow 0.01 = 2\times\sqrt{\frac{\frac{1}{2}\times\frac{1}{2}}{n}} \rightarrow \sqrt{n} = 100 \rightarrow n = 10000 \\ e = 0.01\ , \alpha = 0.05 \rightarrow Z_{0.025} \simeq 2\ , p = q = \frac{1}{2} \end{cases}$$

تأثیر عوامل مختلف بر خطا یا دقت (e) و طول فاصله (2e)

ازآنجاکه توزیع $\bar{p}$ نرمال است، بنابراین طول فاصله و خطای p با طول فاصله و خطای μ برابر است و تمامی موارد بیان‌شده درباره روابط $2e, e, n$ و α برای فاصله p هم صادق است.

برآورد فاصله‌ای تفاضل یا مجموع نسبت دو جامعه

درصورتی‌که دو جامعه میانگین‌پذیر باشند، از تخمین $\mu_1 - \mu_2$ برای مقایسه آن‌ها استفاده می‌کنند، اما درصورتی‌که داده‌ها کیفی باشند، باید از مقایسه نسبت موفقیت در دو جامعه استفاده کرد.

اگر x_1 موفقیت از یک نمونه n_1 تایی از جامعه اول و x_2 موفقیت از یک نمونه n_2 تایی از جامعه دوم را در نظر بگیریم، آن‌گاه $\bar{p}_1 = \frac{x_1}{n_1}$ و $\bar{p}_2 = \frac{x_2}{n_2}$ به ترتیب نسبت نمونه در هر یک از دو جامعه هستند، با فرض:

۱) مستقل بودن نمونه‌ها در دو جامعه

۲) انتخاب نمونه‌های بزرگ

آماره نااریب $(\bar{p}_1 \pm \bar{p}_2)$ با کمترین واریانس بهترین انتخاب و توزیع آن نرمال است، بنابراین:

$$E(\bar{p}_1 \pm \bar{p}_2) = p_1 \pm p_2$$

$$\sigma^2(\bar{p}_1 \pm \bar{p}_2) = \frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2} \longrightarrow \text{تابع محوری} : Z = \frac{(\bar{p}_1 \pm \bar{p}_2) - (p_1 \pm p_2)}{\sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}}$$

بنابراین فاصله اطمینان آن عبارت است از:

$$\boxed{\bar{p}_1 \pm \bar{p}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}}$$

مثال اگر در دو جامعه تخمین‌های $\hat{p}_1 = 0.5$ و $\hat{p}_2 = 0.69$ باشد، فاصله اطمینان 99% برای تفاوت نسبت دو جامعه کدام است؟ به فرض اینکه حجم نمونه‌ها یکسان و برابر با $n_1 = n_2 = 300$ در نظر گرفته شوند. (مدیریت ـ ۷۵)

۱) $(-0.2915\ ,\ -0.0885)$
۲) $(-0.3952\ ,\ -0.086)$
۳) $(-0.3952\ ,\ 0.0865)$
۴) $(-0.2932\ ,\ 0.089)$

حل: گزینه ۱ درست است.

$$\begin{cases} p_1 - p_2 \in (\bar{p}_1 - \bar{p}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}} \\ p_1 - p_2 \in (0.5 - 0.69) \pm 2.58 \sqrt{\frac{(0.5)(0.5)}{300} + \frac{(0.69)(0.31)}{300}} = -0.19 \pm 0.1015 = (-0.2915\ , -0.0885) \\ \hat{p}_1 = 0.5\ ,\ \hat{p}_2 = 0.69\ ,\ n_1 = n_2 = 300\ ,\ \alpha = 0.01 \to Z_{0.005} = 2.58 \end{cases}$$

محاسبه دقت یا خطا (e)

مقدار خطا در برآورد فاصله‌ای تفاضل نسبت دو جامعه عبارت است از:

$$\boxed{e = Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}}$$

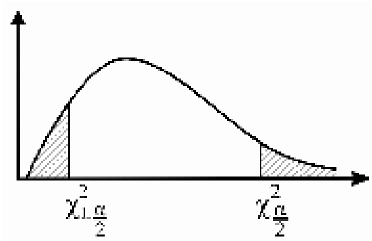
محاسبه طول فاصله (2e)

همان‌طورکه در برآورد فاصله نسبت دیده شد، طول فاصله به صورت زیر مطرح می‌شود:

$$2e = 2Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1}+\frac{\bar{p}_2\bar{q}_2}{n_2}}$$

برآورد فاصله‌ای واریانس جامعه

درصورتی‌که S^2 واریانس یک نمونه n تایی از جامعه‌ای نرمال با میانگین نامعلوم (μ) باشد، آن‌گاه $S^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$ است و فاصله اطمینان برای برآورد واریانس جامعه (σ^2) بر مبنای توزیع نمونه‌گیری $\chi^2_{n-1}=\frac{(n-1)S^2}{\sigma^2}$ (تابع محوری) تعیین می‌شود، بنابراین:



$$\chi^2_{n-1}=\frac{(n-1)S^2}{\sigma^2} \longrightarrow \sigma^2=\frac{(n-1)S^2}{\chi^2_{n-1}}$$

و درنتیجه برآورد فاصله‌ای آن عبارت است از:

$$\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right]$$

معلوم بودن میانگین جامعه

درصورتی‌که میانگین جامعه (μ) معلوم باشد، آن‌گاه $S^2=\frac{\sum(x_i-\mu)^2}{n}$ و تعیین فاصله اطمینان برای واریانس جامعه (σ^2) بر مبنای توزیع نمونه‌گیری $\chi^2_{(n)}=\frac{nS^2}{\sigma^2}$ (تابع محوری) خواهد بود، بنابراین:

$$\chi^2_{(n)}=\frac{nS^2}{\sigma^2} \rightarrow \sigma^2=\frac{nS^2}{\chi^2_{(n)}}$$

و برآورد فاصله‌ای آن عبارت است از:

$$\left(\frac{nS^2}{\chi^2_{\frac{\alpha}{2},n}},\frac{nS^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right)$$

✓ **دقت کنید!**

به طور پیش‌فرض میانگین جامعه (μ) نامعلوم در نظر گرفته می‌شود مگر آنکه در صورت مسئله به صورت صریح ذکر شود که میانگین معلوم است و یا مقدار آن ذکر شود.

مثال فاصله اطمینان $1-\alpha$ درصد برای واریانس جامعه‌ای با توزیع نرمال چیست؟ (α برای دنباله راست توزیع تعریف شده است.) (اقتصاد ـ ۸۵)

۱) $\dfrac{\chi^2_{\frac{\alpha}{2},n-1}}{(n-1)S^2} < \sigma^2 < \dfrac{\chi^2_{1-\frac{\alpha}{2},n-1}}{(n-1)S^2}$

۲) $\dfrac{\chi^2_{1-\frac{\alpha}{2},n-1}}{(n-1)S^2} < \sigma^2 < \dfrac{\chi^2_{\frac{\alpha}{2},n-1}}{(n-1)S^2}$

۳) $\dfrac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$

۴) $\dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}$

حل: گزینه ۴ درست است.

به طور پیش‌فرض میانگین جامعه (μ) نامعلوم فرض می‌شود.

برآورد فاصله‌ای انحراف‌معیار جامعه

هرگاه فاصله اطمینان برای انحراف معیار جامعه خواسته شد، کافی است فاصله اطمینان برای واریانس جامعه $\left(\sigma^2\right)$ را به دست آوریم و سپس از حد بالا و حد پایین فاصله اطمینان به‌دست‌آمده جذر بگیریم بنابراین فاصله اطمینان برای انحراف معیار جامعه عبارت است از:

در شرایطی که میانگین جامعه (μ) **نامعلوم** باشد: (**پیش‌فرض**)

$$\left(\sqrt{\frac{(n-1)\,S^2}{\chi^2_{\frac{\alpha}{2},n-1}}},\sqrt{\frac{(n-1)\,S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}}\right)$$

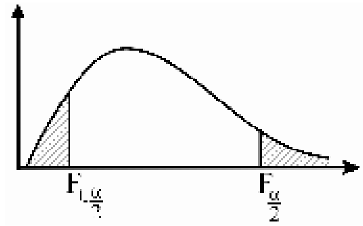
در شرایطی که میانگین جامعه (μ) **معلوم** باشد:

$$\left(\sqrt{\frac{nS^2}{\chi^2_{\frac{\alpha}{2},n}}},\sqrt{\frac{nS^2}{\chi^2_{1-\frac{\alpha}{2},n}}}\right)$$

برآورد فاصله‌ای نسبت واریانس‌های دو جامعه

یکی از راه‌های مقایسه واریانس‌های دو جامعه تشکیل نسبت σ_1^2/σ_2^2 است که اگر دو واریانس با هم مساوی باشد، نسبت مذکور برابر 1 خواهد شد.

درصورتی‌که S_1^2 و S_2^2 به ترتیب واریانس‌های نمونه از نمونه‌های مستقل از هم با اندازه‌های n_1 و n_2 از دو جامعه نرمال با واریانس‌های σ_1^2 و σ_2^2 باشند، فاصله اطمینان برای برآورد نسبت واریانس‌های دو جامعه $\left(\sigma_1^2/\sigma_2^2\right)$ بر مبنای توزیع $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=F_{n_1-1,\,n_2-1}$ (تابع محوری) تعیین می‌شود، بنابراین:



$$F_{n_1-1,n_2-1}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \longrightarrow \frac{\sigma_1^2}{\sigma_2^2}=\frac{1}{F_{n_1-1,n_2-1}}\times\frac{S_1^2}{S_2^2}$$

درنتیجه فاصله اطمینان عبارت است از:

$$\boxed{\left[\frac{1}{F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}}\times\frac{S_1^2}{S_2^2}\ ,\ \frac{1}{F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1}}\times\frac{S_1^2}{S_2^2}\right]}$$

درعین‌حال با توجه به برقراری رابطه $F_{1-\frac{\alpha}{2},m\,,\,n}=\frac{1}{F_{\frac{\alpha}{2},n\,,\,m}}$ حد بالای فاصله اطمینان می‌تواند به صورت زیر نیز بیان شود:

$$\left[\frac{1}{F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}}\times\frac{S_1^2}{S_2^2}\ ,\ F_{\frac{\alpha}{2},\,n_2-1,\,n_1-1}\times\frac{S_1^2}{S_2^2}\right]$$

معلوم بودن میانگین جوامع

همان‌طورکه در توزیع S_1^2/S_2^2 اشاره شد درصورتی‌که میانگین دو جامعه معلوم باشد برای محاسبه واریانس هر نمونه از رابطه $S^2=\frac{\sum\left(x_i-\mu\right)^2}{n}$ استفاده می‌شود. در این حالت فاصله اطمینان برای برآورد نسبت واریانس‌های دو جامعه بر مبنای توزیع $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=F_{n_1,n_2}$ تعیین می‌شود و:

$$F_{n_1,n_2}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \xrightarrow{\text{تابع محوری}} \frac{\sigma_1^2}{\sigma_2^2}=\frac{1}{F_{n_1,n_2}}\frac{S_1^2}{S_2^2}$$

درنتیجه فاصله اطمینان نسبت واریانس دو جامعه عبارت است از:

$$\boxed{\left(\frac{1}{F_{\frac{\alpha}{2},n_1,n_2}}\frac{S_1^2}{S_2^2}\ ,\ \frac{1}{F_{1-\frac{\alpha}{2},n_1,n_2}}\frac{S_1^2}{S_2^2}\right)}$$

مثال فاصله اطمینان $1-\alpha$ برای نسبت واریانس دو جامعه با فرض نرمال بودن جوامع عبارت است از: (اقتصاد ـ ۸۴)

۱) $\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}$

۲) $\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\frac{\alpha}{2},n_1-1,n_2-1}}$

۳) $\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_1^2}{S_2^2}.F_{\frac{\alpha}{2},n_1-1,n_2-1}$

۴) $\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_1^2}{S_2^2}.F_{1-\frac{\alpha}{2},n_2-1,n_1-1}$

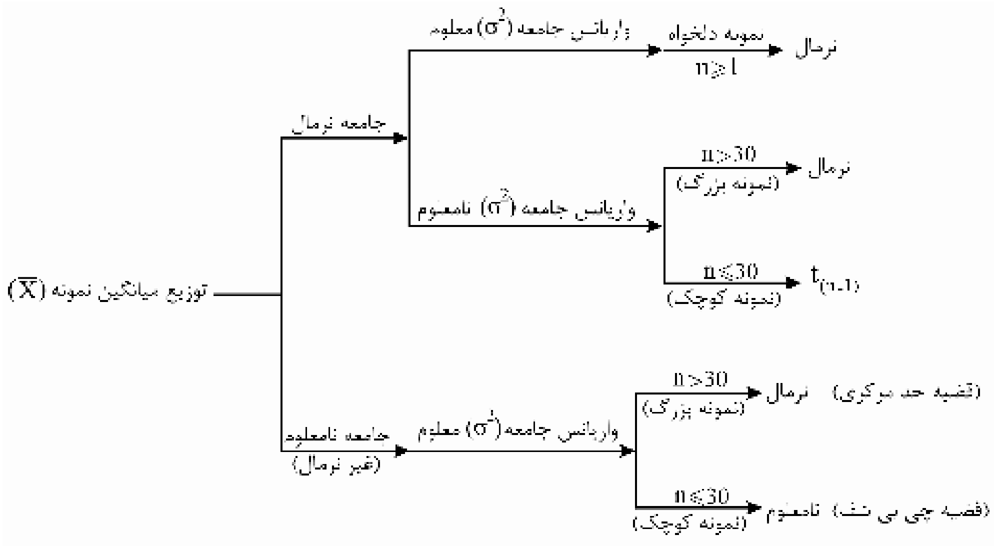
حل: گزینه ۲ درست است.

با توجه به رابطه $\frac{1}{F_{\frac{\alpha}{2},n_2-1,n_1-1}}=F_{1-\frac{\alpha}{2},n_1-1,n_2-1}$ فاصله زیر نیز درست است:

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_1^2}{S_2^2}.F_{\frac{\alpha}{2},n_2-1,n_1-1}$$

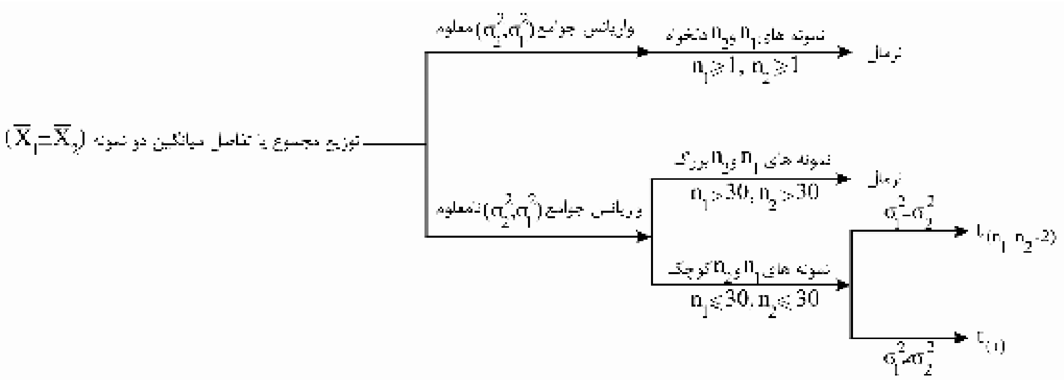
جداول خلاصه توزیع‌های نمونه‌ای

۱- توزیع $\bar{X}$ (میانگین نمونه)



جامعه	واریانس جامعه (σ^2)	تعداد نمونه (n)	توزیع $\bar{X}$	توضیحات
نرمال	معلوم	دلخواه $(n \geq 1)$	نرمال	توزیع $\bar{X}$ مستقل از تعداد نمونه، همیشه نرمال است.
نرمال	نامعلوم	$(n > 30)$ (بزرگ)	نرمال	توزیع $\bar{X}$ وابسته به تعداد نمونه و برای نمونه بزرگ همیشه نرمال است.
نرمال	نامعلوم	$(n \leq 30)$ (کوچک)	t_{n-1}	توزیع $\bar{X}$ وابسته به تعداد نمونه و برای نمونه کوچک t_{n-1} است.
غیرنرمال	معلوم	$(n > 30)$ قضیه حد مرکزی	نرمال	توزیع $\bar{X}$ وابسته به تعداد نمونه (بزرگ) نرمال است.
غیرنرمال	معلوم	$n \leq 30$	نامعلوم (چی‌بی‌شف)	

۲- توزیع $\bar{X}_1 \pm \bar{X}_2$ (تفاضل یا مجموع میانگین دو نمونه)



جامعه	واریانس جامعه (σ_1^2, σ_2^2)	تعداد نمونه (n_1, n_2)	توزیع $\bar{X}_1 \pm \bar{X}_2$	توضیحات
نرمال	معلوم	دلخواه $(n_1 \geq 1), (n_2 \geq 1)$	نرمال	توزیع مستقل از تعداد نمونه همیشه نرمال است.
نرمال	نامعلوم	$n_1 > 30, n_2 > 30$	نرمال	توزیع وابسته به تعداد نمونه $(n > 30)$ نرمال است.
نرمال	نامعلوم	$(n_1 + n_2 \leq 30)$ یا $(n_1 \leq 30, n_2 \leq 30)$	$t_{n_1+n_2-2}$	با فرض برابری واریانس دو جامعه $\sigma_1^2 = \sigma_2^2$
نرمال	نامعلوم	$(n_1 + n_2 \leq 30)$ یا $(n_1 \leq 30, n_2 \leq 30)$	t_r	با فرض عدم برابری واریانس دو جامعه $\sigma_1^2 \neq \sigma_2^2$

۳- توزیع $\bar{p}$ نسبت موفقیت در نمونه

جامعه	توزیع $\bar{p}$	توضیحات
دوجمله‌ای	نرمال	به طور کلی توزیع نمونه‌ای $\bar{p}$ در نمونه‌های بزرگ از تقریب نرمال برخوردار است.

۴- توزیع $\bar{p}_1 \pm \bar{p}_2$ (تفاضل یا مجموع نسبت موفقیت در دو نمونه از دو جامعه)

جامعه	توزیع $\bar{p}_1 \pm \bar{p}_2$	توضیحات
دوجمله‌ای	نرمال	به طور کلی با انتخاب نمونه‌های تصادفی و مستقل و به اندازه کافی بزرگ توزیع آماره $\bar{p}_1 \pm \bar{p}_2$ نرمال است.

۵- توزیع $\frac{(n-1)S^2}{\sigma^2}$

جامعه	تعداد	توزیع $\frac{(n-1)S^2}{\sigma^2}$	توضیحات
نرمال	n	$\chi^2_{(n-1)}$	با فرض برخورداری از جامعه نرمال، آماره $\frac{(n-1)S^2}{\sigma^2}$ برای تعیین حدود اطمینان σ^2 استفاده می‌شود.

۶- توزیع $\left(\frac{S_1^2}{\sigma_1^2}\right)/\left(\frac{S_2^2}{\sigma_2^2}\right)$

جامعه	تعداد نمونه	توزیع $\left(\frac{S_1^2}{\sigma_1^2}\right)/\left(\frac{S_2^2}{\sigma_2^2}\right)$	توضیحات
نرمال	n_2, n_1	F_{n_1-1,n_2-1}	در شرایطی که $\sigma_2^2=\sigma_1^2$ (واریانس دو جامعه برابر) آن‌گاه آماره S_1^2/S_2^2 دارای توزیع F_{n_1-1,n_2-1} است.

توجه: در حالت‌های ۵ و ۶ اگر μ (میانگین جامعه) معلوم باشد آن‌گاه:

$$\frac{nS^2}{\sigma^2}\simeq\chi^2_{(n)}\ ,\ \left(\frac{S_1^2}{\sigma_1^2}\right)/\left(\frac{S_2^2}{\sigma_2^2}\right)\simeq F_{n_1,n_2}$$

تست‌های طبقه‌بندی شده

نمونه‌گیری

۱. مناسب‌ترین روش نمونه‌گیری از اتومبیل‌های سواری که وارد یک بزرگراه می‌شوند، کدام است؟ (اقتصاد ـ ۸۶)

۱) سیستماتیک ۲) ساده ۳) خوشه‌ای ۴) طبقه‌بندی شده

۲. در نمونه‌گیری طبقه‌ای با روش تصادفی ساده، بین طبقه‌ها همگنی وجود .... و درون هر طبقه نیز همگنی وجود .... (محیط زیست ـ ۸۸)

۱) دارد ـ دارد ۲) دارد ـ ندارد ۳) ندارد ـ دارد ۴) ندارد ـ ندارد

توزیع‌های نمونه‌ای

توزیع میانگین نمونه

۳. توزیع نمرات آزمون داوطلبان، نرمال با میانگین 72 و انحراف معیار 12 می‌باشد. از این جامعه یک نمونه n تایی انتخاب می‌کنیم. احتمال اینکه میانگین نمره ارزشیابی آن‌ها حداقل 69 باشد برابر 0.9772 است. می‌دانیم $P(Z \geq 2) = 0.0228$، مقدار n کدام است؟ (حسابداری و مدیریت ـ ۸۸)

۱) 36 ۲) 64 ۳) 81 ۴) 100

۴. احتمال اینکه میانگین به‌دست‌آمده از یک نمونه تصادفی 100تایی از جامعه‌ای که دارای میانگین 50 و انحراف معیار 10 است کمتر از 48 باشد، چند درصد است؟ (اقتصاد ـ ۸۶)

۱) 2.5 ۲) 5 ۳) 45 ۴) 47.5

۵. تعداد اتومبیل‌های فروخته‌شده توسط یک شرکت در ماه دارای میانگین 50 و انحراف معیار 10 دستگاه است. احتمال اینکه میانگین به‌دست‌آمده از یک نمونه تصادفی 100تایی کمتر از 48 دستگاه باشد، چقدر است؟ (اقتصاد ـ ۸۷)

۱) %2.5 ۲) %5 ۳) %45 ۴) %47.5

۶. طول عمر باتری‌های تولیدی کارخانه آلفا دارای توزیع نرمال با میانگین 200 ساعت و انحراف معیار 30 ساعت است. احتمال اینکه میانگین طول عمر باتری‌ها در یک نمونه تصادفی 36 تایی بین 190 تا 210 ساعت باشد، تقریباً چقدر است؟ **(اقتصاد ــ ۸۸)**

۱) %25 ۲) %32 ۳) %68 ۴) %95

۷. در کدام مورد، قانون چی‌بی‌شف برای تخمین μ_x به کار می‌رود؟ **(برنامه‌ریزی شهری ــ ۸۷)**

۱) توزیع $\bar{X}$ نرمال ۲) توزیع $\bar{X}$ غیرنرمال یا نامشخص

۳) توزیع جامعه آماری نرمال ۴) توزیع $\bar{X}$، t استیودنت

۸. نمونه‌های 50 تایی از یک جامعه 626 عضوی با واریانس 200 انتخاب می‌شود. انحراف معیار توزیع میانگین نمونه کدام است؟ **(برنامه‌ریزی شهری ــ ۸۸)**

۱) 1.85 ۲) 1.92 ۳) 2.05 ۴) 2.16

۹. یک نمونه 9 تایی از یک جامعه 65 عضوی با واریانس 14 انتخاب می‌شود. انحراف معیار توزیع میانگین نمونه کدام است؟ **(GIS ــ ۸۸)**

۱) 1.16 ۲) 1.25 ۳) 1.34 ۴) 1.43

۱۰. وزن بسته‌های شکر از توزیع نرمال با میانگین 2 کیلوگرم و انحراف معیار 200 گرم تبعیت می‌کند. از این بسته‌ها تعداد 9 بسته را انتخاب می‌کنیم. احتمال اینکه متوسط وزن بسته‌ها حداقل 2.2 کیلوگرم باشد، چقدر است؟ **(محیط زیست ــ ۸۶)**

۱) 0.0013 ۲) 0.023 ۳) 0.0023 ۴) 0.013

۱۱. کدام عبارت صحیح است؟ **(محیط زیست ــ ۸۶)**

۱) برای نمونه‌های کوچک واریانس میانگین نمونه‌ای بدون جایگذاری از واریانس میانگین نمونه‌ای با جایگذاری کوچک‌تر است.

۲) برای نمونه‌های کوچک واریانس میانگین نمونه‌ای بدون جایگذاری از واریانس میانگین نمونه‌ای با جایگذاری بزرگ‌تر است.

۳) برای نمونه‌های بزرگ واریانس میانگین نمونه‌ای بدون جایگذاری از واریانس میانگین نمونه‌ای با جایگذاری کوچک‌تر است.

۴) برای نمونه‌های بزرگ واریانس میانگین نمونه‌ای بدون جایگذاری از واریانس میانگین نمونه‌ای با جایگذاری بزرگ‌تر است.

۱۲. از یک جامعه متناهی به حجم N یک نمونه به حجم n بدون جایگذاری و از یک جامعه متناهی دیگر به حجم N یک نمونه تصادفی به حجم n با جایگذاری انتخاب می‌کنیم. نسبت واریانس میانگین نمونه‌ای با جایگذاری به بدون جایگذاری عبارت است از: **(محیط زیست ــ ۸۸)**

۱) $\frac{n}{N}$ ۲) $\frac{N-n}{N}$ ۳) $\frac{N-1}{N-n}$ ۴) $\frac{N-n}{N-1}$

۱۳. از یک جامعه آماری نرمال با میانگین 15 و انحراف معیار 3 یک نمونه تصادفی به حجم 9 اختیار شده است. احتمال اینکه میانگین نمونه از 18 بیشتر باشد، کدام است؟ **(محیط زیست ــ ۸۸)**

۱) 0.5 ۲) 0.0013 ۳) 0.4487 ۴) 0.9987

۱۴. فرض کنید متغیر تصادفی X دارای توزیع برنولی به صورت زیر باشد:

$$P(X=x)=\theta^{x}(1-\theta)^{1-x}\ ;\ x=0,1\ ;\ 0<\theta<1$$

طبق قضیه حد مرکزی کدام آماره به سمت توزیع نرمال استاندارد می‌رود؟ (محیط زیست ــ ۸۸)

۱) $\dfrac{\sqrt{n}(\bar{X}-n\theta)}{\sqrt{\theta(1-\theta)}}$ ۲) $\dfrac{\sqrt{n}(\bar{X}-n\theta)}{\sqrt{\dfrac{\theta(1-\theta)}{n}}}$ ۳) $\dfrac{(\bar{X}-\theta)}{\sqrt{\dfrac{\theta(1-\theta)}{n}}}$ ۴) $\dfrac{(\bar{X}-n\theta)}{\sqrt{\dfrac{\theta(1-\theta)}{n-1}}}$

۱۵. حجم نمونه در یک مطالعه چقدر باشد تا توزیع میانگین نمونه‌ای با توزیع داده‌ها یکسان باشد؟ (N حجم جامعه است.) (محیط زیست ــ ۸۸)

۱) $n=1$ ۲) $n=N$ ۳) $n=30$ ۴) بیشتر از 30 باشد.

محاسبه n از روی توزیع $\bar{X}$

۱۶. اگر توزیع نمونه‌گیری $\bar{X}$ دارای واریانس 4 و انحراف معیار جامعه آماری 12 باشد، حجم نمونه کدام است؟ (۸۶ ــ GIS)

۱) 24 ۲) 32 ۳) 36 ۴) 64

۱۷. توزیع نمونه‌ای $\bar{X}$ دارای انحراف معیار 2 است. اگر انحراف معیار جامعه آماری 12 باشد، مقدار n کدام است؟ (محیط زیست ــ ۸۸)

۱) 6 ۲) 36 ۳) 72 ۴) 144

توزیع تفاضل میانگین دو نمونه

۱۸. اگر یک نمونه 100تایی از جامعه اول با واریانس 9 و یک نمونه 25تایی از جامعه دوم با واریانس 4 انتخاب شوند و این دو نمونه مستقل از یکدیگر باشند، انحراف معیار تفاضل میانگین دو جامعه کدام است؟ (حسابداری و مدیریت ــ ۸۶)

۱) 0.25 ۲) 0.5 ۳) 1.25 ۴) 1.5

۱۹. اگر $\bar{X}_1$ میانگین نمونه‌ای به حجم 48 با واریانس 15 و $\bar{X}_2$ میانگین نمونه‌ای به حجم 36 با واریانس 9 از دو جامعه مستقل باشند، انحراف معیار $(\bar{X}_1-\bar{X}_2)$، کدام است؟ (۸۸ ــ GIS)

۱) $\dfrac{3}{2}$ ۲) $\dfrac{2}{3}$ ۳) $\dfrac{4}{3}$ ۴) $\dfrac{3}{4}$

۲۰. فرض کنید $\bar{X}_1$ میانگین نمونه‌ای به حجم n_1 و $\bar{X}_2$ میانگین نمونه‌ای به حجم n_2 از دو جامعه مستقل با واریانس برابر باشند، مقدار $Var(4\bar{X}_1-3\bar{X}_2)$ عبارت است از: (محیط زیست ــ ۸۸)

۱) $\sigma^2\left(\dfrac{4}{n_1}-\dfrac{3}{n_2}\right)$ ۲) $\sigma^2\left(\dfrac{16}{n_1}-\dfrac{9}{n_2}\right)$ ۳) $\sigma^2\left(\dfrac{4}{n_1}+\dfrac{3}{n_2}\right)$ ۴) $\sigma^2\left(\dfrac{16}{n_1}+\dfrac{9}{n_2}\right)$

توزیع نسبت نمونه

۲۱. برای تخمین نسبت موفقیت‌ها در جامعه‌ای دو نمونه تصادفی مستقل به حجم‌های n_1 و n_2 از جامعه گرفته شده و برآوردکننده زیر پیشنهاد شده است که در آن x_1 و x_2 تعداد موفقیت‌ها در نمونه اول و دوم است. کمیت انتظاری این برآوردکننده کدام است؟ **(اقتصاد ـ ۸۷)**

$$\hat{p} = \frac{1}{2}\left(\frac{x_1}{n_1} + \frac{x_2}{n_2}\right)$$

۱) p ۲) μ_x ۳) $\frac{p}{2n_1n_2}$ ۴) $\frac{(n_1+n_2)\mu_X}{n_1+n_2}$

۲۲. در آزمون فرضیه مربوط به یک نسبت خاص در جامعه، توزیع نمونه‌گیری $\bar{p}$ در نمونه‌های کوچک چیست؟ **(اقتصاد ـ ۸۷)**

۱) نمایی ۲) نرمال ۳) دوجمله‌ای ۴) پواسون

توزیع واریانس نمونه

۲۳. احتمال اینکه واریانس یک نمونه تصادفی 36 تایی از جامعه نرمالی کمتر از واریانس مربوط به آن جامعه باشد، کدام است؟ **(اقتصاد ـ ۸۶)**

۱) $P\left(F_{(1,\,35)} < 6\right)$ ۲) $P\left(F_{(1,\,36)} < 5\right)$ ۳) $P\left(\chi^2_{(35)} < 36\right)$ ۴) $P\left(\chi^2_{(35)} < 35\right)$

خواص برآوردکننده‌های نقطه‌ای

نااریبی

۲۴. اگر $\hat{\theta}$ برآوردکننده پارامتر θ با اریب (تورش) $k\theta + 5$ باشد، کدام برآوردکننده زیر نااریب (بدون تورش) است؟ **(اقتصاد ـ ۸۶)**

۱) $\frac{\hat{\theta} - 5}{k}$ ۲) $\frac{\hat{\theta} - 5}{k+1}$ ۳) $\frac{\hat{\theta}}{k} - \frac{5}{k+1}$ ۴) $(k+1)\hat{\theta} + \frac{5}{k+1}$

۲۵. یک برآوردکننده نااریب برای σ_X^2 کدام است؟ **(اقتصاد ـ ۸۸)**

۱) $\frac{\sum(x_i - \mu_x)^2}{n}$ ۲) $\frac{\sum(x_i - \bar{x})^2}{n}$ ۳) $\sum\frac{(x_i - \bar{x})^2}{n-2}$ ۴) $\frac{\sum(x_i - \mu_x)^2}{n-1}$

۲۶. اگر امید یک برآورد کننده بزرگ‌تر از پارامتر آن باشد، کدام عبارت صحیح است؟ **(محیط زیست ـ ۸۶)**

۱) برآوردکننده سازگار است.
۲) برآوردکننده دارای اریب است.
۳) برآوردکننده دارای کمترین واریانس است.
۴) شکل تابع برآوردکننده می‌تواند خاصیت برآوردکننده را بیان کند.

کارایی

۲۷. فرض کنید $\hat{\mu}$ و $\hat{\hat{\mu}}$ دو برآوردگر مستقل نااریب از پارامتر μ می‌باشند. به‌علاوه انحراف معیار $\hat{\hat{\mu}}$ پنج برابر $\hat{\mu}$ است. با ترکیب $\hat{\mu}$ و $\hat{\hat{\mu}}$ سه برآوردگر به صورت زیر برای برآورد μ پیشنهاد شده است:

$$w_1=\frac{1}{2}\left(\hat{\mu}+\hat{\hat{\mu}}\right)\qquad w_2=\frac{4}{5}\hat{\mu}+\frac{1}{5}\hat{\hat{\mu}}\qquad w_3=\hat{\mu}$$

این برآوردگرها به ترتیب کارایی (از راست به چپ) عبارت‌اند از: **(اقتصاد ـ ۸۷)**

۱) w_1, w_2, w_3 ۲) w_2, w_1, w_3 ۳) w_3, w_1, w_2 ۴) w_1, w_3, w_2

۲۸. اگر توزیع X نرمال بوده و دو تخمین‌زننده $S^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$ و $\hat{\sigma}^2=\frac{\sum(x_i-\mu)^2}{n}$ برای تخمین σ_X^2 مورد نظر باشد، به ازای $n=10$ ضریب کارایی (نسبت واریانس $\hat{\sigma}^2$ به S^2) چیست؟ **(اقتصاد ـ ۸۷)**

۱) 0.19 ۲) 0.81 ۳) 0.9 ۴) 1.1

۲۹. سه تخمین‌زننده

$$T_1=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3 \quad ، \quad T_2=\frac{1}{4}X_1+\frac{1}{4}X_2+\frac{2}{4}X_3 \quad و \quad T_3=\frac{1}{8}X_1+\frac{2}{8}X_2+\frac{5}{8}X_3.$$

وجود دارند. کدام رابطه بین واریانس آن‌ها برقرار است؟ **(محیط زیست ـ ۸۶)**

۱) $Var(T_1)<Var(T_2)<Var(T_3)$ ۲) $Var(T_2)<Var(T_3)<Var(T_1)$

۳) $Var(T_2)<Var(T_1)<Var(T_3)$ ۴) $Var(T_3)<Var(T_2)<Var(T_1)$

۳۰. هرگاه θ پارامتر جامعه و T_1 و T_2 دو برآوردکننده نااریب برای θ باشند و داشته باشیم $Var(T_1)=6$ و $Var(T_2)=2$، کدام‌یک برآوردکننده بهتری برای θ است؟ **(محیط زیست ـ ۸۷)**

۱) T_1 ۲) T_2 ۳) هر دو به یک اندازه ۴) هیچ‌کدام

۳۱. در جامعه‌ای، برای برآورد θ سه برآوردکننده زیر مفروض است. با فرض $E(X)=\theta$ کدام برآوردگر نسبت به سایر برآوردگرها کاراتر است؟ **(محیط زیست ـ ۸۸)**

$$T_1=\frac{X_1}{3}+\frac{X_2}{3}+\frac{X_3}{3}\quad ،T_2=\frac{X_1}{8}+\frac{3X_2}{8}+\frac{4}{8}X_3\quad ،\quad T_3=\frac{5X_1}{10}+\frac{4X_2}{10}+\frac{X_3}{10}$$

۱) T_1 ۲) T_2

۳) T_3 ۴) نمی‌توان به طور دقیق تصمیم‌گیری نمود.

میانگین مجذور خطا (MSE)

۳۲. دو برآوردکننده $\hat{\theta}_1$ و $\hat{\theta}_2$ با ویژگی‌های زیر برای برآورد پارامتر θ پیشنهاد شده است: $Var(\hat{\theta}_2)=50$ و $E(\hat{\theta}_2-\theta)=6$ و $Var(\hat{\theta}_1)=90$ و $E(\hat{\theta}_1-\theta)=0$ آن‌گاه: **(اقتصاد ـ ۸۶)**

۱) $\hat{\theta}_2$ مناسب‌تر است، چون کاراتر است.

۲) $\hat{\theta}_1$ مناسب‌تر است، زیرا یک برآوردکننده نااریب (بدون تورش) است.

۳) $\hat{\theta}_1$ مناسب‌تر است، چون میانگین مجذور خطای آن (MSE) کمتر است.

۴) $\hat{\theta}_2$ مناسب‌تر است، چون میانگین مجذور خطای آن (MSE) کمتر است.

۳۳. به منظور برآورد میانگین جامعه بر اساس یک نمونه تصادفی سه‌تایی، دو برآوردکننده A و B زیر پیشنهاد شده است. برای تشخیص آنکه کدام‌یک برآوردکننده بهتری است، چه ملاکی کفایت می‌کند؟ (اقتصاد ــ ۸۷)

$$A=\frac{X_1+X_2+X_3}{2} \qquad B=\frac{X_1+X_2+X_3}{3}+2$$

۱) تورش ۲) واریانس ۳) واریانس + $(\text{تورش})^2$ ۴) واریانس + تورش

۳۴. برای تشخیص مناسب بودن برآوردکننده $\hat{\theta}$ برای پارامتر θ کدام‌یک از موارد زیر ملاک عمل است؟ (اقتصاد ــ ۸۸)

۱) $E(\hat{\theta})$ ۲) $E\left(\hat{\theta}-E(\hat{\theta})\right)^2$ ۳) $E(\hat{\theta}-\theta)^2$ ۴) $E(\hat{\theta})+Var(\hat{\theta})$

برآورد فاصله‌ای

فاصله اطمینان میانگین جامعه

۳۵. نمرات یک نمونه تصادفی 3 تایی از دانشجویان کلاسی که دارای توزیع نرمال است، 16، 15 و 17 بوده است. فاصله اطمینان 90% میانگین نمرات دانشجویان کدام است؟ $(t\approx 3)$ (اقتصاد ــ ۸۶)

۱) $15.3-16.7$ ۲) $14.3-17.7$ ۳) $13.9-18.1$ ۴) $13.7-18.3$

۳۶. وام مسکن پرداختی به یک نمونه تصادفی 3 تایی از مشتریان بانک که دارای توزیع نرمال است. 16، 15 و 17 میلیون تومان بوده است. فاصله اطمینان $1-\alpha$ برای میانگین وام پرداختی بانک چقدر است؟ (اقتصاد ــ ۸۷)

۱) $16\pm Z_{\frac{\alpha}{2}}\sqrt{3}$ ۲) $16\pm\frac{Z_\alpha}{\sqrt{2}}$ ۳) $16\pm\frac{1}{\sqrt{2}}t_{\frac{\alpha}{2},3}$ ۴) $16\pm\frac{1}{\sqrt{3}}t_{\frac{\alpha}{2},2}$

۳۷. از 500 مشتری یک فروشگاه، دو نفر که به تصادف انتخاب شده‌اند، یکی 30 و دیگری 50 هزار تومان خرید کرده است. فاصله اطمینان 95% برای کل فروش این فروشگاه چقدر است؟ $(t=3)$ (اقتصاد ــ ۸۸)

۱) $15000-25000$ ۲) $10000-30000$ ۳) $7000-33000$ ۴) $5000-35000$

۳۸. از میان دانشجویان یک کلاس، 9 نفر را به عنوان نمونه اختیار می‌کنیم. میانگین قد این دانشجویان 160 سانتی‌متر و انحراف معیار 5 سانتی‌متر است. میانگین جامعه با احتمال 95 درصد در کدام فاصله است؟(GIS ــ ۸۸)

۱) $(154.8\ ,\ 165.2)$ ۲) $(155.5\ ,\ 164.5)$
۳) $(156.8\ ,\ 163.2)$ ۴) $(156.4\ ,\ 163.6)$

۳۹. فرض کنید یک نمونه 25 تایی از توزیع نرمال با میانگین μ و واریانس 16 انتخاب شده است و میانگین نمونه‌ای برابر با 10 است. یک فاصله اطمینان 95% برای میانگین جامعه کدام است؟ $(Z_{0.025}=2)$ (محیط زیست ــ ۸۷)

۱) $10\pm\frac{8}{5}$ ۲) $10\pm 2\times\frac{16}{5}$ ۳) $10\pm\frac{32}{25}$ ۴) $\frac{10}{25}\pm\frac{16}{5}$

محاسبه n براساس فاصله μ

۴۰. برای برآورد میانگین یک جامعه نرمال، حجم نمونه چقدر باید باشد تا حداکثر خطای برآورد برابر $\frac{1}{4}$ انحراف معیار جامعه باشد؟ $(Z_{0.025}\simeq 2)$ (اقتصاد ــ ۸۶)

۱) 16 ۲) 32 ۳) 52 ۴) 64

۴۱ . اگر انحراف معیار جامعه‌ای 20 و میزان دقت برآورد 4 باشد حداقل تعداد نمونه لازم برای به دست آوردن فاصله اطمینان 95 درصد میانگین کدام است؟ (برنامه‌ریزی شهری ــ ۸۸)

۱) 56 ۲) 96 ۳) 116 ۴) 120

۴۲ . فردی برای جمع‌آوری اطلاعات و برآورد میانگین صفت در یک جامعه بزرگ با واریانس $\sigma^2 = 25$ چه تعداد نمونه باید انتخاب نماید تا با 95% اطمینان خطای برآورد حداکثر 1 باشد؟ $(Z_{0.025} = 2)$ (محیط زیست ــ ۸۷)

۱) 79 ۲) 85 ۳) 90 ۴) 100

روابط بین 2e , e , n

۴۳ . فرض کنید متغیر تصادفی X دارای توزیع نرمال باشد. اگر واریانس جامعه نصف شود، آن‌گاه طول فاصله اطمینان: (محیط زیست ــ ۸۶)

۱) $\frac{1}{2}$ برابر می‌شود. ۲) $\frac{\sqrt{2}}{2}$ برابر می‌شود. ۳) $\sqrt{2}$ برابر می‌شود. ۴) 2 برابر می‌شود.

فاصله اطمینان تفاضل یا مجموع میانگین دو جامعه

۴۴ . برآورد فاصله‌ای مجموع متوسط وزن مسافر (μ_1) و بار همراه وی (μ_2) در یک پرواز بر اساس یک نمونه تصادفی n_1 تایی از وزن مسافران و یک نمونه تصادفی n_2 تایی مستقل از بار همراه مسافران کدام است؟

(اقتصاد ــ ۸۶)

۱) $(\bar{X}_1 + \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

۲) $(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2},r} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

۳) $(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

۴) $(\mu_1 + \mu_2) \pm t_{\frac{\alpha}{2},r} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

۴۵ . اطلاعات زیر از دو جامعه مستقل از یکدیگر نرمال وجود دارد. یک فاصله 95% برای تفاضل دو میانگین جامعه $(\mu_1 - \mu_2)$ عبارت است از: (محیط زیست ــ ۸۶)

جامعه دوم	جامعه اول
$n_2 = 5$ $\bar{x}_2 = 170$ $\sum (x - \bar{x}_2)^2 = 866$	$n_1 = 5$ $\bar{x}_1 = 174$ $\sum (x - \bar{x}_1)^2 = 944$

۱) $4 \pm 1.96\sqrt{72.4}$

۲) $4 \pm 1.96\sqrt{90.5}$

۳) $4 \pm 2.31\sqrt{72.4}$

۴) $4 \pm 2.31\sqrt{90.5}$

فاصله اطمینان نسبت جامعه

۴۶ . بر اساس یک نمونه تصادفی 100تایی از نوزادان تازه متولدشده، 50 نفر پسر بوده‌اند. فاصله اطمینان 95% برای نسبت واقعی پسران متولدشده کدام است؟ $(Z_{0.025} = 1.96 , Z_{0.05} = 1.64)$ (اقتصاد ــ ۸۶)

۱) $0.418 - 0.582$ ۲) $0.4903 - 0.5097$ ۳) $0.402 - 0.598$ ۴) $0.4918 - 0.5082$

۴۷. یک نمونه تصادفی 100 نفری از بین رأی‌دهندگان یک شهر انتخاب و مشخص شده است که 80% آن‌ها به کاندیدای A رأی می‌دهند. یک فاصله اطمینان 90% برای نسبت افراد در جامعه که به A رأی خواهند داد برابر است با: $(Z_{0.05}=1.65)$ **(اقتصاد ــ ۸۷)**

۱) (0.4 تا 1) ۲) (0.83 تا 0.97) ۳) (0.895 تا 0.905) ۴) (0.88 تا 0.92)

۴۸. در یک نمونه 100تایی از تولیدات کارخانه‌ای 89 کالا سالم و بقیه معیوب هستند. یک بازه 90% برای تولیدات سالم عبارت است از: $(Z_{0.95}=1.64\ ,\ Z_{0.975}=1.96)$ **(محیط زیست ــ ۸۶)**

۱) $0.89\pm1.64\sqrt{\frac{0.89\times0.11}{99}}$ ۲) $0.89\pm1.64\sqrt{\frac{0.89\times0.11}{100}}$

۳) $0.89\pm1.96\sqrt{\frac{0.89\times0.11}{99}}$ ۴) $0.89\pm1.96\sqrt{\frac{0.89\times0.11}{100}}$

۴۹. در یک نمونه تصادفی به حجم $n=125$ از محصولات یک کارخانه دیده شده است که 25 کالای تولیدی به طور غیراستاندارد تولید شده است. یک فاصله اطمینان 95% برای نسبت کالاهای معیوب تولیدی در یک کارخانه عبارت است از: $(Z_{0.025}=2)$ **(محیط زیست ــ ۸۷)**

۱) $\frac{1}{5}\pm\frac{20}{125\sqrt{5}}$ ۲) $\frac{1}{5}\pm\frac{125\times2}{\sqrt{125}}$ ۳) $\frac{1}{5}\pm\frac{20}{5\sqrt{5}}$ ۴) $\frac{1}{5}\pm\frac{125}{\sqrt{5}}$

۵۰. در میان تولیدات یک کارخانه معمولاً تعدادی کالای معیوب یافت می‌شود. اگر θ نسبت کالای معیوب باشد و در یک نمونه 36تایی 6 کالای معیوب موجود باشد یک فاصله اطمینان 95% برای θ عبارت است از:

$\left(Z_{1-\frac{\alpha}{2}}=Z_{0.975}\approx2\right)$

(محیط زیست ــ ۸۸)

۱) $\frac{1}{6}\pm\frac{1}{18}\sqrt{5}$ ۲) $\frac{1}{6}\pm\frac{1}{18}\sqrt{\frac{5}{6}}$ ۳) $\frac{5}{6}\pm\frac{1}{18}\sqrt{5}$ ۴) $\frac{15}{6}\pm\frac{1}{18}\sqrt{\frac{5}{6}}$

محاسبه n براساس فاصله p

۵۱. اگر بخواهیم نسبت افراد باسواد یک جامعه را با خطای ±0.02 و ضریب اطمینان 95% برآوردکنیم، تقریباً چه حجم نمونه‌ای مناسب است؟ **(اقتصاد ــ ۸۷)**

۱) 1000 ۲) 2500 ۳) 5000 ۴) 7500

پاسخ‌های تشریحی

نمونه‌گیری

۱. گزینه ۱ درست است.

در مواردی که چارچوب نمونه‌گیری مناسبی از جامعه مورد بررسی در دسترس نباشد از روش نمونه‌گیری سیستماتیک استفاده می‌شود. در این سؤال نیز جامعه مورد بررسی، اتومبیل‌های سواری واردشده به یک بزرگراه هستند که چارچوب معینی از آن در دسترس نیست.

۲. گزینه ۳ درست است.

در نمونه‌گیری به روش طبقه‌بندی، واریانس بین طبقات جامعه زیاد و واریانس درون هر طبقه کم است. یعنی بین طبقه‌ها همگنی وجود **ندارد** و درون هر طبقه همگنی وجود **دارد**.

توزیع‌های نمونه‌ای

توزیع میانگین نمونه

۳. گزینه ۲ درست است.

نمرات آزمون $X \sim N\left(\mu = 72\ ,\ \sigma^2 = 12^2\right)$

$$(۱)\ \ P\left(\bar{X} \geq 69\right) = 0.9772 \quad \rightarrow \quad P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{69-72}{\frac{12}{\sqrt{n}}}\right) = 0.9772 \quad \rightarrow \quad P\left(Z \geq \frac{-3\sqrt{n}}{12}\right) = 0.9772$$

$$(۲)\ \ P\left(Z \geq 2\right) = 0.0228 \rightarrow P\left(Z < 2\right) = 1 - 0.0228 = 0.9772$$

حال با توجه به $P\left(Z < a\right) = P\left(Z > b\right) \rightarrow a = -b$، داریم:

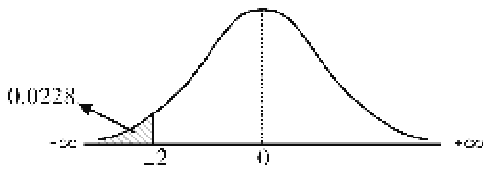
$$\frac{-\sqrt{n}}{4} = -2 \rightarrow \sqrt{n} = 8 \rightarrow \boxed{n = 64}$$

۴ . گزینه ۱ درست است.

جامعه غیرنرمال (چون نرمال بودن ذکر نشده)، واریانس جامعه معلوم و $n = 100 > 30$ است. پس بنا بر قضیه حد مرکزی مورد (۳ ـ الف) از توزیع‌های نمونه‌ای $\bar{X}$ داریم:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ و } Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\begin{cases} P(\bar{X}<48) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{48-50}{\frac{10}{\sqrt{100}}}\right) = P(Z<-2) = 0.0228 \simeq \%2.5 \\ \mu = 50 \ , \ \sigma = 10 \ , \ n = 100 \end{cases}$$



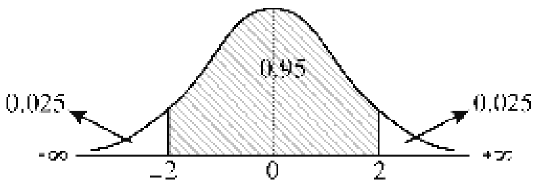
۵ . گزینه ۱ درست است.

توزیع جامعه نامعلوم (چون از نرمال بودن چیزی گفته نشده)، واریانس جامعه معلوم و $n > 30$ است پس بنا بر قضیه حد مرکزی (مورد ۳ـ الف) توزیع $\bar{x}$ نرمال است.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\begin{cases} P(\bar{X} < 48) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{48-50}{\frac{10}{\sqrt{100}}}\right) = P(Z < -2) = 0.025 = \%2.5 \\ \mu = 50 \ , \ \sigma = 10 \ , \ n = 100 \end{cases}$$

یادآوری:



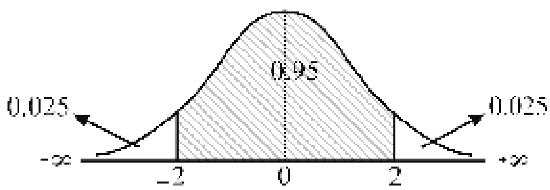
۶ . گزینه ۴ درست است.

طول عمر باتری : $X \sim N(\mu = 200 \ , \ \sigma^2 = 30^2)$

$$P(190 < \bar{X} < 210) = P\left(\frac{190-200}{\frac{30}{\sqrt{36}}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{210-200}{\frac{30}{\sqrt{36}}}\right) = P(-2 < Z < 2) = 0.95$$

یادآوری:

$$P(-2 < Z < 2) = 0.95$$



۷ . گزینه ۲ درست است.

یادآوری: هرگاه توزیع جامعه غیرنرمال (یا توزیع جامعه را ندانیم)، واریانس جامعه معلوم و $n \le 30$ باشد، بنا بر مورد (۳ ـ ب) از توزیع‌های نمونه‌ای $\bar{X}$، توزیع $\bar{X}$ نامعلوم است و در این شرایط از قضیه چی‌بی‌شف برای تخمین $\mu_{\bar{X}}$ استفاده می‌شود.

۸ . گزینه ۲ درست است.

جامعه محدود $N = 626$ و پیش‌فرض نمونه‌گیری بدون جایگذاری است، بنابراین واریانس میانگین نمونه عبارت است از:

$$\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{N-n}{N-1}.\frac{\sigma^2}{n} \rightarrow \sigma_{\bar{X}} = \sqrt{\frac{N-n}{N-1}.\frac{\sigma^2}{n}} = \sqrt{\frac{626-50}{626-1} \times \frac{200}{50}} = \frac{24}{25} \times 2 = 1.92$$

یادآوری:

$$\begin{cases} Var(\bar{X}) = \dfrac{N-n}{N-1}.\dfrac{\sigma^2}{n} & \text{نمونه‌گیری بدون جایگذاری} \\ Var(\bar{X}) = \dfrac{\sigma^2}{n} & \text{نمونه‌گیری با جایگذاری} \end{cases}$$

۹ . گزینه ۱ درست است.

یادآوری:

$$Var(\bar{X}) = \begin{cases} \dfrac{N-n}{N-1}\dfrac{\sigma^2}{n} & \text{نمونه‌گیری بدون جایگذاری} \\ \dfrac{\sigma^2}{n} & \text{نمونه‌گیری با جایگذاری} \end{cases}$$

با توجه به اینکه جامعه محدود است $N = 65$ و پیش‌فرض نمونه‌گیری بدون جایگذاری است داریم:

$$\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{N-n}{N-1}.\frac{\sigma^2}{n} = \frac{65-9}{65-1} \times \frac{14}{9} = \frac{56}{64} \times \frac{14}{9} = \frac{49}{36} \rightarrow \sigma_{\bar{X}} = \sqrt{\frac{49}{36}} = \frac{7}{6} = 1.16$$

۱۰ . گزینه ۱ درست است.

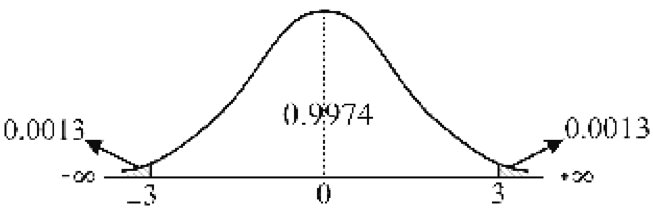
وزن بسته‌های شکر : $X \sim N\left(\mu = 2 \quad , \quad \sigma^2 = (0.2)^2\right)$

دقت کنید که میانگین و انحراف معیار باید بر طبق یک واحد (مقیاس) باشند به عبارت دیگر چون میانگین برحسب کیلوگرم است پس انحراف معیار نیز باید برحسب کیلوگرم باشد:

$$\sigma = 200gr = \frac{200}{1000} = 0.2kg$$

$$P(\bar{X} \ge 2.2) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \ge \frac{2.2-2}{\frac{0.2}{\sqrt{9}}}\right) = P\left(Z \ge \frac{0.2}{\frac{0.2}{3}}\right) = P(Z \ge 3) = 0.0013$$

یادآوری:



۱۱ . گزینه ۱ درست است.

یادآوری:

$$Var(\bar{X})=\begin{cases}\dfrac{N-n}{N-1}\dfrac{\sigma^2}{n} & \text{نمونه‌گیری بدون جایگذاری}\\[2ex] \dfrac{\sigma^2}{n} & \text{نمونه‌گیری با جایگذاری}\end{cases}$$

می‌دانیم که ضریب تصحیح واریانس $\left(\dfrac{N-n}{N-1}\right)$، همواره عددی بین صفر و یک است، بنابراین واریانس میانگین نمونه را کوچک‌تر خواهد کرد: $\dfrac{N-n}{N-1}\cdot\dfrac{\sigma^2}{n}<\dfrac{\sigma^2}{n}$

همچنین می‌دانیم درصورتی‌که تعداد نمونه به نسبت جامعه بزرگ باشد یعنی $\dfrac{n}{N}\le 0.05$، از ضریب تصحیح واریانس $\left(\dfrac{N-n}{N-1}\right)$ در شرایط نمونه‌گیری بدون جایگذاری چشم‌پوشی می‌شود و با حالت با جایگذاری برابر خواهد شد.

بنابراین دو جمله زیر همواره درست است:

برای نمونه‌های کوچک واریانس میانگین نمونه‌ای بدون جایگذاری از واریانس میانگین نمونه‌ای با جایگذاری کوچک‌تر است.

برای نمونه‌های بزرگ واریانس میانگین نمونه‌ای بدون جایگذاری با واریانس میانگین نمونه‌ای با جایگذاری برابر است.

۱۲ .گزینه ۳ درست است.

یادآوری:

$$Var(\bar{X})=\begin{cases}\dfrac{N-n}{N-1}\dfrac{\sigma^2}{n} & \text{نمونه‌گیری بدون جایگذاری}\\[2ex] \dfrac{\sigma^2}{n} & \text{نمونه‌گیری با جایگذاری}\end{cases}$$

$$\frac{\text{واریانس میانگین نمونه‌ای با جایگذاری}}{\text{واریانس میانگین نمونه‌ای بدون جایگذاری}}=\frac{\dfrac{\sigma^2}{n}}{\dfrac{N-n}{N-1}\cdot\dfrac{\sigma^2}{n}}=\frac{N-1}{N-n}$$

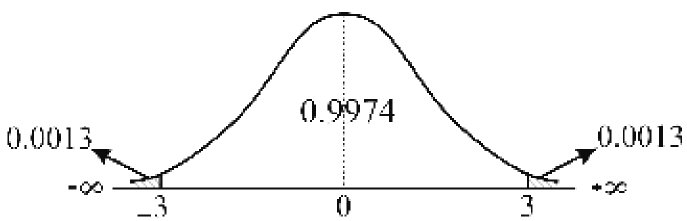
۱۳ . گزینه ۲ درست است.

$$X_i\sim N\left(\mu=15\ ,\ \sigma^2=3^2\right)$$

جامعه نرمال و واریانس جامعه معلوم است، بنابراین n هر چه باشد توزیع $\bar{X}$ نرمال خواهد بود.

$$P\left(\bar{X}>18\right)=P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}>\frac{18-15}{\frac{3}{\sqrt{9}}}\right)=P(Z>3)=0.0013$$

یادآوری:



۱۴. گزینه ۳ درست است.

یادآوری: قضیه حد مرکزی: هرگاه توزیع جامعه غیر نرمال اما واریانس جامعه معلوم و تعداد نمونه (n) بزرگ باشد توزیع $\bar{X}$ نرمال خواهد بود:

$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$X\sim Bin(1,\theta)\to\begin{cases}\mu=E(X)=\theta\\ \sigma^2=Var(X)=\theta(1-\theta)\end{cases}$$

$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$$

۱۵. گزینه ۱ درست است.

توزیع میانگین نمونه:

$$\begin{cases}E(\bar{X})=\mu\ \text{(میانگین جامعه)}\\ Var(\bar{X})=\frac{\sigma^2}{n}\xrightarrow{n=1}Var(\bar{X})=\sigma^2\ \text{(واریانس جامعه)}\end{cases}$$

پس تنها درصورتی‌که تعداد نمونه‌ها $n=1$ باشد توزیع میانگین نمونه با توزیع داده‌ها (جامعه) یکسان است.

محاسبه n از روی توزیع $\bar{X}$

۱۶. گزینه ۳ درست است.

یادآوری: توزیع نمونه‌ای $\bar{X}$ دارای میانگین μ_X و واریانس $\frac{\sigma_X^2}{n}$ است، بنابراین:

$$\sigma_{\bar{X}}^2=\frac{\sigma_X^2}{n}\to\sigma_{\bar{X}}=\frac{\sigma_X}{\sqrt{n}}\xrightarrow[\sigma_X=12]{\sigma_{\bar{X}}=\sqrt{4}=2}2=\frac{12}{\sqrt{n}}\to\sqrt{n}=6\to n=36$$

۱۷. گزینه ۲ درست است.

$$Var(\bar{X})=\frac{\sigma^2}{n}\to\sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}\to 2=\frac{12}{\sqrt{n}}\to\sqrt{n}=6\to n=36$$

$$\sigma_{\bar{X}}=2\ ,\ \sigma=12$$

توزیع تفاضل یا مجموع میانگین دو نمونه

۱۸. گزینه ۲ درست است.

با آنکه ذکر نشده جامعه نرمال است، اما چون در توزیع‌های نمونه‌ای $\bar{X}_1\pm\bar{X}_2$ موردی که جامعه‌ها نرمال نباشد نداریم، پیش فرض، جامعه‌ها را نرمال فرض می‌کنیم؛ واریانس جوامع هم معلوم است و $n_1,n_2\geq 1$ است، پس بنا بر حالت (۱) از توزیع‌های نمونه‌ای $\bar{\mathbf{X}}_1\pm\bar{\mathbf{X}}_2$ داریم:

$$\bar{X}_1\pm\bar{X}_2\sim N\left(\mu_1\pm\mu_2\ ,\ \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

(انحراف معیار تفاضل میانگین دو نمونه) $\sigma^2(\bar{X}_1-\bar{X}_2)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}=\frac{9}{100}+\frac{4}{25}=0.25 \rightarrow \sigma(\bar{X}_1-\bar{X}_2)=0.5$

دقت کنید، به دلیل مستقل بودن دو نمونه، جزء $-2\mathrm{Cov}(\bar{X}_1,\bar{X}_2)$ برابر صفر می‌شود. بنابراین در تمامی موارد توزیع‌های نمونه‌ای $\bar{X}_1\pm\bar{X}_2$ جزء کوواریانس صفر در نظر گرفته شده است.

اشتباه سؤال: طراح باید می‌پرسید انحراف معیار تفاضل میانگین دو نمونه کدام است نه دو جامعه.

توجه: درصورتی‌که سؤال، واریانس تفاضل میانگین دو نمونه را می‌خواست، گزینه ۱ درست بود.

۱۹. گزینه ۴ درست است.

$$\mathrm{Var}(\bar{X}_1-\bar{X}_2)=\mathrm{Var}(\bar{X}_1)+\mathrm{Var}(\bar{X}_2)-\underbrace{\cancelto{0}{2\mathrm{Cov}(\bar{X}_1,\bar{X}_2)}}_{\text{به دلیل استقلال}}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}=\frac{15}{48}+\frac{9}{36}=\frac{9}{16}$$

$$\sqrt{\mathrm{Var}(\bar{X}_1-\bar{X}_2)}=\sqrt{\frac{9}{16}}=\frac{3}{4}$$

۲۰. گزینه ۴ درست است.

یادآوری: $\mathrm{Var}(\bar{X})=\frac{\sigma^2}{n}$

$$\mathrm{Var}(4\bar{X}_1-3\bar{X}_2)=4^2\,\mathrm{Var}(\bar{X}_1)+(-3)^2\,\mathrm{Var}(\bar{X}_2)=16\frac{\sigma^2}{n_1}+9\frac{\sigma^2}{n_2}=\sigma^2\left(\frac{16}{n_1}+\frac{9}{n_2}\right)$$

توجه کنید که چون دو جامعه مستقل از هم هستند، جزء کوواریانس $\mathrm{Cov}(\bar{X}_1,\bar{X}_2)$ در محاسبه واریانس بالا برابر صفر خواهد شد که از نوشتن آن صرف‌نظر کرده‌ایم.

توزیع نسبت نمونه

۲۱. گزینه ۱ درست است.

منظور از کمیت انتظاری همان امید ریاضی برآوردکننده است، بنابراین:

$$\begin{cases} E(\hat{p})=E\left[\frac{1}{2}\left(\frac{X_1}{n_1}+\frac{X_2}{n_2}\right)\right]=\frac{1}{2}\left[\frac{E(X_1)}{n_1}+\frac{E(X_2)}{n_2}\right]=\frac{1}{2}\left(\frac{n_1p}{n_1}+\frac{n_2p}{n_2}\right)=\frac{1}{2}(p+p)=p \\ E(X_1)=n_1p\ ,\ E(X_2)=n_2p \end{cases}$$

دقت کنید که X_1 و X_2 هر دو دارای توزیع دوجمله‌ای هستند.

۲۲. گزینه ۳ درست است.

یادآوری: در نمونه‌های کوچک $(n\le 30)$ توزیع $\bar{p}$ دوجمله‌ای و برای نمونه‌های بزرگ $(n>30)$ توزیع $\bar{p}$ بنا بر قضیه حد مرکزی نرمال است.

توزیع واریانس نمونه

۲۳. گزینه ۴ درست است.

$$P(S^2<\sigma^2)=P\left(\frac{(n-1)S^2}{\sigma^2}<\frac{(n-1)\sigma^2}{\sigma^2}\right)=P\left(\chi^2_{(n-1)}<n-1\right)=P\left(\chi^2_{(35)}<35\right)$$

خواص برآوردکننده‌های نقطه‌ای

نااریبی

۲۴. گزینه ۲ درست است.

ابتدا رابطه مربوط به اریبی را در نظر می‌گیریم:

اریبی: $E\left(\hat{\theta}\right)-\theta=k\theta+5$

حال ازآنجاکه شرط نااریبی برآوردکننده دلخواه T برای پارامتر θ برابر با $E(T)=\theta$ است، کافی است در رابطه بالا θ را به سمت راست تساوی منتقل کرده و بقیه مقادیر را به سمت چپ، داخل امید ریاضی قرار دهیم تا برآوردکننده نااریب T مشخص شود:

$$E\left(\hat{\theta}\right)-\theta=\theta k+5 \ \rightarrow\ E\left(\hat{\theta}\right)-5=\theta k+\theta \ \rightarrow\ E\left(\hat{\theta}-5\right)=(k+1)\theta \ \rightarrow\ E\left(\frac{\hat{\theta}-5}{k+1}\right)=\theta$$

۲۵. گزینه ۱ درست است.

یادآوری: تنها برآوردکننده‌های نااریب برای واریانس جامعه $\left(\sigma_X^2\right)$ عبارت‌اند از:

$S_X^2=\dfrac{\sum(x_i-\mu)^2}{n}$ در صورت معلوم بودن میانگین جامعه

$S_X^2=\dfrac{\sum(x_i-\bar{x})^2}{n-1}$ در صورت مجهول بودن میانگین جامعه

۲۶. گزینه ۲ درست است.

یادآوری:

$E(\text{آماره})=\text{پارامتر}$ $\rightarrow$ برآورد کننده (آماره) ناریب است

$E(\text{آماره})>\text{پارامتر}$ $\rightarrow$ برآورد کننده (آماره) دارای اریبی است

کارایی

۲۷. گزینه ۱ درست است.

$\hat{\mu}$ و $\hat{\hat{\mu}}$ دو برآوردگر نااریب برای μ و مستقل‌اند، بنابراین: $E(\hat{\mu})=E\left(\hat{\hat{\mu}}\right)=\mu$

ابتدا نااریبی برآوردگرها را بررسی می‌کنیم:

$$E(W_1)=E\left(\frac{1}{2}\left(\hat{\mu}+\hat{\hat{\mu}}\right)\right)=\frac{1}{2}E(\hat{\mu})+\frac{1}{2}E\left(\hat{\hat{\mu}}\right)=\frac{1}{2}\mu+\frac{1}{2}\mu=\mu \quad \text{نااریب}$$

$$E(W_2)=E\left(\frac{4}{5}\hat{\mu}+\frac{1}{5}\hat{\hat{\mu}}\right)=\frac{4}{5}E(\hat{\mu})+\frac{1}{5}E\left(\hat{\hat{\mu}}\right)=\frac{4}{5}\mu+\frac{1}{5}\mu=\mu \quad \text{نااریب}$$

$$E(W_3)=E(\hat{\mu})=\mu \quad \text{نااریب}$$

یادآوری: در بین چند برآوردکننده نااریب، برآوردکننده‌ای کاراتر است که واریانس کمتری داشته باشد، بنابراین باید واریانس برآوردگرها را محاسبه کنیم:

$$\sigma_{\hat{\hat{\mu}}}=5\sigma_{\hat{\mu}} \ \rightarrow\ \sigma_{\hat{\hat{\mu}}}^2=25\sigma_{\hat{\mu}}^2$$

$$Var(W_1)=Var\left(\frac{1}{2}\left(\hat{\mu}+\hat{\hat{\mu}}\right)\right)=\frac{1}{4}\left(\sigma_{\hat{\mu}}^2+\sigma_{\hat{\hat{\mu}}}^2\right)=\frac{1}{4}\left(\sigma_{\hat{\mu}}^2+25\sigma_{\hat{\mu}}^2\right)=\frac{26}{4}\sigma_{\hat{\mu}}^2$$

$$Var(W_2)=Var\left(\frac{4}{5}\hat{\mu}+\frac{1}{5}\hat{\mu}\right)=\frac{16}{25}\sigma^2_{\hat{\mu}}+\frac{1}{25}\sigma^2_{\hat{\mu}}=\frac{16}{25}\sigma^2_{\hat{\mu}}+\frac{1}{25}25\sigma^2_{\hat{\mu}}=\frac{41}{25}\sigma^2_{\hat{\mu}}$$

$$Var(W_3)=Var(\hat{\mu})=\sigma^2_{\hat{\mu}}$$

بنابراین به ترتیب $\sigma^2_{\hat{\mu}}<\frac{41}{25}\sigma^2_{\hat{\mu}}<\frac{26}{4}\sigma^2_{\hat{\mu}} \leftarrow \sigma^2(W_3)<\sigma^2(W_2)<\sigma^2(W_1)$ بنابراین کارایی آن‌ها به ترتیب W_1, W_2, W_3 است.

یادآوری: هرگاه X و Y دو متغیر تصادفی مستقل باشند:

$$Var(aX+bY)=a^2\sigma^2_X+b^2\sigma^2_Y+2ab\underbrace{\cancel{Cov(X,Y)}}_{0}=a^2\sigma^2_X+b^2\sigma^2_Y$$

۲۸. گزینه ۳ درست است.

یادآوری:

۱) $$\frac{\sum(x_i-\mu)^2}{\sigma^2}\sim\chi^2_{(n)} \quad ; \quad \frac{\sum(x_i-\bar{x})^2}{\sigma^2}\sim\chi^2_{(n-1)}$$

۲) امید و واریانس متغیر X با توزیع χ^2 با k درجه آزادی عبارت است از:

$$E(X)=k \ , \ Var(X)=2k$$

نسبت واریانس $\hat{\sigma}^2$ به واریانس S^2 برابر است با:

$$\frac{Var(\hat{\sigma}^2)}{Var(S^2)}=\frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}}=\frac{(n-1)}{n}\overset{n=10}{=\!=}\frac{9}{10}=0.9$$

$$\hat{\sigma}^2=\frac{\sum(x_i-\mu)^2}{n} \ ; \ Var(\hat{\sigma}^2)=Var\left(\frac{\sum(x_i-\mu)^2}{n}\right)=Var\left(\frac{\sigma^2\sum(x_i-\mu)^2}{\sigma^2 n}\right)=\frac{\sigma^4}{n^2}Var\left(\frac{\sum(x_i-\mu)^2}{\sigma^2}\right)$$

$$=\frac{\sigma^4}{n^2}Var\left(\chi^2_{(n)}\right)=\frac{\sigma^4}{n^2}2n=\frac{2\sigma^4}{n}$$

$$S^2=\frac{\sum(x_i-\bar{x})^2}{n-1} \ ; \ Var(S^2)=Var\left(\frac{\sum(x_i-\bar{x})^2}{n-1}\right)=Var\left(\frac{\sigma^2\sum(x_i-\bar{x})^2}{\sigma^2(n-1)}\right)=\frac{\sigma^4}{(n-1)^2}Var\left(\frac{\sum(x_i-\bar{x})^2}{\sigma^2}\right)$$

$$=\frac{\sigma^4}{(n-1)^2}Var\left(\chi^2_{(n-1)}\right)=\frac{\sigma^4}{(n-1)^2}2(n-1)=\frac{2\sigma^4}{n-1}$$

۲۹. گزینه ۱ درست است.

نکته تستی: می‌دانیم همواره $Var(\bar{X})=\frac{\sigma^2}{n}$ از واریانس هر برآوردکننده دیگر میانگین کمتر است، بنابراین در این سؤال $T_1=\frac{\sum_{i=1}^{3}x_i}{3}=\bar{X}$ از T_2 و T_3 واریانس کمتری دارد و تنها در گزینه (۱) $Var(T_1)$ از بقیه کوچک‌تر است که حتماً جواب مسئله است.

$$Var(T_1)=\frac{1}{9}\left(Var(X_1)+Var(X_2)+Var(X_3)\right)=\frac{1}{9}\left(\sigma^2+\sigma^2+\sigma^2\right)=\frac{\sigma^2}{3}$$

$$Var(T_2)=\frac{1}{16}Var(X_1)+\frac{1}{16}Var(X_2)+\frac{4}{16}Var(X_3)=\frac{6}{16}\sigma^2=\frac{3}{8}\sigma^2$$

$$Var(T_3)=\frac{1}{64}Var(X_1)+\frac{4}{64}Var(X_2)+\frac{25}{64}Var(X_3)=\frac{30}{64}\sigma^2=\frac{15}{32}\sigma^2$$

$$Var(T_1)=\frac{\sigma^2}{3}<Var(T_2)=\frac{3}{8}\sigma^2<Var(T_3)=\frac{15}{32}\sigma^2$$

دقت کنید که به طور پیش‌فرض در این‌گونه مسائل X_iها مستقل و از یک جامعه با میانگین μ و واریانس σ^2 انتخاب شده‌اند.

۳۰ . گزینه ۲ درست است.

یادآوری: در بین دو برآوردکننده ناریب، برآوردکننده‌ای بهتر (کاراتر) است که واریانس کمتری داشته باشد.

در این سؤال نیز T_1 و T_2 هر دو برای θ ناریب‌اند، درنتیجه : $E(T_1)=E(T_2)=\theta$

بنابراین T_2 که واریانس کمتری دارد برآوردکننده بهتری برای θ است.

$$Var(T_2)=2<Var(T_1)=6$$

۳۱ . گزینه ۱ درست است.

راه حل اول (تستی): در بین برآوردکننده‌های میانگین جامعه همواره $\bar{X}=\frac{\sum_{i=1}^{n}x_i}{n}$ از بقیه کاراتر است (کاراترین)؛ در این سؤال نیز T_1 کاراترین است.

$$T_1=\frac{\sum_{i=1}^{3}x_i}{3}=\bar{X}$$

راه حل دوم: با توجه به اینکه هر سه برآوردکننده ناریب‌اند، برآوردکننده‌ای کاراتر است که واریانس کمتری داشته باشد، بنابراین داریم:

$$Var(T_1)=\frac{1}{3^2}\left(Var(X_1)+Var(X_2)+Var(X_3)\right)=\frac{\sigma^2+\sigma^2+\sigma^2}{9}=\frac{\sigma^2}{3}$$

$$Var(T_2)=\frac{1}{8^2}\left(Var(X_1)+3^2Var(X_2)+4^2Var(X_3)\right)=\frac{\sigma^2+9\sigma^2+16\sigma^2}{64}=\frac{26}{64}\sigma^2$$

$$Var(T_3)=\frac{1}{10^2}\left(5^2Var(X_1)+4^2Var(X_2)+Var(X_3)\right)=\frac{25\sigma^2+16\sigma^2+\sigma^2}{100}=\frac{42}{100}\sigma^2$$

$Var(T_1)=\frac{\sigma^2}{3}<Var(T_2)=\frac{26}{64}\sigma^2<Var(T_3)=\frac{42}{100}\sigma^2 \rightarrow$ T_1 از همه کاراتر است.

توجه: X_iها از یک جامعه و مستقل از هم هستند، بنابراین در محاسبه واریانس برآوردکننده‌ها جزء کوواریانس دو به دوی X_iها برابر صفر می‌شود.

میانگین مجذور خطا (MSE)

۳۲ . گزینه ۴ درست است.

می‌دانیم در بین دو برآوردکننده دلخواه (اریب یا ناریب) برآوردکننده‌ای بهتر (کاراتر) است که MSE کمتری داشته باشد. در این سؤال با توجه به اینکه $\hat{\theta}_1$ ناریب و $\hat{\theta}_2$ اریب است، معیار MSE برای انتخاب مناسب بودن آماره به کار می‌رود.

$$E\left(\hat{\theta}_1-\theta\right)=0 \rightarrow E\left(\hat{\theta}_1\right)=\theta \rightarrow \hat{\theta}_1 \text{ ناریب است } \rightarrow \text{اریبی}\left(\hat{\theta}_1\right)=0$$

$$E\left(\hat{\theta}_2-\theta\right)=6 \rightarrow E\left(\hat{\theta}_2\right)=\theta+6 \rightarrow \text{اریبی}\left(\hat{\theta}_2\right)=E\left(\hat{\theta}_2\right)-\theta=\theta+6-\theta=6$$

$$MSE\left(\hat{\theta}_1\right)=Var\left(\hat{\theta}_1\right)+(\text{اریبی})^2=90+0=90$$

$$MSE\left(\hat{\theta}_2\right)=Var\left(\hat{\theta}_2\right)+(\text{اریبی})^2=50+6^2=86$$

$MSE\left(\hat{\theta}_2\right)$ کمتر است درنتیجه $\hat{\theta}_2$ مناسب‌تر است.

$$MSE\left(\hat{\theta}_2\right)=86<MSE\left(\hat{\theta}_1\right)=90$$

توجه: با اینکه $\hat{\theta}_1$ ناریب است اما کارایی‌اش از $\hat{\theta}_2$ کمتر است چون برآوردکننده‌ای کاراتر است که حداقل واریانس و اریبی را با هم داشته باشد، یعنی MSE کمتری داشته باشد.

۳۳ . گزینه ۳ درست است.

یادآوری: در بین دو برآوردکننده دلخواه (اریب یا ناریب) برآوردکننده‌ای بهتر است که ملاک MSE (میانگین مجذور خطا) کمتری داشته باشد. $MSE = \text{واریانس} + (\text{اریبی})^2$

در این سؤال نیز هر دو برآوردکننده A و B اریب هستند، بنابراین ملاک مقایسه آن‌ها MSE است.

$$\begin{cases} E(A)=E\left(\dfrac{X_1+X_2+X_3}{2}\right)=\dfrac{1}{2}\left(E(X_1)+E(X_2)+E(X_3)\right)=\dfrac{1}{2}(\mu+\mu+\mu)=\dfrac{3}{2}\mu\neq\mu & \text{اریب} \\ E(B)=E\left(\dfrac{X_1+X_2+X_3}{3}+2\right)=\dfrac{1}{3}\left(E(X_1)+E(X_2)+E(X_3)\right)+2=\dfrac{3}{3}\mu+2=\mu+2\neq\mu & \text{اریب} \end{cases}$$

۳۴ . گزینه ۳ درست است.

یادآوری: برای تشخیص مناسب بودن برآورد کننده‌های دلخواه (اریب یا ناریب)، بهترین ملاک سنجش، میانگین مجذور خطا (MSE) است:

$$MSE=E\left(\hat{\theta}-\theta\right)^2=Var\left(\hat{\theta}\right)+\left(\underbrace{E\left(\hat{\theta}\right)-\theta}_{\text{اریبی}}\right)^2$$

در این سؤال نیز برآورد کننده $\hat{\theta}$ بدون هیچ اطلاع دیگر داده شده (دلخواه)، بنابراین بهترین معیار برای تشخیص مناسب بودن آن MSE است.

دقت کنید که گزینه ۲ همان واریانس $\hat{\theta}$ است.

برآورد فاصله‌ای

فاصله اطمینان میانگین جامعه

۳۵. گزینه ۲ درست است.

جامعه نرمال، واریانس نامعلوم و $n \le 30$ است بنا بر مورد (۲ ـ ب) از برآورد فاصله‌ای μ داریم:

$$t_{(n-1)} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \quad \rightarrow \quad \mu : \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

مقدار t داده شده است اما $\bar{X}$ و S باید با توجه به داده‌های نمونه برآورد شوند.

نکته: هرگاه تعداد نمونه $n = 3$ باشد و اعداد نمونه متوالی باشند واریانس نمونه یک است و میانگین اعداد برابر با عدد وسط است.

در این سؤال نیز $n = 3$ و اعداد نمونه متوالی‌اند، پس داریم:

$$\bar{X} = \text{وسط داده} = 16 \,,\, S^2 = 1 \rightarrow S = 1$$

اگر از طریق فرمول نیز محاسبه کنیم به همین اعداد می‌رسیم:

$$\begin{cases} \bar{X} = \dfrac{\sum x_i}{n} = \dfrac{15+16+17}{3} = \dfrac{48}{3} = 16 \\ S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1} = \dfrac{(15-16)^2 + (16-16)^2 + (17-16)^2}{3-1} = \dfrac{1+0+1}{2} = 1 \rightarrow s^2 = 1 \end{cases}$$

$$\mu : \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} = 16 \pm 3 \times \frac{1}{\sqrt{3}} = 16 \pm \sqrt{3} \xrightarrow[\text{بگیرید } 1.7]{\text{همیشه } \sqrt{3} \text{ را برابر}} (14.3\,,17.7)$$

۳۶. گزینه ۴ درست است.

حال جامعه نرمال، واریانس جامعه نامعلوم و $n \le 30$ است پس بنا بر مورد (دوم ـ ب) از برآورد فاصله‌ای μ داریم:

با توجه به داده‌ها $\bar{X}$ و S باید از نمونه برآورد شوند.

$$t_{(n-1)} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \rightarrow \mu \in \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

نکته: هرگاه تعداد نمونه $n = 3$ باشد و اعداد نمونه متوالی باشند میانگین اعداد برابر با عدد وسط و واریانس داده‌ها یک است. در این سؤال نیز $n = 3$ و اعداد نمونه متوالی‌اند پس داریم:

$$n: 15,16,17 \rightarrow \bar{x} = 16 \,,\, S^2 = 1 \rightarrow S = 1$$

$$\mu \in \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \rightarrow \mu \in 16 \pm t_{\frac{\alpha}{2}, 2} \frac{1}{\sqrt{3}}$$

۳۷. گزینه ۴ درست است.

با توجه به راهنمایی مسئله $(t = 3)$، درصورتی‌که پیش‌فرض توزیع جامعه را نرمال در نظر بگیریم؛ واریانس جامعه نامعلوم و تعداد نمونه $n = 2 < 30$ است، بنابراین توزیع $\bar{X}$ (میانگین نمونه) $t-$استیودنت خواهد بود.

$$\text{فاصله اطمینان } \mu : \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} = 40 \pm 3 \frac{\sqrt{200}}{\sqrt{2}} = 40 \pm 30 = (10, 70)$$

$x_i = 30 \quad , \quad 50$

$\bar{X} = \frac{\sum x_i}{n} = \frac{30+50}{2} = \frac{80}{2} = 40$

$S_X^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(30-40)^2 + (50-40)^2}{2-1} = \frac{100+100}{1} = 200$

دقت کنید که فاصله اطمینان میانگین جامعه $\mu = \frac{\sum x_i}{N}$ خواسته نشده بلکه فاصله اطمینان کل فروش جامعه خواسته شده یعنی $\left(\sum x_i\right)$، بنابراین کافی است فاصله اطمینان میانگین جامعه را در $N = 500$ ضرب کنیم تا به فاصله اطمینان کل فروش برسیم:

$\sum x_i : N \times (10, 70) = 500 \times (10, 70) = (5000 \quad , \quad 35000)$

۳۸ . گزینه ۳ درست است.

جامعه غیرنرمال، واریانس جامعه نامعلوم و $n < 30$ است. چنین توزیعی وجود ندارد اما اگر:

۱) اگر واریانس جامعه را معلوم فرض کنیم، توزیع $\bar{X}$ نامشخص است و باید از قضیه چی‌بی‌شف استفاده کنیم:

$\mu : \bar{X} \pm \sqrt{\frac{1}{\alpha}} \frac{S}{\sqrt{n}}$

اما در صورت استفاده از قضیه چی‌بی‌شف جواب درست در گزینه‌ها نیست.

۲) اگرجامعه را نرمال فرض کنیم و از توزیع t–استیودنت نیز استفاده کنیم $\left(t_{0.025,8} = 2.3\right)$ باز هم جواب در گزینه‌ها نیست.

۳) اگر به ناچار، جامعه را نرمال و واریانس را هم معلوم فرض کنیم $\left(Z_{0.025} = 1.96\right)$ جواب، گزینه ۳ یعنی کلید داده شده است:

$\mu : \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 160 \pm 1.96 \frac{5}{\sqrt{9}} = 160 \pm 3.26 = (156.74 \, , 163.26)$

۳۹ . گزینه ۱ درست است.

توزیع جامعه نرمال و واریانس جامعه معلوم است پس n هر چه باشد توزیع $\bar{X}$ نرمال است.

$\mu : \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 10 \pm 2 \frac{\sqrt{16}}{\sqrt{25}} = 10 \pm \frac{8}{5}$

$\bar{x} = 10 \quad , \quad \sigma^2 = 16 \quad , \quad n = 25 \quad , \quad 1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \quad , \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = 2$

محاسبه n براساس فاصله μ

۴۰.گزینه ۴ درست است.

$$\begin{cases} e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \rightarrow \frac{1}{4}\sigma = Z_{0.025} \frac{\sigma}{\sqrt{n}} \rightarrow \frac{1}{4}\cancel{\sigma} = 2 \frac{\cancel{\sigma}}{\sqrt{n}} \rightarrow \sqrt{n} = 8 \rightarrow n = 64 \\ e = \frac{1}{4}\sigma \quad , \quad \alpha = 0.05 \rightarrow Z_{0.025} \simeq 2 \end{cases}$$

۴۱.گزینه ۲ درست است.

$$\begin{cases} e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad 4 = 2 \times \frac{20}{\sqrt{n}} \quad \rightarrow \quad \sqrt{n} = 10 \quad \rightarrow \quad n = 100 \\ e = 4 \quad , \quad \sigma = 20 \quad , \quad 1 - \alpha = 0.95 \quad \rightarrow \quad \alpha = 0.05 \quad \rightarrow \quad \frac{\alpha}{2} = 0.025 \quad , \quad Z_{\frac{\alpha}{2}} = Z_{0.025} \simeq 2 \end{cases}$$

توجه: در جدول توزیع نرمال استاندارد $Z_{0.025}=1.96$ است، اما برای راحتی در محاسبه ما همواره آن را 2 در نظر می‌گیریم و در آخر، گزینه‌ای را انتخاب می‌کنیم که از جواب به‌دست‌آمده کمی کمتر باشد (نزدیک‌ترین و کمتر) یعنی در این سؤال، $n=96$ جواب سؤال است.

۴۲. گزینه ۴ درست است.

$$\begin{cases} e=Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}} \rightarrow 1=2\dfrac{\sqrt{25}}{\sqrt{n}} \rightarrow \sqrt{n}=10 \rightarrow n=100 \\ \sigma^2=25 \ , \ e=1 \ , \ 1-\alpha=0.95 \rightarrow \alpha=0.05 \rightarrow \dfrac{\alpha}{2}=0.025 \ , \ Z_{\frac{\alpha}{2}}=Z_{0.025}=2 \end{cases}$$

روابط بین $2e , e , n$

۴۳. گزینه ۲ درست است.

واریانس نصف می‌شود ; $\sigma^2 \rightarrow \dfrac{\sigma^2}{2} \Rightarrow \sigma \rightarrow \dfrac{\sigma}{\sqrt{2}}$

طول فاصله اطمینان : $2e=2Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}} \xrightarrow{\sigma \rightarrow \frac{\sigma}{\sqrt{2}}} 2e=2\,Z_{\frac{\alpha}{2}}\dfrac{\frac{\sigma}{\sqrt{2}}}{\sqrt{n}}=2\,Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}.\dfrac{\sqrt{2}}{2}$

بنابراین، طول فاصله اطمینان $\dfrac{\sqrt{2}}{2}$برابر می‌شود.

فاصله اطمینان تفاضل یا مجموع میانگین دو جامعه

۴۴. گزینه ۱ درست است.

این سؤال عیناً در سال ۸۳ هم آمده است: پیش‌فرض، جوامع را نرمال، واریانس جوامع را نامعلوم و نابرابر و $n_1,n_2\le 30$ می‌دانیم پس بنا بر مورد (۳ ـ ب) از برآورد فاصله‌ای $\mu_1\pm\mu_2$، توزیع $\bar{X}_1\pm\bar{X}_2$، t_r است. در گزینه‌های ۲ و ۴ از t_r استفاده شده است اما در گزینه ۲ به جای فاصله اطمینان برای مجموع میانگین‌ها فاصله اطمینان برای تفاضل میانگین‌ها آمده و همچنین از S_p استفاده شده است؛در گزینه ۴ نیز به جای $\bar{X}_1+\bar{X}_2$ نمونه، $\mu_1+\mu_2$ جامعه گذاشته شده است. تنها گزینه‌ای که فاصله اطمینان برای مجموع میانگین‌ها است گزینه ۱ است. که ناچار به انتخاب آن هستیم.

۴۵. گزینه ۴ درست است.

دو جامعه نرمال، واریانس جوامع نامعلوم و پیش‌فرض نابرابر و $n_1,n_2<30$ است بنابراین توزیع تفاضل میانگین دو جامعه بنا بر حالت (۳ - ب) $t-$استیودنت با درجه آزادی r است.

$$\mu_1-\mu_2:\bar{X}_1-\bar{X}_2\pm t_{r,\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}=(174-170)\pm t_{r,\frac{\alpha}{2}}\sqrt{\frac{\frac{944}{5-1}}{5}+\frac{\frac{866}{5-1}}{5}}=4\pm2.31\sqrt{\frac{866+944}{20}}=4\pm2.31\sqrt{90.5}$$

$$\begin{cases} S_1^2=\dfrac{\sum(x-\bar{x}_1)^2}{n_1-1}=\dfrac{944}{5-1} \\ S_2^2=\dfrac{\sum(x-\bar{x}_2)^2}{n_2-1}=\dfrac{866}{5-1} \end{cases}$$

توجه: می‌دانیم که مقدار $Z_{\frac{\alpha}{2}}=Z_{0.025}=1.96$ است، پس حتماً 2.31 مقدار $t_{r,0.025}$ است.

فاصله اطمینان نسبت جامعه

۴۶. گزینه ۳ درست است.

$$\begin{cases} \bar{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\dfrac{\bar{p}\,\bar{q}}{n}} = 0.5 \pm Z_{0.025}\sqrt{\dfrac{0.5\times 0.5}{100}} = 0.5 \pm 2\times 0.05 = (0.4, 0.6) \xrightarrow{\text{نزدیک‌ترین گزینه}} (0.402, 0.598) \\ \bar{p} = \dfrac{x}{n} = \dfrac{50}{100} = 0.5 \ , \ \alpha = 0.05 \to \dfrac{\alpha}{2} = 0.025 \to Z_{0.025} \simeq 2 \\ n = 100 \ , \ X = 50 \ \text{(تعداد افراد نمونه که دارای صفت مورد نظر هستند.)} \end{cases}$$

توجه: برای راحتی در محاسبه همیشه $Z_{0.025} \simeq 2$ در نظر بگیرید و در آخر نزدیک‌ترین گزینه به اعداد به‌دست‌آمده را به عنوان جواب انتخاب کنید.

۴۷. گزینه ؟ درست است.

$$\begin{cases} p \in \hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}\hat{q}}{n}} = 0.8 \pm 1.65\sqrt{\dfrac{0.8\times 0.2}{100}} = 0.8 \pm 0.066 = (0.734 \ , \ 0.866) \\ \hat{p} = 0.8 \ , \ n = 100 \ , \ \alpha = 0.1 \ , \ Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.65 \end{cases}$$

البته چون جواب در گزینه‌ها نبود، این سؤال پس از برگزاری کنکور حذف شد.

۴۸. گزینه ۲ درست است.

X : تعداد کالاهای سالم در نمونه

$$\begin{cases} p : \bar{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\dfrac{\bar{p}\,\bar{q}}{n}} = 0.89 \pm 1.64\sqrt{\dfrac{0.89\times 0.11}{100}} \\ \bar{p} = \dfrac{x}{n} = \dfrac{89}{100} = 0.89 \ , \ \bar{q} = 0.11 \\ n = 100 \ , \ \alpha = 0.1 \ \to \ Z_{\frac{\alpha}{2}} = Z_{0.05} = Z_{0.95} = 1.64 \end{cases}$$

۴۹. گزینه ۱ درست است.

X : تعداد کالای دارای صفت مورد نظر

$$\begin{cases} p : \bar{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\dfrac{\bar{p}\,\bar{q}}{n}} = \dfrac{1}{5} \pm 2\sqrt{\dfrac{\frac{1}{5}\times\frac{4}{5}}{125}} = \dfrac{1}{5} \pm 2\dfrac{\frac{2}{5}}{5\sqrt{5}} = \dfrac{1}{5} \pm \dfrac{4}{25\sqrt{5}} = \dfrac{1}{5} \pm \dfrac{20}{125\sqrt{5}} \\ \bar{p} = \dfrac{x}{n} = \dfrac{25}{125} = \dfrac{1}{5} \ , \ \bar{q} = \dfrac{4}{5} \\ n = 125 \ , \ 1-\alpha = 0.95 \ \to \ \alpha = 0.05 \ , \ Z_{\frac{\alpha}{2}} = Z_{0.025} = 2 \end{cases}$$

۵۰. گزینه ۱ درست است.

X: تعداد کالای معیوب در نمونه

$$\begin{cases} \theta = p : \ \bar{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\dfrac{\bar{p}\,\bar{q}}{n}} = \dfrac{1}{6} \pm 2\sqrt{\dfrac{\frac{1}{6}\times\frac{5}{6}}{36}} = \dfrac{1}{6} \pm 2\dfrac{\sqrt{5}}{36} = \dfrac{1}{6} \pm \dfrac{\sqrt{5}}{18} \\ \bar{p} = \dfrac{x}{n} = \dfrac{6}{36} = \dfrac{1}{6} \rightarrow \ \bar{q} = \dfrac{5}{6} \\ n = 36 \ , \ 1-\alpha = 0.95 \rightarrow \ \alpha = 0.05 \ , \ Z_{\frac{\alpha}{2}} = Z_{0.025} = Z_{0.975} = 2 \end{cases}$$

محاسبه n براساس فاصله p

۵۱. گزینه ۲ درست است.

چون هیچ اطلاعی از نسبت افراد باسواد در جامعه نداریم بنابراین به طور پیش‌فرض $p = q = \frac{1}{2}$ در نظر می‌گیریم:

$$\begin{cases} e = Z_{\frac{\alpha}{2}} \sqrt{\dfrac{pq}{n}} \rightarrow 0.02 = 2\sqrt{\dfrac{\frac{1}{2}\times\frac{1}{2}}{n}} \ \rightarrow \ \sqrt{n} = 50 \rightarrow n = 2500 \\ p = q = \dfrac{1}{2} \ (\text{پیش‌فرض}) \ , \ e = 0.02 \ , \ \alpha = 0.05 \ , Z_{\frac{\alpha}{2}} = Z_{0.025} \simeq 2 \end{cases}$$

خودآزمایی

نمونه‌گیری

۱. در کدام روش نمونه‌گیری، نمونه‌ها تجانس بیشتری با جامعه دارند و از گرایش آن‌ها به طرفی خاص، جلوگیری می‌شود؟

۱) اتفاقی ۲) با هدف و منظور خاص ۳) تصادفی ۴) تصادفی طبقه‌ای

۲. در یک جامعه آماری که از نظر صفت مورد نظر ناهمگون است، کدام‌یک از این روش‌ها برای نمونه‌گیری مناسب است؟

۱) تصادفی ساده ۲) منظم ۳) گروهی ۴) هر سه

۳. کدام‌یک از روش‌های نمونه‌گیر زیر مبنای نمونه‌گیری مرحله‌ای است؟

۱) گروهی ۲) منظم ۳) خوشه‌ای ۴) تصادفی ساده

توزیع میانگین نمونه

۴. اگر حجم نمونه 100 و میانگین نمونه 12 و برآورد واریانس داده‌های نمونه 256 باشد، خطای معیار میانگین نمونه کدام است؟

۱) 0.16 ۲) 1.6 ۳) 2.56 ۴) 25.6

۵. تأثیر افزایش اندازه‌های «نمونه» و «انحراف معیار» بر خطای برآورد میانگین، به ترتیب به چه صورت است؟

۱) کاهش ـ افزایش ۲) افزایش ـ افزایش ۳) افزایش ـ کاهش ۴) کاهش ـ کاهش

۶. افزایش حجم نمونه با چه ضریبی، موجب کاهش خطای معیار $\bar{X}$ به $\frac{1}{3}$ مقدار اولیه‌اش می‌شود؟

۱) 3 ۲) 9 ۳) 6 ۴) 12

۷. از جامعه‌ای 6 عضوی با واریانس 10، 2 نمونه تصادفی استخراج شده است. واریانس $\bar{X}$ (میانگین نمونه)، چقدراست؟

۱) 5 ۲) 4 ۳) 3 ۴) 2

۸. میانگین توزیع یک جامعه 90 و انحراف معیار آن 4 محاسبه شده است. برای نمونه‌ای تصادفی به حجم 100، احتمال اینکه میانگین نمونه بین 89 و 91 باشد، در چه بازه‌ای از Z (احتمال‌های نرمال استاندارد) قرار می‌گیرد؟

۱) $-0.4 , 0.4$ ۲) $-2.25 , 2.25$ ۳) $-3.6 , 3.6$ ۴) $-2.5 , 2.5$

۹. کدام‌یک از این رابطه‌ها درست است؟

۱) $E(\bar{X}) = \mu_X$ ۲) $E(\bar{X}) = \mu_{\bar{X}}$ ۳) $\mu_X = \bar{X}$ ۴) ۱ و ۲

۱۰. اگر $\bar{X}$ دارای توزیع نرمال باشد، چند درصد $\bar{X}$ها در فاصله $\bar{X} \pm 2\sigma_{\bar{X}}$، قرار می‌گیرند؟

۱) 68.3 درصد ۲) 99.73 درصد ۳) 95.45 درصد ۴) 90 درصد

۱۱. افزایش حجم نمونه، موجب کاهش کدام مورد می‌شود؟

۱) مخارج عمل نمونه‌گیری ۲) صرف وقت برای عمل نمونه‌گیری
۳) مخارج عمل پردازش‌ها ۴) خطای معیار شاخص آماری

توزیع تفاضل میانگین دو نمونه

۱۲. فرض کنید $\bar{X}_1$ میانگین نمونه‌ای به حجم n_1 و $\bar{X}_2$ میانگین نمونه‌ای به حجم n_2 از دو جامعه مستقل باشند. $\sigma^2_{\bar{X}_1-\bar{X}_2}$ برابر است با:

۱) $\sigma_1^2 + \sigma_2^2$ ۲) $\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}$ ۳) $\frac{\sigma_1^2}{\sqrt{n_1}} + \frac{\sigma_2^2}{\sqrt{n_2}}$ ۴) $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

۱۳. فرض کنید $\bar{X}_1$ میانگین نمونه‌ای به حجم n_1 و $\bar{X}_2$ میانگین نمونه‌ای به حجم n_2 از دو جامعه مستقل با واریانس برابر باشند، مقدار $Var(4\bar{X}_1 - 3\bar{X}_2)$ عبارت است از:

۱) $\sigma^2\left(\frac{4}{n_1} - \frac{3}{n_2}\right)$ ۲) $\sigma^2\left(\frac{16}{n_1} - \frac{9}{n_2}\right)$ ۳) $\sigma^2\left(\frac{4}{n_1} + \frac{3}{n_2}\right)$ ۴) $\sigma^2\left(\frac{16}{n_1} + \frac{9}{n_2}\right)$

توزیع نسبت نمونه

۱۴. از بین نمونه‌ای تصادفی شامل 400 نفر که در فهرست بیش از 8000 فارغ‌التحصیل دوره کاردانی حسابداری بوده‌اند، تعداد 256 نفر تصمیم به ادامه تحصیل دارند؛ خطای معیار برآورد شده برای نسبت آن عده که ادامه تحصیل می‌دهند، چقدر است؟

۱) 0.128 ۲) 0.64 ۳) 0.024 ۴) 0.28

۱۵. رفتار توزیع نمونه‌گیری $\bar{p}$ در یک نمونه تصادفی 20 تایی چه توزیعی دارد؟

۱) یکنواخت ۲) نمایی ۳) نرمال ۴) دوجمله‌ای

توزیع واریانس نمونه

۱۶. اگر متغیر X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد و $X \sim N(\mu,\sigma^2)$، در این صورت $\frac{\sum(x_i-\bar{x})^2}{\sigma^2}$ دارای توزیع:

۱) t است. ۲) χ^2 است. ۳) دوجمله‌ای است. ۴) نرمال است.

۱۷. اگر $S^2 = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}$ تعریف شود، میانگین و واریانس متغیر تصادفی $\frac{(n-1)S^2}{\sigma^2}$ کدام است؟

۱) $(n-1), (n-1)$ ۲) $(n-1), 2(n-1)$ ۳) $2(n-1), (n-1)$ ۴) $\sigma^2, \frac{2\sigma^4}{(n-1)}$

خواص برآوردکننده‌های نقطه‌ای

ناریبی

۱۸. نمونه تصادفی X_1, X_2, X_3 از توزیع دوجمله‌ای با پارامترهای θ و 10 انتخاب می‌شود. مقدار a برای اینکه آماره $aX_1+2aX_2+3aX_3$ یک برآوردگر ناریب برای θ باشد، کدام است؟

۱) $\frac{1}{60}$ ۲) $\frac{1}{6}$ ۳) $\frac{2}{9}$ ۴) $\frac{1}{3}$

کارایی

۱۹. سه برآوردکننده زیر مفروض است؛ X_iها، مستقل از یکدیگرند. چه رابطه‌ای از نظر کارایی بین آنان برقرار است؟

$$T_1=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3 \ , \ T_2=\frac{1}{8}X_1+\frac{5}{8}X_2+\frac{2}{8}X_3 \ , \ T_3=\frac{1}{5}X_1+\frac{1}{5}X_2+\frac{3}{5}X_3$$

۱) $Var(T_2)>Var(T_3)>Var(T_1)$ ۲) $Var(T_1)>Var(T_3)>Var(T_2)$

۳) $Var(T_2)>Var(T_1)>Var(T_3)$ ۴) $Var(T_1)<Var(T_2)<Var(T_3)$

سازگاری

۲۰. در یک برآوردکننده، n را به سمت بی‌نهایت بردیم. اریبی و واریانس به سمت صفر میل کرد برآوردکننده مورد نظر کدام است؟

۱) سازگار ۲) کارا ۳) ناسازگار ۴)ناریب

میانگین مجذور خطا (MSE)

۲۱. برای برآورد میانگین جامعه بر اساس یک نمونه تصادفی دو تایی، دو برآوردکننده A و B زیر پیشنهاد شده است. برای تشخیص آنکه کدام‌یک برآوردکننده بهتری است، چه ملاکی کافی است؟

$$A=X_1+X_2 \qquad B=\frac{X_1+X_2}{2}-2$$

۱) $\left(E(\hat{\theta})-\theta\right)^2$ ۲) $E\left(\hat{\theta}-E(\hat{\theta})\right)^2+\left(E(\hat{\theta})-\theta\right)^2$

۳) $E(\hat{\theta}-\theta)$ ۴) $E(\hat{\theta})+Var(\hat{\theta})$

برآوردفاصله‌ای

فاصله اطمینان میانگین جامعه

۲۲. از بین شیشه‌هایی که با دستگاه پر می‌شوند نمونه تصادفی 64 تایی انتخاب می‌کنیم. میانگین آن‌ها 250 و انحراف معیار 12 است. دلیلی بر نرمال بودن توزیع مایع ریخته‌شده نداریم. با استفاده از قانون چی‌بی‌شف در سطح اطمینان 96 درصد میانگین کل مایع ریخته‌شده در شیشه‌ها در کدام فاصله است؟

۱) $(242\,,\,258)$ ۲) $(242.5\,,\,257.5)$ ۳) $(238,262)$ ۴) $(245.5,254.5)$

۲۳. اگر در یک نمونه تصادفی 9 عضوی از یک جامعه آماری نسبتاً بزرگ میانگین نمونه مساوی 20 و انحراف معیار نمونه برابر 6 باشد، میانگین این جامعه در چه فاصله‌ای قرار دارد؟ (t برای هشت درجه و نه درجه آزادی به ترتیب 3.355 و 3.250 است.)

۱) $23.250\,,\,16.750$ ۲) $26.71\,,\,13.29$ ۳) $24.5\,,\,15.5$ ۴) $23.355\,,\,16.645$

۲۴. تقاضای کالای X با توزیع نرمال برای 30 روز دارای میانگین $\bar{x}=30$ و انحراف معیار $S=5$ است. فاصله اطمینان 99% برای میانگین جامعه μ برابر کدام است؟ $(Z_{0.005}=2.58)$

۱) $30\pm2.58\dfrac{5}{\sqrt{30}}$ ۲) $30\pm2.58\dfrac{\sqrt{5}}{\sqrt{30}}$ ۳) $30\pm2.58\times5$ ۴) $30\pm2.58\dfrac{5}{30}$

۲۵. اگر میانگین و انحراف معیار یک نمونه چندعضوی به ترتیب 20 و 5 و مقدار Z مساوی 2 باشد، برآورد فاصله‌ای میانگین جامعه کدام است؟

۱) $24\,,\,16$ ۲) $22\,,\,18$ ۳) $21\,,\,19$ ۴) $10.5\,,\,19.5$

۲۶. از یک توزیع پواسون یک نمونه تصادفی 36 تایی انتخاب و مشخص شده است. میانگین نمونه‌ای برابر 3.5 است. یک فاصله اطمینان 95% برای میانگین جامعه عبارت است از:

۱) $\lambda:3.5\pm1.96\sqrt{3.5}$ ۲) $\lambda:3.5\pm1.96\dfrac{\sqrt{3.5}}{6}$ ۳) $\lambda:\dfrac{3.5}{36}\pm1.96\dfrac{\sqrt{3.5}}{6}$ ۴) $\lambda:3.5\pm1.96\dfrac{\sqrt{3.5}}{36}$

محاسبه n براساس فاصله μ

۲۷. برآورد میانگین نمرات آزمونی با ضریب اطمینان 90% درصورتی‌که انحراف معیار دارای مقدار تقریباً ثابت 100 باشد و مقدار خطای برآورد از 4 بیشتر نباشد، نیاز به چه تعداد نمونه دارد؟ (میزان Z در سطح اطمینان 90% برابر 1.64 است.)

۱) 1600 ۲) 1681 ۳) 3600 ۴) 3616

۲۸. اگر حجم نمونه کوچک، جامعه غیرنرمال و ε خطای برآورد باشد، حجم نمونه چگونه محاسبه می‌شود؟

۱) $\dfrac{\sigma_X^2}{\varepsilon^2}$ ۲) $\dfrac{\sigma_X^2}{\alpha^2\,\varepsilon^2}$ ۳) $\dfrac{\sigma_X^2}{\alpha\,\varepsilon^2}$ ۴) $\dfrac{\sigma_X^2}{\alpha^2\,\varepsilon}$

روابط بین 2e , e , n

۲۹. یک فاصله اطمینان 95% برای میانگین یک جامعه نرمال که واریانس آن مقدار مشخص σ^2 است، به صورت $\left(\bar{X}-1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right)$ است. اگر بخواهیم طول این فاصله اطمینان را به نصف کاهش دهیم لازم است که:

۱) تعداد نمونه n را دو برابر کنیم.
۲) تعداد نمونه n را 4 برابر کنیم.
۳) σ را نصف کنیم.
۴) σ را دو برابر کنیم.

فاصله اطمینان تفاضل یا مجموع میانگین دو جامعه

۳۰. فرض کنید $X_i \sim N(\mu_1,\sigma_1^2)\ ,\ i=1,2,\ldots,n_1$ و $X_j \sim N(\mu_2,\sigma_2^2)\ ,\ j=1,2,\ldots,n_2$ دو جامعه مستقل از یکدیگرند. یک فاصله اطمینان $1-\alpha$ برای $\mu_1-\mu_2$، درصورتی‌که واریانس‌های جوامع نامعلوم و مساوی باشند، کدام است؟

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\ ,\ S_2^2=\frac{1}{n_2-1}\sum_{j=1}^{n_2}(y_i-\bar{y})^2\ ,\ S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2$$

۱) $\bar{X}_1-\bar{X}_2\pm Z_{1-\frac{\alpha}{2}}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

۲) $\bar{X}_1-\bar{X}_2\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

۳) $\bar{X}_1+\bar{X}_2\pm t_{1-\frac{\alpha}{2},(n_1+n_2-2)}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

۴) $\bar{X}_1-\bar{X}_2\pm t_{1-\frac{\alpha}{2},(n_1+n_2-2)}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

فاصله اطمینان نسبت جامعه

۳۱. تولیدکننده کالایی ادعا می‌کند که بیست درصد مشتریان فروشگاهی، کالای خاص وی را انتخاب می‌کنند. برای تحقیق در درستی ادعای وی به تصادف 225 نفر از مشتریان انتخاب شده‌اند. فاصله اطمینان برای این تجربه کدام است؟ $(\mu\pm 2\sigma)$

۱) $(42\ ,\ 48)$
۲) $(33\ ,\ 57)$
۳) $(36\ ,\ 52)$
۴) $(34\ ,\ 58)$

۳۲. در یک نمونه 192 نفری از دانشجویان 48 نفر شاغل بوده‌اند. فاصله اطمینان نسبت دانشجویان شاغل برای جامعه کدام است؟ $(t=3.14\ ,\ Z=2.56)$

۱) $(0.12\ ,\ 0.28)$
۲) $(0.1\ ,\ 0.4)$
۳) $(0.518\ ,\ 0.348)$
۴) $(0.17\ ,\ 0.33)$

محاسبه n براساس فاصله p

۳۳. هدف بررسی نسبت بیکاری در یک شهر است. چه حجمی از نمونه لازم است تا 95% مطمئن باشیم حداکثر خطای برآورد بیشتر از 5% نخواهد شد؟ $(Z_{0.975}=1.96)$

۱) 380
۲) 385
۳) 290
۴) 400

فاصله اطمینان واریانس جامعه

۳۴ . فرض کنید $0,1,2,3,4,5,6,7$ **یافته‌های یک نمونه تصادفی از توزیع** $N\left(0,\sigma^2\right)$ **باشد. یک برآورد بازه‌ای** 95% **با دم‌های برابر برای** σ^2 **کدام است؟**

۱) $\left(\frac{140}{16.012} \quad , \quad \frac{140}{1.6898}\right)$

۲) $\left(\frac{140}{17.534} \quad , \quad \frac{140}{2.1797}\right)$

۳) $\left(\frac{448}{17.834} \quad , \quad \frac{448}{2.1797}\right)$

۴) $\left(\frac{448}{16.012} \quad , \quad \frac{448}{1.6898}\right)$

فاصله اطمینان نسبت واریانس دو جامعه

۳۵ . برای ساختن فاصله اطمینان برای نسبت واریانس دو جامعه از ملاک .... استفاده می‌شود.

۱) t

۲) χ^2

۳) F

۴) Z

پاسخنامه

۱	۱	۲	۳	●
۲	۱	۲	●	۴
۳	۱	۲	●	۴
۴	۱	●	۳	۴
۵	●	۲	۳	۴
۶	۱	●	۳	۴
۷	۱	●	۳	۴
۸	۱	۲	۳	●
۹	۱	۲	۳	●
۱۰	۱	۲	●	۴
۱۱	۱	۲	۳	●
۱۲	۱	۲	۳	●
۱۳	۱	۲	۳	●
۱۴	۱	۲	●	۴
۱۵	۱	۲	۳	●
۱۶	۱	●	۳	۴
۱۷	۱	۲	●	۴
۱۸	●	۲	۳	۴
۱۹	●	۲	۳	۴
۲۰	●	۲	۳	۴
۲۱	۱	●	۳	۴
۲۲	۱	●	۳	۴
۲۳	۱	●	۳	۴
۲۴	●	۲	۳	۴
۲۵	۱	۲	●	۴
۲۶	۱	●	۳	۴
۲۷	۱	●	۳	۴
۲۸	۱	۲	●	۴
۲۹	۱	●	۳	۴
۳۰	۱	۲	۳	●
۳۱	۱	●	۳	۴
۳۲	۱	۲	۳	●
۳۳	۱	●	۳	۴
۳۴	۱	●	۳	۴
۳۵	۱	۲	●	۴
۳۶	۱	۲	۳	۴
۳۷	۱	۲	۳	۴
۳۸	۱	۲	۳	۴
۳۹	۱	۲	۳	۴
۴۰	۱	۲	۳	۴

برای دریافت پاسخ تشریحی سؤالات به سایت www.Tourani.ir مراجعه کنید.

فصل ۶ آزمون فرض‌های آماری

مقدمه

مقدار واقعی پارامترهای جامعه تنها از طریق سرشماری به دست می‌آید که در صورت سرشماری، دیگر نیازی به آزمودن آن‌ها نیست، اما اگر سرشماری انجام نشود پارامترهای جامعه با آنکه مقادیر ثابتی دارند، مجهول خواهند بود. در این شرایط، همواره حدس‌ها یا ادعاهایی درباره آن‌ها به وجود می‌آید که باید از طریق نمونه‌گیری مورد آزمون قرار گیرد؛ برای مثال «آیا میانگین حقوق کارمندان 125000 تومان است؟»، «ادعا شده است درصد بیکاری در کشور کمتر از 20% است» یا «ریسک شرکت‌های پذیرفته‌شده در بورس حداقل 1000 است».

برای بررسی صحت یا عدم صحت فرضیه‌های مطرح‌شده بر اساس نتایج حاصل از نمونه، از آزمون فرض آماری استفاده می‌شود؛ بنابراین به طور کلی هدف از آزمون فرض آماری، تعیین این موضوع است که آیا با توجه به اطلاعات حاصل از داده‌های نمونه، حدسی که درباره خصوصیتی از یک جامعه می‌زنیم، به طور قوی تأیید می‌شود یا خیر؛ این حدس بنا بر هدف تحقیق، شامل ادعایی درباره مقدار یک پارامتر جامعه است.

مثال ۱ در کدام‌یک از موارد زیر نمی‌توان آزمون فرض آماری را برگزار کرد؟ (مدیریت ـ ۸۳)

۱) از سرشماری استفاده شده باشد.
۲) حجم نمونه کمتر از 30 باشد.
۳) نمونه به روش تصادفی منظم انتخاب شده باشد.
۴) هیچ‌کدام

حل: گزینه ۱ درست است.

مثال ۲ در کدام مورد نباید از آزمون فرض استفاده کرد؟ (حسابداری ـ ۸۳)

۱) $n = 10$
۲) $n < 30$
۳) $n \geq 30$
۴) سرشماری

حل: گزینه ۴ درست است.

✓ **دقت کنید!**

نتایج حاصل از نمونه به هر حجمی $(n \geq 1)$ که برای بررسی صحت یا عدم صحت فرضیه مورد استفاده قرار می‌گیرند، باید آزمون شود.

فرض آماری (Statistical Hypothesis)

هر ادعا یا حدسی درباره یکی از پارامترهای جامعه، فرض آماری نامیده می‌شود که قابل قبول بودن آن باید بر مبنای اطلاعات حاصل از نمونه‌گیری بررسی شود. ازآنجاکه ادعا ممکن است درست یا نادرست باشد، دو فرض مکمل در ذهن به وجود می‌آید: «ادعا درست است»، «ادعا نادرست است». در ادامه به تحلیل و بررسی دو فرض مکمل می‌پردازیم.

فرض صفر و فرض مقابل (Null Hypothesis & Alternative Hypothesis)

فرض صفر (H_0): فرضی که باید آن را اثبات کرد و همواره شامل یکی از علایم «$=$»، «$\geq$» یا «$\leq$» است، فرض صفر (H_0) نامیده می‌شود.

فرض مقابل (H_1): فرض مخالف H_0 که در صورت عدم اثبات H_0 پذیرفته می‌شود و همواره شامل یکی از علایم «$\neq$»، «$<$» یا «$>$» است، فرض مقابل (H_1) نامیده می‌شود.

قوانین فرض‌های H_0 و H_1

۱ـ همواره به دنبال اثبات فرض H_0 هستیم و نه ادعا.

۲- فرض بر درستی H_0 است مگر آنکه خلاف آن با توجه به نتایج حاصل از داده‌های نمونه ثابت شود، مانند شخص متهمی که در دادگاه فرض بر بی‌گناهی اوست مگر آنکه خلاف آن ثابت شود.

۳ـ ادعا یا حدس مطرح‌شده درباره جامعه ممکن است در H_0 یا H_1 باشد؛ اگر شامل«$=$، $\leq$ یا $\geq$» باشد، در H_0 و اگر شامل« $\neq$، $<$ یا $>$» باشد، در H_1 است.

مثال ۳ ادعا شده است ریسک شرکت‌های پذیرفته شده در بورس کمتر از 1000 است. فرضیه‌های آزمون کدام‌اند؟

۱) $\begin{cases} H_0: \sigma > 1000 \\ H_1: \sigma \leq 1000 \end{cases}$
۲) $\begin{cases} H_0: \sigma \geq 1000 \\ H_1: \sigma < 1000 \end{cases}$
۳) $\begin{cases} H_0: \sigma \leq 1000 \\ H_1: \sigma > 1000 \end{cases}$
۴) $\begin{cases} H_0: \sigma < 1000 \\ H_1: \sigma \geq 1000 \end{cases}$

حل: گزینه ۲ درست است.

همان‌طورکه می‌دانیم منظور از ریسک همان انحراف معیار (σ) است. حال اگر بخواهیم فرضیه‌های آزمون را با توجه به ادعا در نظر بگیریم، خواهیم داشت:

$$\begin{cases} H_0: \sigma \geq 1000 \\ H_1: \sigma < 1000 \text{ (ادعا)} \end{cases}$$

ازآنجاکه ادعای مطرح‌شده شامل علامت «$=$» نبود در H_1 قرار گرفت؛ با این حال برای آزمون فرضیه به دنبال اثبات H_0 هستیم. درصورتی‌که H_0 پذیرفته شود ادعای مطرح‌شده رد می‌شود، در غیر این صورت ادعا پذیرفته می‌شود.

مثال ۴ ادعا شده است درصد بیکاری در کشور حداقل 20% است. فرضیه‌های آزمون کدام‌اند؟

۱) $\begin{cases} H_0: p \geq 0.2 \\ H_1: p < 0.2 \end{cases}$
۲) $\begin{cases} H_0: p \leq 0.2 \\ H_1: p > 0.2 \end{cases}$
۳) $\begin{cases} H_0: p > 0.2 \\ H_1: p \leq 0.2 \end{cases}$
۴) $\begin{cases} H_0: p < 0.2 \\ H_1: p \geq 0.2 \end{cases}$

حل: گزینه ۱ درست است.

با توجه به ادعا، فرضیه‌های آزمون به صورت زیر است:

$$\begin{cases} H_0: p \geq 0.2 \quad \textbf{(ادعا)} \\ H_1: p < 0.2 \end{cases}$$

ازآنجاکه ادعای مطرح‌شده شامل علامت «$=$» بود در H_0 قرار گرفت؛ با این حال به دنبال اثبات H_0 هستیم. درصورتی‌که H_0 پذیرفته شود ادعا نیز پذیرفته می‌شود، در غیر این صورت ادعا رد می‌شود.

مثال ۵ کدام عبارت در مورد H_0 و H_1 صحیح نیست؟ (حسابداری ـ ۸۲)

۱) H_0 و H_1 نقیض یکدیگرند.
۲) همواره ادعا در H_1 قرار می‌گیرد.
۳) در H_0 همواره باید $=$ یا $\leq$ یا $\geq$ قرار گیرد.
۴) فرض بر درستی H_0 است مگر خلاف آن ثابت شود.

حل: گزینه ۲ درست است.

سطح معنی‌دار (Significance Level)

پس از تعریف فرضیه‌های آماری، گام بعدی مشخص کردن درجه‌ای برای معنی‌دار بودن تفاوت‌ها یعنی اختلاف بین مقدار پیشنهادی H_0 و مقدار واقعی پارامتر جامعه است؛ حجم نمونه مورد تحقیق (n) نیز باید در این مرحله تعیین شود.

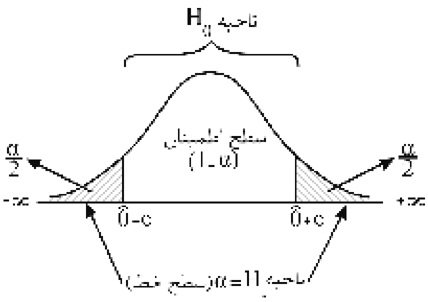
همواره می‌توان فرض H_0 را به نفع H_1 رد کرد، به شرط آنکه از انجام یک آزمون آماری با توجه به نتایج حاصل از نمونه مقداری به دست آوریم که احتمال وقوع آن مقدار با توجه به H_0، کوچک‌تر یا مساوی یک احتمال بسیار کوچک باشد. این احتمال کوچک را «سطح معنی‌دار یا سطح خطا یا خطای نوع اول» می‌گویند و آن را با « α » نمایش می‌دهند. مقادیر مرسوم برای α، 0.01 و 0.05 است.

ازآنجاکه مقدار α در تعیین اینکه H_0 باید رد شود یا خیر دخالت مستقیم دارد، باید آن را پیش از شروع جمع‌آوری اطلاعات نمونه مشخص کنیم. برای مثال، اگر فرض آماری درباره تأثیر عمل جراحی روی قلب باشد، محقق باید مقدار α را بسیار کم در نظر بگیرد، زیرا خطرهای رد کردن نادرست فرض صفر (H_0) و قبول فرض مقابل (H_1) ممکن است به یک توصیه غلط منجر شود.

نتیجه: وقتی یک آزمون فرضیه در سطح معنی‌دار α بررسی شود، به آن مفهوم است که:

«حداکثر α درصد احتمال وجود دارد که H_0 رد شود.»

«حداقل $1-\alpha$ درصد احتمال وجود دارد که H_0 پذیرفته شود.»



هنگام اتخاذ تصمیم درباره H_0 ممکن است دو نوع خطا پیش آید؛ یکی «خطای نوع اول» که در آن H_0 رد می‌شود درحالی‌که فرض درست است و دیگری «خطای نوع دوم» که در آن H_0 پذیرفته می‌شود در حالی‌که فرض غلط است.

برای درک بهتر خطای نوع اول و دوم به مثال شخص متهم برمی‌گردیم. می‌دانیم که فرض بر بی‌گناهی متهم است مگر آنکه خلاف آن ثابت شود، پس اگر بی‌گناه بودن را فرضیه صفر تلقی کنیم، این فرضیه باید در دادگاه رد یا قبول شود. اعضای هیئت منصفه به اظهارات در دادگاه گوش می‌دهند و درنهایت نظر خود را صادر می‌کنند.

حال فرض کنید که شخص متهم واقعاً بی‌گناه است ولی هیئت منصفه وی را گناهکار تشخیص دهد. در این صورت هیئت منصفه یک فرضیه صفر درست را رد کرده است، یعنی خطای نوع اول را مرتکب شده است. از طرف دیگر اگر متهم گناهکار باشد و هیئت منصفه وی را بی‌گناه تشخیص دهد، یعنی فرضیه صفر را که درست نیست قبول کرده و مرتکب خطای نوع دوم شده است.

خطاهای آماری (Statistical Errors)

هر یک از خطاهای آماری نوع اول و دوم درواقع یک احتمال هستند که در ادامه به بررسی آن‌ها می‌پردازیم.

خطای نوع اول

احتمال رد کردن H_0 وقتی H_0 به‌واقع درست است. به عبارتی:

$$\boxed{\alpha = P(\text{خطای نوع اول}) = P(H_0 \text{ رد} \mid \text{درست } H_0)}$$

ازآنجاکه فرضیه‌های H_0 و H_1 خلاف یکدیگرند، سه تعریف دیگر نیز برای خطای نوع اول (α) می‌تواند درست باشد:

$$\alpha = P(H_0 \text{ رد} \mid \text{نادرست } H_1) = P(H_1 \text{ پذیرش} \mid \text{نادرست } H_1) = P(H_1 \text{ پذیرش} \mid \text{درست } H_0)$$

✓ **دقت کنید!**

همواره بحث بر روی رد کردن یا رد نکردن فرض H_0 است، بنابراین تعریفی که بر مبنای H_0 باشد ارجح است.

نکته:

۱- خطای نوع اول (α) همان سطح معنی‌دار است.

۲- همان‌طورکه در فصل قبل هم مطرح شد، α (سطح خطا، سطح معنی‌دار) با e (خطا یا دقت برآورد) رابطه معکوس دارد.

یادآوری:

$$e = t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \quad \text{یا} \quad e = Z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} \quad \text{یا} \quad e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \longrightarrow \begin{cases} \alpha\uparrow \xrightleftharpoons{\text{رابطه معکوس}} e\downarrow \\ \alpha\downarrow \xrightleftharpoons{\text{رابطه معکوس}} e\uparrow \end{cases}$$

مثال ۱ در آزمون معنی‌دار بودن، احتمال خطای نوع اول عبارت است از: (مدیریت ـ ۷۱)

۱) رد فرضیه H_0، به شرط درست بودن فرضیه H_1

۲) پذیرش فرضیه H_0، به شرط نادرست بودن فرضیه H_1

۳) رد فرضیه H_0، به شرط نادرست بودن فرضیه H_1

۴) پذیرش فرضیه H_0، به شرط درست بودن فرضیه H_1

حل: گزینه ۳ درست است.

مثال ۲ رابطه بین خطای نوع اول (α) و دقت برآورد در ساختن یک فاصله اطمینان چگونه است؟ (حسابداری ـ ۸۰)

۱) خطی ۲) معکوس ۳) مستقیم ۴) خطی مستقیم

حل: گزینه ۲ درست است.

منظور از دقت برآورد همان خطا (e) است.

خطای نوع دوم

احتمال قبول کردن (رد نکردن) فرض H_0 است وقتی H_0 به‌واقع غلط باشد؛ به عبارت دیگر:

$$\boxed{\beta = P(\text{خطای نوع دوم}) = P(H_0 \text{ پذیرش} \mid H_0 \text{ غلط})}$$

ازآنجاکه فرضیه‌های H_0 و H_1 خلاف یکدیگرند، سه تعریف دیگر نیز برای خطای نوع دوم (β) می‌تواند درست باشد:

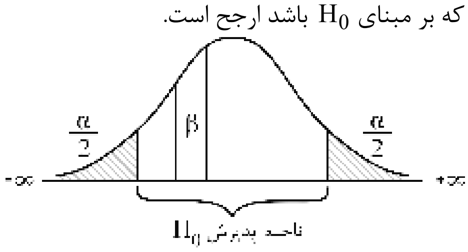
$$\beta = P(H_0 \text{ پذیرش} \mid H_1 \text{درست}) = P(H_1 \text{ رد} \mid H_1 \text{درست}) = P(H_1 \text{ رد} \mid H_0 \text{ نادرست})$$

✓ **دقت کنید!**

همواره بحث بر روی رد کردن یا رد نکردن فرض H_0 است، بنابراین تعریفی که بر مبنای H_0 باشد ارجح است.

مفهوم β

احتمال خطای نوع دوم (β) همیشه قسمتی از سطح اطمینان (ناحیه پذیرش H_0) است، زمانی که H_0 به‌واقع غلط است.



مثال ۳ در آزمون معنی‌دار بودن، احتمال خطای نوع دوم عبارت است از احتمال: (مدیریت ـ ۷۰)

۱) رد فرضیه H_0، به شرط درست بودن فرضیه H_1

۲) پذیرش فرضیه H_0، به شرط نادرست بودن فرضیه H_1

۳) رد فرضیه H_0، به شرط نادرست بودن فرضیه H_1

۴) پذیرش فرضیه H_0، به شرط درست بودن فرضیه H_1

حل: گزینه ۴ درست است.

توان آزمون (Power of a Test)

در هر استنباط آماری احتمال وقوع یکی از دو خطای نوع اول و دوم (α یا β) وجود دارد و لازم است آزمون‌کننده به نوعی تعادل بین این دو نوع خطا برسد. در رسیدن به این تعادل، موضوع تابع توان آزمون مطرح می‌شود.

$$\text{توان آزمون}: \beta^* = 1-\beta = 1-P(H_0 \text{ پذیرش} \mid \text{غلط } H_0)$$
$$= P(H_0 \text{ رد} \mid \text{غلط } H_0)$$

یادآوری: $1-P(A'|B)=P(A|B)$

ازآنجاکه فرضیه‌های H_0 و H_1 خلاف یکدیگرند، سه تعریف دیگر نیز برای توان آزمون $(1-\beta)$ می‌تواند درست باشد:

$$\beta^* = 1-\beta = P(H_0 \text{ رد} \mid \text{درست } H_1) = P(H_1 \text{ پذیرش} \mid \text{درست } H_1) = P(H_1 \text{ پذیرش} \mid \text{نادرست } H_0)$$

✓ دقت کنید!

همواره بحث بر روی رد کردن یا رد نکردن فرض H_0 است، بنابراین تعریفی که بر مبنای H_0 باشد ارجح است.

مثال ۴ توان آزمون **(Power of the test)** یعنی: (اقتصاد ـ ۷۵)

۱) احتمال پذیرش فرضیه نادرست

۲) احتمال رد فرضیه نادرست

۳) احتمال پذیرش فرضیه درست

۴) رد فرضیه نادرست

حل: گزینه ۲ درست است.

یادآوری: هر سه مقدار خطای نوع اول (α)، خطای نوع دوم (β) و توان آزمون $(1-\beta)$، احتمال هستند به همین دلیل در مثال بالا گزینه ۴ نادرست است.

آزمون توانمند

آزمونی توانمند است که با توجه به مقدار مشخصی از α، دارای خطای نوع دوم (β) کمتر یا توان آزمون بیشتری باشد.

جدول خطاها و احتمال آن‌ها

غلط H_0	درست H_0	نتیجه‌گیری
خطای نوع دوم (β)	تصمیم درست $(1-\alpha)$	H_0 پذیرفته می‌شود.
تصمیم درست (توان آزمون) $(1-\beta)$	خطای نوع اول (α)	H_0 رد می‌شود.

روابط بین α و β و β^*

۱- درصورتی‌که اندازه نمونه (n) ثابت باشد، α و β با هم رابطه **معکوس** دارند. به عبارت دیگر، با افزایش α، β کاهش می‌یابد و برعکس.

$$\alpha\uparrow \xrightleftharpoons{\text{رابطه معکوس}} \beta\downarrow$$

۲- لزوماً **جمع** احتمال α و β **یک نمی‌شود** $(\alpha+\beta\le 1)$.

۳- با **افزایش** n (اندازه نمونه) احتمال **هر دو** نوع خطای α و β **کاهش** می‌یابد.

۴- با توجه به اینکه α و β مقدار احتمال‌اند، **حداکثر مقدارشان یک** و **حداقل مقدارشان صفر** است $(0\le\alpha\ ,\ \beta\le 1)$.

۵- خطای نوع دوم (β) و توان آزمون $\left(\beta^*=1-\beta\right)$ رابطه **معکوس** دارند و **مجموعشان** همیشه برابر **یک** است.

۶- خطای نوع اول (α) و توان آزمون $\left(\beta^*=1-\beta\right)$ رابطه **مستقیم** دارند.

۷- ازآنجاکه α و β رابطه معکوس دارند، β (احتمال خطای نوع دوم) با e (خطا یا دقت برآورد) رابطه **مستقیم** دارد و $\left(\beta^*=1-\beta\right)$ (توان آزمون) با e رابطه **معکوس** دارد.

$$\begin{cases}\uparrow e \xrightleftharpoons{\text{رابطه معکوس}} \alpha\downarrow\\ \uparrow e \xrightleftharpoons{\text{رابطه مستقیم}} \beta\uparrow\\ \uparrow e \xrightleftharpoons{\text{رابطه معکوس}} \beta^*\downarrow\end{cases}$$

مثال ۱ با ثابت نگه داشتن اندازه نمونه، احتمال اشتباه نوع اول فقط در صورتی کاهش می‌یابد که: (مدیریت ـ ۷۴)

۱) طول ناحیه رد H_0 افزایش یابد.
۲) طول ناحیه پذیرش H_0 کاهش یابد.
۳) احتمال خطای نوع دوم افزایش یابد.
۴) احتمال خطای نوع دوم کاهش یابد.

حل: گزینه ۳ درست است.

بنا بر رابطه (۱)، α و β رابطه معکوس دارند.

یادآوری: طول ناحیه رد H_0 همان α است و طول ناحیه پذیرش H_0 همان $1-\alpha$ است. بنابراین اگر گزینه (۱) به صورت «طول ناحیه رد H_0 کاهش یابد» باشد، درست است و اگر گزینه (۲) به صورت «طول ناحیه پذیرش H_0 افزایش یابد» باشد، درست است.

مثال ۲ احتمال خطای نوع اول و احتمال خطای نوع دوم: (مدیریت ـ ۷۵)

۱) با یکدیگر رابطه معکوس دارند.
۲) مجموعشان بیشتر از یک است.
۳) مجموعشان مساوی با یک است.
۴) مساوی‌اند.

حل: گزینه ۱ درست است.

بنا بر رابطه (۱)، α و β رابطه معکوس دارند.

مثال ۳ چگونه می‌توان اشتباه نوع اول را کاهش داد به طوری که اشتباه نوع دوم افزایش نیابد؟

۱) تعداد نمونه را افزایش دهیم.
۲) سطح اطمینان را کاهش دهیم.
۳) تعداد نمونه را کاهش دهیم.
۴) سطح معنی‌دار را افزایش دهیم.

حل: گزینه ۱ درست است.

بنا بر رابطه (۳)، با افزایش n هر دو نوع خطا کاهش می‌یابد.

مثال ۴ در آزمون فرضیه، افزایش خطای نوع اول به شرط ثابت بودن سایر عوامل، موجب می‌شود که: (اقتصاد ـ ۷۳)

۱) توان آزمون افزایش یابد. ۲) خطای نوع دوم افزایش یابد.

۳) توان آزمون کاهش یابد. ۴) خطای نوع دوم ثابت باقی بماند.

حل: گزینه ۱ درست است.

بنا بر رابطه (۶)، α و β^* رابطه مستقیم دارند.

مثال ۵ کارخانه‌ای ادعا کرده است که میانگین محصولاتش بیش از 40 کیلوگرم است. خطای نوع اول (α) به معنی آن است که نتیجه بگیریم:

۱) میانگین محصولات کارخانه بیشتر از 40 کیلوگرم است درحالی‌که چنین نیست.

۲) میانگین محصولات کارخانه حداکثر 40 کیلوگرم است درحالی‌که چنین هم هست.

۳) میانگین محصولات کارخانه بیشتر از 40 کیلوگرم است درحالی‌که چنین هم هست.

۴) میانگین محصولات کارخانه حداکثر 40 کیلوگرم است درحالی‌که چنین نیست.

حل: گزینه ۱ درست است.

دقت کنید، عبارت «درحالی‌که چنین هم هست» بر **درست بودن** فرضیه و عبارت «درحالی‌که چنین نیست» بر **غلط بودن** فرضیه دلالت دارد.

در این مسئله، فرضیه‌های H_0 و H_1 به صورت زیر هستند:

H_0 : میانگین محصولات کارخانه کوچک‌تر یا مساوی (حداکثر) 40 کیلوگرم است $(\mu \leq 40)$.

H_1 : میانگین محصولات کارخانه بیشتر از 40 کیلوگرم است $(\mu > 40)$.

با توجه به این فرضیه‌ها مشخص می‌شود که ادعا در فرض H_1 قرار گرفته است.

حال به بررسی گزینه‌ها می‌پردازیم:

گزینه ۱: میانگین محصولات کارخانه بیشتر از 40 کیلوگرم است | درحالی‌که چنین نیست.

پذیرش H_1 (رد H_0) | H_1 غلط (H_0 درست)

رد H_0 وقتی H_0 درست است $= \alpha =$ خطای نوع اول

گزینه ۲: میانگین محصولات کارخانه حداکثر 40 کیلوگرم است | درحالی‌که چنین هم هست.

پذیرش H_0 (رد H_1) | H_1 غلط (H_0 درست)

گزینه ۳: میانگین محصولات کارخانه بیشتر از 40 کیلوگرم است | درحالی‌که چنین هم هست.

پذیرش H_1 (رد H_0) | H_0 غلط (H_1 درست)

رد H_0 وقتی H_0 غلط است $= 1-\beta =$ توان آزمون

گزینه ۴: $\underbrace{\underbrace{\text{میانگین محصولات کارخانه حداکثر 40 کیلوگرم است}}_{\text{پذیرش } H_0 \text{ (رد } H_1)} \quad \underbrace{\text{درحالی‌که چنین نیست.}}_{H_0 \text{ غلط (} H_1 \text{ درست)}}}$

پذیرش H_0 وقتی H_0 غلط است $= \beta =$ خطای نوع دوم

توجه: برای محاسبه خطاها ابتدا نیاز به بررسی آزمون فرض‌های مختلف داریم. ازاین‌رو محاسبه آن‌ها را به بعد موکول می‌کنیم.

ملاک (آماره) آزمون (Test Statistic)

برای هر آزمون فرضیه می‌توان یک ملاک آزمون مناسب، بر اساس برآوردی از یک نمونه n تایی از جامعه محاسبه کرد و از طریق آن ملاک، میزان نزدیکی نتیجه حاصل را به فرض پیشنهادی H_0 به دست آورد.

ملاک (آماره) آزمون همان متغیر استانداردی است که در برآورد فاصله‌ای بر اساس توزیع نمونه‌ای به عنوان تابع محوری مورد استفاده قرار می‌گرفت با این تفاوت که در ملاک آزمون، با پذیرش فرض پیشنهادی H_0 درباره پارامتر جامعه، تمام اجزای تابع معلوم است.

برای مثال، تابع محوری برآورد فاصله‌ای میانگین جامعه (μ) در یک نمونه n تایی با میانگین $\bar{X}$ از جامعه‌ای نرمال با واریانس σ^2 را در نظر بگیرید:

$$\text{تابع محوری} : Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

که در آن μ نامعلوم است.

در این صورت ملاک (آماره) آزمون برای آزمون فرضیه $H_0 : \mu \overset{\geq}{\underset{\leq}{=}} \mu_0$ به شکل زیر است:

$$\text{ملاک آزمون} : Z = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

که در آن μ_0 (مقدار پیشنهادی H_0) معلوم است.

مثال در یک آزمون فرضیه، آماره آزمون (تابع نمونه‌ای آزمون) همواره: (اقتصاد ـ ۷۹)

۱) بر اساس نحوه برآوردکننده پارامتر جامعه و با فرض صحیح بودن H_1 ساخته می‌شود.

۲) بر اساس نحوه برآوردکننده پارامتر جامعه و با فرض صحیح بودن H_0 ساخته می‌شود.

۳) بر اساس نحوه متغیر مورد نظر در جامعه و با فرض صحیح بودن H_1 ساخته می‌شود.

۴) بر اساس نحوه متغیر مورد نظر در جامعه و با فرض صحیح بودن H_0 ساخته می‌شود.

حل: گزینه ۲ درست است.

ملاک آزمون‌های مختلف که در این فصل مورد بررسی قرار می‌گیرند به شرح زیر است:

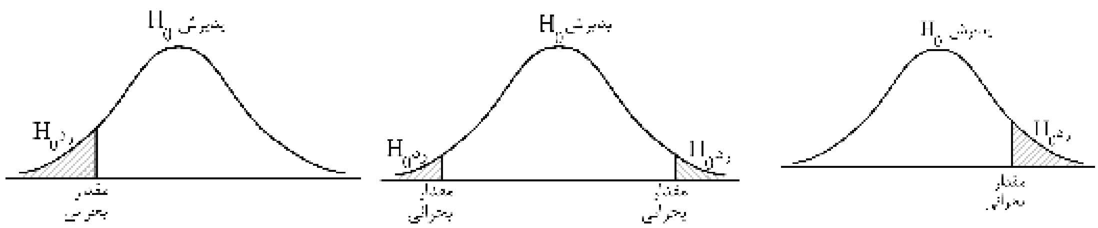
جدول فرضیه‌های آماری و ملاک‌های آن

فرضیه	توزیع نمونه‌گیری	آماره آزمون (استاندارد)
میانگین جامعه	نرمال یا t–استیودنت	Z یا t
مقایسه میانگین دو جامعه	نرمال یا t–استیودنت	Z یا t
نسبت موفقیت	نرمال یا دوجمله‌ای	Z
مقایسه نسبت موفقیت دو جامعه	نرمال یا دوجمله‌ای	Z
واریانس جامعه	کای اسکور	χ^2
مقایسه واریانس دو جامعه	فیشر	F
نیکویی برازش	کای‌اسکور	χ^2
استقلال متغیرهای کیفی	کای‌اسکور	χ^2
مقایسه میانگین چند جامعه	فیشر	F
رگرسیون و ضریب همبستگی	t–استیودنت یا فیشر	t یا F

سطح H_0 و H_1 و مقادیر بحرانی

توزیع آماره هر آزمون با توجه به فرض‌های H_0 و H_1 به دو ناحیه تقسیم می‌شود؛ ناحیه پذیرش H_0 و ناحیه رد آن. مقادیر بحرانی، ناحیه رد را از ناحیه پذیرش جدا می‌کنند. این مقادیر از جدول استاندارد ملاک آزمون در پیوست کتاب با توجه به سطح معنی‌دار (α) به دست می‌آیند.

برای مثال، این نواحی در توزیع نرمال با توجه به فرض‌های H_0 و H_1 به صورت زیر هستند:



$$\begin{cases} H_0: \theta \le \theta_0 \\ H_1: \theta > \theta_0 \end{cases} \qquad \begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \ne \theta_0 \end{cases} \qquad \begin{cases} H_0: \theta \ge \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$

✓ **دقت کنید!**

با توجه به اینکه در فرض H_1 دامنه نامساوی به چپ است یا به راست و یا به هر دو طرف، ناحیه بحرانی نیز به همان طرف خواهد بود.

برای نواحی پذیرش و رد H_0 اصطلاحات معادل زیر به کار می‌رود:

ناحیه پذیرش H_0 (سطح H_0):

سطح اطمینان $(1-\alpha)$، ناحیه فرض آزمون، ناحیه‌ای که اختلاف معنی‌دار نیست.

ناحیه رد H_0 (سطح H_1):

سطح معنی‌دار (α)، سطح خطا، ناحیه بحرانی، خطای نوع اول، ناحیه‌ای که اختلاف معنی‌دار است.

درصورتی‌که مقدار آماره (ملاک) آزمون حاصل از نتایج نمونه در ناحیه پذیرش H_0 قرار گیرد، فرض H_0 را نمی‌توان رد کرد، اما اگر مقدار آن در ناحیه رد H_0 قرار گیرد فرض H_0 رد می‌شود.

مثال سطح زیر منحنی مربوط به تابع نمونه‌ای (آماره) آزمون فرضیه H_0 در آزمون فرضیه آماری همواره برابر است با:
(مدیریت ـ ۷۳ ، ۷۴)

۱) درصد اطمینان آزمون فرضیه
۲) خطای نوع اول (α)
۳) ناحیه پذیرش H_1
۴) سطح معنی‌دار

حل: گزینه ۱ درست است.
گزینه‌های ۲، ۳ و ۴ یک مفهوم دارند.

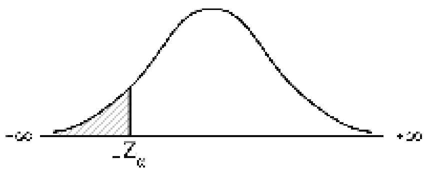
انواع آزمون‌های آماری

ازآنجاکه ناحیه H_1 (سطح معنی‌دار) دامنه‌ای را مشخص می‌کند که فرض H_0 در آن رد می‌شود، آزمون‌های آماری با توجه به وضعیت فرض H_1 به دو حالت یک‌دامنه و دودامنه به شرح زیر تقسیم می‌شوند:

۱- آزمون یک‌دامنه (One - Tail (Sided) Test)

الف) آزمون دامنه به چپ

هرگاه فرض آزمون به صورت $\begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ باشد، آزمون، دامنه به چپ است؛ ناحیه بحرانی (معنی‌دار) و ناحیه اطمینان برای مثال برای توزیع Z به صورت زیر است:



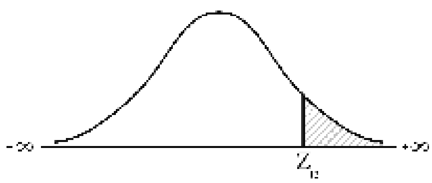
$$\begin{cases} \text{ملاک آزمون}: U \geq -Z_\alpha \rightarrow \text{فرض } H_0 \text{ پذیرفته می‌شود.} \\ \text{ملاک آزمون}: U < -Z_\alpha \rightarrow \text{فرض } H_0 \text{ رد می‌شود.} \end{cases}$$

✓ **دقت کنید!**

همیشه در تعیین ناحیه بحرانی توجه ما به فرض H_1 است.

ب) آزمون دامنه به راست

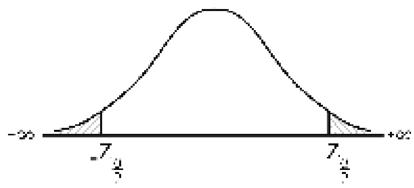
هرگاه فرض آزمون به صورت $\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ باشد، آزمون، دامنه به راست است؛ ناحیه بحرانی (معنی‌دار) و ناحیه اطمینان برای مثال برای توزیع Z به صورت زیر است:



$$\begin{cases} \text{ملاک آزمون}: U \leq Z_\alpha \rightarrow \text{فرض } H_0 \text{ پذیرفته می‌شود.} \\ \text{ملاک آزمون}: U > Z_\alpha \rightarrow \text{فرض } H_0 \text{ رد می‌شود.} \end{cases}$$

۲- آزمون دودامنه (Two - Tail (Sided) Test)

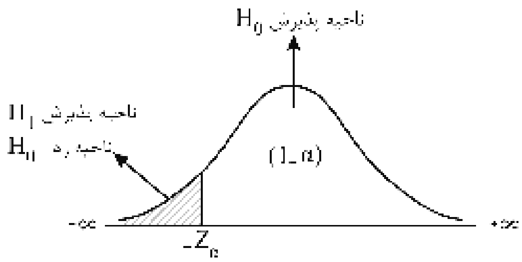
هرگاه فرض آزمون به صورت $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$ باشد، آزمون، دودامنه است؛ ناحیه بحرانی (معنی‌دار) و ناحیه اطمینان برای مثال برای توزیع Z به صورت زیر است:



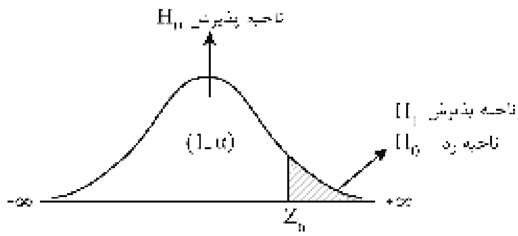
$$\begin{cases} U : \text{ملاک آزمون} \quad ; \quad -Z_{\frac{\alpha}{2}} \leq U \leq Z_{\frac{\alpha}{2}} \rightarrow \text{فرض } H_0 \text{ پذیرفته می‌شود} \\ U : \text{ملاک آزمون} \quad ; \quad U > Z_{\frac{\alpha}{2}} \\ U : \text{ملاک آزمون} \quad ; \quad U < -Z_{\frac{\alpha}{2}} \end{cases} \rightarrow \text{فرض } H_0 \text{ رد می‌شود.}$$

سطح معنی‌دار و سطح اطمینان در آزمون‌های یک‌دامنه و دودامنه

آزمون یک‌دامنه

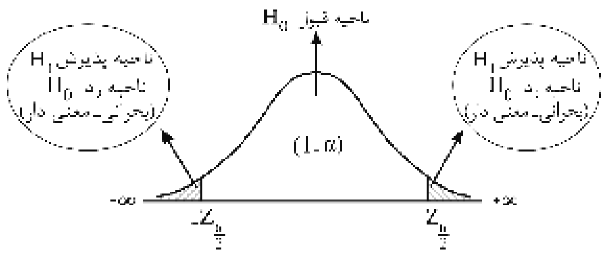


$\begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$ دامنه به چپ



$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$ دامنه به راست

آزمون دودامنه



$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$

توجه:

۱- $\alpha = 0.10 , 0.05 , 0.01$ مقادیر مرسوم برای سطوح اطمینان 0.99، 0.95 و 0.90 هستند.

۲- بسته به آماره آزمون، ممکن است در تعیین سطوح اطمینان و معنی‌دار به جای Z از توزیع‌های t، χ^2 یا F استفاده شود.

مراحل عمومی آزمون فرض آماری

از جمع‌بندی مبانی آزمون فرض که تا اینجا بیان شد، می‌توان نتیجه گرفت که برای انجام یک آزمون فرضیه مراحل زیر باید طی شود:

۱- تعریف فرض صفر (H_0) و فرض مقابل (H_1)

۲- تعیین سطح معنی‌دار (α) بر اساس اهمیت موضوع آزمون

۳- تعیین حجم نمونه

۴- تعیین آماره (ملاک) آزمون

۵- تقسیم توزیع آماره به دو ناحیه (رد و پذیرش) و مشخص کردن مقادیر بحرانی

۶- جمع‌آوری اطلاعات از طریق نمونه‌گیری و محاسبه آماره آزمون بر اساس H_0 و نمونه گرفته‌شده

۷- تعیین اینکه آیا آماره آزمون در ناحیه رد قرار می‌گیرد یا ناحیه پذیرش

۸- تصمیم‌گیری درباره H_0 و H_1

با خلاصه کردن موارد ذکرشده می‌توان مراحل چهارگانه زیر را برای انجام هر آزمون آماری تدوین کرد:

مرحله اول: تعریف فرضیه‌های آماری H_0 و H_1

با توجه به قواعدی که برای H_0 و H_1 تعریف شد، در این مرحله با توجه به ادعای مطرح‌شده درباره پارامتر جامعه، باید فرض‌های H_0 و H_1 را مشخص کنیم.

مرحله دوم: تعیین توزیع نمونه‌گیری و ملاک آزمون

همان‌طورکه گفته شد، توزیع نمونه‌گیری و ملاک آزمون با توجه به شرایط برآورد پارامتر جامعه مشخص می‌شوند. ملاک یا آماره آزمون همان تابع محوری برآورد فاصله‌ای است، با این تفاوت که با پذیرش پیشنهاد H_0 تمام اجزای آن معلوم می‌شود.

مرحله سوم: تعیین سطح زیر منحنی H_0 و H_1 و مقادیر بحرانی

سطح زیر منحنی H_0 و H_1 به توزیع نمونه‌گیری و مقدار α (سطح معنی‌دار) بستگی دارد؛ در عین حال یک‌دامنه یا دودامنه بودن آزمون نیز بر سطح H_0 و H_1 تأثیر مستقیم دارد.

مقدار بحرانی عددی است که از مقدار استاندارد توزیع نمونه‌گیری با توجه به سطح معنی‌دار (α) از جداول پیوست کتاب استخراج می‌شود.

مرحله چهارم: تصمیم‌گیری درباره H_0 و H_1

در این مرحله مقدار آماره آزمون تعیین‌شده در مرحله دوم را با مقدار بحرانی مرحله سوم مقایسه می‌کنیم، اگر مقدار آماره در ناحیه H_0 (سطح اطمینان) قرار گیرد، فرض H_0 پذیرفته می‌شود، اما اگر این مقدار در ناحیه H_1 (سطح معنی‌دار) قرار گیرد، فرض H_0 رد می‌شود.

این مراحل چهارگانه برای تمام آزمون‌ها تا انتهای کتاب انجام خواهد شد.

آزمون میانگین جامعه

اگر فرضیه‌ای درباره مقدار میانگین یک جامعه آماری مطرح شود، با استفاده از آزمون فرض مناسب می‌توان درباره صحت یا عدم صحت آن فرضیه در سطح اطمینان $(1-\alpha)$ یا سطح خطای α تصمیم‌گیری کرد. مراحل چهارگانه آزمون فرض برای میانگین جامعه به شرح زیر است:

۱- فرض‌های آماری

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X \neq \mu_0 \end{cases} , \quad \begin{cases} H_0 : \mu_X \geq \mu_0 \\ H_1 : \mu_X < \mu_0 \end{cases} , \quad \begin{cases} H_0 : \mu_X \leq \mu_0 \\ H_1 : \mu_X > \mu_0 \end{cases}$$

در این فرضیه‌ها، μ_0 همان مقدار میانگین مورد آزمون است.

۲- آماره آزمون (ملاک آزمون)

درصورتی‌که $X_1, X_2, \dots, X_n$ نمونه مستقل n تایی از جامعه باشد، آن‌گاه آماره $\bar{X}$ با توجه به حالات مختلف برآورد، دارای توزیع Z یا t است. حال بر اساس هر یک از این حالات، ملاک (آماره) آزمون به صورت زیر تعریف می‌شود:

۱) اگر جامعه، نرمال و σ^2 معلوم باشد، توزیع $\bar{X}$ صرف‌نظر از n (n دلخواه) نرمال است و ملاک آزمون عبارت است از:

$$\boxed{Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} \quad ; \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

۲) اگر جامعه نرمال و σ^2 نامعلوم باشد، توزیع $\bar{X}$ به n وابسته است و دو حالت به وجود می‌آید:

الف) اگر $n > 30$ باشد، توزیع $\bar{X}$ نرمال است و ملاک آزمون عبارت است از:

$$\boxed{Z = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} \quad ; \quad S_{\bar{X}} = \frac{S}{\sqrt{n}}}$$

ب) اگر $n \leq 30$ باشد، توزیع $\bar{X}$، t با $n-1$ درجه آزادی است و ملاک آزمون عبارت است از:

$$\boxed{t_{n-1} = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} \quad ; \quad S_{\bar{X}} = \frac{S}{\sqrt{n}}}$$

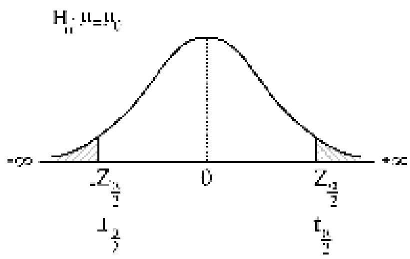
۳) اگر جامعه غیرنرمال (نامعلوم)، σ^2 معلوم و تعداد نمونه بزرگ باشد $(n > 30)$، توزیع $\bar{X}$ بنا بر قضیه حد مرکزی نرمال است و ملاک آزمون عبارت است از:

$$\boxed{Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} \quad ; \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

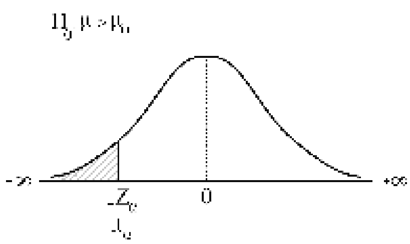
۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

در این مرحله ابتدا با توجه به حالت‌های مرحله اول، نوع آزمون را از نظر یک‌دامنه یا دودامنه بودن مشخص می‌کنیم، سپس با توجه به سطح خطای α و توزیع آماره، مقدار بحرانی (مرز H_0 و H_1) را از جدول مربوط به توزیع‌ها در پیوست کتاب، استخراج می‌کنیم.

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X \neq \mu_0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود}]{U > Z_{\frac{\alpha}{2}} \ \text{or} \ U < -Z_{\frac{\alpha}{2}}}$$

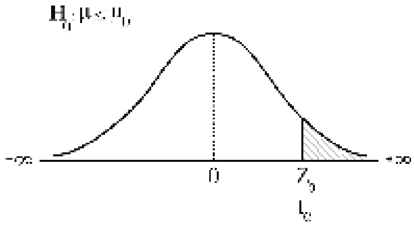


$$\begin{cases} H_0 : \mu_X \geq \mu_0 \\ H_1 : \mu_X < \mu_0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود}]{U < -Z_{\alpha}}$$



$$\begin{cases} H_0 : \mu_X \le \mu_0 \\ H_1 : \mu_X > \mu_0 \end{cases} \xrightarrow{U > Z_\alpha}$$

فرض H_0 رد می‌شود.



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $(1-\alpha)$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

توجه: در مرحله تصمیم‌گیری جملاتی مانند «دلیلی برای رد H_0 وجود ندارد»، « H_0 را نمی‌توان رد کرد»، « H_0 اثبات می‌شود»، « H_0 پذیرفته می‌شود» و « H_0 تأیید می‌شود» معادل‌اند. اما در شرایط مشابه، دو جمله اول که با فعل منفی به کار رفته‌اند ارجح هستند.

مثال ۱ ادعا شده است میانگین نمره مسئولیت‌پذیری در کشور، حداقل 50 است. برای بررسی فرضیه، یک نمونه 64 تایی از بین مدیران کشور به طور تصادفی انتخاب شده و میانگین و انحراف معیار 45 و 16 به دست آمده است. با فرض نرمال بودن جامعه در سطح خطای 5% صحت این فرضیه را بررسی کنید.

۱) ادعا پذیرفته می‌شود. ۲) فرض H_1 رد می‌شود. ۳) ادعا رد می‌شود. ۴) فرض H_0 پذیرفته می‌شود.

حل: گزینه ۳ درست است.

راه حل اول:

۱) فرض‌های آزمون

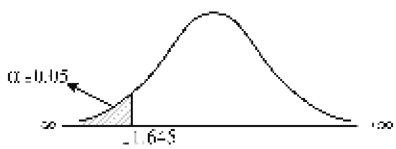
$$\begin{cases} H_0 : \mu \ge 50 \\ H_1 : \mu < 50 \end{cases}$$

۲) آماره آزمون

$$\begin{cases} Z = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}} = \dfrac{45 - 50}{\dfrac{16}{\sqrt{64}}} = -2.5 \\ \overline{x} = 45 \text{ , } S = 16 \text{ , } \mu_0 = 50 \text{ , } n = 64 > 30 \end{cases}$$

۳) مقادیر بحرانی

$$\alpha = 0.05 \text{ , } Z_{0.05} = -1.645$$



۴) تصمیم‌گیری

با توجه به قرار گرفتن $Z = -2.5$ در ناحیه بحرانی، فرض H_0 رد می‌شود و چون ادعا در H_0 قرارگرفته است، ادعا رد می‌شود.

راه حل دوم: با کمی دقت در گزینه‌ها مشخص می‌شود که گزینه‌های ۱، ۲ و ۴ مفاهیم یکسانی را بیان می‌کنند؛ بنابراین، بدون حل سؤال نیز می‌توانستیم گزینه ۳ را به عنوان گزینه درست انتخاب کنیم.

مثال ۲ در آزمون $H_0 : \mu = \mu_0$ برای جامعه نرمال با انحراف معیار نامشخص و درجه آزادی کمتر از 30، تابع نمونه‌ای آزمون (آماره آزمون) عبارت است از: (مدیریت ـ ۷۲)

۱) $t = \dfrac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$ ۲) $Z = \dfrac{\bar{X} - \mu_0}{S_{\bar{X}}}$ ۳) $t = \dfrac{\bar{X} - \mu_0}{S_{\bar{X}}}$ ۴) $Z = \dfrac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$

حل: گزینه ۳ درست است.

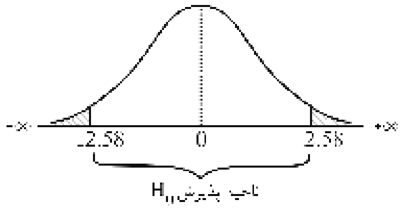
با توجه به حالت (۲ـب) آماره آزمون t است.

مثال ۳ آماره آزمون، نرمال صفر و یک است و $H_0 : \mu_X = 60$ تعریف شده است. مقدار آماره آزمون 3.5 است. در سطح اطمینان 99 درصد کدام گزینه درست است؟

۱) H_0 تأیید می‌شود. ۲) H_1 تأیید می‌شود.

۳) H_1 تأیید نمی‌شود. ۴) اطلاعات بیشتری برای تصمیم‌گیری مورد نیاز است.

حل: گزینه ۲ درست است.



$$(1)\ \begin{cases} H_0 : \mu_X = 60 \\ H_1 : \mu_X \neq 60 \end{cases}$$

(2) $Z = 3.5$ آماره آزمون

(3) $\alpha = 0.01$, $Z_{0.005} = 2.58 \rightarrow$

۴) با توجه به اینکه مقدار آماره آزمون در ناحیه بحرانی است، فرض H_0 رد شده و H_1 تأیید می‌شود.

آزمون مقایسه میانگین دو جامعه

فرضیه‌های مربوط به مقایسه دو جامعه آماری به «فرضیه‌های تطبیقی» معروف هستند که از مراحل آزمون فرض آماری میانگین دو جامعه برای این مقایسه استفاده می‌شود.

هرگاه نمونه n_1 از جامعه اول با میانگین μ_1 و واریانس σ_1^2 و نمونه n_2 از جامعه دوم مستقل از نمونه اول با میانگین μ_2 و واریانس σ_2^2 انتخاب شوند، داریم:

۱- فرض‌های آماری

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} , \quad \begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} , \quad \begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$$

۲- آماره آزمون (ملاک آزمون)

با فرض آنکه می‌دانیم آماره $\bar{X}_1 - \bar{X}_2$ برای $\mu_1 - \mu_2$ ناریب است، آماره آزمون عبارت است از:

۱) وقتی نمونه‌ها از دو جامعه نرمال با انحراف معیارهای معلوم (σ_2, σ_1) انتخاب شوند و یا دو جامعه غیرنرمال با انحراف معیارهای معلوم (σ_2, σ_1) و $n_1, n_2 > 30$ باشد (قضیه حد مرکزی)، توزیع $\bar{X}_1 - \bar{X}_2$ نرمال است و داریم:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \xrightarrow{H_0 : \mu_1 = \mu_2} Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

✓ **دقت کنید!**

تحت فرض $\mu_1-\mu_2=0 \quad \rightarrow \quad H_0:\mu_1=\mu_2$ قسمت $\mu_1-\mu_2$ در صورت کسر آماره برابر صفر می‌شود.

۲) وقتی نمونه‌ها از دو جامعه نرمال با انحراف معیار نامعلوم، انتخاب شوند $n_1>30$ و $n_2>30$ باشد، توزیع $\bar{x}_1-\bar{x}_2$ نرمال است و داریم:

$$\boxed{Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \xrightarrow{H_0:\mu_1=\mu_2} Z=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}}$$

۳) اگر نمونه‌ها از دو جامعه نرمال با انحراف معیار نامعلوم انتخاب شوند و $n_1\le 30\,,\,n_2\le 30$ یا $n_1+n_2\le 30$ باشد، آماره t است، اما در این حالت خطای استاندارد تحت تأثیر فرض تساوی یا عدم تساوی واریانس دو جامعه است.

الف) با فرض تساوی واریانس دو جامعه $\left(\sigma_1^2=\sigma_2^2\right)$، آماره t با n_1+n_2-2 درجه آزادی به صورت زیر است:

$$\boxed{t_{n_1+n_2-2}=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \xrightarrow{H_0:\mu_1=\mu_2} t=\frac{(\bar{X}_1-\bar{X}_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}$$

که در آن:

$$S_P^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

اگر $n_1=n_2=n$ باشد:

$$S_p^2=\frac{(n-1)S_1^2+(n-1)S_2^2}{\underbrace{\cancel{n+n-2}}_{2(n-1)}}=\frac{S_1^2+S_2^2}{2}$$

ب) با فرض عدم برابری واریانس دو جامعه $\left(\sigma_1^2\neq\sigma_2^2\right)$، آماره t و درجه آزادی آن به صورت زیر است:

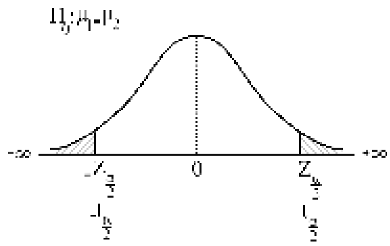
$$\boxed{t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \xrightarrow{H_0:\mu_1=\mu_2} t=\frac{(\bar{X}_1-\bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}}$$

$$\text{درجه آزادی}=\frac{\left(S_1^2/n_1+S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1}+\frac{\left(S_2^2/n_2\right)^2}{n_2-1}}$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

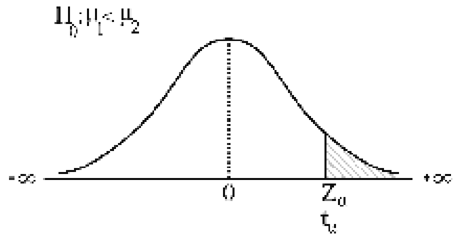
$$\begin{cases} H_0:\mu_1=\mu_2 \\ H_1:\mu_1\neq\mu_2 \end{cases} \xrightarrow{U>Z_{\frac{\alpha}{2}} \quad \text{or} \quad U<-Z_{\frac{\alpha}{2}}}$$

فرض H_0 رد می‌شود.



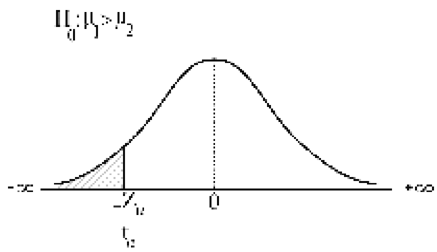
$$\begin{cases} H_0:\mu_1\leq\mu_2 \\ H_1:\mu_1>\mu_2 \end{cases} \xrightarrow{U>Z_{\alpha}}$$

فرض H_0 رد می‌شود.



$$\begin{cases} H_0:\mu_1\geq\mu_2 \\ H_1:\mu_1<\mu_2 \end{cases} \xrightarrow{U<-Z_{\alpha}}$$

فرض H_0 رد می‌شود.



۴- تصمیم‌گیری

در این مرحله، مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

توجه:

۱- درصورتی‌که از تعداد نمونه صحبتی نشود، به طور پیش‌فرض $n_1, n_2 \leq 30$ است.

۲- درصورتی‌که از واریانس دو جامعه صحبتی نشود، به طور پیش‌فرض، واریانس دو جامعه نابرابر و نامعلوم است.

۳- در آزمون مقایسه میانگین دو جامعه، به طور پیش‌فرض، دو جامعه نرمال است.

مثال ۱ اگر میانگین یک نمونه 100تایی از X مساوی 30 و انحراف معیار آن 5 باشد و میانگین یک نمونه 200 تایی از Y مساوی 25 و انحراف معیار آن 10 باشد، مقدار ملاک آزمون صفر بودن تفاوت میانگین‌ها برابر است با: (مدیریت ـ ۷۳)

۱) 15.81 ۲) 1.96 ۳) 5.77 ۴) 3.44

حل: گزینه ۳ درست است.

$$\begin{cases} H_0:\mu_X-\mu_Y=0 \\ H_1:\mu_X-\mu_Y\neq0 \end{cases}$$

با توجه به حالت (۲) داریم:

$$\begin{cases} Z=\dfrac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\dfrac{S_X^2}{n_X}+\dfrac{S_Y^2}{n_Y}}} \xrightarrow{H_0:\mu_X-\mu_Y=0} Z=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_X^2}{n_X}+\dfrac{S_Y^2}{n_Y}}}=\dfrac{(30-25)}{\sqrt{\dfrac{5^2}{100}+\dfrac{10^2}{200}}}=5.77 \\ \bar{x}=30\ ,\ S_X=5\ ,\ n_X=100>30 \\ \bar{y}=25\ ,\ S_Y=10\ ,\ n_Y=200>30 \end{cases}$$

مثال ۲ برای آزمون برابری میانگین‌های دو جامعه نرمال مستقل با واریانس‌های برابر، نتایج زیر از نمونه‌های انتخاب‌شده از این دو جامعه به دست آمده‌اند. مقدار آماره آزمون برابر است با: (اقتصاد ـ ۸۶)

جامعه اول	$n_1=18$	$\bar{x}_1=170$	$S_1^2=15$
جامعه دوم	$n_2=18$	$\bar{x}_2=153$	$S_2^2=17$

۱) 12.75 ۲) $\dfrac{17\sqrt{17}}{4}$ ۳) $\dfrac{34\sqrt{2}}{3}$ ۴) $\dfrac{17}{4}$

حل: گزینه ۱ درست است.

$$1)\begin{cases} H_0:\mu_1=\mu_2 \\ H_1:\mu_1\neq\mu_2 \end{cases}$$

با توجه به حالت (۳ـ الف) داریم:

$$2)\begin{cases} t_{n_1+n_2-2}=\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}=\dfrac{(170-153)-0}{4\sqrt{\dfrac{1}{18}+\dfrac{1}{18}}}=\dfrac{17}{4\sqrt{\dfrac{1}{9}}}=\dfrac{51}{4}=12.75 \\ n_1=n_2=18\ \rightarrow\ S_p^2=\dfrac{S_1^2+S_2^2}{2}=\dfrac{15+17}{2}=16\ \rightarrow\ S_p=4 \\ \bar{x}_1=170\ ,\ S_1^2=15\ ,\ \bar{x}_2=153\ ,\ S_2^2=17 \end{cases}$$

آزمون مقایسه زوج‌ها (فرضیه نمونه‌های جفت‌شده)

در آزمون تفاضل میانگین دو جامعه (بخش قبل)، فرض اساسی، مستقل بودن نمونه‌ها در دو جامعه نسبت به هم بود.

در مواردی مانند
(۱) بررسی تفاوت نمرات دانشجویان، قبل و بعد از یک دوره کلاس تقویتی
(۲) بررسی تفاوت وزن افراد، قبل و بعد از یک دوره رژیم غذایی

نمونه‌های انتخاب‌شده برای ارزشیابی، قبل و بعد از دوره، یکسان هستند و مستقل نیستند، بنابراین نمی‌توان از روش آزمون تفاضل میانگین برای این نمونه‌ها استفاده کرد؛ در چنین مواردی از «آزمون مقایسه زوج‌ها» به صورت زیر استفاده می‌شود:

الف) با فرض انتخاب یک نمونه n تایی که رفتار هر عضو آن، قبل و بعد از دوره مشخص، به صورت جفت (x_i, y_i) باشد، جدول زیر را تشکیل می‌دهیم:

قبل x_i	بعد y_i	$d_i=x_i-y_i$
x_1	y_1	$d_1=x_1-y_1$
x_2	y_2	$d_2=x_2-y_2$
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	$d_n=x_n-y_n$

ب) با توجه به جدول مقادیر(الف) داریم:

میانگین تفاضل رفتار	$\bar{d}=\dfrac{\sum_{i=1}^{n} d_i}{n}=\bar{x}-\bar{y}$
واریانس تفاضل رفتار	$S_d^2=\dfrac{\sum\left(d_i-\bar{d}\right)^2}{n-1}$
انحراف معیار تفاضل رفتار	$S_d=\sqrt{\dfrac{\sum\left(d_i-\bar{d}\right)^2}{n-1}}$
واریانس میانگین تفاضل رفتار	$S_{\bar{d}}^2=\dfrac{S_d^2}{n}$
انحراف معیار میانگین تفاضل رفتار	$S_{\bar{d}}=\dfrac{S_d}{\sqrt{n}}$

۱- فرض‌های آماری

$$\begin{cases} H_0: \text{تفاوتی بین قبل و بعد از دوره وجود ندارد} \ (d=0) \\ H_1: \text{تفاوت بین قبل و بعد از دوره وجود دارد} \ (d\neq 0) \end{cases}$$

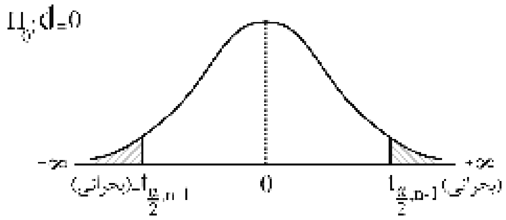
۲- آماره (ملاک آزمون)

ملاک آزمون زوجی از توزیع t با $n-1$ درجه آزادی به شرح زیر تبعیت می‌کند:

$$t_{n-1}=\dfrac{\bar{d}-0}{\dfrac{S_d}{\sqrt{n}}}$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

$$\begin{cases} H_0: d=0 \\ H_1: d\neq 0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{U>t_{\frac{\alpha}{2},n-1} \quad \text{or} \quad U<-t_{\frac{\alpha}{2},n-1}}$$



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، در سطح اطمینان $1-\alpha$ فرض H_0 پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال ۱ در کدام‌یک از موارد زیر از آزمون زوجی استفاده می‌شود؟

۱) گروه‌های مستقل　　۲) گروه‌های همبسته
۳) زوج‌های مستقل از هم　　۴) ۲ و ۳

حل: گزینه ۴ درست است.

مثال ۲ برای بررسی میزان تأثیر یک دوره کوتاه‌مدت حسابداری، دوره‌ای برای یک نمونه تصادفی شش نفره برگزار شده است و نمره‌های دانشجویان قبل و بعد از دوره در جدول زیر آمده است. در سطح خطای 5%، آماره آزمون یکسان بودن نمرات دانشجویان قبل و بعد از دوره کدام است؟

نمره قبل از دوره	40	70	45	50	68	55
نمره بعد از دوره	45	72	56	50	72	63

۱) -2.41 ۲) -3.06 ۳) -1.61 ۴) -2.01

حل: گزینه ۲ درست است.

$$\begin{cases} H_0 : d = 0 \\ H_1 : d \neq 0 \end{cases}$$

$$t_{n-1} = \frac{\bar{d} - 0}{\frac{S_d}{\sqrt{n}}} \rightarrow t_5 = \frac{-5}{\frac{4}{\sqrt{6}}} = -3.06$$

$$\begin{cases} d_i = x_i - y_i = \text{بعد} - \text{قبل} = -5, -2, -11, 0, -4, -8 \\ \bar{d} = \frac{\sum d_i}{n} = \frac{-30}{6} = -5 \\ S_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{(-5+5)^2 + (-2+5)^2 + (-11+5)^2 + (0+5)^2 + (-4+5)^2 + (-8+5)^2}{6-1} \\ \quad = \frac{0+9+36+25+1+9}{5} = 16 \rightarrow S_d = 4 \end{cases}$$

نکته: در فرضیه نمونه‌های جفت‌شده، چون نمونه‌ها از هم مستقل نیستند، امکان تعیین فاصله اطمینان برای تفاوت قبل و بعد از دوره وجود ندارد.

مثال ۳ برای بررسی میزان تأثیر داروی کاهش فشار خون، این دارو به 15 نفر تجویز گردیده است و فشار خون آنان قبل از مصرف دارو و سپس دو ساعت بعد از مصرف دارو اندازه‌گیری شده است که نتایج حاصل به شرح جدول زیر است. آیا بر اساس این نتایج می‌توان یک برآوردکننده فاصله‌ای برای میزان تأثیر دارو در کاهش فشار خون به دست آورد؟ (اقتصاد ـ ۸۳)

	$\bar{x}$	S^2
قبل از مصرف	18	4.5
دو ساعت بعد از مصرف	16	3.1

۱) بله به شرط آنکه بدانیم واریانس‌ها با هم برابرند.
۲) بله، به شرط آنکه افراد نمونه با جایگذاری انتخاب شده باشند.
۳) بله، به شرط آنکه بدانیم توزیع نرمال است.
۴) خیر، چون دو نمونه مستقل نیستند.

حل: گزینه ۴ درست است.

شرط اساسی در آزمون تفاضل میانگین دو جامعه، مستقل بودن نمونه‌هاست. با توجه به اینکه در این سؤال نمونه‌ها به هم وابسته‌اند، نمی‌توانیم از برآوردکننده‌های تفاضل میانگین دو جامعه استفاده کنیم و باید با استفاده از آزمون مقایسات زوجی مسئله را حل کنیم.

آزمون نسبت جامعه

گاهی علاقه‌مندیم درصد یا نسبت خاصی را برای یک صفت یا ویژگی از جامعه در نظر بگیریم؛ برای مثال، آیا درصد بیکاری در کشور حداقل 70% است $(p \geq 70\%)$؟ در این موارد لازم است آزمونی در خصوص نسبت یا درصد انجام شود. مراحل چهارگانه آزمون نسبت یک جامعه به شرح زیر است:

۱- فرض‌های آماری

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases} \qquad \begin{cases} H_0 : p \geq p_0 \\ H_1 : p < p_0 \end{cases} \qquad \begin{cases} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{cases}$$

در فرض‌های بالا p_0 نسبتی است که باید مورد آزمون قرار گیرد.

۲- آماره آزمون (ملاک آزمون)

درصورتی‌که در یک نمونه n تایی از جامعه، x تا صفت یا ویژگی مورد نظر را داشته باشیم، آن‌گاه:

$$\bar{p} = \frac{x}{n} : \text{نسبت موفقیت در نمونه}$$

حال با توجه به آنکه توزیع X (تعداد موفقیت در نمونه) دوجمله‌ای است، برای تعیین توزیع آماره $\bar{p}$ و ملاک آزمون دو حالت به وجود می‌آید:

حالت اول: تعداد نمونه کوچک $(n \leq 30)$

در این وضعیت آماره $\bar{p}$ نیز دارای توزیع دوجمله‌ای است ولی هیچ ملاک استانداردی برای آزمون وجود ندارد.

مثال در آزمون فرضیه مربوط به یک نسبت خاص در جامعه، توزیع نمونه‌گیری $\bar{p}$ در نمونه‌های کوچک چیست؟ (اقتصاد ـ ۸۷)

۱) نمایی ۲) نرمال ۳) دوجمله‌ای ۴) پواسون

حل: گزینه ۳ درست است.

حالت دوم: تعداد نمونه بزرگ $(n > 30)$

در این وضعیت بنا بر قضیه حد مرکزی، آماره $\bar{p}$ دارای توزیع نرمال است و ملاک آزمون برای آزمون این فرضیه که آیا نسبت مورد نظر در جامعه معادل p_0 هست یا خیر، دارای مقدار استاندارد Z به صورت زیر است:

$$\boxed{Z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}}$$

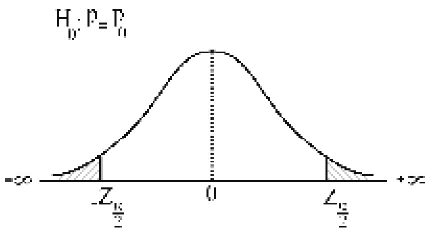
✓ **دقت کنید!**

با فرض آنکه $\bar{p} = \frac{x}{n}$ است، درصورتی‌که صورت و مخرج این ملاک را در n ضرب کنیم، به ملاک جدیدی می‌رسیم که از نظر مقدار، هیچ تفاوتی با آن ندارد:

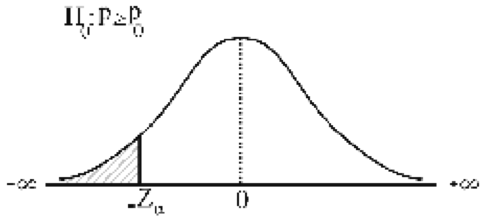
$$Z = \frac{x - np_0}{\sqrt{np_0 q_0}}$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

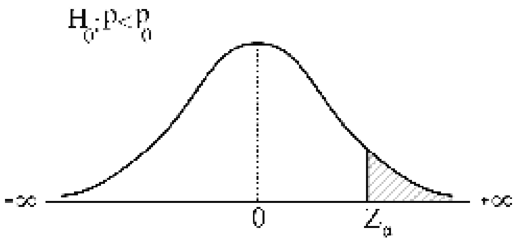
$$\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{U > Z_{\frac{\alpha}{2}} \text{ or } U < -Z_{\frac{\alpha}{2}}}$$



$$\begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{U < -Z_{\alpha}}$$



$$\begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{U > Z_{\alpha}}$$



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال ۱ اگر $\begin{cases} H_0: p = \frac{1}{2} \\ H_1: p \neq \frac{1}{2} \end{cases}$ و از بین 100 آزمایش 59 موفقیت مشاهده شده باشد، آماره آزمون و نتیجه آن در سطح $\alpha = 0.05$ کدام است؟ (اقتصاد ـ ۸۱)

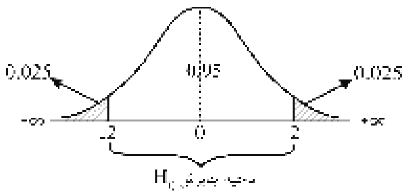
۱) 1.84 و H_0 رد می‌شود.
۲) 1.8 و H_0 رد می‌شود.
۳) 1.84 و H_0 رد نمی‌شود.
۴) 1.8 و H_0 رد نمی‌شود.

حل: گزینه ۴ درست است.

$$(1)\begin{cases} H_0: p = \frac{1}{2} \\ H_1: p \neq \frac{1}{2} \end{cases}$$

$$(2)\begin{cases} Z=\dfrac{\overline{p}-p_0}{\sqrt{\dfrac{p_0q_0}{n}}}=\dfrac{0.59-0.5}{\sqrt{\dfrac{0.5\times 0.5}{100}}}=\dfrac{0.09}{\dfrac{0.5}{10}}=1.8 \\ n=100\ ,\ x=59 \rightarrow \overline{p}=\dfrac{x}{n}=\dfrac{59}{100}=0.59 \\ \alpha=0.05 \rightarrow Z_{0.025}\simeq 2 \end{cases}$$

(3)



(4) چون مقدار عددی آماره آزمون در ناحیه اطمینان (پذیرش H_0) قرار می‌گیرد، فرض H_0 رد نمی‌شود.

مثال ۲ صاحب یک کارخانه ادعا می‌کند حداقل 80% مردم، محصول کارخانه او را ترجیح می‌دهند. در یک نمونه تصادفی 100تایی، بیش از چند نفر باید کالای وی را ترجیح دهند تا در سطح $\alpha = 0.05$ ادعای صاحب کارخانه پذیرفته شود؟

$(Z_{0.05}=1.65)$ (اقتصاد ـ ۸۵)

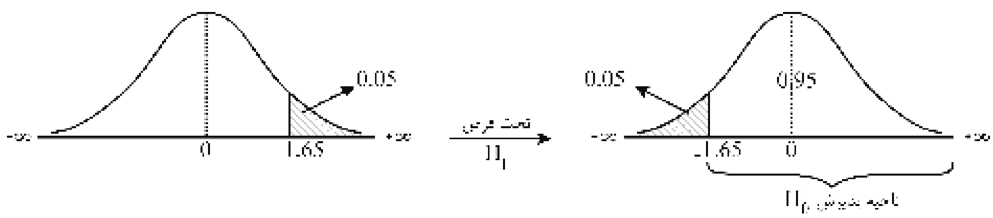
۱) 64 ۲) 73 ۳) 79 ۴) 86

حل: گزینه ۲ درست است.

$$(1)\begin{cases} H_0: p \geq 0.8 \\ H_1: p < 0.8 \end{cases}$$

$$(2)\begin{cases} Z=\dfrac{\overline{p}-p_0}{\sqrt{\dfrac{p_0q_0}{n}}}=\dfrac{\dfrac{x}{n}-p_0}{\sqrt{\dfrac{p_0q_0}{n}}}=\dfrac{x-np_0}{\sqrt{np_0q_0}}=\dfrac{x-100\times 0.8}{\sqrt{100\times 0.8\times 0.2}}=\dfrac{x-80}{4} \\ \overline{p}=\dfrac{x}{n}, n=100 \end{cases}$$

$$(3)\ \alpha=0.05\ , Z_{0.05}=1.65 \rightarrow$$



با توجه به شکل برای پذیرفتن ادعا یعنی همان فرض H_0:

$$(4)\ \frac{x-80}{4} \geq -1.65 \rightarrow x \geq 73.4 \quad \rightarrow \quad x > 73$$

باید بیش از 73 نفر کالای او را ترجیح دهند تا ادعا پذیرفته شود.

آزمون مقایسه نسبت دو جامعه

درصورتی‌که مقایسه دو جامعه آماری با استفاده از داده‌های کیفی انجام شود، آزمون تفاضل نسبت در دو جامعه می‌تواند به صورت زیر مطرح شود. p_1 و p_2 به ترتیب نسبت‌های موفقیت در دو جامعه می‌باشند.

۱- فرض‌های آماری

$$\begin{cases}H_0: p_1 = p_2\\ H_1: p_1 \neq p_2\end{cases} \qquad \begin{cases}H_0: p_1 \leq p_2\\ H_1: p_1 > p_2\end{cases} \qquad \begin{cases}H_0: p_1 \geq p_2\\ H_1: p_1 < p_2\end{cases}$$

۲- آماره آزمون (ملاک آزمون)

هرگاه دو نمونه مستقل n_1 و n_2 از دو جامعه با نسبت‌های p_1 و p_2 انتخاب شوند و به ترتیب x_1 و x_2 موفقیت در هر یک از نمونه‌ها دیده شود، $\bar{p}_1 = \frac{x_1}{n_1}$ و $\bar{p}_2 = \frac{x_2}{n_2}$ به ترتیب نسبت‌های نمونه هستند. ملاک آزمون در این وضعیت بر اساس تابع محوری برآورد به شرح زیر از **توزیع** Z تبعیت می‌کند:

$$\text{تابع محوری}: Z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}}} \xrightarrow{H_0: p_1 = p_2} Z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}}}$$

ازآنجاکه ملاک آزمون بر اساس توزیع برآورد و با پذیرش فرض H_0 ساخته می‌شود، می‌توانیم فرض کنیم نسبت‌ها با هم مساوی‌اند؛ درنتیجه چنین تصور کنیم که دو نمونه گرفته‌شده متعلق به یک جامعه هستند در این صورت $(x_1 + x_2)$ تعداد موفقیت‌ها در مجموع دو نمونه $(n_1 + n_2)$ خواهند بود تا $\bar{p}$ (برآوردی از نسبت مشترک در دو جامعه) به صورت زیر حاصل شود:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

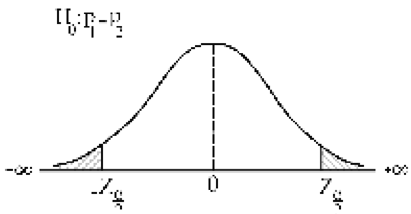
در این وضعیت آماره آزمون به صورت زیر است:

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

درنهایت هر دو رابطه زیر برای ملاک آزمون برابری نسبت دو جامعه می‌توانند استفاده شوند و از نظر مقدار تفاوتی بین آن‌ها وجود ندارد:

$$Z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}}} = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

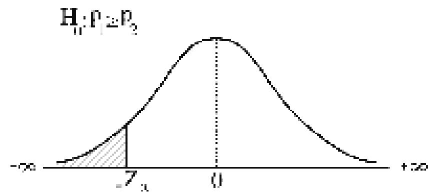
۳- تعیین سطح H_0 و H_1 و مقدار بحرانی



$$\begin{cases}H_0: p_1 = p_2\\ H_1: p_1 \neq p_2\end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{U > Z_{\frac{\alpha}{2}} \text{ or } U < -Z_{\frac{\alpha}{2}}}$$

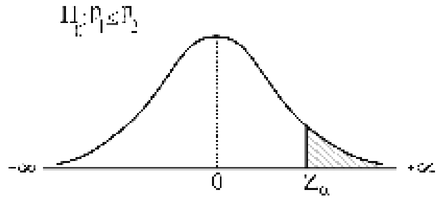
$$\begin{cases} H_0: p_1 \geq p_2 \\ H_1: p_1 < p_2 \end{cases} \xrightarrow{U<-Z_\alpha}$$

فرض H_0 رد می‌شود.



$$\begin{cases} H_0: p_1 \leq p_2 \\ H_1: p_1 > p_2 \end{cases} \xrightarrow{U>Z_\alpha}$$

فرض H_0 رد می‌شود.



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال ۱ اگر $n_1 = 120$، $n_2 = 100$، $\bar{p}_1 = 0.60$ و $\bar{p}_2 = 0.5$ باشد، آماره آزمون $H_0: p_1 \leq p_2$ کدام است؟ (مدیریت ـ ۸۳)

۱) -2.1 ۲) -1.347 ۳) 2.575 ۴) 1.491

حل: گزینه ۴ درست است.

$$Z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}} = \frac{0.6 - 0.5}{\sqrt{\frac{0.6 \times 0.4}{120} + \frac{0.5 \times 0.5}{100}}} = 1.491$$

مثال ۲ برای بررسی عدم تفاوت در نسبت طرفداران نظریه‌های کینزی بین دانشجویان کارشناسی و دانشجویان کارشناسی ارشد رشته اقتصاد، دو نمونه تصادفی مستقل به حجم‌های 64 و 36 از هر یک انتخاب گردیده است. تعداد طرفداران در هر کدام از نمونه‌ها به ترتیب 14 و 6 نفر بوده است. اگر $\alpha = 0.05$ باشد، کمیت آماره آزمون و نتیجه آزمون چیست؟ (اقتصاد ـ ۸۴)

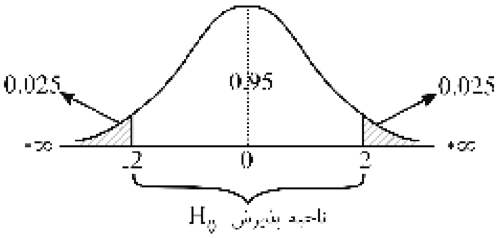
۱) 2.06 و H_0 رد می‌شود. ۲) 1.8 و H_0 را نمی‌توان رد کرد.

۳) 1.78 و H_0 رد می‌شود. ۴) 0.6 و H_0 را نمی‌توان رد کرد.

حل: گزینه ۴ درست است.

$$(1)\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$$

$$(2)\begin{cases} Z = \dfrac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{(0.22 - 0.17)}{\sqrt{0.2 \times 0.8\left(\frac{1}{64} + \frac{1}{36}\right)}} = \dfrac{0.05}{0.4\frac{10}{6 \times 8}} = 0.6 \\ \bar{p}_1 = \dfrac{x_1}{n_1} = \dfrac{14}{64} \simeq 0.22 \ , \ \bar{p}_2 = \dfrac{x_2}{n_2} = \dfrac{6}{36} \simeq 0.17 \ \to \ \bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2} = \dfrac{14+6}{64+36} = 0.2 \\ \alpha = 0.05 \ \to \ Z_{0.025} \simeq 2 \end{cases}$$

(3)


(4) چون مقدار آماره آزمون به‌دست‌آمده در ناحیه اطمینان (پذیرش H_0) قرار دارد، فرض H_0 را نمی‌توان رد کرد.

مثال ۳ ادعا شده است که در شهر (الف) افرادی که از فروشگاه‌های زنجیره‌ای خرید می‌کنند، بیشتر از افراد خریدکننده از فروشگاه‌های زنجیره‌ای در شهر (ب) هستند. آماره آزمون این فرضیه کدام است؟ (اقتصاد ـ ۸۱)

۱) $\dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{pq\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ ۲) $\dfrac{n_1-n_2}{\sqrt{\dfrac{p_1q_1}{n_1}+\dfrac{p_2q_2}{n_2}}}$ ۳) $\dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{S_p^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ ۴) $\dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p(1-p)}{n}}}$

حل: گزینه ۱ درست است.

آزمون واریانس جامعه

در بعضی موارد می‌خواهیم بدانیم آیا پراکندگی جامعه دارای حدود قابل قبولی هست یا خیر؛ برای مثال، آیا ریسک شرکت‌های پذیرفته‌شده در بورس برابر 1000 است $(\sigma=1000)$؟

در این موارد لازم است آزمونی برای واریانس یا انحراف معیار انجام شود. مراحل چهارگانه آزمون واریانس جامعه به شرح زیر است:

۱- فرض‌های آماری

$$\begin{cases}H_0:\sigma_X^2=\sigma_0^2\\H_1:\sigma_X^2\neq\sigma_0^2\end{cases}\qquad\begin{cases}H_0:\sigma_X^2\geq\sigma_0^2\\H_1:\sigma_X^2<\sigma_0^2\end{cases}\qquad\begin{cases}H_0:\sigma_X^2\leq\sigma_0^2\\H_1:\sigma_X^2>\sigma_0^2\end{cases}$$

۲- آماره آزمون (ملاک آزمون)

درصورتی‌که نمونه مستقل n تایی از جامعه‌ای نرمال انتخاب شود، برای تعیین توزیع و ملاک آزمون برای این فرضیه که آیا واریانس جامعه معادل مقدار σ_0^2 هست یا خیر، دو حالت به وجود می‌آید:

حالت اول: میانگین جامعه (μ) نامعلوم (پیش فرض)

در این وضعیت واریانس نمونه از رابطه $S^2=\dfrac{\sum(x-\bar{x})^2}{n-1}$ به دست می‌آید و آماره $\dfrac{(n-1)S^2}{\sigma^2}$ دارای توزیع χ^2 با $n-1$ درجه آزادی است، درنتیجه ملاک آزمون به صورت زیر دارای مقدار استاندارد χ^2_{n-1} است:

$$\chi^2_{(n-1)}=\frac{(n-1)S_X^2}{\sigma_0^2}$$

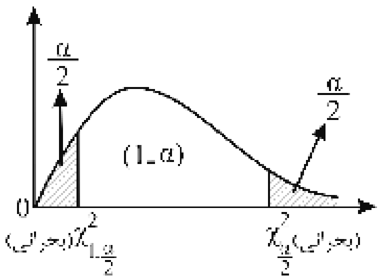
حالت دوم: میانگین جامعه (μ) معلوم

در این وضعیت واریانس نمونه از رابطه دقیق‌تری به صورت $S^2 = \frac{\sum (x-\mu)^2}{n}$ به دست می‌آید و آماره $\frac{nS^2}{\sigma^2}$ دارای توزیع χ^2 با n درجه آزادی است، درنتیجه ملاک آزمون به صورت زیر دارای مقدار استاندارد χ^2_n است:

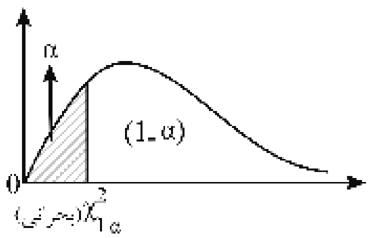
$$\boxed{\chi^2_{(n)} = \frac{nS^2}{\sigma_0^2}}$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

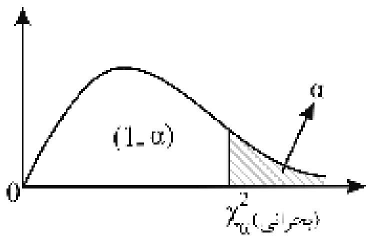
$$\begin{cases} H_0 : \sigma_X^2 = \sigma_0^2 \\ H_1 : \sigma_X^2 \neq \sigma_0^2 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{\chi^2_{n-1} < \chi^2_{1-\frac{\alpha}{2}} \text{ or } \chi^2_{n-1} > \chi^2_{\frac{\alpha}{2}}}$$



$$\begin{cases} H_0 : \sigma_X^2 \geq \sigma_0^2 \\ H_1 : \sigma_X^2 < \sigma_0^2 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{\chi^2_{n-1} < \chi^2_{1-\alpha}}$$



$$\begin{cases} H_0 : \sigma_X^2 \leq \sigma_0^2 \\ H_1 : \sigma_X^2 > \sigma_0^2 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{\chi^2_{n-1} > \chi^2_{\alpha}}$$



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال ۱ برای $n = 14$ و $S^2 = 75$ فرضیه $H_0 : \sigma^2 \geq 100$ را در مقابل فرضیه $H_1 : \sigma^2 < 100$ در سطح $\alpha = 0.01$ آزمون می‌کنیم. درصورتی‌که فرض شود جامعه‌ای که نمونه تصادفی از آن انتخاب می‌شود نزدیک به نرمال باشد، کدام‌یک از موارد درست است؟ (کمیت بحرانی از جدول 4.107) (اقتصاد ـ ۸۳)

۱) $\chi^2 = 0.75$ و درنتیجه فرضیه H_0 رد نمی‌شود.

۲) $\chi^2 = 9.75$ و درنتیجه فرضیه H_0 رد می‌شود.

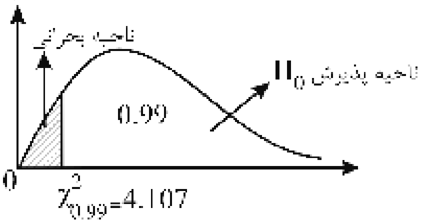
۳) $\chi^2 = 9.75$ و درنتیجه فرضیه H_0 رد نمی‌شود.

۴) $F = 0.75$ و درنتیجه فرضیه H_0 رد می‌شود.

حل: گزینه ۳ درست است.

$$(1)\begin{cases} H_0 : \sigma^2 \geq 100 \\ H_1 : \sigma^2 < 100 \end{cases} \qquad (2)\ \chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma_0^2} \rightarrow \chi^2_{13} = \frac{13 \times 75}{100} = 9.75$$

(3)



(4) چون مقدار 9.75 در ناحیه پذیرش H_0 قرار می‌گیرد، فرض H_0 پذیرفته می‌شود.

با توجه به اینکه مقادیر χ^2 همواره مثبت است و از طرفی در مثال بالا χ^2 به دست آمده (9.75) بعد از نقطه بحرانی (4.107) قرار گرفته و درنتیجه در ناحیه اطمینان یا همان پذیرش H_0 است.

مثال ۲ در یک جامعه نرمال با میانگین معلوم μ می‌خواهیم فرضیه $H_0 : \sigma^2 = 0.05$ را در مقابل $H_1 : \sigma^2 \neq 0.05$ بر اساس اطلاعات $n = 19$ و $S = \sqrt{\frac{\sum (x_i - \mu)^2}{n}} = 0.3$ آزمون کنیم. مقدار عددی آماره آزمون عبارت است از: (اقتصاد ـ ۸۴)

۱) 2.1 ۲) 2.3 ۳) 32.4 ۴) 34.2

حل: گزینه ۴ درست است.

$$(1)\begin{cases} H_0 : \sigma^2 = 0.05 \\ H_1 : \sigma^2 \neq 0.05 \end{cases}$$

با توجه به اینکه μ جامعه معلوم است، آماره آزمون $\chi^2_{(n)} = \frac{nS^2}{\sigma_0^2}$ است:

$$(2)\begin{cases} \chi^2_{(n)} = \frac{nS^2}{\sigma_0^2} \rightarrow \chi^2_{(19)} = \frac{19 \times (0.3)^2}{0.05} = 34.2 \\ n = 19, S^2 = (0.3)^2, \sigma_0^2 = 0.05 \end{cases}$$

آزمون مقایسه واریانس دو جامعه

در بسیاری از موارد می‌خواهیم بدانیم آیا واریانس دو جامعه برابر است یا خیر. برای این کار همواره می‌توانیم نسبت واریانس دو جامعه را مورد توجه قرار دهیم چرا که اگر فرض کنیم هر دو جامعه نرمال هستند، آن‌گاه وقتی واریانس‌های دو جامعه برابر باشد نسبت $\frac{S_1^2}{S_2^2}$ دارای توزیع فیشر (F) خواهد بود که ملاک مناسبی برای آزمون است.

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \xrightarrow{\sigma_1^2=\sigma_2^2} F=\frac{S_1^2}{S_2^2}$$

مراحل چهارگانه آزمون فرض برای واریانس دو جامعه به شرح زیر است:

۱- فرض‌های آماری

$\begin{cases} H_0:\sigma_1^2\le\sigma_2^2 \quad \text{یا} \quad \frac{\sigma_1^2}{\sigma_2^2}\le 1 \\ H_1:\sigma_1^2>\sigma_2^2 \quad \text{یا} \quad \frac{\sigma_1^2}{\sigma_2^2}>1 \end{cases}$	$\begin{cases} H_0:\sigma_1^2\ge\sigma_2^2 \quad \text{یا} \quad \frac{\sigma_1^2}{\sigma_2^2}\ge 1 \\ H_1:\sigma_1^2<\sigma_2^2 \quad \text{یا} \quad \frac{\sigma_1^2}{\sigma_2^2}<1 \end{cases}$	$\begin{cases} H_0:\sigma_1^2=\sigma_2^2 \quad \text{یا} \quad \frac{\sigma_1^2}{\sigma_2^2}=1 \\ H_1:\sigma_1^2\ne\sigma_2^2 \quad \text{یا} \quad \frac{\sigma_1^2}{\sigma_2^2}\ne 1 \end{cases}$

۲- آماره آزمون (ملاک آزمون)

اگر نمونه‌های مستقل n_1 و n_2 از دو جامعه نرمال انتخاب شده باشند، برای تعیین توزیع آماره و ملاک آزمون برای برابری واریانس دو جامعه، دو حالت به وجود می‌آید:

حالت اول: میانگین جوامع (μ_1 و μ_2) نامعلوم (پیش‌فرض)

در این حالت با فرض برابری واریانس دو جامعه، آماره S_1^2/S_2^2 دارای توزیع فیشر (F) با n_1-1 و n_2-1 درجه آزادی و ملاک آزمون به صورت زیر است:

$$\boxed{F_{n_1-1,n_2-1}=\frac{S_1^2}{S_2^2}}$$

✓ **دقت کنید!**

در این حالت واریانس‌های نمونه در دو جامعه از روابط $S_1^2=\frac{\sum(x-\bar{x}_1)^2}{n_1-1}$ و $S_2^2=\frac{\sum(x-\bar{x}_2)^2}{n_2-1}$ محاسبه می‌شوند.

حالت دوم: میانگین جوامع (μ_1 و μ_2) معلوم

در این وضعیت با فرض برابری واریانس دو جامعه، آماره S_1^2/S_2^2 دارای توزیع فیشر (F) با n_1 و n_2 درجه آزادی و ملاک آزمون به صورت زیر است:

$$\boxed{F_{n_1,n_2}=\frac{S_1^2}{S_2^2}}$$

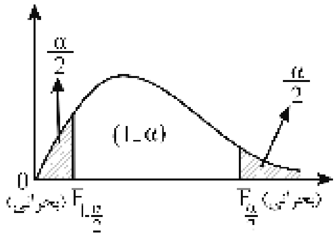
✓ **دقت کنید!**

در این حالت واریانس‌های نمونه در دو جامعه از روابط دقیق‌تری به صورت $S_1^2=\dfrac{\sum(x-\mu_1)^2}{n_1}$ و $S_2^2=\dfrac{\sum(x-\mu_2)^2}{n_2}$ محاسبه می‌شوند.

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

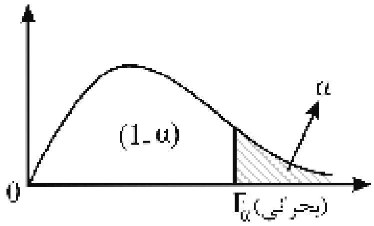
$$\begin{cases}H_0:\sigma_1^2=\sigma_2^2\\ H_1:\sigma_1^2\neq\sigma_2^2\end{cases}\xrightarrow{F_{n_1-1,n_2-1}>F_{\frac{\alpha}{2}}\ \text{or}\ F_{n_1-1,n_2-1}<F_{1-\frac{\alpha}{2}}}$$

فرض H_0 رد می‌شود.



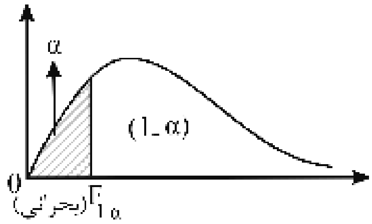
$$\begin{cases}H_0:\sigma_1^2\le\sigma_2^2\\ H_1:\sigma_1^2>\sigma_2^2\end{cases}\xrightarrow{F_{n_1-1,\,n_2-1}>F_{\alpha}}$$

فرض H_0 رد می‌شود.



$$\begin{cases}H_0:\sigma_1^2\ge\sigma_2^2\\ H_1:\sigma_1^2<\sigma_2^2\end{cases}\xrightarrow{F_{n_1-1,\,n_2-1}<F_{1-\alpha}}$$

فرض H_0 رد می‌شود.



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

یادآوری: در توزیع F با m و n درجه آزادی رابطه زیر برقرار است:

$$\boxed{F_{1-\alpha,m,n}=\frac{1}{F_{\alpha,n,m}}}$$

مثال اطلاعات زیر مربوط به دو شرکت «الف» و «ب» است، با فرض نرمال بودن توزیع X در دو شرکت، مقدار آماره آزمون برای بررسی فرضیه «تساوی واریانس X در دو شرکت» کدام است؟

شرکت «الف»	$n_1=10$	$\bar{x}_1=100$	$S_1=8$
شرکت «ب»	$n_2=15$	$\bar{x}_2=120$	$S_2=10$

۱) 0.80
۲) 0.41
۳) 0.64
۴) 0.53

حل: گزینه ۳ درست است.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

$$F_{9,14} = \frac{S_1^2}{S_2^2} = \frac{8^2}{10^2} = 0.64$$

بررسی آزمون با استفاده از فاصله اطمینان

یکی از روش‌های تحلیل آزمون فرضیه‌های آماری، بررسی فرضیه‌های آزمون با استفاده از فاصله اطمینان است. به این صورت که ابتدا فاصله اطمینان پارامتر دلخواه با استفاده از داده‌های نمونه به دست می‌آید، سپس با توجه به فرض آزمون یکی از تحلیل‌های زیر انجام می‌شود:

۱) آزمون میانگین جامعه

مثال رد H_0	مثال پذیرش H_0	فاصله رد H_0	فاصله پذیرش H_0	فرضیه
$\mu_0 = 4$ $(-4,3)$ $(0,2)$ $(1,3)$	$\mu_0 = 4$ $(1,5)$ $(4,6)$ $(0,4)$	باید فاقد μ_0 باشد.	باید شامل μ_0 باشد.	$H_0 : \mu = \mu_0$ $H_0 : \mu \le \mu_0$ $H_0 : \mu \ge \mu_0$

۲) آزمون برابری میانگین دو جامعه $\left(\mu_1 - \mu_2 \overset{\ge}{\underset{\le}{=}} 0\right)$

مثال رد H_0	مثال پذیرش H_0	فاصله رد H_0	فاصله پذیرش H_0	فرضیه
$(2\ ,\ 3.5)$ $(1,\ 3)$ $(-4, -1)$	$(-1, 0.5)$ $(0,\ 5)$ $(-2,\ 0)$	باید فاقد صفر باشد.	باید شامل صفر باشد.	$H_0 : \mu_1 = \mu_2$ $H_0 : \mu_1 \le \mu_2$ $H_0 : \mu_1 \ge \mu_2$

۳) آزمون نسبت جامعه

مثال رد H_0	مثال پذیرش H_0	فاصله رد H_0	فاصله پذیرش H_0	فرضیه
$p_0 = 0.5$ $(-1\ ,\ 0.25)$ $(-0.6\ ,\ 0)$ $(0.8\ ,\ 0.9)$	$p_0 = 0.5$ $(0\ , 0.6)$ $(-0.6,\ 0.5)$ $(0.5,\ 0.8)$	باید فاقد p_0 باشد.	باید شامل p_0 باشد.	$H_0 : p = p_0$ $H_0 : p \le p_0$ $H_0 : p \ge p_0$

۴) آزمون برابری نسبت دو جامعه $\left(p_1 - p_2 \overset{\ge}{\underset{\le}{=}} 0\right)$

مثال رد H_0	مثال پذیرش H_0	فاصله رد H_0	فاصله پذیرش H_0	فرضیه
$(0.3\ ,\ 0.8)$ $(0.2\ ,\ 0.5)$ $(-0.75\ ,\ -0.1)$	$(-0.5\ ,\ 0.5)$ $(0,\ 0.8)$ $(-0.75\ ,\ 0)$	باید فاقد صفر باشد.	باید شامل صفر باشد.	$H_0 : p_1 = p_2$ $H_0 : p_1 \le p_2$ $H_0 : p_1 \ge p_2$

۵) آزمون واریانس جامعه

فرضیه	فاصله پذیرش H_0	فاصله رد H_0	مثال پذیرش H_0	مثال رد H_0
$H_0: \sigma^2 = \sigma_0^2$ $H_0: \sigma^2 \le \sigma_0^2$ $H_0: \sigma^2 \ge \sigma_0^2$	باید شامل σ_0^2 باشد.	باید فاقد σ_0^2 باشد.	$\sigma_0^2 = 2$ $(1, 8)$ $(2, 16)$ $(0.5, 2)$	$\sigma_0^2 = 2$ $(4, 9)$ $(0.5, 1.6)$ $(3, 16)$

۶) آزمون برابری واریانس دو جامعه $\left(\dfrac{\sigma_1^2}{\sigma_2^2} \begin{smallmatrix}\ge\\=\\\le\end{smallmatrix} 1\right)$

فرضیه	فاصله پذیرش H_0	فاصله رد H_0	مثال پذیرش H_0	مثال رد H_0
$H_0: \sigma_1^2 = \sigma_2^2$ $H_0: \sigma_1^2 \le \sigma_2^2$ $H_0: \sigma_1^2 \ge \sigma_2^2$	باید شامل 1 باشد.	باید فاقد 1 باشد.	$(0, 3)$ $(1, 4)$ $(0.5, 2)$	$(2,8)$ $(0, 0.8)$ $(4, 16)$

مثال ۱ اگر حدود $(-4, 1.5)$ یک فاصله اطمینان 95% برای پارامتر μ باشد:

۱) فرضیه $\mu > 1$ را نمی‌توانیم رد کنیم.
۲) فرضیه $\mu = 1$ را نمی‌توانیم بپذیریم.
۳) فرضیه $\mu \le 1$ را نمی‌توانیم رد کنیم.
۴) فرضیه $\mu \ge 1$ را نمی‌توانیم بپذیریم.

حل: گزینه ۳ درست است.

با توجه به حالت (۱)، چون فاصله اطمینان داده‌شده شامل نقطه $\mu = 1$ است، فرضیه‌های $\mu = 1$، $\mu \le 1$ و $\mu \ge 1$ پذیرفته می شوند و فرضیه‌های $\mu \ne 1$، $\mu > 1$ و $\mu < 1$ پذیرفته نمی‌شوند.

مثال ۲ می‌خواهیم فرض $H_0: \mu_1 = \mu_2$ را در مقابل فرض $H_1: \mu_1 \ne \mu_2$ آزمون کنیم. بر اساس کدام‌یک از فاصله‌های اطمینان 0.95 مربوط به $\mu_1 - \mu_2$، می‌توان فرض H_0 را با $\alpha = 0.05$ رد کرد؟ (اقتصاد ـ ۸۰)

۱) $(0, 15)$ ۲) $(-4, 1.2)$ ۳) $(-6, 6)$ ۴) $(2.5, 3.8)$

حل: گزینه ۴ درست است.

با توجه به حالت (۲)، فاصله اطمینان برای عدم پذیرش H_0 باید فاقد نقطه صفر باشد که تنها فاصله اطمینان گزینه (۴) چنین است.

مثال ۳ در فرضیه $\begin{cases} H_0: p = 0.5 \\ H_1: p \ne 0.5 \end{cases}$ کدام گزینه ناحیه پذیرش H_0 است؟

۱) $(0.6, 1)$ ۲) $(-0.2, 1)$ ۳) $(0.1, 0.45)$ ۴) $(-0.4, 0.1)$

حل: گزینه ۲ درست است.

با توجه به حالت (۳)، فاصله اطمینان برای پذیرش H_0 باید شامل نقطه $p_0 = 0.5$ باشد و تنها فاصله اطمینان گزینه ۲ شامل نقطه $p_0 = 0.5$ است.

مثال ۴ یک فاصله اطمینان 95% برای پذیرش فرضیه $p_1 > p_2$ کدام است؟

۱) $\left(-\frac{1}{2},\frac{1}{4}\right)$ ۲) $\left(0,\frac{3}{4}\right)$ ۳) $\left(-1,\frac{1}{2}\right)$ ۴) $\left(-\frac{1}{2},-\frac{1}{6}\right)$

حل: گزینه ۴ درست است.

با توجه به حالت (۴)، فاصله اطمینان برای عدم پذیرش فرضیه $p_1 \le p_2$ و یا به عبارت دیگر پذیرش فرضیه $p_1 > p_2$، باید فاقد نقطه صفر باشد، یعنی یا کاملاً در ناحیه مثبت باشد و یا کاملاً در ناحیه منفی باشد؛ فاصله اطمینان گزینه ۴ تماماً در ناحیه منفی است.

مثال ۵ اگر $P\left(2.53 < \frac{\sigma_1^2}{\sigma_2^2} < 4.15\right) = 0.99$ یک فاصله اطمینان در سطح خطای $\alpha = 0.01$ برای نسبت واریانس دو جامعه نرمال باشد:

۱) فرضیه $\sigma_1^2 \ge \sigma_2^2$ پذیرفته می‌شود. ۲) فرضیه $\sigma_1^2 = \sigma_2^2$ را نمی‌توان رد کرد.

۳) فرضیه $\sigma_1^2 < \sigma_2^2$ را نمی‌توان پذیرفت. ۴) فرضیه $\sigma_1^2 \le \sigma_2^2$ را نمی‌توان پذیرفت.

حل: گزینه ۴ درست است.

با توجه به حالت (۶)، فاصله اطمینان برای پذیرش فرضیه‌های $\sigma_1^2 \le \sigma_2^2$، $\sigma_1^2 \ge \sigma_2^2$ و $\sigma_1^2 = \sigma_2^2$ باید شامل نقطه یک باشد. فاصله اطمینان داده‌شده فاقد نقطه یک است، بنابراین در سطح اطمینان 99% نمی‌توان این فرضیه‌ها را پذیرفت و نمی‌توان فرضیه‌های $\sigma_1^2 \ne \sigma_2^2$، $\sigma_1^2 > \sigma_2^2$ و $\sigma_1^2 < \sigma_2^2$ را رد کرد.

تأثیر اندازه سطوح در تصمیم‌گیری

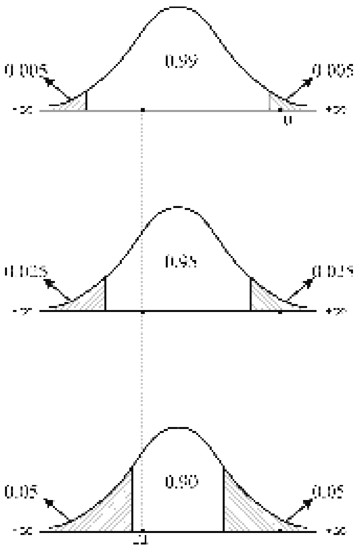
قضیه ۱:

الف) هرگاه فرضیه H_0 در سطح اطمینان کوچک‌تر پذیرفته شود، آن‌گاه حتماً در سطح اطمینان بزرگ‌تر هم پذیرفته می‌شود.

ب) هرگاه فرضیه H_0 در سطح اطمینان بزرگ‌تر رد شود (معنی‌دار شود)، آن‌گاه حتماً در سطح اطمینان کوچک‌تر هم رد می‌شود.

تبصره: عکس این قضیه ممکن است درست باشد یا نباشد، بنابراین در مورد بقیه حالات نمی‌توان بحث کرد.

نمایش قضیه بالا در توزیع نرمال به شکل روبه‌رو است:



مثال ۱ اگر فرضیه‌ای در سطح معنی‌دار $\alpha = \%10$ رد نشود، در مورد فرضیه مذکور در سطح معنی‌دار $\alpha = \%5$ می‌توان گفت:

۱) حتماً رد می‌شود.
۲) حتماً رد نمی‌شود.
۳) نمی‌توان اظهار نظر کرد.
۴) بستگی به حجم نمونه دارد.

حل: گزینه ۲ درست است.

با توجه به حالت (الف) از قضیه۱، آزمون در سطح اطمینان $1-\alpha = \%90$ پذیرفته شده است، بنابراین در سطح اطمینان بزرگ‌تر یعنی $1-\alpha = \%95$ نیز پذیرفته می‌شود.

مثال ۲ کدام‌یک از موارد زیر درباره آزمون فرض درست نیست ؟

۱) اگر H_0 در سطح احتمال $\%5$ رد شود در سطح احتمال $\%1$ نیز رد می‌شود.
۲) اگر H_0 در سطح احتمال $\%5$ رد شود ممکن است در سطح $\%1$ نیز رد شود.
۳) اگر H_0 در سطح احتمال $\%1$ رد شود ممکن نیست در سطح احتمال $\%5$ نیز رد شود.
۴) اگر H_0 در سطح احتمال $\%1$ رد شود در سطح احتمال $\%5$ حتماً رد می‌شود.

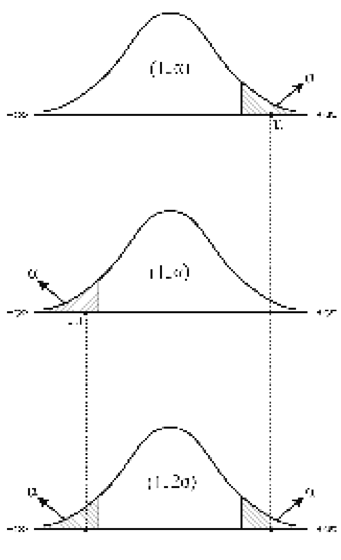
حل: گزینه ۳ جواب است.

با توجه به حالت (ب) از قضیه۱، اگر آزمون در سطح اطمینان $1-\alpha = \%99$ رد شود، حتماً در سطح اطمینان کوچک‌تر یعنی $1-\alpha = \%95$ نیز رد می‌شود، بنابراین گزینه ۴ حتماً درست و گزینه ۳ حتماً نادرست است. درعین‌حال با توجه به تبصره مطرح‌شده، گزینه ۲ درست است و گزینه ۱ ممکن است درست باشد.

قضیه ۲: هرگاه فرضیه H_0 یک‌دامنه در سطح خطای α رد شود (معنی‌دار شود)، آن‌گاه فرضیه دودامنه آن در سطح خطای 2α نیز رد می شود.

تبصره: عکس این قضیه ممکن است درست باشد یا نباشد، بنابراین در مورد بقیه حالات نمی‌توان بحث کرد.

نمایش قضیه بالا در توزیع نرمال به شکل زیر است :

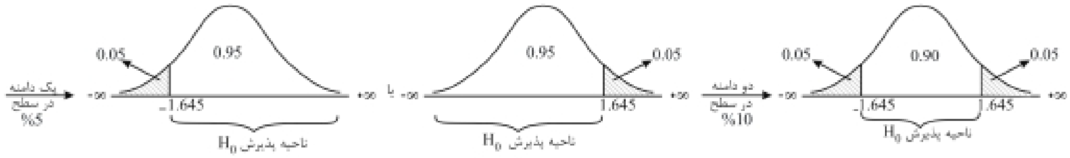


مثال کمیت آماره آزمون Z (تابع نمونه‌ای Z) که بر اساس یک‌دامنه در سطح $\%5$ معنی‌دار است: (اقتصاد ـ ۷۷)

۱) به احتمال $\%95$ متمایز از صفر است.
۲) با افزایش حجم نمونه، سطح معنی‌دار بودن آن افزایش خواهد یافت.
۳) بر اساس یک آزمون دودامنه در سطح $\%10$ معنی‌دار است.
۴) می‌باید بزرگ‌تر از 2 یا کوچک‌تر از -2 باشد.

حل: گزینه ۳ جواب است.

با توجه به قضیه۲، اگر آزمون در سطح $\alpha = \%5$ رد شود (معنی‌دار باشد)، آن‌گاه حتماً آزمون دودامنه آن در سطح $2\alpha = \%10$ رد می‌شود. برای درک بیشتر به شکل زیر توجه کنید:



تأثیر نتایج نمونه در تصمیم‌گیری

درصورتی‌که چند نمونه برای بررسی آزمون فرض جامعه داشته باشیم، پس از محاسبه ملاک (آماره) آزمون با توجه به وضعیت‌های مختلف فرض‌های آماری، می‌توانیم تحلیل زیر را انجام دهیم:

I) فرض $\begin{cases} H_0: \theta \le \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ (دامنه به راست)

هرچه مقدار آماره (ملاک) آزمون بزرگ‌تر باشد، احتمال رد شدن فرض H_0 و هرچه این مقدار کوچک‌تر باشد، احتمال پذیرفتن فرض H_0 بیشتر می‌شود.

II) فرض $\begin{cases} H_0: \theta \ge \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ (دامنه به چپ)

هرچه مقدار آماره (ملاک) آزمون کوچک‌تر باشد، احتمال رد شدن فرض H_0 و هرچه این مقدار بزرگ‌تر باشد، احتمال پذیرفتن فرض H_0 بیشتر می‌شود.

III) فرض $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \ne \theta_0 \end{cases}$ (دودامنه)

هرچه مقدار آماره (ملاک) آزمون از صفر دورتر باشد، احتمال رد شدن فرض H_0 و هرچه این مقدار به صفر نزدیک‌تر باشد، احتمال پذیرفتن فرض H_0 بیشتر می‌شود.

مثال برای آزمون $H_0: \mu \ge 30$ و $H_1: \mu < 30$ کدام‌یک از گزینه‌های زیر بر مبنای حجم نمونه 36 قوی‌ترین شواهد را برای رد H_0 ارائه می‌دهد؟

۱) $\bar{x} = 28\ ,\ S = 6$ ۲) $\bar{x} = 32\ ,\ S = 2$ ۳) $\bar{x} = 27\ ,\ S = 4$ ۴) $\bar{x} = 26\ ,\ S = 9$

حل: گزینه ۳ درست است.

با محاسبه آماره آزمون برای تمام گزینه‌ها، با توجه به فرض H_1، آماره‌ای احتمال رد را بیشتر می‌کند که مقدار کمتری داشته باشد:

$$Z = \frac{\bar{X} - \cancel{\mu_0}^{30}}{\frac{S}{\sqrt{n}}}$$

گزینه ۱ : $Z = \frac{28-30}{\frac{6}{\sqrt{36}}} = -2$ گزینه ۲ : $Z = \frac{32-30}{\frac{2}{\sqrt{36}}} = 6$

گزینه ۳ : $Z = \frac{27-30}{\frac{4}{\sqrt{36}}} = -4.5$ گزینه ۴ : $Z = \frac{26-30}{\frac{9}{\sqrt{36}}} = -2.67$

محاسبه خطاهای آماری

محاسبه خطای نوع اول (α)

درصورتی‌که یک آزمون در سطح اطمینان $1-\alpha$ انجام شود، α احتمال خطای نوع اول است که همان سطح بحرانی یا سطح معنی‌دار (ناحیه رد H_0) است.

به عبارت ریاضی می‌توان گفت:

$\alpha = P($ در ناحیه بحرانی باشد $X \mid H_0$ برقراری فرض$)$

مثال ۱ با آزمون $H_0: \mu = \mu_0$ در سطح 95% احتمال خطای نوع اول کدام است؟

۱) 5%　　۲) 95%　　۳) 100%　　۴) 2.5%

حل: گزینه ۱ درست است.

محاسبه خطای نوع دوم

همان‌طورکه می‌دانیم، خطای نوع دوم عبارت است از احتمال شرطی «پذیرش فرض H_0 زمانی که H_0 نادرست است»، بنابراین برای محاسبه آن به مقدار واقعی پارامتر جامعه نیاز داریم، زیرا فرض H_0 تنها یک مقدار پیشنهادی برای پارامتر جامعه است و فقط با داشتن مقدار واقعی پارامتر می‌توان فهمید که احتمال پذیرش H_0 در شرایط نادرست بودن آن چقدر است.

مراحل به دست آوردن β

۱- تشکیل فرضیه‌های آزمون (H_1, H_0)

۲- محاسبه ملاک (آماره) آزمون (U) با این تفاوت که به‌جای مقدار آماره $\hat{\theta}$ از مقدار واقعی θ در آن استفاده می‌کنیم.

برای مثال:

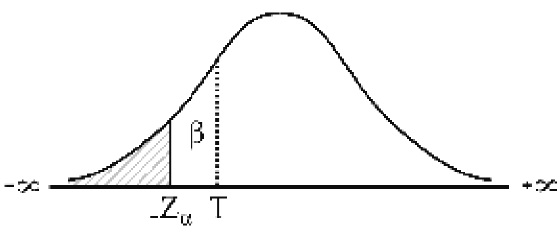
آماره آزمون میانگین $\quad T = \dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \longrightarrow T = \dfrac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ آماره آزمون برای محاسبه β

۳- با توجه به وضعیت آزمون از نظر یک‌دامنه (راست یا چپ) یا دودامنه بودن، مقدار β به صورت زیر محاسبه می‌شود:

۱) آزمون دامنه به چپ

$$\begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$

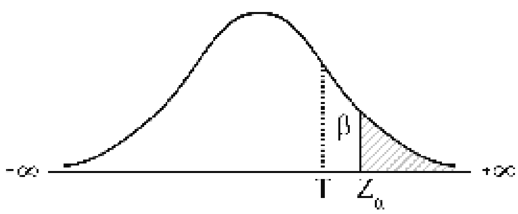
$$\begin{cases} \beta = P\big(Z < T-(-Z_\alpha)\big) = P\big(Z < T + Z_\alpha\big) \\ T = \text{(ملاک آزمون) مقدار آماره آزمون} \end{cases}$$



۲) آزمون دامنه به راست

$$\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

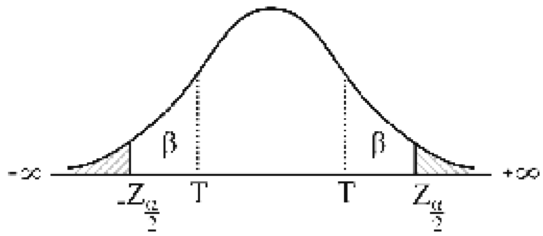
$$\begin{cases} \beta = P\big(Z > T - Z_\alpha\big) \\ T = \text{(ملاک آزمون) مقدار آماره آزمون} \end{cases}$$



۳) آزمون دودامنه

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

$$\begin{cases} \beta = P\left(T - Z_{\alpha/2} < Z < T - \left(-Z_{\alpha/2}\right)\right) \\ \quad = P\left(T - Z_{\alpha/2} < Z < T + Z_{\alpha/2}\right) \\ T = \text{(مقدار آماره آزمون (ملاک آزمون} \end{cases}$$



توجه: بسته به توزیع، آماره آزمون به‌جای Z می‌تواند t، χ^2 و یا F باشد.

نکته: درصورتی‌که توزیع جامعه گسسته باشد مقدار β را به شکل زیر نیز می‌توان محاسبه کرد:

$\beta = P($ برقراری فرض $X \mid H_1$ در ناحیه اطمینان باشد $)$

محاسبه توان آزمون

برای محاسبه توان آزمون بهتر است ابتدا خطای نوع دوم β را به دست آوریم و سپس مقدار حاصل را از یک کم کنیم تا توان آزمون به دست آید.

البته در توزیع‌های گسسته می‌توان توان آزمون را به صورت زیر نیز محاسبه کرد:

$\beta^* = 1 - \beta = P($ برقراری فرض $X \mid H_1$ در ناحیه بحرانی باشد $)$

مثال ۱ اگر میانگین واقعی مقدار نوشابه ریخته‌شده به درون شیشه‌ها در کارخانه‌ای 324 با انحراف معیار 12 سی‌سی باشد، خطای نوع دوم آزمون زیر بر اساس یک نمونه تصادفی 36 تایی با خطای نوع اول 2.5 درصد تقریباً کدام است؟ $\left(Z_{0.025} = 1.96 \simeq 2\right)$ (اقتصاد ـ ۸۳)

$$\begin{cases} H_0 : \mu \geq 330 \\ H_1 : \mu < 330 \end{cases}$$

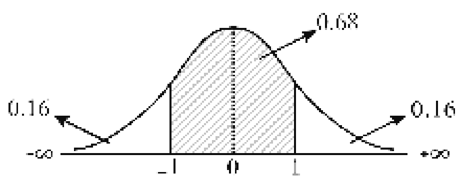
۱) 0.025 ۲) 0.16 ۳) 0.975 ۴) 0.84

حل: گزینه ۲ درست است.

با توجه به حالت (۱) داریم:

$$\begin{cases} \beta = P\left(Z < T - \left(-Z_\alpha\right)\right) = P\left(Z < T + Z_\alpha\right) = P\left(Z < -3 + 2\right) = P\left(Z < -1\right) = 0.16 \\ T = \dfrac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{324 - 330}{\frac{12}{\sqrt{36}}} = -3 \\ \mu = 324 \,,\, \mu_0 = 330 \,,\, n = 36 \,,\, \sigma = 12 \\ \alpha = 0.025 \quad \rightarrow \quad Z_{0.025} \simeq 2 \end{cases}$$

یادآوری:



مثال ۲ اگر در آزمون فرضیه $H_0 : \mu = 10$ در برابر $H_1 : \mu \neq 10$ مقدار $\alpha = 0.05$ باشد، آن‌گاه مقدار β: (اقتصاد ـ ۷۴)

۱) به μ واقعی بستگی دارد. ۲) 0.95 است.

۳) کمتر از 0.05 است. ۴) بیش از 0.05 است.

حل: گزینه ۱ درست است.

با توجه به مرحله۲ محاسبه β، به مقدار واقعی μ و σ جامعه نیاز داریم.

اختلاف مقدار پیشنهادی H_0 و مقدار واقعی آن

همان‌طورکه می‌دانیم برای محاسبه خطای نوع دوم (β) و درنتیجه توان آزمون $(1-\beta)$، درکنار فرض H_0 به مقدار واقعی جامعه نیز نیاز داریم. حال فرض کنید می‌خواهیم بدانیم اختلاف میان مقدار پیشنهادی H_0 و مقدار واقعی پارامتر جامعه، چه تأثیری بر سطح H_0، $1-\beta$ و حتی α دارد. برای بررسی صحت مبحث بالا، از یک مثال استفاده می‌کنیم:

برای مثال، در آزمون فرض مربوط به میانگین جامعه بر اساس یک نمونه 100تایی، اگر بدانیم مقدار میانگین واقعی و انحراف معیار، به‌ترتیب 240 و 10 است، می‌خواهیم مقدار خطای نوع دوم (β) و توان آزمون $(1-\beta)$ را برای سه مقدار پیشنهادی H_0 با فرض ثابت بودن خطای نوع اول $\alpha = \%5$ به صورت زیر بررسی کنیم:

الف) $H_0 : \mu \geq 240$

$$\begin{cases} \beta = P(Z < T + Z_\alpha) = P(Z < 0+2) = P(Z<2) = 0.9772 \rightarrow \qquad \text{توان آزمون} = 1-\beta = 0.0228 \\ T = \dfrac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{240-240}{\frac{10}{\sqrt{100}}} = 0 \end{cases}$$

ب) $H_0 : \mu \geq 241$

$$\begin{cases} \beta = P(Z < T + Z_\alpha) = P(Z < -1+2) = P(Z<1) = 0.68 \rightarrow \qquad \text{توان آزمون} = 1-\beta = 0.32 \\ T = \dfrac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{240-241}{\frac{10}{\sqrt{100}}} = -1 \end{cases}$$

ج) $H_0 : \mu \geq 244$

$$\begin{cases} \beta = P(Z < T + Z_\alpha) = P(Z < -4+2) = P(Z<-2) = 0.0228 \rightarrow \qquad \text{توان آزمون} = 1-\beta = 0.9772 \\ T = \dfrac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{240-244}{\frac{10}{\sqrt{100}}} = -4 \end{cases}$$

همان‌طورکه مشاهده می‌شود:

هرچه میانگین مطرح‌شده در فرض H_0 به میانگین واقعی جامعه نزدیک‌تر باشد، احتمال مرتکب شدن خطای نوع دوم (β) بیشتر می‌شود، درنتیجه توان آزمون $(1-\beta)$ کاهش می‌یابد. در مقابل، هرچه میانگین مطرح‌شده در فرض H_0 از میانگین واقعی جامعه دورتر باشد احتمال مرتکب شدن خطای نوع دوم (β) کمتر است، درنتیجه توان آزمون $(1-\beta)$ بیشتر می‌شود.

نتیجه:

وضعیت	α	β	$1-\beta$
۱) اختلاف H_0 و مقدار واقعی پارامتر کم (H_0 به واقعیت نزدیک)	کاهش $\to 0$	افزایش $\to 1$	کاهش $\to 0$
۲) اختلاف H_0 و مقدار واقعی پارامتر زیاد (H_0 از واقعیت دور)	افزایش $\to 1$	کاهش $\to 0$	افزایش $\to 1$

قضیه:

درصورتی‌که فرضیه H_0 به‌واقع درست باشد (H_1 به‌واقع نادرست)، آن‌گاه $1-\beta=0\,,\beta=1\,,\alpha=0$ و برعکس.

درصورتی‌که فرضیه H_0 به‌واقع نادرست باشد (H_1 به‌واقع درست)، آن‌گاه $1-\beta=1\,,\beta=0\,,\alpha=1$ و برعکس.

مقایسه آزمون‌های یک‌دامنه و دودامنه

در صورت امکان بهتر است آزمون فرضیه را به صورت یک‌دامنه تنظیم کنیم، زیرا آزمون‌های یک‌دامنه توان بیشتری نسبت به آزمون‌های دودامنه دارند. دلیل این امر آن است که در آزمون‌های یک‌دامنه، تمام ناحیه رد فرض در یک‌دامنه توزیع قرار دارد و تکلیف روشن است اما در آزمون‌های دودامنه، ناحیه رد در دو دامنه چپ و راست قرار دارد؛ به عبارت دیگر، مشخص نیست از کدام دامنه برای رد فرض استفاده می‌شود.

برای بررسی صحت این مطلب از یک مثال استفاده می‌کنیم:

برای مثال، در آزمون فرض مربوط به میانگین جامعه بر اساس یک نمونه 100تایی، مقدار میانگین واقعی و انحراف معیار به‌ترتیب 230 و 20 است. می‌خواهیم مقدار توان آزمون $(1-\beta)$ را برای دو آزمون یک‌دامنه و دودامنه با فرض ثابت بودن خطای نوع اول $\alpha=0.05$، به صورت زیر بررسی کنیم:

الف) $H_0:\mu\ge 233.3$

$$\begin{cases}\beta=P\left(Z<T+Z_\alpha\right)=P\left(Z<-1.65+1.65\right)=P\left(Z<0\right)=0.5\to \text{توان آزمون}=1-\beta=0.5\\ T=\dfrac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\dfrac{230-233.3}{\frac{20}{\sqrt{100}}}=-1.65\\ Z_\alpha=Z_{0.05}=1.65\end{cases}$$

ب) $H_0:\mu=233.3$

$$\begin{cases}\beta=P\left(T-Z_{\frac{\alpha}{2}}<Z<T+Z_{\frac{\alpha}{2}}\right)=P\left(\underbrace{-1.65-2}_{-3.65}<Z<\underbrace{-1.65+2}_{0.35}\right)>0.5\to \text{توان آزمون}=1-\beta<0.5\\ T=\dfrac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\dfrac{230-233.3}{\frac{20}{\sqrt{100}}}=-1.65\\ Z_{\frac{\alpha}{2}}=Z_{0.025}=1.96\simeq 2\end{cases}$$

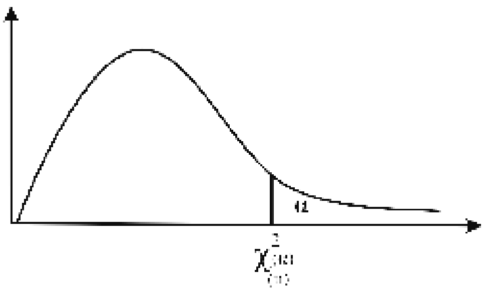
همان‌طورکه مشاهده می‌شود در حالت (الف) که آزمون یک‌دامنه است، توان آزمون بیشتر از حالت (ب) یعنی آزمون دودامنه است.

قضیه:

در شرایط مساوی، در حالتی که امکان هر دو آزمون یک‌دامنه و دودامنه وجود داشته باشد، آزمون یک‌دامنه توان بیشتری دارد و مناسب‌تر است.

کاربردهای آزمون کای اسکور

همان‌طورکه می‌دانیم، توزیع χ^2 (کای- مربع) از مقادیر مجذورشده Z تشکیل شده است، درنتیجه بین صفر تا بی‌نهایت تغییر می‌کند و هیچ‌گاه نمی‌تواند مقادیر منفی را به خود اختصاص دهد. منحنی این توزیع (چوله به راست) به صورت روبرو است:



در مبحث برآورد و آزمون فرض آماری برای واریانس جامعه $\left(\sigma^2\right)$ از توزیع χ^2 (کای - مربع) استفاده کردیم.

توزیع کای- مربع علاوه بر تخمین و آزمون آماری واریانس جامعه $\left(\sigma^2\right)$کاربردهای دیگری نیز دارد. از این توزیع برای آزمون فرضیه‌هایی که داده‌های مورد تجزیه و تحلیل به صورت فراوانی ارائه شده‌اند می‌توان استفاده کرد. این آزمون‌ها که در این بخش به بررسی آن‌ها می‌پردازیم، به شرح زیر هستند:

۱- آزمون نیکویی برازش (χ^2 ساده)

۲- آزمون استقلال (χ^2 مضاعف)

آزمون نیکویی برازش (χ^2 ساده) (Goodness – of - fit Test)

فرض کنید یک سکه را $n=100$ بار پرتاب کرده‌ایم و نتایج زیر برای آن به دست آمده است:

X_i	شیر	خط	
F_{o_i}	60	40	$n=\sum F_{o_i}=100$

حال می‌خواهیم بدانیم آیا سکه مورد نظر سالم است یا خیر. اگر سکه سالم باشد، آن‌گاه توزیع، یکنواخت است و هر یک از وجوه آن در هر بار پرتاب با احتمال $p_i=\frac{1}{2}$ ظاهر می‌شود، درنتیجه انتظار داریم فراوانی مشاهده‌شده به صورت زیر باشد:

X_i	شیر	خط
$F_{e_i}=np_i$	50	50

گاهی در مسایل پرسیده می‌شود که آیا مجموعه‌ای از داده‌های آماری دارای توزیع احتمال خاصی مانند یکنواخت، پواسون، دوجمله‌ای یا نرمال هستند (به طور پیش‌فرض یکنواخت)؟ در چنین مواردی برای تعیین خوبی برازش داده‌های آماری به توزیع احتمالی که عنوان شده است، **فراوانی مشاهده شده** **(تجربی)** در هر گروه (طبقه) را با **فراوانی مورد انتظار** **(تئوریک)** که از توزیع احتمال عنوان‌شده به دست می‌آید، مقایسه می‌کنیم.

مراحل انجام آزمون نیکویی برازش

فرض کنید در یک جامعه، فراوانی‌های مشاهده‌شده $\left(F_{o_i}\right)$ برای k طبقه مختلف از n مشاهده به صورت زیر باشد:

X_i	x_1	x_2	...	x_k	
F_{o_i}	F_{o_1}	F_{o_2}	...	F_{o_k}	$n=\sum F_{o_i}$

آن‌گاه مراحل انجام آزمون نیکویی برازش به شرح زیر است:

۱- عنوان کردن تابع احتمالی که انتظار داریم مشاهدات با آن تطبیق داشته باشند (به طور پیش‌فرض، یکنواخت $p_i=\frac{1}{k}$)

۲- محاسبه مقادیر احتمال هر یک از طبقات (p_i) و سپس فراوانی‌های مورد انتظار هر یک از طبقات $(F_{e_i} = np_i)$ بر اساس تابع احتمال عنوان‌شده

برای مثال، اگر تابع احتمال را به طور پیش‌فرض یکنواخت در نظر بگیریم، آن‌گاه احتمال وقوع هر طبقه، $p_i = \frac{1}{k}$ و فراوانی مورد انتظار هر طبقه، $F_{e_i} = np_i = \frac{n}{k}$ است، بنابراین داریم:

X_i	x_1	x_2	...	x_k
$F_{e_i} = np_i$	$\frac{n}{k}$	$\frac{n}{k}$	...	$\frac{n}{k}$

۳- استفاده از توزیع χ^2 برای مشخص کردن اینکه آیا توزیع احتمال عنوان‌شده برازش خوبی برای داده‌ها هست یا خیر

درجه آزادی توزیع χ^2

اگر در آزمون نیکویی برازش، k، تعداد طبقات و r، تعداد پارامتر برآوردشده در جامعه باشد، آن‌گاه درجه آزادی آماره χ^2 برابر با $k-r-1$ است.

برای مثال، در توزیع نرمال با پارامترهای μ و σ^2 یا توزیع دوجمله‌ای با پارامترهای n و p، درجه آزادی برابر است با:

$$k-2-1=k-3$$

در توزیع پواسون با پارامتر λ، درجه آزادی برابر است با:

$$k-1-1=k-2$$

در توزیع یکنواخت که در **پیش‌فرض** به عنوان توزیع جامعه در نظر گرفته می‌شود، پارامتری برآورد نمی‌شود و درجه آزادی برابر است با:

$$k-0-1=k-1$$

مثال ۱ تعداد طبقات، در یک آزمون نیکویی برازش، 10تاست که در آن μ_X و σ_X^2 برآورد شده‌اند. تعداد درجات آزادی این آزمون چقدر است؟

۱) 7 ۲) 8 ۳) 9 ۴) 10

حل: گزینه ۱ درست است.

$$\text{درجه آزادی} = k - r - 1 = 7 \quad ; \quad k = 10 , r = 2$$

حال با توجه به مباحث مطرح‌شده، مراحل عمومی چهارگانه آزمون نیکویی برازش به شرح زیر است:

۱- فرض‌های آماری

$$\begin{cases} H_0 : \text{فراوانی‌های مشاهده‌شده } (F_o) \text{ با فراوانی‌های مورد انتظار } (F_e) \text{ یکسان هستند} \\ H_1 : \text{فراوانی‌های مشاهده‌شده } (F_o) \text{ با فراوانی‌های مورد انتظار } (F_e) \text{ یکسان نیستند} \end{cases}$$

۲- آماره آزمون (ملاک آزمون)

ملاک آزمون نیکویی برازش برای یک آزمایش با k حالت، دارای توزیع χ^2_{k-1} به صورت زیر است:

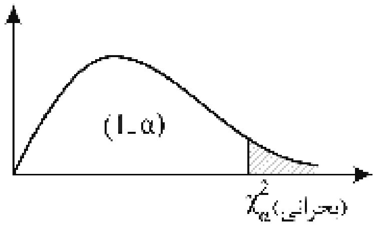
$$\chi^2_{(k-1)} = \sum_{i=1}^{k} \frac{\left(F_{o_i} - F_{e_i}\right)^2}{F_{e_i}}$$

$$\text{درجه آزادی} = k - 1$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

آزمون نیکویی برازش (χ^2ساده)، یک آزمون یک‌دامنه (دامنه به راست) به شکل زیر است:

$$\begin{cases} H_0: F_{o_i} = F_{e_i} \\ H_1: F_{o_i} \neq F_{e_i} \end{cases} \quad \xrightarrow{U > \chi^2_\alpha} \text{فرض } H_0 \text{ رد می‌شود}$$



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال ۲ آزمون نیکویی برازش، کدام‌یک از آزمون‌های زیر است؟

۱) یک دنباله به چپ
۲) یک دنباله به راست
۳) دو دنباله
۴) به ترتیب H_0 و مقدار α بستگی دارد.

حل: گزینه ۲ درست است.

نکته:

۱- در آزمون نیکویی برازش، H_0 نشان‌دهنده نوع توزیع مورد نظر است.

۲- هرچه تفاوت فراوانی‌های مشاهده‌شده و فراوانی‌های مورد انتظار کمتر باشد، مقدار ملاک آزمون $\left(\chi^2\right)$ کمتر و به صفر نزدیک‌تر می‌شود؛ بنابراین، احتمال پذیرش H_0 بیشتر است.

مثال ۳ بر اساس اطلاعات زیر، آماره آزمون برای آزمون این ادعا که فروش کالا در سه نوع بسته‌بندی الف و ب و ج دارای احتمال یکسان است کدام است؟ (اقتصاد ـ ۸۰)

نوع بسته‌بندی	الف	ب	ج
تعداد کالای فروش رفته	100	80	120

۱) $\chi_2^2 = 8$
۲) $\chi_2^2 = 4.7$
۳) $\chi_3^2 = 4.7$
۴) $\chi_3^2 = 8$

حل: گزینه ۱ درست است.

در این مسئله با توجه به نمونه $n = 300$تایی برای $k = 3$ (گروه)، فراوانی‌های مشاهده‌شده $\left(F_{o_i}\right)$ که همان تعداد کالای فروش‌رفته است، داده شده است. درعین‌حال فراوانی‌های مورد انتظار $\left(F_{e_i}\right)$ با در نظر گرفتن احتمال یکسان $\left(p_i = \frac{1}{k} = \frac{1}{3}\right)$ برای هر دسته به صورت $F_{e_i} = np_i = \frac{n}{k} = \frac{300}{3} = 100$ محاسبه می‌شود:

نوع بسته‌بندی	الف	ب	ج	
F_{o_i}	100	80	120	$n=\sum F_{o_i}=300$
F_{e_i}	100	100	100	

$$\chi^2_{(k-1)}=\chi^2_{(2)}=\sum_{i=1}^{3}\frac{\left(F_{o_i}-F_{e_i}\right)^2}{F_{e_i}}=\frac{(100-100)^2}{100}+\frac{(80-100)^2}{100}+\frac{(120-100)^2}{100}=8$$

مثال ۴ یک قنادی، آجیل مورد نیاز خود را با ترکیب 40 درصد بادام $\left(p_1=0.4\right)$ و 40 درصد پسته $\left(p_2=0.4\right)$ و 20 درصد گردو $\left(p_3=0.2\right)$ سفارش داده است. خریدار به هنگام تحویل محموله آجیل، نمونه‌ای به حجم $n=200$ را به طور تصادفی انتخاب می‌کند. این نمونه حاوی 60 بادام، 120 پسته و 20 گردو است. مقدار عددی آماره آزمون برای قبول محموله (خوبی برازش) عبارت است از: (اقتصاد ـ ۸۴)

۱) 9 ۲) 20 ۳) 35 ۴) 50

حل: گزینه ۳ درست است.

در این مسئله در یک نمونه $n=200$ تایی، فراوانی‌های مشاهده‌شده $\left(F_{o_i}\right)$ برای $k=3$ دسته بادام، پسته و گردو داده شده است. اما فراوانی مورد انتظار $\left(F_{e_i}\right)$ برای هر دسته با توجه به احتمال‌های داده‌شده به صورت زیر به دست می‌آید:

$$F_{e_i}=np_i=\begin{cases}\text{بادام}: F_e=200\times 0.4=80\\ \text{پسته}: F_e=200\times 0.4=80\\ \text{گردو}: F_e=200\times 0.2=40\end{cases}$$

X_i	بادام	پسته	گردو	
F_{o_i}	60	120	20	$n=\sum F_{o_i}=200$
F_{e_i}	80	80	40	

$$\chi^2_{(k-1)}=\chi^2_{(2)}=\sum_{i=1}^{3}\frac{\left(F_{o_i}-F_{e_i}\right)^2}{F_{e_i}}=\frac{(60-80)^2}{80}+\frac{(120-80)^2}{80}+\frac{(20-40)^2}{40}=35$$

تصحیح یتس (χ^2 تصحیح شده)

زمانی که در آزمون نیکویی برازش دو طبقه داشته باشیم، درجه آزادی 1 می‌شود. در بعضی منابع در این حالت از تصحیح یتس استفاده می‌شود و رابطه χ^2 به صورت زیر در می‌آید:

$$\chi^2_{\text{تصحیح‌شده}}=\sum\frac{\left(\left|F_{o_i}-F_{e_i}\right|-0.5\right)^2}{F_{e_i}}$$

✓ **دقت کنید!**

تنها در صورتی از تصحیح یتس (رابطه بالا) استفاده می‌کنیم که در صورت سؤال ذکر کنند از این تصحیح استفاده کنید در غیر این صورت در همه سؤالات از همان رابطه قبلی استفاده می‌کنیم.

مثال سکه‌ای را 100 بار پرتاب می‌کنیم. 55 بار شیر و 45 بار خط ظاهر می‌شود. می‌خواهیم آزمون کنیم که آیا این سکه با احتمال 95% سالم است یا خیر. آماره آزمون با استفاده از تصحیح یتس کدام است؟

۱) 0.81 ۲) 2 ۳) 5 ۴) 1.9

حل: گزینه ۱ درست است.

فرضیه‌های آماری به صورت زیر هستند:

$$\begin{cases} H_0: \text{سکه سالم است} \\ H_1: \text{سکه سالم نیست} \end{cases}$$

چون درجه آزادی $df = 2-1 = 1$ است، از تصحیح یتس استفاده می‌کنیم و جدول زیر را تشکیل می‌دهیم:

سکه	F_{o_i}	F_{e_i}	$F_{o_i}-F_{e_i}$	$\left\|F_{o_i}-F_{e_i}\right\|$	$\left(\left\|F_{o_i}-F_{e_i}\right\|-0.5\right)^2$	$\frac{\left(\left\|F_{o_i}-F_{e_i}\right\|-0.5\right)^2}{F_{e_i}}$
شیر	55	50	5	5	20.25	0.405
خط	45	50	-5	5	20.25	0.405
$\sum$	100	100	0			0.81

بنابراین آماره آزمون عبارت است از:

$$\chi^2 = \sum \frac{\left(\left|F_{o_i}-F_{e_i}\right|-0.5\right)^2}{F_{e_i}} = 0.81$$

آزمون استقلال (χ^2 مضاعف) (Chi- Square Test of Independence)

هرگاه بخواهیم وجود رابطه (وابستگی) یا استقلال بین دو متغیر X و Y را که حداقل یکی از آن‌ها کیفی هستند، بررسی کنیم، از آزمون استقلال (χ^2 مضاعف) استفاده می‌کنیم.

برای مثال برای تعیین فرضیه «وجود ارتباط بین تحصیل و درآمد» می‌توانیم متغیر X و Y را به صورت زیر در نظر بگیریم:

$$\begin{cases} X: \text{سطوح تحصیل (کیفی)} \\ Y: \text{درآمد برحسب ریال (کمی)} \end{cases}$$

اگر دو متغیر مستقل باشند، نتیجه می‌گیریم که توزیع درآمد بدون توجه به تحصیلات اشخاص انجام می‌شود؛ در غیر این صورت در شرایطی که X و Y وابسته باشند، توزیع درآمد بر اساس سطوح تحصیل انجام خواهد شد.

برای انجام آزمون استقلال دو متغیر X و Y درصورتی‌که متغیر X را با r دسته (سطر) به صورت $x_1, \dots, x_r$ و متغیر Y را با c دسته (ستون) به صورت $y_1, \dots, y_c$ در نظر بگیریم، فراوانی‌های مشاهده‌شده $\left(F_{o_{ij}}\right)$ در یک نمونه n تایی برای دو متغیر در یک «جدول توافقی» با r سطر و c ستون به صورت زیر در نظر گرفته می‌شوند:

$X \diagdown Y$	y_1	y_2	....	y_c	$F_{o_{i.}}$
x_1	$F_{o_{11}}$	$F_{o_{12}}$	....	$F_{o_{1c}}$	$\sum_j F_{o_{1j}} = F_{o_{1.}}$
x_2	$F_{o_{21}}$	$F_{o_{22}}$	....	$F_{o_{2c}}$	$\sum_j F_{o_{2j}} = F_{o_{2.}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_r	$F_{o_{r1}}$	$F_{o_{r2}}$	....	$F_{o_{rc}}$	$\sum_j F_{o_{rj}} = F_{o_{r.}}$
$F_{o_{.j}}$	$\sum_i F_{o_{i1}} = F_{o_{.1}}$	$\sum_i F_{o_{i2}} = F_{o_{.2}}$	...	$\sum_i F_{o_{ic}} = F_{o_{.c}}$	$n = \sum_{i=1}^{r} \sum_{j=1}^{c} F_{o_{ij}} = F_{o_{..}}$

جدول فراوانی‌های مورد انتظار $\left(F_{e_{ij}}\right)$ با توجه به جدول بالا به صورت زیر محاسبه می‌شود:

$$F_{e_{ij}} = \frac{F_{o_{i.}} \times F_{o_{.j}}}{n} \quad ; \quad i = 1, 2, \dots, r \ , \ j = 1, 2, \dots, c$$

برای مثال، در جدول فراوانی مشاهده‌شده $\left(F_{o_{ij}}\right)$ در یک نمونه n تایی به صورت زیر:

$X \diagdown Y$	y_1	y_2
x_1	20	10
x_2	40	30

برای محاسبه جدول فراوانی‌های مورد انتظار $\left(F_{e_{ij}}\right)$ به صورت زیر عمل می‌کنیم:

$X \diagdown Y$			$F_{o_{i.}}$
	20	10	30
	40	30	70
$F_{o_{.j}}$	60	40	$\sum\sum = 100$

$F_{o_{ij}}$

$$\xrightarrow{F_{e_{ij}} = \frac{F_{o_{i.}} \times F_{o_{.j}}}{n}}$$

$X \diagdown Y$		
	$\frac{30 \times 60}{100} = 18$	$\frac{30 \times 40}{100} = 12$
	$\frac{70 \times 60}{100} = 42$	$\frac{70 \times 40}{100} = 28$

$F_{e_{ij}}$

مراحل انجام آزمون استقلال

با توجه به مباحث بالا، مراحل چهارگانه آزمون استقلال متغیرهای کیفی به شرح زیر است:

۱- فرض‌های آماری

$$\begin{cases} H_0 : \text{دو متغیر } X \text{ و } Y \text{ مستقل هستند} \\ H_1 : \text{دو متغیر } X \text{ و } Y \text{ وابسته‌اند} \end{cases}$$

۲- آماره آزمون (ملاک آزمون)

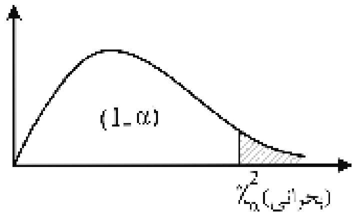
ابتدا $F_{e_{ij}}$ را با توجه به $F_{o_{ij}}$ از رابطه $F_{e_{ij}}=\dfrac{F_{o_{i.}}\times F_{o_{.j}}}{n}$ به ازای تمام مقادیر i و j محاسبه می‌کنیم؛ در این صورت ملاک آزمون استقلال توزیع χ^2 با $(r-1)(c-1)$ درجه آزادی به صورت زیر است:

$$\boxed{\begin{array}{l}\chi^2_{(r-1)(c-1)}=\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(F_{o_{ij}}-F_{e_{ij}}\right)^2}{F_{e_{ij}}}\\ \text{درجه آزادی}=(r-1)(c-1)\end{array}}$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

آزمون استقلال (χ^2 مضاعف)، یک آزمون یک‌دامنه (دامنه به راست) به صورت زیر است:

$$\begin{cases}H_0: \text{دو متغیر X و Y مستقل‌اند}\\ H_1: \text{دو متغیر X و Y وابسته‌اند}\end{cases}\qquad \xrightarrow{U>\chi^2_\alpha} \text{ فرض } H_0 \text{ رد می‌شود.}$$



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:
اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

نکته:

۱- ضریب توافقی (C)، شدت ارتباط میان X و Y را نشان می‌دهد.

$$\boxed{C=\sqrt{\frac{\chi^2}{\chi^2+n}}}$$

۲- جدول توافقی آزمون استقلال (χ^2 مضاعف) دو حاشیه تصادفی دارد.

۳- در آزمون استقلال (χ^2 مضاعف)، H_0 نشان‌دهنده صفر بودن کوواریانس X و Y است.

مثال ۵ می‌خواهیم ببینیم آیا بین متغیرهای X و Y رابطه وجود دارد یا خیر. متغیر کیفی X دارای 3 سطح و متغیر Y دارای 4 سطح است. درجه آزادی آزمون چقدر است؟ (مدیریت ـ ۸۳)

۱) 1　　۲) 7　　۳) 12　　۴) 6

حل: گزینه ۴ درست است.

$$r=3\ ,\ c=4\ \rightarrow\ \text{درجه آزادی}=(r-1)(c-1)=2\times3=6$$

مثال ۶ برای آزمون استقلال سطوح A و B از دو متغیر در یک آزمایش، جدول توافقی زیر حاصل شده است. مقدار آماره آزمون برابر است با: (اقتصاد ـ ۸۲)

A \ B	B_1	B_2
A_1	30	20
A_2	20	30

۱) 20
۲) 5
۳) 4
۴) 0.8

حل: گزینه ۳ درست است.

A \ B	B_1	B_2	$F_{o_{i.}}$
A_1	30	20	50
A_2	20	30	50
$F_{o_{.j}}$	50	50	100

$F_{o_{ij}}$

$$\xrightarrow{F_{e_{ij}}=\frac{F_{o_{i.}}\times F_{o_{.j}}}{n}}$$

A \ B	B_1	B_2
A_1	$\frac{50\times 50}{100}=25$	$\frac{50\times 50}{100}=25$
A_2	$\frac{50\times 50}{100}=25$	$\frac{50\times 50}{100}=25$

$F_{e_{ij}}$

$$\begin{cases}\chi^2_{(2-1)(2-1)}=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(F_{o_{ij}}-F_{e_{ij}}\right)^2}{F_{e_{ij}}}=\frac{(30-25)^2}{25}+\frac{(20-25)^2}{25}+\frac{(30-25)^2}{25}+\frac{(20-25)^2}{25}=4\\ r=2\ ,\ c=2\end{cases}$$

ملاک χ^2 و Z^2

می‌توان نشان داد که برای یک جدول توافقی 2×2 یا آزمون نیکویی برازش با $k=2$ طبقه، آماره Z^2 به طور دقیق برابر χ^2 است.

مثال در نظر است فرضیه سالم بودن یک سکه را آزمون کنیم. سکه را 50 بار آزمایش می‌کنیم که 30 بار شیر مشاهده می‌کنیم، در سطح معنی‌دار 5 درصد درباره فرضیه سالم بودن سکه چه می‌توان گفت؟ (کمیت بحرانی 2.706)

۱) رد می‌شود، چون $\chi^2=4.25$ است.
۲) رد می‌شود، چون $F=3.75$ است.
۳) قبول می‌شود، چون $\chi^2=2$ است.
۴) قبول می‌شود، چون $\chi^2=2.57$ است.

حل: گزینه ۳ درست است.

راه حل اول:

در این مسئله با توجه به $n=50$ بار پرتاب یک سکه برای $k=2$ دسته (شیر و خط) فراوانی‌های مشاهده‌شده $\left(F_{o_i}\right)$ 30 شیر و 20 خط است، درعین‌حال فراوانی‌های مورد انتظار $\left(F_{e_i}\right)$ با در نظر گرفتن سالم بودن سکه با احتمال یکسان $\left(p_i=\frac{1}{k}=\frac{1}{2}\right)$ به صورت $F_{e_i}=np_i=\frac{n}{k}=\frac{50}{2}=25$ به دست می‌آید. بنابراین داریم:

	شیر	خط	
F_{o_i}	30	20	$n=\sum F_{o_i}=50$
F_{e_i}	25	25	

$$\begin{cases}H_0: \text{سکه سالم است}\\ H_1: \text{سکه سالم نیست}\end{cases}$$

$$\chi^2_{(k-1)} = \chi^2_{(1)} = \sum_{i=1}^{2} \frac{\left(F_{o_i} - F_{e_i}\right)^2}{F_{e_i}} = \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} = \frac{50}{25} = 2$$

راه حل دوم: استفاده از آزمون نسبت

$$\begin{cases} H_0: p = \dfrac{1}{2} \\ H_1: p \neq \dfrac{1}{2} \end{cases}$$

$$\begin{cases} \text{ملاک آزمون}: Z = \dfrac{\overline{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{0.60 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{50}}} = \dfrac{0.1}{\sqrt{0.005}} \rightarrow \chi^2_{(1)} = z^2 = \dfrac{0.01}{0.005} = 2 \\ p_0 = \dfrac{1}{2} \quad , \ \overline{p} = \dfrac{x}{n} = \dfrac{30}{50} = 0.6 \end{cases}$$

تحلیل واریانس (Analysis of Variance)

در آزمون مقایسه میانگین دو جامعه، تفاوت‌های مشاهده‌شده بین میانگین دو نمونه $\left(\bar{X}_1 - \bar{X}_2\right)$ از دو جامعه مورد تحلیل و بررسی قرار گرفت. برای مقایسه میانگین بیش از دو جامعه، دو روش وجود دارد:

روش اول: مقایسه میانگین جوامع به صورت دوبه‌دو (با استفاده از آماره آزمون t یا Z)

روش دوم: مقایسه میانگین‌ها به صورت همزمان (تحلیل واریانس با استفاده از آماره آزمون F)

حال فرض کنید می‌خواهیم میانگین سه جامعه را دوبه‌دو با هم مقایسه کنیم (روش اول)، بنابراین باید سه آزمون زیر را انجام دهیم:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \qquad \begin{cases} H_0 : \mu_2 = \mu_3 \\ H_1 : \mu_2 \neq \mu_3 \end{cases} \qquad \begin{cases} H_0 : \mu_1 = \mu_3 \\ H_1 : \mu_1 \neq \mu_3 \end{cases}$$

اگر هر سه فرضیه H_0 پذیرفته شود، نتیجه می‌گیریم $\mu_1 = \mu_2 = \mu_3$ است، اما اگر برای مثال، احتمال مرتکب شدن خطای نوع اول (α) برای هر آزمون 0.1 باشد، آن‌گاه:

$$\alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست}) = 0.1$$

$$1-\alpha = P(H_0 \text{ پذیرش} \mid H_0 \text{ درست}) = 0.9$$

احتمال پذیرش فرضیه H_0 برابر با 0.9 است، درنتیجه احتمال پذیرش هر سه فرضیه H_0 برابر با $0.9 \times 0.9 \times 0.9 = 0.73$ و احتمال وقوع خطای نوع اول برای آن‌ها $1-0.73 = 0.27$ خواهد بود که نسبتاً مقدار بالایی است. درعین‌حال هرچه تعداد جوامع بیشتر شود، این مقدار افزایش می‌یابد؛ ازاین‌رو روش اول مناسب نیست. در این وضعیت نیاز به آزمونی داریم که بتواند خطای نوع اول را ثابت نگه دارد. روشی که برای این منظور پیشنهاد می‌شود روش توانمند آنالیز واریانس با آماره آزمون F (فیشر) است که خطای نوع اول (α) در آن ثابت و کمتر از خطای نوع اول آماره t یا Z در روش اول است.

مثال دلیل آنکه برای آزمون فرضیه برابری میانگین در سه جامعه از توزیع t استفاده نمی‌کنیم چیست؟ (اقتصاد ـ ۷۸)

۱) آزمون t غیرممکن است.

۲) توزیع χ^2 دارای توان آزمون بیشتری است.

۳) توزیع F دارای خطای نوع اول کمتری از توزیع t است.

۴) روش آنالیز واریانس دارای خطای نوع دوم کمتری است.

حل: گزینه ۳ درست است.

با توجه به توضیحات بالا آماره آزمون فیشر در روش تحلیل واریانس خطای نوع اول کمتری نسبت به آماره t یا Z در روش مقایسه دوبه‌دوی میانگین‌ها دارد.

نتیجه آنکه برای صرفه‌جویی در زمان و هزینه و به‌دلیل داشتن دقت بالاتر، مقایسه میانگین‌ها را به صورت همزمان انجام می‌دهیم. در این حالت هدف از مقایسه میانگین‌های بیش از دو جامعه این است که بدانیم آیا بین میانگین‌های نمونه‌ای جوامع مختلف تفاوت معنی‌داری وجود دارد یا خیر؟

تعریف: تحلیل واریانس روشی برای مقایسه میانگین‌های چند جامعه (بیش از دو جامعه) است که از طریق آن تفاوت‌های مشاهده‌شده بین میانگین نمونه‌های چند جامعه را با استفاده از آماره (F) (فیشر) آزمون می‌کنیم.

شرح روش تحلیل واریانس

فرض کنید می‌خواهیم اثر سه نوع کود شیمیایی (تیمار) مختلف را بر روی میوه پرتقال تولیدی کشاورزان با یکدیگر مقایسه کنیم. ایده اصلی آنالیز واریانس برای این مقایسه، نمایش میزان کل پراکندگی مشاهدات جوامع به صورت مجموع چند عبارت است که بتوان آن عبارت را به یک منشأ پراکندگی نسبت داد.

در مثال بالا، دو منشأ پراکندگی عبارت‌اند از:

الف) انحرافات ناشی از اثرهای سه کود شیمیایی (اثرهای تیماری یا انحرافات بین گروهی)

ب) انحرافات ناشی از سایر عوامل غیرقابل کنترل مانند نوع خاک، درجه حرارت (اثرهای باقی‌مانده یا خطا)

انواع تحلیل واریانس

درصورتی‌که تحلیل واریانس مانند مثال بالا بر اساس مشاهدات یک منبع (اثر کودهای شیمیایی) باشد، به آن **تحلیل واریانس یک‌عامله** و در شرایطی که تحلیل بر اساس مشاهدات دو منبع باشد (کود و آب)، به آن **تحلیل واریانس دوعامله** می‌گویند.

در ادامه به شرح و بررسی تحلیل واریانس یک‌عامله می‌پردازیم؛ بررسی تحلیل واریانس دوعامله خارج از مباحث این کتاب است.

تحلیل واریانس یک‌عامله (Analysis of Variance With Single Factor)

اگر $\mu_1 , \mu_2 , \dots , \mu_k$ میانگین‌های k جامعه باشند، می‌خواهیم بررسی کنیم آیا تفاوت معنی‌داری بین آن‌ها وجود دارد یا خیر. برای مقایسه میانگین k جامعه $(k > 2)$ مانند مقایسه میانگین دو جامعه، باید از هر جامعه نمونه‌گیری کنیم و از روی داده‌های نمونه، میانگین و واریانس هر جامعه را برآورد کنیم. توجه کنید که لزومی ندارد نمونه‌های گرفته‌شده از k جامعه یکسان باشند، اما در اینجا حالت خاص آن را که $n_1 = n_2 = \dots = n_k = n$ باشند، بررسی می‌کنیم.

فرض کنید k جامعه $(i = 1 , 2 , \dots , k)$ داشته باشیم و از هر جامعه یک نمونه n تایی $(j = 1 , 2 , \dots , n)$ انتخاب کنیم، در این صورت با توجه به جدول زیر:

					میانگین نمونه	واریانس نمونه
					$\bar{x}_i = \dfrac{\sum_{j=1}^{n} x_{ij}}{n}$	$S_i^2 = \dfrac{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}{n-1}$
نمونه اول	x_{11}	x_{12}	$\dots$	x_{1n}	$\bar{x}_1$	S_1^2
نمونه دوم	x_{21}	x_{22}	$\dots$	x_{2n}	$\bar{x}_2$	S_2^2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
نمونه kام	x_{k1}	x_{k2}	$\dots$	x_{kn}	$\bar{x}_k$	S_k^2
$N = kn$					$\bar{\bar{x}} = \dfrac{\sum_{i=1}^{k} \bar{x}_i}{k}$	$\dfrac{\sum_{i=1}^{k} S_i^2}{k}$
					میانگین کل داده‌ها	**میانگین واریانس نمونه‌ها**

نتایج زیر به دست می‌آید:

x_{ij} : مشاهده jام از جامعه iام

$\bar{x}_i$: میانگین نمونه از جامعه iام

S_i^2 : واریانس نمونه از جامعه iام

$\bar{\bar{x}}$: میانگین نمونه کل k جامعه

$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{\bar{x}})^2$: مجموع کل مجذور انحرافات از میانگین نمونه کل

ما در اینجا مجموع کل مجذور انحرافات از میانگین کل (SST) را به عنوان مقیاسی از کل پراکندگی در k جامعه در نظر می‌گیریم که می‌توانیم آن را به دو مجموعه زیر تجزیه کنیم:

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{\bar{x}}\right)^2}_{SST}=\underbrace{n\sum_{i=1}^{k}\left(\bar{x}_i-\bar{\bar{x}}\right)^2}_{SS(Tr)}+\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2}_{SSE}$$

همان‌طورکه مشاهده می‌شود:

$SS(Tr)$: پراکندگی **بین میانگین نمونه‌های** جوامع را نشان می‌دهد.

SSE: پراکندگی **درون نمونه‌ها** را نشان می‌دهد.

درنهایت می‌توانیم وضعیت نتایج به‌دست‌آمده از تجزیه مجموع کل مجذور انحرافات از میانگین کل (SST) را به صورت زیر در نظر بگیریم:

مجموع مجذور انحرافات از میانگین کل

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{\bar{x}}\right)^2$$

مجموع مجذور انحرافات بین گروهی

« ناشی از منابع یا اثرهای تیماری (رویه‌ای)»

$$SS(Tr)=n\sum_{i=1}^{k}\left(\bar{x}_i-\bar{\bar{x}}\right)^2$$

+

مجموع مجذور انحرافات درون گروهی

«ناشی ازسایر عوامل (باقی‌مانده یا خطا)»

$$SSE=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2$$

برای اثبات درستی مجموع بالا کافی است در مجموع مربعات کل، مقدار $\bar{x}_i$ را یک بار کم و یک بار جمع کنیم.

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{\bar{x}}\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}\underbrace{-\bar{x}_i+\bar{x}_i}_{0}-\bar{\bar{x}}\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\left(\bar{x}_i-\bar{\bar{x}}\right)^2+\left(x_{ij}-\bar{x}_i\right)^2+2\left(\bar{x}_i-\bar{\bar{x}}\right)\left(x_{ij}-\bar{x}_i\right)\right)$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\bar{x}_i-\bar{\bar{x}}\right)^2+\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2+0=n\sum_{i=1}^{k}\left(\bar{x}_i-\bar{\bar{x}}\right)^2+\sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2$$

برای سادگی محاسبات در روابط ذکرشده، تعداد نمونه‌ها در هریک از k جامعه برابر با n فرض شد؛ اما در حالت کلی، اگر n_i تعداد نمونه در جامعه iام باشد، آن‌گاه در تمام روابط کافی است به جای n، از n_i استفاده کنیم؛ بنابراین:

$$\begin{cases}SS(Tr)=n_i\sum_{i=1}^{k}\left(\bar{x}_i-\bar{\bar{x}}\right)^2\\ SSE=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_i\right)^2=\sum_{i=1}^{k}(n_i-1)S_i^2\\ SST=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{\bar{x}}\right)^2\end{cases}$$

مراحل آزمون تحلیل واریانس

۱- فرض‌های آماری

فرض‌های H_0 و H_1 برای آزمون برابری میانگین بیش از دو جامعه به شرح زیر است.

$$\begin{cases} H_0: \mu_1=\mu_2=\ldots=\mu_k & \text{«همه اثرهای تیماری برابر صفر هستند» یا «میانگین k جامعه با هم برابر است»} \\ H_1: & \text{«حداقل یکی از اثرهای تیماری مخالف صفر است» یا «حداقل میانگین دو جامعه با هم برابر نیستند»} \end{cases}$$

✓ **دقت کنید!**

علت استفاده از اثرهای تیماری به‌جای میانگین k جامعه آن است که بیشتر تکنیک‌های تحلیل واریانس در ابتدا با آزمایش‌های کشاورزی مطرح شدند و در آن تحلیل‌ها، کودهای مختلفی را که به خاک اضافه می‌کردند، تیمار می‌خواندند و میزان تأثیر این تیمارها (کودها) بر خاک را مقایسه می‌کردند.

۲- آماره آزمون(ملاک آزمون)

آماره آزمون مقایسه میانگین k جامعه $(k>2)$، فیشر F است که نحوه محاسبه آن با توجه به جدول آنالیزواریانس به شرح زیر است:

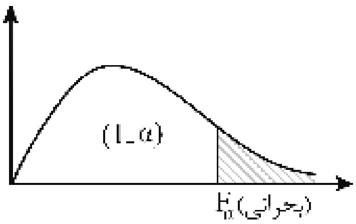
آماره آزمون $F_{k-1,k(n-1)}$	میانگین مجذور انحرافات (واریانس $MS=S^2$)	مجموع مجذور انحرافات SSD	درجه آزادی df	منشأ (منبع) تغییرات
$\dfrac{MS(Tr)}{MSE}$	$MS(Tr)=\dfrac{SS(Tr)}{k-1}$	$SS(Tr)=n\sum\left(\bar{x}_i-\bar{\bar{x}}\right)^2$	$k-1$	رویه یا تیمار (بین گروهی B)
	$MSE=\dfrac{SSE}{k(n-1)}$	$SSE=\sum\sum\left(x_{ij}-\bar{x}_i\right)^2=\sum(n-1)S_i^2$	$k(n-1)$	خطا یا باقی‌مانده (درون گروهی E)
		$SST=\sum\sum\left(x_{ij}-\bar{\bar{x}}\right)^2$	$kn-1$	جمع کل (T)

با توجه به جدول بالا تمام روابط زیر برای آماره آزمون مناسب هستند:

$$F_{k-1,k(n-1)}=\frac{MS(Tr)}{MSE}=\frac{\frac{SS(Tr)}{k-1}}{\frac{SSE}{k(n-1)}}=\frac{\frac{n\sum\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{k-1}}{\frac{\sum\sum\left(x_{ij}-\bar{x}_i\right)^2}{k(n-1)}}=\frac{\frac{n\sum\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{k-1}}{\frac{\sum S_i^2}{k}}$$

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

آزمون مقایسه میانگین k جامعه $(k>2)$ (آنالیز واریانس)، یک آزمون یک‌دامنه (دامنه به راست) به صورت زیر است:



$$\xrightarrow[\text{فرض } H_0 \text{ رد می‌شود}]{U>F_\alpha}$$

$$\begin{cases} H_0: \mu_1=\mu_2=\cdots=\mu_k \\ H_1: \text{حداقل دو میانگین متفاوت‌اند.} \end{cases}$$

۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:

اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

نکته:

۱- در تحلیل واریانس لازم نیست که اندازه نمونه‌ها برابر باشد.

۲- در تحلیل واریانس یک‌عامله، فرض H_1 بیانگر آن است که حداقل یکی از میانگین‌ها با دیگری متفاوت است.

۳- علت انتخاب ملاک آزمون F برای آزمون برابری چند میانگین به‌جای ملاک t آن است که مقدار خطای نوع اول (α) در آن کمتر است.

مثال ۱ به منظور مقایسه هزینه مواد غذایی خانوارها در 4 منطقه، از هر یک از این مناطق نمونه‌ای به حجم $n=5$ خانوار به طور تصادفی انتخاب شده است. بر اساس نتایج مشاهدات، میانگین مجذورات (تصحیح‌شده) بین گروه‌ها $\left(S_B^2=MS_B\right)$ مساوی با 6.34 و میانگین مجذورات (تصحیح‌شده) داخل گروه‌ها $\left(S_E^2=MS_E\right)$ مساوی با 3.5 به دست آمده‌اند. کدام‌یک از گزاره‌های زیر صحیح است؟ $(\alpha=0.05)$. (اقتصاد ـ ۷۲)

۱) میانگین‌های 4 جامعه فوق با هم یکسان هستند.

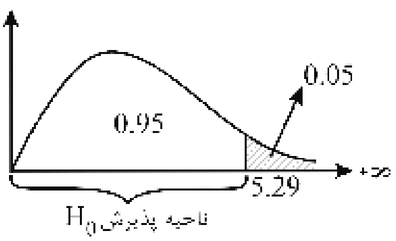
۲) یکسان بودن میانگین‌های 4 جامعه فوق را نمی‌توانیم رد کنیم.

۳) میانگین 4 جامعه فوق با هم تفاوت دارند.

۴) اقلاً یکی از میانگین‌ها و میانگین‌های دیگر تفاوت دارد.

حل: گزینه ۲ درست است.

$$\begin{cases} F_{k-1,\,k(n-1)}=\dfrac{MS(Tr)}{MSE} \quad \rightarrow \quad F_{3,16}=\dfrac{6.34}{3.5}=1.81 \\ k=4 \ , \ n=5 \ , \ F_{0.05,3,16}=5.29 \\ MS(Tr)=6.34 \ , \ MSE=3.5 \end{cases}$$



ابتدا مقدار F بحرانی را از جدول پیدا می‌کنیم. حال ازآنجاکه مقدار F آماره آزمون در ناحیه اطمینان (پذیرش H_0) است، یکسان بودن میانگین چهار جامعه را نمی‌توانیم رد کنیم.

مثال ۲ برای آزمون برابری متوسط هزینه‌های خانوار در 5 منطقه از هر یک از مناطق، نمونه‌ای به حجم 6 خانوار به طور تصادفی انتخاب شده و بر اساس آن $SSE=62.5$ و $SST=68.9$ به دست آمده است. مقدار عددی آماره آزمون عبارت است از: (اقتصاد ـ ۸۴)

۱) 0.1 ۲) 0.64 ۳) 1.56 ۴) 9.76

حل: گزینه ۲ درست است.

$$\begin{cases} \text{آماره آزمون}: F_{k-1,k(n-1)} = \dfrac{\dfrac{SS(Tr)}{k-1}}{\dfrac{SSE}{k(n-1)}} = \dfrac{\dfrac{6.4}{4}}{\dfrac{62.5}{5(5)}} = 0.64 \\ SST = SS(Tr) + SSE \quad \rightarrow \quad SS(Tr) = 68.9 - 62.5 = 6.4 \\ n = 6\,,\, k = 5 \end{cases}$$

مثال ۳ برای آزمون برابری میانگین بین 5 جامعه آماری، یک نمونه 5 تایی مستقل از هر یک گرفته شده و اطلاعات زیر محاسبه شده است. آماره آزمون تساوی میانگین‌ها چه عددی است؟ (اقتصاد ـ ۸۴)

$$\sum_{i=1}^{K}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2 = 60 \quad , \quad n\sum_{i=1}^{K}\left(\bar{x}_i-\bar{\bar{x}}\right)^2 = 80$$

۱) 0.14 ۲) 8.9 ۳) 7.6 ۴) 6.6

حل: گزینه ۴ درست است.

$$\begin{cases} F_{k-1,k(n-1)} = \dfrac{\dfrac{n\sum\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{k-1}}{\dfrac{\sum\sum\left(x_{ij}-\bar{x}_i\right)^2}{k(n-1)}} \quad \rightarrow \quad F_{4,20} = \dfrac{\dfrac{80}{4}}{\dfrac{60}{5(4)}} = 6.6 \\ k=5\,,\ n=5 \end{cases}$$

مثال ۴ چهار خط تولید مهره‌های یکسان می‌سازند که قطر آن‌ها بر طبق نرمال توزیع می‌شوند. از هر خط تولید نمونه‌هایی به حجم $n_1 = n_2 = n_3 = n_4 = 11$ انتخاب می‌کنیم و بر اساس نتایج مشاهدات کمیت‌های زیر به دست آمده است:

$$\sum\left(x_{1j}-\bar{x}_1\right)^2 = 120 \text{ و } \sum\left(x_{2j}-\bar{x}_2\right)^2 = 110$$

$$\sum\left(x_{3j}-\bar{x}_3\right)^2 = 90 \text{ و } \sum\left(x_{4j}-\bar{x}_4\right)^2 = 80 \text{ و } \sum_i\sum_j\left(x_{ij}-\bar{\bar{x}}\right)^2 = 550$$

مقدار عددی تابع نمونه‌ای (F) در آزمون فرضیه یکسان بودن میانگین‌های 4 جامعه نرمال کدام است؟ (اقتصاد ـ ۷۴)

۱) 10 ۲) 5 ۳) 2 ۴) 6

حل: گزینه ۲ درست است.

$$\begin{cases} F_{k-1,k(n-1)} = \dfrac{\dfrac{SS(Tr)}{k-1}}{\dfrac{\sum\sum\left(x_{ij}-\bar{x}_i\right)^2}{k(n-1)}} \quad \rightarrow \quad F_{3,40} = \dfrac{\dfrac{150}{4-1}}{\dfrac{400}{4(11-1)}} = \dfrac{50}{10} = 5 \\ SSE = \sum_i\sum_j\left(x_{ij}-\bar{x}_i\right)^2 = 120+110+90+80 = 400 \\ SST = \sum_i\sum_j\left(x_{ij}-\bar{\bar{x}}\right)^2 = 550 \rightarrow SST = SS(Tr)+SSE \rightarrow SS(Tr) = 550-400 = 150 \\ k = 4\,,\, n = 11 \end{cases}$$

مثال ۵ بر اساس نمونه‌های تصادفی 4 تایی از 3 جامعه که دارای توزیع نرمال‌اند، اطلاعات زیر به دست آمده است:

$$\begin{cases} \bar{x}_1 = 110 \ , \ \bar{x}_2 = 100 \ , \ \bar{x}_3 = 120 \\ S_1^2 = 180 \ , \ S_2^2 = 220 \ , \ S_3^2 = 200 \end{cases}$$

کمیت (ملاک) آماره آزمون، برای آزمون برابری میانگین در این سه جامعه چیست؟ (اقتصاد ـ ۸۳)

۱) $\chi^2_{1,1} = 2$ ۲) $F_{2,9} = 2$ ۳) $F_{2,9} = 3$ ۴) $F_{9,2} = 3$

حل: گزینه ۲ درست است.

$$\begin{cases} F_{k-1,k(n-1)} = \dfrac{\dfrac{n\sum\left(\bar{x}_i - \bar{\bar{x}}\right)^2}{k-1}}{\dfrac{\sum\limits_{i=1}^{k} S_i^2}{k}} \ \rightarrow \ F_{2,9} = \dfrac{\dfrac{4}{3-1}\left[(110-110)^2 + (100-110)^2 + (120-110)^2\right]}{\dfrac{180+220+200}{3}} = 2 \\ \bar{\bar{x}} = \dfrac{\sum\limits_{i=1}^{k} \bar{x}_i}{k} = \dfrac{110+100+120}{3} = 110 \\ k = 3 \ , \ n = 4 \end{cases}$$

مثال ۶ به منظور مقایسه هزینه ممکن خانوارها در 5 منطقه از هر یک از این مناطق، نمونه‌ای به حجم $n = 6$ خانوار به طور تصادفی انتخاب شده و بر اساس نتایج مشاهدات جدول تجزیه واریانس به صورت زیر به دست آمده است:

نوع تغییرات	SSD	df	$MS = S^2$	F
بین گروه‌ها (B)	6.4			
داخل گروه‌ها یا خطا (E)				
تغییرات کل (T)	68.9			

مقدار عددی تابع آزمون (F) کدام است؟ (اقتصاد ـ ۷۱)

۱) 1.6 ۲) 1.64 ۳) 2.5 ۴) 0.64

حل: گزینه ۴ درست است.

$$\begin{cases} F_{k-1,k(n-1)} = \dfrac{\dfrac{SS(Tr)}{k-1}}{\dfrac{SSE}{k(n-1)}} \ \rightarrow \ F_{4,25} = \dfrac{\dfrac{6.4}{5-1}}{\dfrac{62.5}{5(6-1)}} = \dfrac{1.6}{2.5} = 0.64 \\ SST = SSE + SS(Tr) \ \rightarrow \ SSE = 68.9 - 6.4 = 62.5 \\ SST = 68.9 \ , \ SS(Tr) = 6.4 \ , \ k = 5 \ , \ n = 6 \end{cases}$$

آزمون رگرسیون

هرگاه بتوانیم به کمک یک تابع، از روی تک تک مقادیر چند متغیر تصادفی مستقل $(X_1, X_2, \ldots, X_{k-1})$، مقدار یک متغیر تصادفی وابسته به آن‌ها (Y) را پیش‌بینی کنیم، می‌گوییم یک **تابع رگرسیون** یا **تابع برگشت** k **متغیره** داریم. اصطلاح رگرسیون را نخستین بار فرانسیس گالتن انگلیسی هنگام مطالعه و مقایسه اندازه قد پدر و پسر به کار برد.

بهترین تابع پیش‌بینی‌کننده دومتغیره

فرض کنید X و Y دو متغیر تصادفی با چگالی توأم $f(x,y)$ و توابع کناره‌ای $f(x)$ و $f(y)$ باشند. می‌خواهیم مقدار Y را به عنوان یک متغیر وابسته از روی مقداری که برای X به عنوان یک متغیر آزاد داده شده است، پیش‌بینی کنیم؛ برای مثال، اگر Y و X به ترتیب هزینه (وابسته) و درآمد (مستقل) خانوارهای یک شهر باشند می‌خواهیم مقدار هزینه را با دانستن درآمد پیش‌بینی کنیم.

اگر $h(x)$ را به عنوان تابع پیش‌بینی کننده y در نظر بگیریم، میانگین مربع خطای پیش‌بینی به صورت زیر خواهد بود:

$$\text{میانگین مربع خطا} = E\left[\left(y-h(x)\right)^2\right]$$

این مقدار با انتخاب $h(x)=E(y|x)$ به ازای هر x مینیمم می‌شود.

یادآوری: مجموع مجذور انحرافات از میانگین همواره مینیمم است.

$$\text{حداقل}: \sum\left(y-E(y)\right)^2$$

قضیه: بهترین تابع پیش‌بینی کننده برای پیش‌بینی y همواره امید شرطی تابع رگرسیون y روی x است:

$$\text{بهترین تابع پیش‌بینی کننده}: E(y|x)$$

مثال ۱ توزیع احتمال مشترک دو متغیر X و Y به صورت جدول زیر است:

X \ Y	0	1	2
0	0.20	0.15	0.05
1	0.05	0.20	0.05
6	0.05	0.05	0.20

۱) 2
۲) 1
۳) 0.5
۴) 1.5

تابع رگرسیون y روی x (امید y به شرط x) به ازای $x=6$ کدام است؟

حل: گزینه ۴ درست است.

X \ Y	0	1	2	$f(x)$
0	0.20	0.15	0.05	0.4
1	0.05	0.20	0.05	0.3
6	0.05	0.05	0.20	$f(X=6)=0.3$
$f(y)$	0.3	0.4	0.3	1

$$E(y|x=6)=\frac{\sum y f(x=6,y)}{f(x=6)}=\frac{0\times f(X=6,Y=0)+1\times f(X=6,Y=1)+2\times f(X=6,Y=2)}{f(X=6)}$$

$$=\frac{0\times 0.05+1\times 0.05+2\times 0.2}{0.3}=\frac{0.45}{0.3}=1.5$$

مثال ۲ تابع چگالی توأم $f(x,y)=\frac{1}{3}(x+y)\ ;\ 0<x<1\ ,\ 0<y<2$ را در نظر بگیرید. تابع رگرسیون y روی x (امید y به شرط x) به ازای $x=0.5$ کدام است؟

۱) $\frac{1}{3}$ ۲) $\frac{8}{9}$ ۳) $\frac{11}{9}$ ۴) $\frac{13}{9}$

حل: گزینه ۳ درست است.

$$E(y\,|\,x=0.5)=\int_y y\cdot f(y\,|\,x)dy=\int_0^2 y\cdot\frac{0.5+y}{2(0.5+1)}dy=\int_0^2\left(\frac{1}{6}y+\frac{1}{3}y^2\right)dy=\left[\frac{1}{12}y^2+\frac{1}{9}y^3\right]_0^2=\frac{1}{3}+\frac{8}{9}=\frac{11}{9}$$

$$\begin{cases} f(y\,|\,x)=\dfrac{f(x,y)}{f(x)}=\dfrac{\frac{1}{3}(x+y)}{\frac{1}{3}(2x+2)}=\dfrac{(x+y)}{2(x+1)} \\ f(x)=\displaystyle\int_0^2\frac{1}{3}(x+y)dy=\frac{1}{3}\left[xy+\frac{y^2}{2}\right]_0^2=\frac{1}{3}(2x+2) \end{cases}$$

بهترین تابع پیش‌بینی کننده خطی

همان‌طورکه گفته شد، برای تعیین بهترین تابع پیش‌بینی‌کننده، باید تابع چگالی مشترک $f(x,y)$ در دست باشد، اما اغلب چنین نیست؛ به همین دلیل در عمل معمولاً تابع پیش‌بینی کننده را به صورت خطی $E(y|x)=\beta x+\alpha$ اختیار می‌کنند تا نیازی به تابع چگالی توأم نباشد. به کمک این تابع، میانگین مربع خطای پیش‌بینی، $E(y-\alpha-\beta x)^2$ است که به ازای مقادیر زیر مینیمم می‌شود:

$$\begin{cases} \beta=\dfrac{Cov(X,Y)}{\sigma_x^2}=\rho_{x,y}\times\dfrac{\sigma_y}{\sigma_x} \\ \alpha=\mu_y-\beta\mu_x \end{cases}$$

از آنجاکه هرگاه x و y دارای توزیع توأم نرمال باشند، $E(y|x)$ دارای فرم خطی بالا است، انتخاب فرم خطی با فرض نرمال بودن جامعه به عنوان تابع پیش‌بینی‌کننده خطی قابل توجیه است.

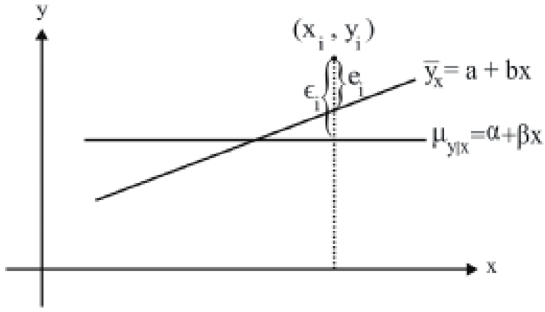
نتیجه: هرگاه میانگین (μ)، واریانس (σ^2) و ضریب همبستگی (ρ) جامعه نرمال موجود باشد، بدون نیاز به تابع چگالی توأم می‌توانیم بهترین تابع خطی را به دست آوریم.

برآورد بهترین تابع پیش‌بینی کننده خطی

تابع $E(y|x)=\beta x+\alpha$ به عنوان بهترین تابع پیش‌بینی کننده خطی، در صورتی مفید است که پارامترهای α و β در آن موجود باشند، اما در عمل این پارامترها مجهول هستند و باید آن‌ها را برآورد کنیم.

برازش خط رگرسیون

فرض کنید $y=\beta x+\alpha+\varepsilon$ معادله خط رگرسیون جامعه باشد که از طریق مشاهدات واقعی جامعه به دست آمده است، در این صورت:



β : شیب معادله خط

α : عرض از مبدأ (ثابت معادله)

ε : خطای (باقی‌مانده) معادله

با توجه به آنکه مجموع باقی‌مانده‌ها صفر است $(E(\varepsilon)=0)$، امید ریاضی معادله خط به صورت زیر خواهد بود:

$$\mu(y|x)=\beta x+\alpha$$

که همان بهترین تابع پیش‌بینی‌کننده خطی است.

حال برای برآورد پارامترهای α و β از معادله خط برآوردی $(\hat{y}=bx+a+e)$ حاصل، از یک نمونه n تایی $(x_n, y_n), \dots, (x_1, y_1)$ استفاده می‌کنیم.

با توجه به آنکه a و b برآوردکننده‌های α و β هستند، می‌توانیم خطای حاصل از برآورد (e) را به صورت زیر نشان دهیم:

$$e=y-\hat{y}$$

می‌خواهیم میزان خطا را به حداقل برسانیم اما چون $\sum e$ برای تمام خط‌های برآوردی به صورت $\hat{y}=bx+a+e$ برابر با صفر است، از مجموع مربعات خطا یعنی $\sum e^2$ استفاده می‌کنیم که همواره حداقل است:

$$\sum e^2=\sum(y-\hat{y})^2=\sum(y-bx-a)^2$$

روابط زیر در مورد خطا (باقی‌مانده) در خط رگرسیون برآورد شده برقرارند:

مجموع مقادیر باقی‌مانده‌های خط رگرسیون برآوردشده صفر است. $1)\ \sum e_i=0$

مجموع حاصل‌ضرب مقادیر باقی‌مانده‌ها در x های متناظرشان صفر است. $2)\ \sum e_i x_i=0$

مجموع حاصل‌ضرب مقادیر باقی‌مانده‌ها در مقادیر پیش‌بینی‌شده $(\hat{y}_i)$ صفر است. $3)\ \sum e_i \hat{y}_i=0$

مجموع مجذور باقی‌مانده‌ها در معادله خط رگرسیون برآوردشده حداقل است. $4)\ \sum e_i^2=\sum(y-\hat{y})^2=\min$

مثال ۱ در مدل رگرسیونی جامعه به صورت $y_i=\beta_0+\beta_1 x_i+\varepsilon_i$ داریم: (اقتصاد ـ ۷۰)

۱) $E(\varepsilon_i)=0$ ۲) $E(\varepsilon_i)=E(y_i)$ ۳) $E(\varepsilon_i)<0$ ۴) $E(\varepsilon_i)>0$

حل: گزینه ۱ درست است.

به دلیل شباهت رفتار e_i و ε_i در تحلیل رگرسیونی برای بررسی رفتار خطا از e_i استفاده می‌شود.

مثال ۲ در رگرسیون خطی $y=ax+b$ پراکندگی مشاهدات حول خط رگرسیون به وسیله کدام‌یک از عبارت‌های زیر بهتر توصیف می‌شود؟ $\bar{y}$: میانگین نمونه‌ها، $\hat{y}$: برآورد مقدار y (محیط زیست ـ ۸۸)

۱) $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ ۲) $\sum_{i=1}^{n}(y_i-\bar{y})^2$ ۳) $\sum_{i=1}^{n}(y_i+\bar{y})^2$ ۴) $\sum_{i=1}^{n}(y_i+\hat{y}_i)^2$

حل: گزینه ۱ درست است.

مفهوم اقتصادی پارامترهای خط رگرسیون

فرض کنید می‌خواهیم میزان هزینه ماهیانه یک خانوار را پیش‌بینی کنیم. متغیرهای مستقلی که می‌توانند بر « y : هزینه» (متغیر وابسته) تأثیرگذار باشند عبارت‌اند از درآمد خانوار، تعداد افراد خانواده و ... اما در رگرسیون دومتغیره تنها به یک متغیر مستقل نیاز داریم، بنابراین مهم‌ترین عامل، یعنی درآمد را در نظر می‌گیریم. حال فرض کنید پس از جمع‌آوری داده‌ها و برازش خط رگرسیون، معادله خط به صورت زیر به دست آمده است:

$$y = 2.1 + 5.6x + \varepsilon$$

مفهوم ثابت معادله (a): $a = 2.1$ به این مفهوم است که حتی اگر درآمد خانوار صفر باشد، یک هزینه ثابت به میزان 2.1 ماهیانه برای این خانوار وجود دارد؛ به عبارت دیگر، در ازای صفر بودن متغیر مستقل $(x = 0)$، متغیر وابسته به میزان a واحد است.

مفهوم شیب خط (b): $b = 5.6$ به این مفهوم است که به ازای افزایش یک واحد در درآمد خانوار، هزینه آن‌ها به میزان 5.6 افزایش می‌یابد؛ به عبارت دیگر، به ازای یک واحد افزایش متغیر مستقل (x)، متغیر وابسته به میزان b واحد افزایش یا کاهش می‌یابد که البته بسته به اینکه شیب خط مثبت یا منفی است، متغیر وابسته (y) افزایش یا کاهش می‌یابد.

برآورد پارامترهای α و β به روش حداقل مربعات (Ordinary Least Squares)

مقادیر برآورد a و b بهترین مقادیر در خط برآوردی رگرسیون $\hat{y} = bx + a$ هستند اگر مجموع مربعات خطا حداقل شود. این روش برآورد a و b، به روش حداقل مربعات معمولی یا OLS معروف است.

در این وضعیت a و b **برآوردکننده‌های نااریبی** برای پارامترهای α و β بوده و درعین‌حال واریانس آن‌ها از هر برآوردکننده دیگر کمتر و **کارایی‌شان بیشتر** است. در ادامه، تخمین مقادیر a و b را در روش حداقل مربعات برای مدل‌های مختلف رگرسیون جامعه، محاسبه می‌کنیم.

بسته به آنکه ضرایب α و β در مدل خطی دومتغیره، صفر یا مخالف صفر باشند، سه مدل مختلف برای معادله خط وجود دارد:

حالت اول: $\beta \neq 0 , \alpha \neq 0$ (پیش‌فرض)

در این وضعیت معادله رگرسیون جامعه به صورت $y = \beta x + \alpha + \varepsilon$ است، بنابراین برای محاسبه مقادیر تخمین a و b کافی است از معادله $\sum e^2 = \sum (y - bx - a)^2$ یک بار نسبت به a و یک بار نسبت به b مشتق بگیریم. در این صورت مقادیر a و b به صورت زیر خواهند بود:

$$(۱)\quad b = \frac{SP_{xy}}{SS_x} = r_{x,y}\frac{S_y}{S_x}$$

$$(۲)\quad a = \bar{y} - b\bar{x}$$

که در آن:

$$\begin{cases} SP_{xy} = \sum (x-\bar{x})(y-\bar{y}) = \sum xy - \dfrac{\sum x \sum y}{n} = \sum x(y-\bar{y}) = \sum y(x-\bar{x}) \\ SSx = \sum (x-\bar{x})^2 = \sum x^2 - \dfrac{\left(\sum x\right)^2}{n} = \sum x^2 - n\bar{x}^2 \\ SSy = \sum (y-\bar{y})^2 = \sum y^2 - \dfrac{\left(\sum y\right)^2}{n} = \sum y^2 - n\bar{y}^2 \end{cases}$$

مدل رگرسیون $y = \beta x + \alpha + \varepsilon$

معادله خط رگرسیون $y = \beta x + \alpha + \varepsilon$ به کمک نمونه‌ای به صورت $\{(x_i, y_i) ; i = 1,2,\ldots,n\}$ با خط $\hat{y} = bx + a$ برآورد می‌شود که در آن a و b به عنوان تخمین α و β از روابط (۱) و (۲) محاسبه می‌شوند.

اثبات روابط (۱) و (۲) خارج از بحث کتاب است، برای درک بهتر می‌توانید به منابع معرفی‌شده در انتهای کتاب مراجعه کنید.

مثال در معادله رگرسیون $E(y|x)=\alpha+\beta x$ تخمین حداقل مربعات از پارامتر β عبارت است از: (اقتصاد ـ ۸۷)

۱) $b=\dfrac{\sum x_i y_i}{\sum x_i^2}$

۲) $b=\dfrac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum x_i^2}$

۳) $b=\dfrac{\sum x_i y_i}{\sum (x_i-\bar{x})^2}$

۴) $b=\dfrac{\sum_{i=1}^{n} x_i (y_i-\bar{y})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$

حل: گزینه ۴ درست است.

حالت دوم : $\beta\neq 0, \alpha=0$

در این وضعیت معادله رگرسیون جامعه به صورت $y=\beta x+\varepsilon$ است؛ به عبارت دیگر، خط رگرسیون از مبدأ مختصات می‌گذرد، بنابراین مقدار تخمین $a=0$ است.

درعین‌حال با مشتق‌گیری از معادله $\sum e^2=\sum (y-bx)^2$ نسبت به b می‌توانیم مقدار تخمین b را به صورت زیر محاسبه کنیم:

$$(1)\quad b=\frac{\sum xy}{\sum x^2}$$

$$(2)\quad a=0$$

مدل رگرسیون $y=\beta x+\varepsilon$

معادله خط رگرسیون $y=\beta x+\varepsilon$ به کمک نمونه‌ای به صورت $\{(x_i,y_i)\,;\, i=1,2,\dots,n\}$ با خط $\hat{y}=bx$ برآورد می‌شود که در آن b به عنوان تخمین β از رابطه (۱) محاسبه می‌شود.

اثبات رابطه (۱):

$$\frac{\partial \sum_{i=1}^{n}(y-bx_i)^2}{\partial b}=0 \rightarrow -2\sum_{i=1}^{n} x_i(y_i-bx_i)=0 \rightarrow \sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} x_i^2=0 \rightarrow b=\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

مثال در معادله رگرسیون $E(y|x)=\beta x$ تخمین حداقل مربعات از پارامتر β عبارت است از:

۱) $b=\dfrac{\sum x_i y_i}{\sum x_i^2}$

۲) $b=\dfrac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum x_i^2}$

۳) $b=\dfrac{\sum x_i y_i}{\sum (x_i-\bar{x})^2}$

۴) $b=\dfrac{\sum_{i=1}^{n} x_i (y_i-\bar{y})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$

حل: گزینه ۱ درست است.

حالت سوم : $\beta=0, \alpha\neq 0$

در این وضعیت معادله رگرسیون جامعه به صورت $y=\alpha+\varepsilon$ است، بنابراین مقدار تخمین $b=0$ است و با مشتق‌گیری از معادله $\sum e^2=\sum(y-a)^2$ نسبت به a می‌توانیم مقدار تخمین a را به صورت زیر محاسبه کنیم:

(۱) $b=0$

(۲) $a=\bar{y}$

> **مدل رگرسیون $y=\alpha+\varepsilon$**
>
> معادله خط رگرسیون $y=\alpha+\varepsilon$ به کمک نمونه‌ای به صورت $\{(x_i,y_i)\,;\, i=1,2,\ldots,n\}$ با خط $\hat{y}=a$ برآورد می‌شود که در آن a به عنوان تخمین α از رابطه (۲) محاسبه می‌شود.

اثبات رابطه (۲):

$$\frac{\partial\sum_{i=1}^{n}(y_i-a)^2}{\partial a}=0 \quad\rightarrow\quad -2\sum_{i=1}^{n}(y_i-a)=0 \quad\rightarrow\quad \sum_{i=1}^{n}y_i-na=0 \quad\rightarrow\quad a=\frac{\sum_{i=1}^{n}y_i}{n}=\bar{y}$$

مثال در یک مدل رگرسیون به صورت $Y_i=a+\varepsilon_i\ ;\ i=1,\ldots,n$ مقدار برآورد a از روش کمترین مربعات خطا، کدام است؟
(محیط زیست ـ ۸۷)

۱) $\bar{y}$ ۲) $\sum(y_i-\bar{y})^2$ ۳) $\dfrac{\sum y_i}{\sum y_i^2}$ ۴) $\dfrac{\sum(y_i-\bar{y})}{\sum(y_i-\bar{y})^2}$

حل: گزینه ۱ درست است.

انحراف معیار خطای برآورد

برای انجام استنباط‌های آماری (برآورد و آزمون) در خط رگرسیون، ابتدا باید پراکندگی نقاط را در بالا و پایین خط اندازه‌گیری کنیم. برای این کار از انحراف معیار خطای تخمین (S_e) استفاده می‌کنیم که پراکندگی نقاط را در بالا و پایین خط رگرسیون نمونه اندازه‌گیری می‌کند؛ به عبارت دیگر، S_e انحراف معیار خطاها یعنی e است که از رابطه زیر محاسبه می‌شود:

$$S_e=\sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}=\sqrt{\frac{\sum e_i^2}{n-2}}$$

✓ دقت کنید!

چون دو پارامتر a و b در معادله خط رگرسیون برآورد شده است، به تعداد پارامترهای برآوردشده، از درجه آزادی توزیع واریانس کم می‌شود و مخرج واریانس $n-2$ است؛ بنابراین درصورتی‌که مدل رگرسیون طبق حالت‌های دوم وسوم باشد تنها یک پارامتر برآورد می‌شود و درجه آزادی آن $n-1$ است که در مخرج کسر خطای معیار برآورد نیز باید $n-1$ قرار گیرد. درواقع انحراف معیار برآورد، حاصل جذر مجموع مربعات خطا بر درجه آزادی آن است.

با بسط رابطه بالا، به روابط محاسباتی زیر می‌رسیم:

$$\begin{cases} S_e=\sqrt{\dfrac{\sum y^2-a\sum y-b\sum xy}{n-2}}=\sqrt{\dfrac{SS_y-b^2SS_x}{n-2}}=\sqrt{\dfrac{SS_y-bS_{xy}}{n-2}} \\ SS_x=\sum(x_i-\bar{x})^2=\sum x^2-\dfrac{(\sum x)^2}{n} \\ SS_y=\sum(y_i-\bar{y})^2=\sum y^2-\dfrac{(\sum y)^2}{n} \\ S_{xy}=Sp_{xy}=\sum(x_i-\bar{x})(y_i-\bar{y})=\sum x_iy_i-\dfrac{\sum x\sum y}{n} \end{cases}$$

مثال برای ارزیابی بستگی بین دو صفت x و y در یک جامعه نرمال دوبعدی، نمونه‌ای به حجم $n = 20$ انتخاب و کمیت‌های مقابل به دست آمده‌اند: $\sum x = 40 \ , \ \sum x^2 = 90 \ , \ \sum y = 100 \ , \ \sum y^2 = 528 \ , \ \sum xy = 210$

کمیت تخمین خطای (خطای تخمین S_e) معادله رگرسیون y برحسب x کدام است؟ (اقتصاد ـ ۷۷)

۱) 0 ۲) 0.58 ۳) 1 ۴) 1.58

حل: گزینه ۳ درست است.

$$S_e = \sqrt{\frac{SS_y - b^2 SS_x}{n-2}} = \sqrt{\frac{28 - 1^2 \times 10}{20-2}} = 1$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 528 - \frac{(100)^2}{20} = 28 \quad , \quad SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 90 - \frac{(40)^2}{20} = 10$$

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{210 - \frac{40 \times 100}{20}}{90 - \frac{(40)^2}{20}} = \frac{10}{10} = 1$$

امید و واریانس برآوردکننده‌های a و b

الف) با توجه به نااریبی برآوردکننده‌های a و b برای پارامترهای α و β از خط رگرسیون جامعه، امید آماره‌های a و b به صورت زیر است:

$$E(a) = \alpha$$
$$E(b) = \beta$$

ب) برای محاسبه واریانس برآوردکننده‌های a و b با استفاده از S_e^2 (واریانس خطای برآورد) از روابط زیر استفاده می‌کنیم:

$$\boxed{S_b^2 = \frac{S_e^2}{SS_x} \quad \rightarrow \quad S_b = \frac{S_e}{S_x}}$$

$$\boxed{S_a^2 = \frac{S_e^2 \, \overline{x^2}}{SS_x} \quad \rightarrow \quad S_a = \frac{S_e \sqrt{\overline{x^2}}}{S_x}}$$

یادآوری:

$$SSx = \sum \left(x_i - \bar{x}\right)^2 = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$S_x = \sqrt{SSx}$$

$$\overline{x^2} = \frac{\sum x^2}{n}$$

توزیع برآوردکننده‌های a و b

ازآنجاکه برآوردکننده‌های a و b با میانگین و انحراف معیار به‌دست‌آمده از مرحله قبل دارای توزیع نرمال هستند، آماره‌های زیر برای ساختن برآورد فاصله‌ای و ملاک آزمون ضرایب α و β مورد استفاده قرار می‌گیرند:

$$t = \frac{b-\beta}{S_b}$$

$$t = \frac{a-\alpha}{S_a}$$

برآورد فاصله‌ای پارامترهای خط رگرسیون

با توجه به آماره t به عنوان تابع محوری، برای ساختن فاصله اطمینان $(1-\alpha)$ درصدی برای پارامترهای α و β به شکل زیر عمل می‌کنیم:

فاصله اطمینان شیب خط رگرسیون : $b \pm t_{\frac{\alpha}{2}} S_b$

فاصله اطمینان عرض از مبدأ خط رگرسیون : $a \pm t_{\frac{\alpha}{2}} S_a$

آزمون‌های معنی‌داری در رگرسیون

برای هر معادله خط رگرسیون به صورت $y = \beta x + \alpha + \varepsilon$، آزمون‌های معنی‌داری زیر بررسی می‌شوند:

۱ـ آزمون معنی‌دار بودن شیب خط یا رابطه خطی $(\rho \neq 0, \beta \neq 0)$

۲ـ آزمون معنی‌دار بودن ثابت معادله $(\alpha \neq 0)$

۳ـ آزمون معنی‌دار بودن معادله خط رگرسیون ($\alpha \neq 0$ یا $\beta \neq 0$ یا هردو)

۱ـ آزمون معنی‌دار بودن شیب خط یا رابطه خطی

همان‌طورکه می‌دانیم در حالت‌های ۱ و ۲ از مدل‌های خط رگرسیون جامعه به فرم $y = \beta x + \alpha + \varepsilon$ یا $y = \beta x + \varepsilon$، مقدار β (شیب خط) که از رابطه زیر محاسبه می‌شود حتماً مخالف صفر است:

$$\beta = \frac{Cov(x,y)}{\sigma_x^2} = \rho_{x,y} \times \frac{\sigma_y}{\sigma_x}$$

با توجه به مبحث همبستگی، زمانی‌که رابطه خطی میان دو متغیر x و y وجود دارد (یعنی رابطه خطی معنی‌دار است)، کوواریانس، ضریب همبستگی و درنتیجه شیب خط مخالف صفر هستند.

فرض کنید $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ مقادیر مشاهده‌شده یک نمونه تصادفی از دو متغیر X و Y در جامعه‌ای نرمال با ضریب همبستگی ρ باشند که ضریب همبستگی نمونه (r) از طریق آن‌ها به صورت زیر محاسبه می‌شود:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sqrt{\left(\sum x^2 - n\bar{x}^2\right)\left(\sum y^2 - n\bar{y}^2\right)}}$$

بدیهی است که این ضریب همبستگی از یک نمونه به نمونه دیگر تغییر می‌کند. حال سؤال اینجاست که آیا بین دو متغیر x و y در جامعه مربوط به نمونه‌ها همبستگی معنی‌داری وجود دارد یا خیر؟ به عبارت دیگر، آیا بین دو متغیر رابطه خطی وجود دارد $(\rho \neq 0, \beta \neq 0)$ یا ضریب همبستگی به‌دست‌آمده از نمونه، به طور تصادفی مخالف صفر شده است و بین دو متغیر x و y رابطه خطی وجود ندارد $(\rho = 0, \beta = 0)$؟

برای پاسخ دادن به این سؤال کافی است فرضیه استقلال دو متغیر تصادفی x و y را آزمون کنیم، زیرا:

x , y نرمال و مستقل $\rightarrow \rho = 0 , \beta = 0$

x , y نرمال و وابسته خطی $\rightarrow \rho \neq 0 , \beta \neq 0$

در ادامه به شرح مراحل چهارگانه این آزمون می‌پردازیم.

۱- فرضیه‌های آزمون

$$\begin{cases} H_0 : (\rho = 0 \ \text{یا}\ \beta = 0) & \text{«ضریب (شیب) رگرسیون معنی‌دار نیست» یا «رابطه خطی معنی‌دار نیست»} \\ H_1 : (\rho \neq 0 \ \text{یا}\ \beta \neq 0) & \text{«ضریب یا (شیب) رگرسیون معنی‌دار است» یا «رابطه خطی معنی‌دار است»} \end{cases}$$

۲- آماره آزمون (ملاک آزمون)

آماره (ملاک) مناسب برای آزمون دارای توزیع t_{n-2} است که با در نظر گرفتن فرض $H_0 : \rho = 0$ یا $H_0 : \beta = 0$ به صورت زیر خواهد بود:

$$\boxed{t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{b}{S_b}}$$

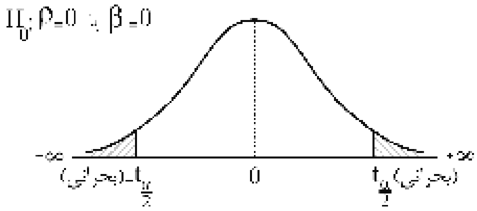
درجه آزادی آماره آزمون

الف) اگر مدل رگرسیون خطی جامعه به فرم $y=\beta x+\alpha+\varepsilon$ (پیش فرض) باشد، درجه آزادی آماره آزمون t، $n-2$ است.

ب) اگر مدل رگرسیون خطی جامعه به فرم $y=\beta x+\varepsilon$ باشد (خط رگرسیون از مبدأ مختصات می‌گذرد)، درجه آزادی آماره آزمون t، $n-1$ است که در این حالت مخرج کسر آماره در زیر رادیکال نیز $n-1$ می‌شود.

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

با توجه به فرض H_1، آزمون دودامنه است، همچنین بر اساس توزیع آماره t با $n-2$ درجه آزادی در سطح خطای α، سطوح H_0 و H_1 به شکل زیر خواهد بود:



$$U>t_{\frac{\alpha}{2},n-2} \quad \text{or} \quad U<-t_{\frac{\alpha}{2},n-2} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{}$$

$$\begin{cases} H_0:\rho=0 \\ H_1:\rho\neq 0 \end{cases} \qquad \begin{cases} H_0:\beta=0 \\ H_1:\beta\neq 0 \end{cases}$$

۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:

اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال ۱ بر اساس یک نمونه تصادفی $n=18$ تایی، ضریب همبستگی بین x و y برابر $r=0.8$ محاسبه شده است. مقدار آماره آزمون برای آزمون صفر بودن ضریب همبستگی جامعه برابر است با: (اقتصاد ـ ۸۵)

۱) $\frac{1}{2}$ ۲) 1 ۳) 1.39 ۴) 5.3

حل: گزینه ۴ درست است.

$$\begin{cases} H_0:\rho=0 \\ H_1:\rho\neq 0 \end{cases}$$

$$\begin{cases} t_{n-2}=\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} \rightarrow t_{16}=\dfrac{0.8}{\sqrt{\dfrac{1-0.64}{18-2}}}=5.3 \\ n=18 \ , \ r=0.8 \end{cases}$$

مثال ۲ با توجه به آمار مربوط به واردات کل کشور (y) و درآمدهای ارزی حاصل از صادرات نفت (x) در طی 21 سال گذشته رگرسیون زیر برآورد شده است. آماره آزمون برای آزمون معنی‌دار بودن رابطه خطی بین واردات و درآمدهای ارزی کدام است؟ (اقتصاد ـ ۸۳)

$$y=3.2+0.9x+e \ , \ R^2=0.81$$

۱) $z=2.5$ ۲) $\chi^2_{20}=4.6$ ۳) $F_{1,19}=82.2$ ۴) $t_{19}=9$

حل: گزینه ۴ درست است.

راه حل اول:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

$$\begin{cases} t_{(n-2)} = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} \rightarrow t_{19} = \dfrac{0.9}{\sqrt{\dfrac{1-0.81}{21-2}}} = 9 \\ n = 21, R^2 = 0.81 \rightarrow |r| = 0.9 \rightarrow r = +0.9 \end{cases}$$ (با توجه به مثبت بودن شیب خط رگرسیون)

راه حل دوم: ازآنجاکه در آزمون معنی‌دار بودن فقط از آماره t استفاده می‌کنیم، بدون حل سؤال نیز می‌توانستیم گزینه ۴ را به عنوان گزینه درست انتخاب کنیم.

مثال ۳ برای برآورد رابطه بین فروش (y) و هزینه تبلیغات (x) رابطه زیر بر اساس یک نمونه تصادفی 51 تایی برآورد شده است:

$y = 36 + 4.7x + e \ , \ R^2 = 0.5$

مقدار عددی آماره آزمون برای آزمون معنی‌دار بودن ضریب x یعنی β چقدر است؟ (اقتصاد ـ ۷۸)

۱) 5 ۲) 6 ۳) 7 ۴) 8

حل: گزینه ۳ درست است.

$$\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

$$\begin{cases} t_{(n-2)} = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} \rightarrow t_{49} = \dfrac{\sqrt{0.5}}{\sqrt{\dfrac{1-0.5}{51-2}}} = 7 \\ R^2 = 0.5 \rightarrow |r| = \sqrt{0.5} \rightarrow r = +\sqrt{0.5}, n = 51 \end{cases}$$ (با توجه به مثبت بودن شیب خط رگرسیون)

مثال ۴ در یک الگوی رگرسیون خطی ساده $y = \alpha + \beta x + \varepsilon$ اگر $n = 22$، $R^2 = 0.2$ (ضریب تعیین یا تشخیص) و $t_{0.975, 20} = 2.101$ باشد، در سطح معنی‌دار بودن $\alpha = \%5$، کدام‌یک از گزاره‌ها صحیح است؟ (اقتصاد ـ ۷۱)

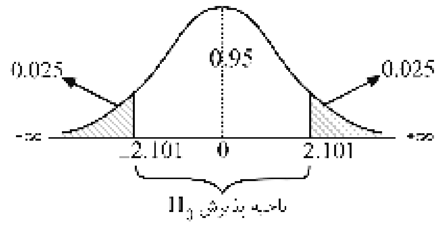
۱) فرضیه $\rho = 0$ را در مقابل $\rho \neq 0$ رد نمی‌کنیم.
۲) فرضیه $\alpha = 0$ را در مقابل $\alpha \neq 0$ رد می‌کنیم.
۳) فرضیه $\beta = 0$ را در مقابل $\beta \neq 0$ رد می‌کنیم.
۴) فرضیه $\alpha = 0$ را در مقابل $\alpha \neq 0$ رد نمی‌کنیم.

حل: گزینه ۳ درست است.

$$(1)\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases} \sim \begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

$$(2)\begin{cases} t_{(n-2)} = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} \rightarrow t_{(20)} = \dfrac{\pm\sqrt{0.2}}{\sqrt{\dfrac{1-0.2}{22-2}}} = \dfrac{\pm\sqrt{0.2}}{0.2} = \pm\dfrac{1}{\sqrt{0.2}} = \pm\sqrt{\dfrac{10}{2}} = \pm\sqrt{5} = \pm 2.24 \\ R^2 = 0.2 \rightarrow |r| = \sqrt{0.2} \rightarrow r = \pm\sqrt{0.2} \\ n = 22 \ , \ t_{0.975, 20} = 2.101 \end{cases}$$

(3)



(4) چون مقدار عددی آماره آزمون در ناحیه بحرانی (رد H_0) قرار دارد، فرضیه‌های $\begin{cases} H_0 : \rho = 0 \\ H_0 : \beta = 0 \end{cases}$ را در مقابل $\begin{cases} H_1 : \rho \neq 0 \\ H_1 : \beta \neq 0 \end{cases}$ رد می‌کنیم.

توجه: اگر گزینه «فرضیه $\rho = 0$ را در مقابل $\rho \neq 0$ رد می‌کنیم» در بین گزینه‌ها بود، این گزینه هم درست بود، چون آماره آزمون و نتایج آزمون معنی‌داری ضریب همبستگی و شیب خط رگرسیون یکسان‌اند.

مثال ۵ اگر در رگرسیون دومتغیره $y = b_0 + b_1 x + u$، شیب منحنی $\hat{b}_1 = 0.5$ و واریانس آن $S^2_{\hat{b}_1} = 0.04$ باشد، t محاسبه‌شده پارامتر برابر است با: (اقتصاد ـ ۷۴)

۱) $t = 12.5$ ۲) $t = 2.5$ ۳) $t = 0.125$ ۴) $t = 0.04$

حل: گزینه ۲ درست است.

$$\begin{cases} H_0 : b_1 = 0 \\ H_1 : b_1 \neq 0 \end{cases}$$

$$\begin{cases} t_{(n-2)} = \dfrac{\hat{b}_1}{S_{\hat{b}_1}} = \dfrac{0.5}{0.2} = 2.5 \\ \hat{b}_1 = 0.5 \,,\, S^2_{\hat{b}_1} = 0.04 \;\rightarrow\; S_{\hat{b}_1} = 0.2 \end{cases}$$

محاسبه انحراف معیار شیب خط با استفاده از آماره آزمون

با استفاده از آماره آزمون معنی‌دار بودن رابطه خطی، می‌توان رابطه دیگری برای محاسبه انحراف معیار شیب خط رگرسیون به دست آورد:

$$\frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \frac{b}{S_b} \;\rightarrow\; \boxed{S_b = \frac{b}{r}\sqrt{\frac{1-r^2}{n-2}}}$$

مثال رابطه بین مصرف (c) و درآمد قابل تصرف (y_d) بر اساس نمونه‌ای به حجم $n = 66$ به صورت روبه‌رو برآورد شده است:

$\hat{c}_t = 6 + 0.72 y_{dt}$

اگر ضریب همبستگی بین مصرف و درآمد قابل تصرف در نمونه 0.6 باشد، خطای معیار میل نهایی به مصرف $\left(S_{\hat{\beta}}\right)$ برابر است با: (اقتصاد ـ ۷۶)

۱) 0.075 ۲) 0.12 ۳) 1.2 ۴) 0.09

حل: گزینه ۲ درست است.

$$\begin{cases} t_{(n-2)} = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{\hat{\beta}}{S_{\hat{\beta}}} \rightarrow \dfrac{0.6}{\sqrt{\dfrac{1-0.36}{66-2}}} = \dfrac{0.72}{S_{\hat{\beta}}} \rightarrow S_{\hat{\beta}} = \dfrac{0.72}{6} = 0.12 \\ r = 0.6 \ , \ n = 66 \ , \ \hat{\beta} = 0.72 \end{cases}$$

۲ـ آزمون معنی‌دار بودن ثابت معادله

همان‌طورکه می‌دانیم در حالت‌های ۱ و ۳ از مدل‌های خط رگرسیون جامعه به فرم $y = \beta x + \alpha + \varepsilon$ یا $y = \alpha + \varepsilon$، مقدار α (عرض از مبدأ) حتماً مخالف صفر است.

حال هرگاه بخواهیم معنی‌دار بودن عرض از مبدأ $(\alpha \neq 0)$ را آزمون کنیم، از آماره آزمون t استفاده می‌کنیم.

در ادامه به شرح مراحل چهارگانه این آزمون می‌پردازیم.

۱- فرضیه‌های آزمون

$$\boxed{\begin{cases} H_0 : \alpha = 0 & \text{(عرض از مبدأ معنی‌دار نیست)} \\ H_1 : \alpha \neq 0 & \text{(عرض از مبدأ معنی‌دار است)} \end{cases}}$$

۲- آماره آزمون (ملاک آزمون)

آماره (ملاک) مناسب برای این آزمون دارای توزیع t است که با در نظر گرفتن فرض $H_0 : \alpha = 0$ به صورت زیر خواهد بود:

$$\boxed{t = \frac{a}{S_a}}$$

درجه آزادی آماره آزمون

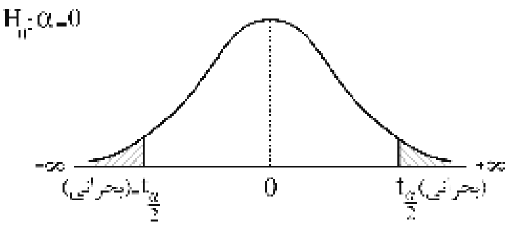
الف) اگر مدل رگرسیون خطی جامعه به فرم $y = \beta x + \alpha + \varepsilon$ (پیش‌فرض) باشد، درجه آزادی آماره آزمون t، $n-2$ است.

ب)اگر مدل رگرسیون خطی جامعه به فرم $y = \alpha + \varepsilon$ باشد، درجه آزادی آماره آزمون t، $n-1$ است.

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی

با توجه به فرض H_1، آزمون دوطرفه است، همچنین بر اساس توزیع آماره t با n-2 درجه آزادی در سطح خطای α، سطوح H_0 و H_1 به صورت زیر خواهد بود:

$$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases} \xrightarrow[\text{فرض } H_0 \text{ رد می‌شود.}]{U > t_{\frac{\alpha}{2}, n-2} \ \text{or} \ U < -t_{\frac{\alpha}{2}, n-2}}$$



۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:

اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

۳ـ آزمون معنی‌دار بودن معادله خط رگرسیون

همان‌طورکه می‌دانیم در حالت ۱ از مدل‌های خط رگرسیون جامعه به فرم $y=\beta x+\alpha+\varepsilon$، مقدار β (شیب خط) یا α (عرض از مبدأ) که از روابط زیر به دست می‌آیند، باید مخالف صفر باشند تا مدل رگرسیون $y=\beta x+\alpha+\varepsilon$ معنی‌دار باشد:

$$\begin{cases}\beta=\dfrac{\mathrm{Cov}(x,y)}{\sigma_x^2}=\rho_{x,y}\times\dfrac{\sigma_y}{\sigma_x}\\ \alpha=\mu_y-\beta\mu_x\end{cases}$$

حال هرگاه بخواهیم وجود معادله خط رگرسیون $(y=\beta x+\alpha+\varepsilon)$ را آزمون کنیم، از آنالیز واریانس رگرسیون (که بعداً توضیح داده خواهد شد) استفاده می‌کنیم. در این مورد آماره آزمون معنی‌داری رگرسیون، F با درجه آزادی 1 و $n-2$ است.

۱- فرضیه‌های آزمون

در آزمون معنی‌دار بودن معادله خط رگرسیون، معنی‌داری شیب خط و عرض از مبدأ همزمان بررسی می‌شود، بنابراین داریم:

$$\begin{cases}H_0:\alpha=\beta=0\\ H_1:\text{حداقل یکی از مقادیر } \alpha \text{ و } \beta \text{ مخالف صفر}\end{cases}$$

۲- آماره آزمون (ملاک آزمون)

ملاک آزمون معنی‌دار بودن معادله رگرسیون بر اساس تحلیل واریانس با دو تیمار (β,α) دارای توزیع فیشر با 1 و $(n-2)$ درجه آزادی است:

$$F_{1,n-2}=\left(t_{(n-2)}\right)^2=\frac{r^2}{\dfrac{1-r^2}{n-2}}$$

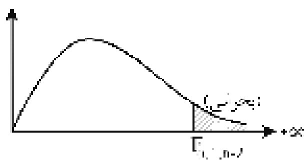
✓ **دقت کنید!**

الف) همان‌طورکه در انتهای فصل چهارم (توزیع‌های گسسته و پیوسته) بیان شد، توزیع فیشر با درجه آزادی 1 در صورت، برابر با مجذور توزیع t است. در اینجا توزیع t با $n-2$ درجه آزادی، همان ملاک آزمون معنی‌دار بودن شیب خط رگرسیون است.

ب) r ضریب همبستگی نمونه و $R^2=r^2$ ضریب تعیین است، بنابراین تنها با داشتن ضریب همبستگی یا ضریب تعیین و تعداد نمونه مقدار آماره آزمون معنی‌دار بودن معادله رگرسیون به دست می‌آید.

در مبحث آنالیز واریانس رگرسیون، نمایش دیگری از ملاک این آزمون را خواهیم دید.

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی



$$\begin{cases}H_0:\alpha=\beta=0\\ H_1:\text{حداقل یکی از } \alpha \text{ و } \beta \text{ مخالف صفر}\end{cases}$$

$\xrightarrow{U>F_{\alpha,1,n-2}}$ فرض H_0 رد می‌شود.

۴- تصمیم‌گیری

در این مرحله مقدار ملاک (آماره) آزمون محاسبه‌شده در مرحله دوم را با مقادیر بحرانی در مرحله سوم مقایسه می‌کنیم:

اگر ملاک آزمون در ناحیه اطمینان (پذیرش H_0) قرار گیرد، فرض H_0 در سطح اطمینان $1-\alpha$ پذیرفته می‌شود، اما اگر ملاک آزمون در ناحیه بحرانی (معنی‌دار) قرار گیرد، فرض H_0 رد شده و فرض H_1 در سطح خطای α پذیرفته می‌شود و اصطلاحاً می‌گوییم اختلاف معنی‌دار است.

مثال رابطه بین مصرف (c) و درآمد قابل تصرف (y_d) بر اساس یک نمونه تصادفی 32 تایی به صورت روبه‌رو برآورد شده است:

$\hat{c} = 2.6 + 0.85\, y_d \;\; , \;\; R^2 = 0.9$

آماره آزمون برای آزمون معنی‌دار بودن رگرسیون کدام است؟ (اقتصاد ـ ۸۴)

۱) $F_{1,30} = 270$ ۲) $F_{1,30} = 150$ ۳) $\chi^2_{(30)} = 15$ ۴) $t_{(30)} = 7.2$

حل: گزینه ۱ درست است.

$$\begin{cases} F_{1,n-2} = \left(t_{(n-2)}\right)^2 = \left(\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}\right)^2 = \dfrac{r^2}{\dfrac{1-r^2}{n-2}} \rightarrow F_{1,30} = \dfrac{0.9}{\dfrac{1-0.9}{32-2}} = \dfrac{27}{0.1} = 270 \\ r^2 = 0.9 \;\; , n = 32 \end{cases}$$

اگر $t_{(30)} = \sqrt{270}$ در میان گزینه‌ها وجود داشت، درست نبود، چون آماره آزمون باید F باشد.

دقت کنید، با توجه به مقدار آماره آزمون، واضح است که فرض H_0 رد می‌شود (چون مقدار آن خیلی بزرگ است) این بدان مفهوم است که معادله رگرسیون معنی‌دار است؛ به عبارت دیگر، می‌توان به یکی از نتایج زیر رسید:

اولاً، به ازای یک واحد افزایش در درآمد قابل تصرف، مصرف به میزان 0.85 واحد افزایش می‌یابد.

ثانیاً، در صورت صفر بودن درآمد قابل تصرف، مصرف به میزان ثابت 2.6 است.

آماره آزمون معادله رگرسیون k متغیره

در رگرسیون k متغیره که شامل $(k-1)$ متغیر مستقل x و یک متغیر وابسته y است، رابطه بین آماره F و R^2 (ضریب تعیین) عبارت است از:

$$\boxed{F_{k-1,n-k} = \dfrac{\dfrac{R^2}{k-1}}{\dfrac{1-R^2}{n-k}}}$$

نتیجه: درصورتی‌که رگرسیون خطی ساده با یک متغیر مستقل x و یک متغیر وابسته داشته باشیم $(\hat{y} = a + bx + e)$ با قرار دادن $k = 2$ در معادله بالا به همان آماره معروف می‌رسیم:

$$F_{1,n-2} = \dfrac{r^2}{\dfrac{1-r^2}{n-2}} = \left(\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}\right)^2 = \left(t_{(n-2)}\right)^2$$

مثال درصورتی‌که در یک رگرسیون با دو متغیر مستقل که دارای 23 مورد است $(N = 23)$، همبستگی متغیرهای مستقل با متغیر وابسته (0.8) باشد، مقدار عددی F برابر است با: (برنامه‌ریزی شهری ـ ۸۳)

۱) 1.782 ۲) 8.78 ۳) 12.78 ۴) 17.8

حل: گزینه ۴ درست است.

در این سؤال، $k = 3$ است. دو متغیر مستقل و یک متغیر وابسته داریم و $R^2 = (0.8)^2 = 0.64$ است، بنابراین:

$$F_{k-1,n-k} = \dfrac{\dfrac{R^2}{k-1}}{\dfrac{1-R^2}{N-k}} \rightarrow F_{2,20} = \dfrac{\dfrac{0.64}{3-1}}{\dfrac{1-0.64}{23-3}} = \dfrac{0.32}{0.018} = 17.78$$

یادآوری:

رگرسیون خطی چندمتغیره

هرگاه $x_1, x_2, \dots, x_{k-1}$ ، $k-1$ متغیر مستقل (آزاد) و y متغیر وابسته (تابع) به آن‌ها باشد، معادله زیر یک رگرسیون خطی k متغیره است:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} + \varepsilon$$

که ما در این بخش مدل رگرسیون دو متغیره خطی ساده را به شکل زیر مورد بررسی قرار دادیم:

$$y = \beta x + \alpha + \varepsilon$$

ارتباط ملاک‌های t و F در خط رگرسیون

ابتدا یک بار جدول مربوط به آزمون‌های مختلف خط رگرسیون را مرور می‌کنیم:

آماره (ملاک) آزمون	فرض‌ها	آزمون
t	$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$ (معنی‌دار بودن)	معنی‌دار بودن شیب خط (رابطه خطی)
t	$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases}$ (معنی‌دار بودن)	معنی‌دار بودن ثابت معادله (عرض از مبدأ)
$F_{1,\ n-2} = (t_{n-2})^2$	$H_0 : \alpha = \beta = 0$ H_1 : حداقل یکی مخالف صفر (معنی‌دار بودن)	معنی‌دار بودن معادله رگرسیون

نتیجه ۱:

هرگاه حداقل یکی از دو آزمون شیب خط یا ثابت معادله معنی‌دار شود (آماره t_{n-2} در ناحیه بحرانی و H_1 پذیرفته شود)، آن‌گاه حتماً آزمون معادله رگرسیون معنی‌دار می‌شود (آماره $F_{1,\ n-2}$ در ناحیه بحرانی و H_1 پذیرفته می‌شود)؛ به عبارت دیگر:

t_{n-2} در ناحیه بحرانی باشد. $\rightarrow \begin{cases} H_1 : \beta \neq 0 \quad \text{پذیرفته می‌شود.} \\ H_1 : \alpha \neq 0 \quad \text{پذیرفته می‌شود.} \end{cases} \rightarrow$ $F_{1,\ n-2}$ در ناحیه بحرانی خواهد بود.

نتیجه ۲:

هرگاه آزمون معادله رگرسیون معنی‌دار شود (آماره $F_{1,\ n-2}$ در ناحیه بحرانی و H_1 پذیرفته شود)، آن‌گاه حداقل یکی از مقادیر α یا β مخالف صفر است، درنتیجه حداقل یکی از دو آزمون شیب خط یا ثابت معادله معنی‌دار خواهد شد (آماره t_{n-2} در ناحیه بحرانی و H_1 پذیرفته می‌شود)؛ به عبارت دیگر:

$F_{1, n-2}$ در ناحیه بحرانی باشد. $\rightarrow$ حداقل یکی از α و β مخالف صفر $\rightarrow$ حداقل یکی از آماره‌های t_{n-2} در ناحیه بحرانی است.

مثال در یک رگرسیون دومتغیره، اگر آماره t مربوط به شیب در ناحیه بحرانی قرار گیرد، در خصوص آماره F و نتیجه آزمون کدام گزینه صحیح است؟ (اقتصاد ـ ۸۴)

۱) مقدار آماره F نیز در ناحیه بحرانی قرار گرفته و H_0 رد می‌شود.

۲) آماره F در ناحیه بحرانی قرار گرفته و H_a رد می‌شود.

۳) آماره F در ناحیه بحرانی قرار نگرفته و بین متغیرهای مستقل و وابسته رابطه وجود دارد.

۴) آماره F در ناحیه بحرانی قرار نگرفته و بین متغیرهای مستقل و وابسته رابطه وجود ندارد.

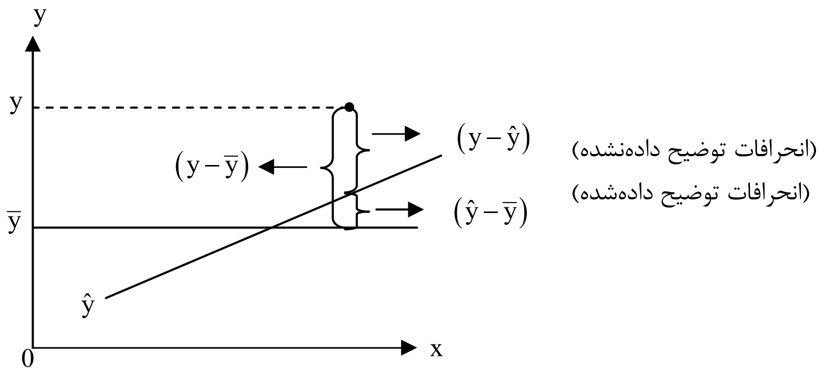
حل: گزینه ۱ درست است.

بنا بر نتیجه۱، ازآنجاکه آماره t در ناحیه بحرانی قرار گرفته است، فرضیه $H_0: \beta = 0$ رد شده و شیب خط رگرسیون معنی‌دار است، پس معادله رگرسیون نیز معنی‌دار شده و آماره F در ناحیه بحرانی قرار می‌گیرد. (منظور از H_a همان فرض H_1 است.)

آنالیز واریانس رگرسیون

اگر $y_1, y_2, \dots, y_n$ مشاهدات نمونه و $\bar{y}$ میانگین مشاهدات باشد، آن‌گاه مجموع کل مربعات تصحیح‌شده مشاهدات حول میانگین، $\sum (y-\bar{y})^2$ است. حال اگر خط رگرسیون برآوردی $(\hat{y})$ را در نظر بگیریم، می‌توانیم مجموع مربعات کل حول میانگین $\sum (y-\bar{y})^2$ را به دو قسمت تجزیه کنیم:

شکل زیر انحراف یکی از مشاهدات (y) را از میانگین مشاهدات $\bar{y}$ و خط رگرسیون $\hat{y}$ نشان می‌دهد:



مشاهده می‌شود:

$$SST = SS_t = \sum (y-\bar{y})^2 \quad \text{(بیان تغییرپذیری کل مشاهدات)}$$

$$SSE = SS_{res} = \sum (y-\hat{y})^2 \quad + \quad SS_{reg} = \sum (\hat{y}-\bar{y})^2$$

نماد	مفهوم	توضیحات
$SS_{reg} = \sum (\hat{y}-\bar{y})^2$	بیان تغییرپذیری مشاهداتی است که توسط خط رگرسیون $(\hat{y})$ توضیح داده می‌شوند (مجموع مربعات رگرسیونی)	مقداری بین صفر و SS_t دارد. $0 \le SS_{reg} \le SS_t$
$SS_{res} = \sum (y-\hat{y})^2$	بیان تغییرپذیری باقی‌مانده‌ها (خطا)، است که توسط خط رگرسیون $(\hat{y})$ توضیح داده نشده است (پراکندگی نقاط حول خط رگرسیون)	مقداری بین صفر و SS_t دارد. $0 \le SS_{res} \le SS_t$

$$SS_t = \sum (y-\bar{y})^2 = SS_y = S_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

$$SS_{reg} = \sum (\hat{y}-\bar{y})^2 = \sum b^2 (x-\bar{x})^2 = b^2 SS_x = b\sum (x-\bar{x})(y-\bar{y}) = bSp_{xy}$$

$$SS_{res} = \sum (y-\hat{y})^2 = SS_y - b^2 SS_x = SS_y - bSp_{xy}$$

محاسبه انحراف معیار خطای برآورد

$$S_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} = \sqrt{\frac{\sum e_i^2}{n-2}} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{SST-SSR}{n-2}}$$

محاسبه ضریب تعیین

$$\text{ضریب تعیین } R^2 = \left(r_{x,y}\right)^2 = \frac{SS_{reg}}{SS_t} = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{bSp_{xy}}{SS_y} = \frac{Sp_{xy}^2}{SS_x SS_y}$$

البته می‌توان ضریب تعیین R^2 را به صورت زیر نیز نوشت:

$$R^2 = \frac{SS_{reg}}{SS_t} = \frac{SS_t - SS_{res}}{SS_t} = 1 - \frac{SS_{res}}{SS_t} = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

حال با توجه به مفهوم ضریب تعیین داریم:

« R^2 : نسبت مشاهدات توضیح داده‌شده توسط خط رگرسیون» برابر است با:

$$R^2 = \frac{SS_{reg}}{SS_t} = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}$$

« $1-R^2$: نسبت مشاهدات توضیح داده‌نشده توسط خط رگرسیون» برابر است با:

$$1-R^2 = 1-\frac{SS_{reg}}{SS_t} = \frac{SS_{res}}{SS_t} = \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

حالات خاص

۱) $SS_t = SS_{reg} \quad \rightarrow \quad \sum(y-\hat{y})^2 = SS_{res} = 0$

در این حالت تمام مشاهدات توسط خط رگرسیون توضیح داده شده‌اند، درنتیجه $R^2 = 1$ است؛ به عبارت دیگر، خط رگرسیون توانسته تغییرات y را به طور کامل به تغییرات متغیر x نسبت دهد.

۲) $SS_{res} = SS_t \quad \rightarrow \quad \sum(\hat{y}-\bar{y})^2 = SS_{reg} = 0$

در این حالت هیچ مشاهده‌ای توسط خط رگرسیون توضیح داده نشده است، درنتیجه $R^2 = 0$ است؛ به عبارت دیگر، خط رگرسیون هرگز نتوانسته است تغییرات y را به تغییرات متغیر مستقل x نسبت دهد (هیچ مشاهده‌ای توسط خط رگرسیون توضیح داده نشده است)، بنابراین دو متغیر، رابطه غیرخطی دارند یا مستقل‌اند.

مثال اگر رگرسیون $y = \alpha + \beta x + \varepsilon$ را در نظر بگیریم، اگر در نمونه $n = 12$ تایی از جامعه نرمال دوبعدی، اطلاعات زیر در اختیار باشد:

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = 300 \ , \ \sum(x_i - \bar{x})^2 = 120 \ , \ \sum e_i^2 = SSD_E = 240$$

تخمین‌های β و ρ^2 کدام است؟ (اقتصاد ـ ۷۱)

۱) $(2.5 , 0.76)$ ۲) $(-2.5, 0.76)$ ۳) $(2.5 , 0.85)$ ۴) $(1.5 , 0.76)$

حل: گزینه ۱ درست است.

$$\begin{cases} \text{شیب خط رگرسیون}: \hat{\beta} = \dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \dfrac{SP_{xy}}{SS_x} = \dfrac{300}{120} = 2.5 \\ \text{ضریب تعیین}: \rho^2 = R^2 = \dfrac{SS(reg)}{SST} = \dfrac{750}{990} = 0.7575 \simeq 0.76 \\ SS(reg) = \sum(\hat{y}-\bar{y})^2 = b^2 SS_x = b.SP_{xy} = 2.5\times 300 = 750 \\ \sum e_i^2 = SSE = 240 \\ SST = \sum(y-\bar{y})^2 = SS_y = SS(reg) + SS_E \to SS_y = 750 + 240 = 990 \end{cases}$$

آزمون معنی‌داری معادله رگرسیون با استفاده از آنالیز واریانس

همان‌طورکه قبلاً نیز اشاره شد، آماره آزمون معنی‌دار بودن وجود معادله رگرسیون $(y=\alpha+\beta x)$ F با 1 و $n-2$ درجه آزادی است که از آنالیز واریانس رگرسیون به دست می‌آید. البته از طریق مجذور آزمون t نیز به همین جواب خواهیم رسید.

۱- فرضیه‌های آزمون

$$\begin{cases} H_0: \alpha = \beta = 0 \\ H_1: \text{حداقل یکی مخالف صفر} \end{cases}$$

۲- آماره آزمون (ملاک آزمون)

$$F_{1,n-2} = \dfrac{\sum(\hat{y}-\bar{y})^2}{\dfrac{\sum(y-\hat{y})^2}{n-2}} = \dfrac{\dfrac{SS_{reg}}{1}}{\dfrac{SS_{res}}{n-2}} = \dfrac{\dfrac{\chi^2_{(1)}}{1}}{\dfrac{\chi^2_{(n-2)}}{n-2}} = \dfrac{MS_{reg}}{MS_{res}}$$

می‌دانیم که از تقسیم دو توزیع χ^2 مستقل از هم به درجه آزادی‌هایشان توزیع F به دست می‌آید.

توزیع مجموع مربعات رگرسیون (SS_{reg})، χ^2 با یک درجه آزادی است.

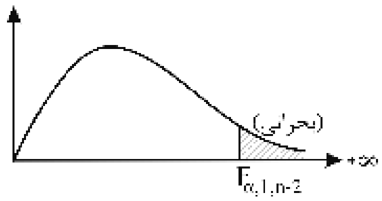
توزیع مجموع مربعات باقی‌مانده (SS_{res})، χ^2 با $n-2$ درجه آزادی است.

بنابراین حاصل تقسیم میانگین مربعات رگرسیون به میانگین مربعات خطا، توزیع F با 1 و $n-2$ درجه آزادی است.

جدول آنالیز واریانس رگرسیون

منشأ (منبع) تغییرات	درجه آزادی df	مجموع مجذور انحرافات SS	میانگین مجذور انحرافات (واریانس $MS=S^2$)	آماره آزمون $F_{1,n-2}$
رگرسیون	1	$SS(Tr)=\sum(\hat{y}-\bar{y})^2$	$MS(Tr)=SS(Tr)$	$\dfrac{MS(Tr)}{MSE}$
خطا یا باقی‌مانده	$n-2$	$SSE=\sum(y-\hat{y})^2$	$MSE=\dfrac{SSE}{n-2}$	
جمع کل (T)	$n-1$	$SST=\sum(y-\bar{y})^2$		

۳- تعیین سطح H_0 و H_1 و مقدار بحرانی



فرض H_0 رد می‌شود. $\xrightarrow{U>F_{\alpha,1,n-2}}$

$$\begin{cases} H_0: \alpha = \beta = 0 \\ H_1: \text{حداقل یکی از } \alpha \text{ و } \beta \text{ مخالف صفر} \end{cases}$$

مثال ۱ در تجزیه واریانس (پراش) رگرسیونی می‌توان معنی‌دار بودن: (اقتصاد ـ ۷۱)

۱) پارامترها را بررسی کرد. ۲) وجود معادله رگرسیون را بررسی کرد.

۳) پارامترها و رگرسیون را بررسی کرد. ۴) هیچ‌کدام

حل: گزینه ۲ درست است.

مثال ۲ برای آزمون معنی‌دار بودن معادله رگرسیون در یک جامعه نرمال دوبعدی نمونه‌ای به حجم $n = 15$ خانوار انتخاب و بر اساس آن‌ها درآمد (x) و هزینه خانوارها (y) اندازه‌گیری شده است و بر اساس نتایج مشاهدات کمیت‌های زیر به دست آمده است:
$b_1 = 0.8\ ,\ \sum(x-\bar{x})^2 = SSD_x = 25\ ,\ \sum(y-\bar{y})^2 = SSD_y = 36$

مقدار عددی ملاک آزمون‌کننده (آماره آزمون) برای آزمون معنی‌دار بودن ضرایب معادله رگرسیون تابع مصرف کوتاه‌مدت کدام است؟ (اقتصاد ـ ۷۷)

۱) 0.44 ۲) 5.44 ۳) 10.45 ۴) 13.07

حل: گزینه ۳ درست است.

$$\begin{cases} F_{1,13} = \dfrac{SS(reg)}{\dfrac{SSE}{n-2}} = \dfrac{16}{\dfrac{20}{13}} = 10.4 \\ SS(reg) = b^2 SS_x = (0.8)^2 \times 25 = 16 \\ SSE = SS_y - SS(reg) = 36 - 16 = 20 \end{cases}$$

جدول خلاصه آزمون فرض‌های آماری

آزمون	فرض H_0	آماره (ملاک) آزمون
برابری میانگین جامعه با یک عدد خاص	$H_0: \mu \underset{\geq}{\overset{\leq}{=}} \mu_0$	$Z = \dfrac{\bar{x}-\mu}{\frac{\sigma \text{ یا } S}{\sqrt{n}}}$ یا t
برابری میانگین دو جامعه یا تفاضل میانگین دو جامعه	$H_0: \mu_1 \underset{\geq}{\overset{\leq}{=}} \mu_2$ $H_0: \mu_1-\mu_2 \underset{\geq}{\overset{\leq}{=}} 0$ یا عدد خاص	$Z = \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2 \text{ یا } S_1^2}{n_1}+\frac{\sigma_2^2 \text{ یا } S_2^2}{n_2}}}$ یا t $t = \dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ یا
برابری نسبت جامعه با یک عدد خاص	$H_0: p \underset{\geq}{\overset{\leq}{=}} p_0$	$Z = \dfrac{\bar{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$
برابری نسبت دو جامعه یا تفاضل نسبت دو جامعه	$H_0: p_1 \underset{\geq}{\overset{\leq}{=}} p_2$ $H_0: p_1-p_2 \underset{\geq}{\overset{\leq}{=}} 0$ یا عدد خاص	$Z = \dfrac{(\bar{p}_1-\bar{p}_2)-(p_1-p_2)}{\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1}+\frac{\bar{p}_2\bar{q}_2}{n_2}}}$
برابری واریانس جامعه با یک عدد خاص	$H_0: \sigma^2 \underset{\geq}{\overset{\leq}{=}} \sigma_0^2$	μ نامعلوم $\chi_{n-1}^2 = \dfrac{(n-1)S^2}{\sigma^2}$ یا μ معلوم $\chi_n^2 = \dfrac{nS^2}{\sigma^2}$
برابری واریانس دو جامعه یا نسبت واریانس دو جامعه	$H_0: \sigma_1^2 \underset{\geq}{\overset{\leq}{=}} \sigma_2^2$ $H_0: \dfrac{\sigma_1^2}{\sigma_2^2} \underset{\geq}{\overset{\leq}{=}} 1$ یا عدد خاص	$F = \dfrac{S_1^2}{S_2^2}$ یا $F = \dfrac{S_1^2}{S_2^2}\cdot\dfrac{\sigma_2^2}{\sigma_1^2}$

آزمون	فرض H_0	آماره (ملاک) آزمون
نیکویی برازش	آیا توزیع مشاهدات ... است : H_0	$\chi^2_{(k-1)}=\sum_{i=1}^{k}\frac{\left(F_{o_i}-F_{e_i}\right)^2}{F_{e_i}}$
استقلال دو متغیر	آیا X و Y مستقل‌اند : H_0	$\chi^2_{(r-1)(c-1)}=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(F_{o_{ij}}-F_{e_{ij}}\right)^2}{F_{e_{ij}}}$
برابری میانگین $k>2$ جامعه (بیش از دو جامعه)	$H_0:\mu_1=\mu_2=\ldots=\mu_k$	$F_{k-1,k(n-1)}=\frac{\frac{SS(Tr)}{k-1}}{\frac{SSE}{k(n-1)}}$
معنی‌دار بودن ضریب همبستگی صفر بودن ضریب همبستگی معنی‌دار بودن شیب خط صفر بودن شیب خط معنی‌دار بودن ضریب معادله	$H_0:\rho=0$ یا $H_0:\beta=0$	$t_{n-2}=\frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ یا $t_{n-2}=\frac{b}{S_b}$
معنی‌دار بودن ضریب ثابت خط صفر بودن عرض از مبدأ	$H_0:\alpha=0$	$t_{n-2}=\frac{a}{S_a}$
معنی‌دار بودن معادله (خط) صفر بودن شیب خط و عرض از مبدأ همزمان	$H_0:\alpha=\beta=0$	$F_{1,n-2}=t^2=\frac{r^2}{\frac{1-r^2}{n-2}}$ یا $F_{1,n-2}=\frac{SS_{reg}}{\frac{SSE}{n-2}}$ فیشر

مباحث اضافی در آزمون فرض

محاسبه خطاهای نوع اول، نوع دوم و توان آزمون در سر فصل‌های کنکوری رشته‌های علوم انسانی وجود ندارد، اما این مباحث در کنکور رشته محیط زیست مطرح شده است؛ ازاین‌رو تنها برای داوطلبان محیط زیست بیان می‌شود.

محاسبه خطای نوع اول

همان‌طورکه می‌دانیم، خطای نوع اول عبارت است از احتمال رد فرض H_0 (وقوع ناحیه بحرانی) در شرایطی که H_0 به‌واقع درست است، بنابراین برای محاسبه احتمال مورد نظر کافی است احتمال وقوع ناحیه بحرانی (C) را برای H_0 به دست آوریم، درنتیجه:

$$\alpha = P(\text{رد } H_0 \mid \text{درست } H_0) = P(\text{ناحیه بحرانی} \mid \text{درست } H_0)$$

به عبارت دیگر:

$$\boxed{\alpha = P_{H_0}(\text{ناحیه بحرانی})}$$

مثال ۱ متغیر تصادفی X دارای تابع چگالی احتمال $0 \le x \le 1 \;,\; \theta > 0 \;;\; f(x, \theta) = (1+\theta)x^{\theta}$ است. فرض صفر (یا خنثی) را به صورت $H_0 : \theta = 2$ و ناحیه بحرانی (یعنی مکمل ناحیه پذیرش) را به صورت $x < 0.25$ در نظر بگیرید. خطای نوع اول برابر است با:

۱) $\frac{1}{16}$ ۲) $\frac{1}{64}$ ۳) $\frac{1}{4}$ ۴) $\frac{3}{4}$

حل: گزینه ۲ درست است.

با توجه به ناحیه بحرانی $C = \left\{x < 0.25 = \frac{1}{4}\right\}$ و فرض صفر $H_0 : \theta = 2$ برای محاسبه احتمال خطای نوع اول (α) به صورت زیر عمل می‌کنیم:

$$\begin{cases} \alpha = P_{H_0}(\text{ناحیه بحرانی}) = P_{\theta=2}\left(x < \frac{1}{4}\right) = \int_0^{\frac{1}{4}} (1+2)x^2\,dx = \left[x^3\right]_0^{\frac{1}{4}} = \frac{1}{64} \\ f(x,\theta) = (1+\theta)x^{\theta} \;;\; 0 \le x \le 1 \end{cases}$$

مثال ۲ فرض کنید X دارای توزیع دوجمله‌ای با پارامترهای 4 و p باشد. برای آزمون فرض $H_0 : p = \frac{1}{3}$ در برابر $H_1 : p = \frac{1}{5}$، اگر $x = 0$ یا $x = 1$ ملاک رد کردن فرض H_0 باشد، یعنی $C = \{0 \,,\, 1\}$، احتمال خطای نوع اول چقدر است؟

۱) $\left[\frac{2}{3}\right]^4$ ۲) $\left[\frac{2}{3}\right]^3$ ۳) $3\left[\frac{2}{3}\right]^3$ ۴) $3\left[\frac{2}{3}\right]^4$

حل: گزینه ۴ درست است.

با توجه به ناحیه بحرانی $C = \{x = 0 \text{ یا } x = 1\}$ و فرض صفر $H_0 : p = \frac{1}{3}$، برای محاسبه احتمال خطای نوع اول (α) به صورت زیر عمل می‌کنیم:

$$P(x) = \binom{4}{x} p^x q^{4-x}$$

$$\alpha = P_{H_0}(\text{ناحیه بحرانی}) = P_{p=\frac{1}{3}}(X=0) + P_{p=\frac{1}{3}}(X=1) = \binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^4 + \binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^3 = \left(\frac{2}{3}\right)^4 + 2\left(\frac{2}{3}\right)^4 = 3\left(\frac{2}{3}\right)^4$$

مثال ۳ متغیر تصادفی X توزیع پواسون با میانگین نامعلوم θ دارد. برای آزمون فرض $H_0: \theta = 0.5$ در برابر $H_1: \theta > 0.5$، نمونه تصادفی X_1, X_2 را می‌گیریم و معدل نمونه یعنی $\bar{X}$ را تعریف می‌کنیم تا اگر $\bar{x} \le 1$ باشد فرض صفر را بپذیریم. در این‌صورت احتمال ارتکاب خطای نوع اول چقدر است؟

۱) $\frac{2}{e}$ ۲) $\frac{5}{2e}$ ۳) $\frac{1}{e}(e-2)$ ۴) $\frac{1}{2}\left(2-\frac{5}{e}\right)$

حل: گزینه ۴ درست است.

۱- درصورتی‌که X دارای توزیع پواسون با پارامتر θ باشد، آن‌گاه:

$$\begin{cases} f(x) = \dfrac{e^{-\theta}\theta^x}{x!} & ; \; x = 0,1,2,\ldots \\ E(X) = \theta & ; \sigma_X^2 = \theta \end{cases}$$

۲- درصورتی‌که x_1, x_2 نمونه‌های مستقل از جامعه‌ای پواسون با پارامتر θ باشند، آن‌گاه $X = X_1 + X_2$ دارای توزیع پواسون با پارامتر 2θ است:

$$X_1, X_2, \ldots, X_n \sim P(\lambda = \theta) \rightarrow \sum_{i=1}^{n} x_i \sim P(\lambda = n\theta)$$

در این مثال با فرض آنکه $\bar{x} = \frac{x_1 + x_2}{2}$ میانگین یک نمونه دوتایی مستقل از جامعه پواسون با پارامتر $\theta = \frac{1}{2}$ باشد، آن‌گاه $X = X_1 + X_2$ نیز دارای توزیع پواسون با پارامتر $2\theta = 1$ است:

$$\alpha = P_{H_0}(\text{ناحیه بحرانی}) = P_{\theta=\frac{1}{2}}(\bar{X} > 1) = P\left(\frac{X_1 + X_2}{2} > 1\right) = P(X_1 + X_2 > 2) = P_{\theta=1}(X > 2) = 1 - P(X \le 2)$$

$$= 1 - P(X=0) - P(X=1) - P(X=2) = 1 - \frac{e^{-1}1^0}{0!} - \frac{e^{-1}1^1}{1!} - \frac{e^{-1}1^2}{2!} = 1 - \frac{5}{2}e^{-1} = \frac{1}{2}\left(2 - \frac{5}{e}\right)$$

محاسبه خطای نوع دوم

همان‌طورکه می‌دانیم خطای نوع دوم عبارت است از احتمال پذیرش H_0 (وقوع ناحیه اطمینان) در شرایطی که H_0 به‌واقع نادرست است (H_1 به‌واقع درست است)، بنابراین برای محاسبه احتمال مورد نظر کافی است احتمال وقوع ناحیه اطمینان (C') را برای H_1 به دست آوریم، درنتیجه:

$$\beta = P(H_0 \text{ پذیرش} \mid H_0 \text{ نادرست}) = P(\text{ناحیه اطمینان} \mid H_1 \text{ درست})$$

به عبارت دیگر:

$$\boxed{\beta = P_{H_1}(\text{ناحیه اطمینان})}$$

مثال ۱ تابع توزیع احتمال برای متغیر تصادفی X به صورت زیر است:

X	-1	b	4
$P(x)$	0.25	0.35	0.4

برای آزمون $H_0: b = 0$ در مقابل $b > 0$ درصورتی‌که بدانیم $b = 3$ است، مقدار β (احتمال خطای نوع دوم) را به دست آورید. (ناحیه بحرانی $x > 2$ است.)

۱) 0.25 ۲) 0.4 ۳) 0.75 ۴) 0.05

حل: گزینه ۱ درست است.

با توجه به ناحیه اطمینان به صورت $C' = \{x \le 2\}$ و فرض $H_1 : b > 0$ که در اینجا $b = 3$ در نظر گرفته شده است، برای محاسبه احتمال خطای نوع دوم (β) به صورت زیر عمل می‌کنیم:

$$\beta = P_{H_1}(\text{ناحیه اطمینان}) = P_{b=3}(X \le 2) = P(X = -1) = 0.25$$

X	-1	$\overset{3}{\cancel{b}}$	4
$P(x)$	0.25	0.35	0.4

توجه کنید، اگر در این سؤال محاسبه احتمال خطای نوع اول (α) مطرح بود، آن‌گاه با توجه به ناحیه بحرانی $C = \{x > 2\}$ و فرض صفر $H_0 : b = 0$ گزینه ۴ درست بود، زیرا:

$$P_{H_0}(\text{ناحیه بحرانی}) = P_{b=0}(X > 2) = P(X = 4) = 0.4$$

X	-1	$\overset{0}{\cancel{b}}$	4
$P(x)$	0.25	0.35	0.4

مثال ۲ مدل آماری

X	1	2	3
$f(x)$	$\frac{1-\theta}{3}$	$\frac{\theta}{3}$	$\frac{2}{3}$

مفروض است. بر اساس یک مشاهده می‌خواهیم فرض $H_0 : \theta = \frac{1}{3}$ را در مقابل $H_1 : \theta = \frac{2}{3}$ آزمون کنیم. احتمال خطای نوع دوم وقتی $x = 2\ ,\ 3$ فرض H_0 را رد می‌کند، برابر است با:

۱) $\frac{1}{3}$ ۲) $\frac{2}{3}$ ۳) $\frac{1}{9}$ ۴) $\frac{5}{9}$

حل: گزینه ۳ درست است.

با توجه به ناحیه اطمینان به صورت $C' = \{x = 1\}$ و فرض $H_1 : \theta = \frac{2}{3}$، برای محاسبه احتمال خطای نوع دوم (β) به صورت زیر عمل می‌کنیم:

$$\beta = P_{H_1}(\text{ناحیه اطمینان}) = P_{\theta=\frac{2}{3}}(X = 1) = \frac{1 - \overset{2/3}{\cancel{\theta}}}{3} = \frac{1}{9}$$

توجه: اگر در این سؤال محاسبه احتمال خطای نوع اول (α) مطرح بود، آن‌گاه با توجه به ناحیه بحرانی $C = \{x = 2,3\}$ و فرض صفر $H_0 : \theta = \frac{1}{3}$ گزینه ۴ درست بود، زیرا:

$$\alpha = P_{H_0}(\text{ناحیه بحرانی}) = P_{\theta=\frac{1}{3}}(X = 2) + P_{\theta=\frac{1}{3}}(X = 3) = \frac{\theta}{3} + \frac{2}{3} = \frac{\frac{1}{3}}{3} + \frac{2}{3} = \frac{1}{9} + \frac{4}{9} = \frac{5}{9}$$

مثال ۳ فرض کنید X_2, X_1 یک نمونه تصادفی دوتایی از توزیع برنولی $B(1,p)$ باشد. می‌خواهیم فرض زیر را آزمون کنیم:

$$H_0 : p = \frac{1}{4} \ \text{VS}\ H_1 : p = \frac{3}{4}$$

اگر ناحیه پذیرش ۱ یا $x_1 + x_2 = 0$، α نمایانگر احتمال خطای نوع اول و β احتمال خطای نوع دوم باشد، مقدار $(\alpha\ ,\ \beta)$ کدام است؟

۱) $\left(\frac{10}{16}, \frac{7}{16}\right)$ ۲) $\left(\frac{1}{16}, \frac{7}{16}\right)$ ۳) $\left(\frac{1}{16}, \frac{9}{16}\right)$ ۴) $\left(\frac{10}{16}, \frac{9}{16}\right)$

حل: گزینه ۲ درست است.

یادآوری: اگر $X_1, X_2, ..., X_n$ به صورت مستقل از هم دارای توزیع برنولی $B(1,P)$ باشند، آن‌گاه $x = \sum_{i=1}^{n} x_i$ دارای توزیع دوجمله‌ای به صورت $Bin(n,p)$ است:

$$X = \sum_{i=1}^{n} x_i = X_1 + x_2 + ... + X_n \sim Bin(n,p)$$

$$f(x) = P(x) = \binom{n}{x} p^x q^{n-x} \quad ; \quad x = 0,1,2,..,n$$

پس در این سؤال $x = x_1 + x_2$ دارای توزیع دوجمله‌ای $Bin(2,p)$ است.

$$f(x) = P(x) = \binom{2}{x} p^x q^{2-x} \quad ; \quad x = 0,1,2$$

با توجه به ناحیه بحرانی (رد) $c = \{x = 2\}$، ناحیه اطمینان (پذیرش) $c' = \{x = 0,1\}$ و نیز فرض صفر $H_0 : p = \frac{1}{4}$ و فرض مقابل $H_1 : p = \frac{3}{4}$ داریم:

$$\alpha = P_{H_0}(\text{ناحیه بحرانی}) = P_{p=\frac{1}{4}}(X = 2) = \binom{2}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^0 = \frac{1}{16}$$

$$\beta = P_{H_1}(\text{ناحیه اطمینان}) = P_{p=\frac{3}{4}}(X = 0,1) = \binom{2}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^2 + \binom{2}{1}\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^1 = \frac{1}{16} + \frac{6}{16} = \frac{7}{16}$$

محاسبه توان آزمون

ازآنجاکه توان آزمون (β^*) برابر است با $1-\beta$، برای محاسبه آن می‌توانیم یکی از روش‌های زیر را انتخاب کنیم:

روش اول: ابتدا خطای نوع دوم (β) را محاسبه می‌کنیم، سپس طبق تعریف توان آزمون آن را به صورت زیر به دست می‌آوریم:

$$\text{توان آزمون}: \beta^* = 1 - \beta$$

روش دوم: با توجه به تعریف توان آزمون (β^*) به صورت زیر، آن را محاسبه می‌کنیم:

$$\beta^* = P(H_0 \text{ رد} \mid H_1 \text{ درست}) = P(\text{ناحیه بحرانی} \mid H_1 \text{ درست})$$

به عبارت دیگر:

$$\boxed{\beta^* = P_{H_1}(\text{ناحیه بحرانی})}$$

مثال متغیر تصادفی X دارای تابع چگالی احتمال زیر است:

$$f(x, \theta) = (1+\theta)x^{\theta} \quad ; \quad \theta > 0 \quad , \quad 0 \le x \le 1$$

فرض صفر (یا خنثی) را به صورت $H_0 : \theta = 2$ و ناحیه بحرانی (یعنی مکمل ناحیه پذیرش) را به صورت $x < 0.25$ در نظر بگیرید. احتمال غلط بودن H_0 مشروط بر اینکه θ واقعاً مساوی 3 باشد برابر است با:

۱) $\frac{1}{64}$ ۲) $\frac{1}{256}$ ۳) $\frac{2}{3}$ ۴) $\frac{1}{16}$

حل: گزینه ۲ درست است.

با توجه به ناحیه بحرانی $C = \left\{x < \frac{1}{4}\right\}$ و فرض مقابل $H_1 : \theta = 3$، برای محاسبه توان آزمون (β^*) به صورت زیر عمل می‌کنیم:

$$\text{توان آزمون}: \beta^* = P_{H_1}(\text{ناحیه بحرانی}) = P_{\theta=3}\left(X < \frac{1}{4}\right) = \int_0^{\frac{1}{4}} (1+3)x^3 dx = \left[x^4\right]_0^{\frac{1}{4}} = \frac{1}{256}$$

سؤالات محیط زیست

۱. در کیسه‌ای 7 توپ وجود دارد که n تای آن قرمز و بقیه سیاه می‌باشند. برای آزمون فرض $H_0 : n = 4$ سه توپ استخراج کرده و فرض H_0 را تنها هنگامی رد می‌کنیم که هر سه سیاه باشند. احتمال خطای نوع اول کدام است؟

(محیط زیست ـ ۸۴)

۱) $\frac{41}{42}$　　۲) $\frac{1}{42}$　　۳) $\frac{4}{35}$　　۴) $\frac{1}{35}$

۲. فرض کنید شانس بارندگی در هر یک از چهار هفته بهمن ماه برابر q و بارندگی در هر هفته مستقل از هفته‌های دیگر باشد. برای آزمون فرض $H_0 : q = \frac{1}{2}$ مقدار $P\left(X \le 1 \middle| H_0\right)$ کدام است؟

(محیط زیست ـ ۸۴)

۱) $\frac{1}{16}$　　۲) $\frac{3}{16}$　　۳) $\frac{5}{16}$　　۴) $\frac{7}{16}$

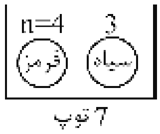
۳. در یک آزمایش دو وضعیتی می‌خواهیم فرض $H_0 : \theta = \frac{1}{2}$ را در مقابل $H_1 : \theta = \frac{3}{4}$ آزمون کنیم. آزمایش را 10 بار تکرار می‌کنیم و چنانچه حداکثر با یک موفقیت مواجه شویم، فرض H_0 رد می‌شود. در این حالت احتمال خطای نوع اول چقدر است؟

(محیط زیست ـ ۸۸)

۱) $10\left(\frac{1}{2}\right)^{10}$　　۲) $10\left(\frac{3}{4}\right)^{10}$　　۳) $11\left(\frac{1}{2}\right)^{10}$　　۴) $11\left(\frac{3}{4}\right)^{10}$

پاسخ‌های تشریحی

۱. گزینه ۴ درست است.



$$\begin{cases} H_0 : n = 4 \\ H_1 : n \neq 4 \end{cases}$$

$$\text{خطای نوع اول}: \alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست}) = P_{H_0}(\text{ناحیه رد}) = P_{n=4}(\text{هر سه سیاه}) = \frac{\binom{3}{3}}{\binom{7}{3}} = \frac{1}{35}$$

توجه: در کیسه 7 توپ وجود دارد که تحت فرض H_0، تعداد $n = 4$ توپ، قرمز و بقیه یعنی 3 تا سیاه‌اند. حال برای محاسبه احتمال خطای نوع اول باید احتمال آن را به دست آوریم که 3 توپ از 7 توپ خارج می‌کنیم $\binom{7}{3}$ و می‌خواهیم هر سه از بین 3 سیاه کیسه باشند، یعنی $\binom{3}{3}$.

۲. گزینه ۳ درست است.

X : تعداد موفقیت (بارندگی) در $n = 4$ هفته تحت فرض H_0 دارای توزیع دو جمله‌ای است با:

$$n = 4 \quad , \quad p = q = \frac{1}{2}$$

$$P\left(X \leq 1 \mid H_0\right) = P_{H_0}(X = 0) + P_{H_0}(X = 1) = \binom{4}{0}\left(\frac{1}{2}\right)^4 + \binom{4}{1}\left(\frac{1}{2}\right)^4 = \frac{1+4}{16} = \frac{5}{16}$$

۳. گزینه ۳ درست است.

$$\text{احتمال خطای نوع اول}: \alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست}) = P_{H_0}(\text{ناحیه بحرانی}) = P_{\theta=\frac{1}{2}}(X \leq 1)$$

$$= P_{\theta=\frac{1}{2}}(X = 0) + P_{\theta=\frac{1}{2}}(X = 1) = \binom{10}{0}\left(\frac{1}{2}\right)^{10} + \binom{10}{1}\left(\frac{1}{2}\right)^{10} = 11\left(\frac{1}{2}\right)^{10}$$

دقت کنید که X دارای توزیع دو جمله‌ای با احتمال موفقیت θ و تعداد نمونه $n = 10$ است، زیرا در صورت سؤال به دو وضعیتی (برنولی) بودن آزمایش اشاره شده که 10 بار هم تکرار شده است.

تست‌های طبقه‌بندی شده

تعاریف

۱. در یک فرآیند بسته‌بندی زعفران، برای آزمون اینکه میانگین وزن بسته‌ها کمتر از 3 گرم است. فرضیه صفر و مقابل کدام است؟ **(اقتصاد ــ ۸۷)**

۱) $\begin{cases} H_0 : \mu \le 3 \\ H_1 : \mu > 3 \end{cases}$
۲) $\begin{cases} H_0 : \mu \ge 3 \\ H_1 : \mu < 3 \end{cases}$
۳) $\begin{cases} H_0 : \mu = 3 \\ H_1 : \mu < 3 \end{cases}$
۴) $\begin{cases} H_0 : \mu > 3 \\ H_1 : \mu < 3 \end{cases}$

۲. سطح زیر منحنی H_0 در آزمون فرض آماری همواره برابر کدام است؟ **(برنامه ریزی شهری ــ ۸۸)**

۱) خطای نوع اول
۲) خطای نوع دوم
۳) وابسته به تعریف H_0
۴) سطح اطمینان آزمون

خطاهای آماری

۳. در یک آزمون فرض خطای نوع اول عبارت است از : **(محیط زیست ــ ۸۶)**

۱) احتمال رد فرض H_1 درصورتی‌که فرض H_1 درست باشد.
۲) احتمال رد فرض H_0 درصورتی‌که فرض H_0 درست باشد.
۳) احتمال پذیرش فرض H_0 درصورتی‌که فرض H_0 غلط باشد.
۴) احتمال پذیرش فرض H_1 درصورتی‌که فرض H_1 غلط باشد.

توان آزمون

۴. بر اساس تعریف، توان یک آزمون عبارت است از .... **(محیط زیست ــ ۸۷)**

۱) احتمال خطای نوع اول
۲) احتمال خطای نوع دوم
۳) یک منهای احتمال خطای نوع اول
۴) یک منهای احتمال خطای نوع دوم

۵. در آزمون فرض آماری، توان یک آزمون یعنی .... **(محیط زیست ــ ۸۸)**

۱) یک منهای احتمال خطای نوع اول
۲) یک منهای احتمال خطای نوع دوم
۳) احتمال خطای نوع اول
۴) احتمال خطای نوع دوم

روابط خطاها

۶. در آزمون فرضیه‌های آماری هر چه اختلاف بین مقدار پیشنهادی در H_0 و مقدار واقعی آن بیشتر باشد، توان آزمون: **(اقتصاد ــ ۸۸)**

۱) کمتر است. ۲) بیشتر است.

۳) فقط در آزمون‌های دوطرفه کمتر است. ۴) فقط در آزمون‌های یک‌طرفه بیشتر است.

۷. کدام عبارت، صحیح است؟ **(محیط زیست ــ ۸۶)**

۱) جمع خطای نوع اول و دوم مساوی یک است.

۲) خطای نوع دوم برابر است با یک منهای توان آزمون.

۳) خطای نوع اول برابر است با یک منهای توان آزمون.

۴) خطای نوع دوم برابر است با یک منهای خطای نوع اول.

آزمون فرض میانگین جامعه

۸. معمولاً در یک کارخانه، وزن بسته‌های شکر دارای توزیع نرمال با میانگین 2 کیلوگرم و انحراف معیار 200 گرم بوده است. شخصی ادعا می‌کند بر اساس فرسودگی دستگاه، وزن بسته‌های شکر بیشتر از 2 کیلوگرم شده است. برای بررسی در یک نمونه 25 تایی دیده شده است $\bar{x} = 2.3$ کیلوگرم است. مقدار آماره آزمون برابر است با .... **(محیط زیست ــ ۸۷)**

۱) -1 ۲) صفر ۳) 1 ۴) 7.5

آزمون مقایسه میانگین دو جامعه

۹. برای آزمون برابری میانگین‌های دو جامعه نرمال مستقل با واریانس‌های برابر، نتایج زیر از نمونه‌های انتخاب‌شده از این دو جامعه به دست آمده‌اند.

$S_1^2 = 15$	$\bar{x}_1 = 170$	$n_1 = 18$	جامعه اول
$S_2^2 = 17$	$\bar{x}_2 = 153$	$n_2 = 18$	جامعه دوم

مقدار آماره آزمون برابر است با: **(اقتصاد ــ ۸۶)**

۱) 12.75 ۲) $\frac{17\sqrt{17}}{4}$ ۳) $\frac{34\sqrt{2}}{3}$ ۴) $\frac{17}{4}$

۱۰. برای آزمون برابری میانگین‌های دو جامعه مستقل نرمال با واریانس‌های نامعلوم، نتایج زیر از نمونه‌های مستقل به دست آمده است. مقدار آماره این آزمون برابر است با: **(اقتصاد ــ ۸۸)**

$n_1 = 5$ $\quad n_2 = 7$

$\bar{x}_1 = 171$ $\quad \bar{x}_2 = 162$

$S_1^2 = 70$ $\quad S_2^2 = 120$

۱) $\frac{9\sqrt{35}}{\sqrt{1090}}$ ۲) $\frac{9}{10\sqrt{12}}$

۳) $\frac{9}{\sqrt{1090}}$ ۴) $\frac{9\sqrt{35}}{10\sqrt{12}}$

۱۱. در دو توزیع نرمال مستقل از هم به صورت $i=1,2,...,n$ و $X_i \sim N\left(\mu_1, \sigma_1^2\right)$ و $j=1,2,...,m$ و $Y_j \sim N\left(\mu_2, \sigma_2^2\right)$ آماره آزمون برای آزمون فرض $H_0: \mu_1 = \mu_2$ در مقابل $H_1: \mu_1 - \mu_2 > 0$ کدام است؟ (انحراف معیارهای σ_1 و σ_2 نامعلوم و مساوی هستند، حجم نمونه‌ها کوچک است و $S_p^2 = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$ است.)
(محیط زیست ـ ۸۸)

۱) $Z = \dfrac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$ (توزیع نرمال)

۲) $Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$ (توزیع نرمال)

۳) $t = \dfrac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$ (توزیع t−استیودنت)

۴) $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$ (توزیع t−استیودنت)

آزمون نسبت جامعه

۱۲. مدیر یک بانک ادعا کرده است 50 درصد مشتریان او علاوه بر حساب پس‌انداز، دارای حساب‌های دیگری نیز هستند. در نمونه‌ای تصادفی به حجم $n=100$، 45 درصد مشتریان حساب‌های دیگر داشته‌اند. مقدار آماره آزمون برابر است با:
(اقتصاد ـ ۸۷)

۱) -10 ۲) -1 ۳) 1 ۴) 10

۱۳. در یک نمونه‌گیری به حجم 30 محل دیده شده است که 18 محل به موقع زباله‌های خود را در خارج از منزل قرار می‌دهند. برای آزمون $H_0: p = \frac{1}{2}$، مقدار آماره آزمون کدام است؟
(محیط زیست ـ ۸۶)

۱) $\sqrt{\frac{6}{5}}$ ۲) $-\sqrt{\frac{5}{6}}$ ۳) $-\sqrt{\frac{6}{5}}$ ۴) $\sqrt{\frac{5}{6}}$

آزمون مقایسه نسبت دو جامعه

۱۴. در آزمون فرض مقایسه نسبت موفقیت در دو جامعه آماری اگر $n_1 = 120$، $n_2 = 100$، $\bar{p}_1 = 0.6$ و $\bar{p}_2 = 0.5$ باشند، مقدار آماره آزمون برای آزمون فرض $H_0: p_1 \le p_2$ کدام است؟
(حسابداری و مدیریت ـ ۸۷)

۱) $\frac{2\sqrt{3}}{3}$ ۲) $\frac{4\sqrt{2}}{5}$ ۳) $\frac{3\sqrt{5}}{4}$ ۴) $\frac{2\sqrt{5}}{3}$

آزمون مقایسه واریانس دوجامعه

۱۵. فرض کنید شما به عنوان یک تحلیل‌گر مالی بخواهید سود سهام دو بورس زیر را با توجه به داده‌های آن‌ها مقایسه کنید. برای آزمون تفاوت بین واریانس‌ها، مقدار آماره آزمون چقدر است؟
(اقتصاد ـ ۸۸)

بورس	تهران	مشهد
تعداد سهام	5	10
میانگین	3.0	3.25
انحراف معیار	5.0	3.0

۱) $\chi^2 = 0.36$
۲) $F = 2.8$
۳) $F = 1.6$
۴) $\chi^2 = 2.12$

۱۶. اطلاعات زیر مربوط به دو جامعه می‌باشد. با فرض نرمال بودن توزیع X در دو جامعه، آماره آزمون مربوط به فرض تساوی واریانس دو جامعه کدام است؟ **(محیط زیست ــ ۸۶)**

جامعه اول	جامعه دوم
$n_1=20$	$n_2=30$
$\bar{x}_1=50$	$\bar{x}_2=60$
$S_1=6$	$S_2=10$

۱) 0.36
۲) 0.5
۳) 0.6
۴) 0.83

آزمون نیکویی برازش ($\left(\chi^2\right)$ ساده)

۱۷. در 60 بار پرتاب یک تاس نتایج زیر حاصل شده است:

عدد	1	2	3	4	5	6
فراوانی	12	8	11	13	9	7

آماره آزمون برای آزمون فرضیه H_0 مبنی بر همگن بودن تاس کدام است؟ **(اقتصاد ــ ۸۶)**

۱) 0.17　۲) 0.28　۳) 1.7　۴) 2.8

آزمون استقلال ($\left(\chi^2\right)$ مضاعف)

۱۸. یک جدول توافقی از 3 ردیف و 4 ستون تشکیل شده است. تعداد درجات آزادی آن کدام است؟ **(GIS ــ ۸۸)**

۱) 3　۲) 4　۳) 6　۴) 12

۱۹.جدول پیشایندی زیر مفروض است. تحت فرض استقلال، جدول مورد انتظار کدام است؟ **(محیط زیست ــ ۸۶)**

B \ A	1	2	
1	10	10	
2	20	60	
			100

۱)

B \ A	1	2
1	10	10
2	25	55

۲)

B \ A	1	2
1	14	6
2	56	24

۳)

B \ A	1	2
1	10	15
2	25	50

۴)

B \ A	1	2
1	6	14
2	24	56

آنالیز واریانس (آزمون برابری میانگین چند جامعه)

۲۰. به منظور مقایسه هزینه خوراک خانوارها در 3 منطقه، از هر یک از این مناطق نمونه‌ای به حجم 10 خانوار به طور تصادفی انتخاب می‌شود و بر اساس نتایج مشاهدات جدول تحلیل واریانس به صورت زیر به دست می‌آید. مقدار عددی آماره آزمون کدام است؟ **(اقتصاد ــ ۸۶)**

منبع تغییرات	درجه آزادی	مجموع مجذور انحرافات
رویه	–	–
خطا	–	54
جمع	–	60.4

۱) 1.3
۲) 1.8
۳) 1.6
۴) 2.5

۲۱. به منظور آزمون برابری میانگین هزینه‌های مصرفی خانوارها در سه شهر مختلف، یک نمونه تصادفی 4 تایی از هر شهر انتخاب شده و اطلاعات زیر به دست آمده است. کمیت آماره آزمون چقدر است؟ (اقتصاد ـ ۸۷)

$\bar{x}_1 = 110 \quad \bar{x}_2 = 100 \quad \bar{x}_3 = 120$

$S_1^2 = 180 \quad S_2^2 = 220 \quad S_3^2 = 200$

۱) $\chi^2_{(2)} = 3$
۲) $F_{2,9} = 3$
۳) $F_{2,9} = 2$
۴) $F_{2,3} = 2$

۲۲. اطلاعات زیر در مورد طول عمر سه نوع روغن موتور اتومبیل به هزار کیلومتر بر اساس نمونه‌های 6 تایی در دست است. کمیت آماره آزمون برای آزمون برابری میانگین طول عمر این روغن‌ها کدام است؟ (اقتصاد ـ ۸۸)

روغن موتور	میانگین طول عمر	واریانس طول عمر
الف	7	3
ب	6	2
ج	8	4

۱) $\frac{1}{3}$
۲) $\frac{1}{2}$
۳) 2
۴) 3

مبانی رگرسیون و آزمون فرض‌های رگرسیون و همبستگی

۲۳. در یک نمونه‌گیری به حجم $n = 4$، نتایج زیر حاصل شده‌اند:

$\sum x_i y_i = 43, \sum x_i = 8, \sum y_i = 16, \sum x_i^2 = 30, \sum y_i^2 = 74$

در معادله رگرسیون $y = \beta_0 + \beta_1 x + \varepsilon$، برآوردهای حداقل مربعات β_0 و β_1 به ترتیب عبارت‌اند از:

(اقتصاد ـ ۸۶)

۱) $\frac{23}{7}, \frac{11}{14}$
۲) $\frac{17}{7}, \frac{11}{14}$
۳) $\frac{35}{26}, \frac{17}{13}$
۴) $\frac{17}{13}, \frac{52}{13}$

۲۴. رابطه بین x و y بر اساس یک نمونه تصادفی 18 تایی به صورت زیر برآورد شده است:

$\hat{y} = 15 + 1.7x \ , \ R^2 = 0.64$

آماره آزمون t برای آزمون فرضیه $\beta = 0$، یعنی عدم وجود رابطه بین x و y کدام است؟ (اقتصاد ـ ۸۶)

۱) 2.46
۲) 3.12
۳) 5.33
۴) 6.05

۲۵. در معادله رگرسیون $E(y|x) = \alpha + \beta x$ تخمین حداقل مربعات از پارامتر β عبارت است از: (اقتصاد ـ ۸۷)

۱) $b = \frac{\sum x_i y_i}{\sum x_i^2}$

۲) $b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum x_i^2}$

۳) $b = \frac{\sum x_i y_i}{\sum (x_i - \bar{x})^2}$

۴) $b = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

۲۶.تابع تولید $\underset{(0.02)}{\log Q_t = 0.12} + \underset{(0.9)}{0.7\log L_t} + e_t$ **، در کوتاه‌مدت با استفاده از 22 مشاهده برآورده شده است. اعداد داخل پرانتز انحراف معیار ضرایب است. در این صورت می‌توان گفت:** **(اقتصاد ـ ۸۸)**

۱) یک درصد افزایش در L سبب 0.7 درصد افزایش در تولید می‌شود.

۲) یک واحد افزایش در L سبب 0.7 واحد افزایش در تولید می‌شود.

۳) با اطمینان 95% رگرسیون برآوردشده معنی‌دار نیست.

۴) با اطمینان 95% سطح تولید متأثر از مقدار L است.

۲۷.منظور از هم‌خطی در رگرسیون این است که: **(محیط زیست ـ ۸۶)**

۱) مطابقت دو خط رگرسیونی با یکدیگر برقرار است.

۲) بین خطاهای رگرسیونی وابستگی وجود دارد.

۳) بین عوامل رگرسیونی وابستگی وجود دارد.

۴) تطابق خط رگرسیون واقعی جامعه با برآورد آن خط وجود دارد.

۲۸.در یک مدل رگرسیون به صورت $y_i = a + \varepsilon_i \ , i = 1,\dots,n$ **مقدار برآورد a از روش کمترین مربعات خطا، کدام است؟** **(محیط زیست ـ ۸۷)**

۱) $\bar{y}$ ۲) $\sum(y_i - \bar{y})^2$ ۳) $\dfrac{\sum y_i}{\sum y_i^2}$ ۴) $\dfrac{\sum(y_i - \bar{y})}{\sum(y_i - \bar{y})^2}$

۲۹.فرض کنید مدل رگرسیونی به صورت $y_i = a + bx_i + e_i \ , i = 1,2,\dots,n$ **و شرایط زیر برقرار است. مقدار b چقدر است؟** $\bar{x} = 0 \ , \ \sum x_i^2 = n \ , \ \sum y_i = 3n \ , \ \sum x_i y_i = 2n$ **(محیط زیست ـ ۸۷)**

۱) 2 ۲) 1 ۳) صفر ۴) -1

۳۰. داده‌های جدول مقابل مفروض است. برآورد شیب خط رگرسیون کدام است؟ **(محیط زیست ـ ۸۸)**

x	5	0	7
y	2	15	4

۱) $\dfrac{23}{13}$ ۲) $-\dfrac{23}{13}$ ۳) 1 ۴) -1

۳۱. در رگرسیون خطی $y = ax + b$ **پراکندگی مشاهدات حول خط رگرسیون به وسیله کدام‌یک از عبارت‌های زیر بهتر توصیف می‌شود؟** $\bar{y}$ **: میانگین نمونه‌ها،** $\hat{y}$ **: برآورد مقدار y** **(محیط زیست ـ ۸۸)**

۱) $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ ۲) $\sum_{i=1}^{n}(y_i - \bar{y})^2$ ۳) $\sum_{i=1}^{n}(y_i + \bar{y})^2$ ۴) $\sum_{i=1}^{n}(y_i + \hat{y}_i)^2$

پاسخ‌های تشریحی

تعاریف

۱. گزینه ۲ درست است.

با توجه به اینکه:

۱- فرض H_0 باید شامل تساوی باشد،

۲- ادعا در هر دو فرض می‌تواند باشد،

۳- H_1, H_0 نقیض یکدیگرند،

داریم:

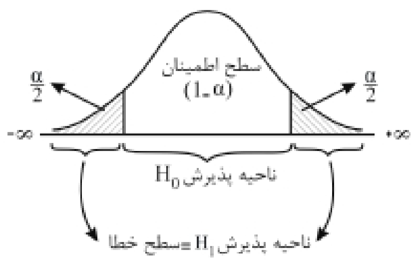
ادعا: میانگین وزن بسته‌ها کمتر از 3 گرم است $(\mu < 3)$.

چون ادعا شامل تساوی نیست در H_1 قرار گرفته و نقیض آن $(\mu \geq 3)$ در H_0 است.

$$\begin{cases} H_0 : \mu \geq 3 \\ H_1 : \mu < 3 \end{cases}$$

۲. گزینه ۴ درست است.

یادآوری:



ناحیه قبول H_0 = سطح اطمینان آزمون = $1-\alpha$ = سطح زیر منحنی H_0

سطح معنی‌دار = ناحیه قبول H_1 = ناحیه رد H_0 = خطای نوع اول = سطح بحرانی آزمون = α = سطح زیر منحنی H_1

خطاهای آماری

۳. گزینه ۲ درست است.

احتمال رد فرض H_0 درصورتی‌که فرض H_0 درست باشد = $\alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست})$: خطای نوع اول

دقت کنید که همواره بحث بر روی رد کردن و یا پذیرفتن H_0 است، بنابراین گزینه‌ای در انتخاب ارجح است که در مورد H_0 صحبت کرده باشد.

توان آزمون

۴. گزینه ۴ درست است.

یادآوری: توان آزمون = احتمال خطای نوع دوم − 1 $\rightarrow \beta^* = 1-\beta$

۵. گزینه ۲ درست است.

یادآوری: توان آزمون = احتمال خطای نوع دوم − 1 $\rightarrow \beta^* = 1-\beta$

روابط خطاها

۶. گزینه ۲ درست است.

با توجه به اینکه اختلاف بین مقدار پیشنهادی و واقعی H_0 زیاد است یعنی H_0 به‌واقع نادرست است.

بنابراین داریم:

$$\beta^* = P(H_0 \text{ رد} \mid H_0 \text{ به واقع غلط}) = 1$$

۷. گزینه ۲ درست است.

نکات زیر در مورد احتمال خطای نوع اول (α)، خطای نوع دوم (β) و توان آزمون (β^*) برقرار است:

۱- جمع خطای نوع اول و نوع دوم لزوماً یک نمی شود $(\alpha+\beta \neq 1)$.

۲- جمع خطای نوع دوم و توان آزمون برابر یک است $(\beta+\beta^* = 1)$، بنابراین $\begin{cases} \beta = 1-\beta^* \\ \beta^* = 1-\beta \end{cases}$.

با توجه به دو نکته بالا، تنها گزینه ۲ درست است.

آزمون فرض میانگین جامعه

۸. گزینه ۴ درست است.

$$\text{وزن بسته‌های شکر} : X \sim N\left(\mu = 2\text{kg} \ , \ \sigma^2 = (0.2\text{kg})^2\right)$$

$$(1)\begin{cases} H_0 & : \ \mu \le 2 \\ H_1 & : \ \mu > 2 \leftarrow \text{ادعا: وزن بسته‌ها بیش از 2 کیلوگرم} \end{cases}$$

جامعه، نرمال و واریانس جامعه معلوم است پس n هر چه باشد توزیع $\bar{X}$ نرمال خواهد بود.

$$(2)\ \text{آماره آزمون} : Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.3-2}{\frac{0.2}{\sqrt{25}}} = \frac{0.3\times 5}{0.2} = 7.5$$

دقت کنید که میانگین و انحراف معیار باید براساس یک مقیاس باشند، بنابراین انحراف معیار 200 گرم را 0.2 کیلوگرم در نظر می‌گیریم که با میانگین هم‌مقیاس باشد.

آزمون مقایسه میانگین دو جامعه

۹. گزینه ۱ درست است.

$$(1)\begin{cases} H_0: \mu_1 = \mu_2 \leftarrow \text{برابری میانگین‌های دوجامعه} \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

با توجه به اینکه دو جامعه نرمال، واریانس‌ها نامعلوم اما برابر و $n_1, n_2 \le 30$ است، بنا بر مورد(۳ - الف) آماره آزمون $\mu_1 = \mu_2$، توزیع $\bar{X}_1 - \bar{X}_2$، t با درجه آزادی $n_1 + n_2 - 2$ است و داریم:

$$(2)\begin{cases} t_{n_1+n_2-2} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{(170-153)-0}{4\sqrt{\dfrac{1}{18}+\dfrac{1}{18}}} = \dfrac{17}{4\sqrt{\dfrac{1}{9}}} = \dfrac{51}{4} = 12.75 \\ n_1 = n_2 = 18 \rightarrow S_p^2 = \dfrac{S_1^2 + S_2^2}{2} = \dfrac{15+17}{2} = 16 \rightarrow S_p = 4 \\ \bar{x}_1 = 170 \ , \ S_1^2 = 15 \ , \ \bar{x}_2 = 153 \ , \ S_2^2 = 17 \end{cases}$$

اگر $n_1 = n_2 = n$ باشد،

$$S_p^2 = \frac{(n-1)S_1^2 + (n-1)S_2^2}{\underset{2(n-1)}{\cancel{n+n-2}}} = \frac{S_1^2 + S_2^2}{2}$$

۱۰. گزینه ۱ درست است.

دو جامعه نرمال، واریانس جوامع نامعلوم و پیش فرض نابرابر و $n_1, n_2 < 30$ است، بنابراین آماره آزمون برابری میانگین‌های دو جامعه، t–استیودنت با درجه آزادی r است.

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$t_r = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \frac{171-162}{\sqrt{\dfrac{70}{5} + \dfrac{120}{7}}} = \frac{9}{\sqrt{\dfrac{490+600}{35}}} = \frac{9\sqrt{35}}{\sqrt{1090}}$$

۱۱. گزینه ۳ درست است.

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$$

دو جامعه نرمال، واریانس جوامع نامعلوم اما برابر و تعداد نمونه‌ها کوچک است. بنابراین توزیع تفاضل میانگین‌های دو نمونه $(\bar{X}_1 - \bar{X}_2)$، t– استیودنت با درجه آزادی $n+m-2$ است.

$$\begin{cases} t_{n+m-2} = \dfrac{(\bar{X}_1 - \bar{X}_2) - 0}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \\ S_p^2 = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} \end{cases}$$

آزمون فرض نسبت جامعه

۱۲. **گزینه ۲ درست است.**

آماره آزمون نسبت برای $n > 30$ نرمال است:

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p \neq 0.5 \end{cases}$$

$$\begin{cases} \text{آماره آزمون} : Z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{0.45 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{100}}} = \dfrac{-0.05}{0.05} = -1 \\ \bar{p} = \dfrac{x}{n} = \dfrac{45}{100} = 0.45,\ p_0 = q_0 = 0.5 \\ x = 45\ (\text{تعداد افراد دارای صفت خاص})\ \text{و}\ n = 100 \end{cases}$$

۱۳. **گزینه ۱ درست است.**

$$\begin{cases} H_0 : p = \dfrac{1}{2} \\ H_1 : p \neq \dfrac{1}{2} \end{cases}$$

$$\begin{cases} \text{آماره آزمون} : Z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{0.6 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{30}}} = \dfrac{0.1\sqrt{30}}{0.5} = \dfrac{\sqrt{30}}{5} = \sqrt{\dfrac{6}{5}} \\ \bar{p} = \dfrac{x}{n} = \dfrac{18}{30} = \dfrac{6}{10} = 0.6 \\ x : \text{تعداد محل های مورد نظر}\ ,\ n = 30 \end{cases}$$

آزمون مقایسه نسبت دو جامعه

۱۴. **گزینه ۴ درست است.**

$$\begin{cases} H_0 : p_1 \leq p_2 \\ H_1 : p_1 > p_2 \end{cases}$$

$$\begin{cases} Z = \dfrac{\bar{p}_1 - \bar{p}_2}{\sqrt{\dfrac{\bar{p}_1 \bar{q}_1}{n_1} + \dfrac{\bar{p}_2 \bar{q}_2}{n_2}}} = \dfrac{0.6 - 0.5}{\sqrt{\dfrac{0.6 \times 0.4}{120} + \dfrac{0.5 \times 0.5}{100}}} = \dfrac{0.1}{\sqrt{\dfrac{0.02}{10} + \dfrac{0.25}{100}}} = \dfrac{0.1}{\dfrac{\sqrt{0.45}}{10}} = \dfrac{1}{0.3\sqrt{5}} = \dfrac{2\sqrt{5}}{3} \\ \bar{p}_1 = 0.6\ ,\ \bar{p}_2 = 0.5 \\ \bar{q}_1 = 0.4\ ,\ \bar{q}_2 = 0.5 \\ n_1 = 120\ ,\ n_2 = 100 \end{cases}$$

آزمون مقایسه واریانس دو جامعه

۱۵. **گزینه ۲ درست است.**

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

$$\text{آماره آزمون} : F = \frac{S_1^2}{S_2^2} = \frac{5^2}{3^2} = \frac{25}{9} = 2.77 \simeq 2.8$$

۱۶. گزینه ۱ درست است.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

آماره آزمون : $F = \dfrac{S_1^2}{S_2^2} = \dfrac{6^2}{10^2} = \dfrac{36}{100} = 0.36$

آزمون نیکویی برازش ($\left(\chi^2\right)$ساده)

۱۷. گزینه ۴ درست است.

برای بررسی همگن بودن وجوه یک تاس، از آزمون نیکویی برازش (χ^2 ساده) به صورت $\chi^2_{(k-1)} = \sum_{i=1}^{k} \dfrac{\left(F_{o_i} - F_{e_i}\right)^2}{F_{e_i}}$ استفاده می‌کنیم. در این مسئله با توجه به $n = 60$ بار پرتاب یک تاس برای $k = 6$ دسته (6 وجه تاس)، فراوانی‌های مشاهده‌شده $\left(F_{o_i}\right)$ داده شده است؛ فراوانی‌های موردانتظار $\left(F_{e_i}\right)$ با در نظر گرفتن همگن بودن وجوه تاس با احتمال یکسان $\left(p_i = \dfrac{1}{k} = \dfrac{1}{6}\right)$ به صورت $F_{e_i} = np_i = \dfrac{n}{k} = \dfrac{60}{6} = 10$ به دست می‌آید، بنابراین داریم:

عدد	1	2	3	4	5	6	
F_{o_i}	12	8	11	13	9	7	$n = \sum F_{o_i} = 60$
F_{e_i}	10	10	10	10	10	10	

$$\chi^2_{(k-1)} = \chi^2_{(5)} = \sum_{i=1}^{6} \frac{\left(F_{o_i} - F_{e_i}\right)^2}{F_{e_i}} = \frac{(12-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(7-10)^2}{10} = 2.8$$

آزمون استقلال ($\left(\chi^2\right)$ مضاعف)

۱۸. گزینه ۳ درست است.

یادآوری: درجه آزادی یک جدول توافقی با r سطر و c ستون عبارت است از:

درجه آزادی $= (r-1)(c-1) \xrightarrow{r=3, c=4}$ درجه آزادی $= (3-1)(4-1) = 2 \times 3 = 6$

۱۹. گزینه ۴ درست است.

یادآوری: در آزمون استقلال دو پیشامد (χ^2 مضاعف) جدول داده‌های مشاهده‌شده $F_{o_{ij}}$ (جدول پیشایندی) داده می‌شود که پس از به دست آوردن جدول $F_{e_{ij}}$ (جدول مورد انتظار)، از روی آن آماره آزمون را به‌دست می‌آوریم:

آماره آزمون : $\chi^2 = \sum\sum \dfrac{\left(F_{o_{ij}} - F_{e_{ij}}\right)^2}{F_{e_{ij}}}$

$$F_{e_{ij}} = \frac{\text{جمع سطر } i\text{ام} \times \text{جمع ستون } j\text{ام}}{\text{جمع کل}}$$

$$F_{o_{ij}}\left\{\begin{array}{|c|c|c|c|}\hline {}^{B}\diagdown^{A} & 1 & 2 & \sum \\ \hline 1 & 10 & 10 & 20 \\ \hline 2 & 20 & 60 & 80 \\ \hline \sum & 30 & 70 & 100 \\ \hline \end{array}\right. \rightarrow F_{eij}\left\{\begin{array}{|c|c|c|}\hline {}^{B}\diagdown^{A} & 1 & 2 \\ \hline 1 & \frac{20\times 30}{100}=6 & \frac{20\times 70}{100}=14 \\ \hline 2 & \frac{80\times 30}{100}=24 & \frac{80\times 70}{100}=56 \\ \hline \end{array}\right.$$

آنالیز واریانس (آزمون برابری میانگین چندجامعه)

۲۰. گزینه ۳ درست است.

برای آزمون مقایسه میانگین چند جامعه (3 منطقه) از آنالیز واریانس (آماره آزمون F) استفاده می‌کنیم و با توجه به داده‌های مسئله، رابطه اول آنالیز واریانس مورد استفاده قرار می‌گیرد:

$$\begin{cases} F_{k-1,k(n-1)}=\dfrac{\frac{SS(Tr)}{k-1}}{\frac{SSE}{k(n-1)}} \rightarrow F_{2,27}=\dfrac{\frac{6.4}{3-1}}{\frac{54}{3(10-1)}}=1.6 \\ SST=SSE+SS(Tr)\rightarrow SS(Tr)=60.4-54=6.4 \\ SSE=54\ ,\ SST=60.4\ ,\ k=3\ ,\ n=10 \end{cases}$$

۲۱. گزینه ۳ درست است.

چون مقایسه میانگین بین سه شهر (بیش از دو جامعه) است از آنالیز واریانس و آماره آزمون F استفاده می‌کنیم. با توجه به داده‌های مسئله، رابطه دوم آنالیز واریانس مورد استفاده قرار می‌گیرد:

$$\begin{cases} F_{k-1,k(n-1)}=\dfrac{\frac{n\sum\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{k-1}}{\frac{\sum s_i^2}{k}} \rightarrow F_{2,9}=\dfrac{\frac{4}{3-1}\left[(110-110)^2+(100-110)^2+(120-110)^2\right]}{\frac{180+220+200}{3}}=2 \\ \bar{\bar{x}}=\dfrac{\sum x_i}{k}=\dfrac{110+100+120}{3}=110\ ,\ k=3\ ,\ n=4 \end{cases}$$

۲۲. گزینه ۳ درست است.

$$\begin{cases} H_0:\mu_1=\mu_2=\mu_3 \\ H_1:\text{حداقل یکی متفاوت است} \end{cases}$$

$k=3\ ,\ n=6$

$$F_{k-1,k(n-1)}=\dfrac{\frac{n\sum\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{K-1}}{\frac{\sum S_i^2}{k}} \rightarrow F_{2,15}=\dfrac{\frac{6}{3-1}\left[(7-7)^2+(6-7)^2+(8-7)^2\right]}{\frac{1}{3}(3+2+4)}=\dfrac{3(0+1+1)}{3}=2$$

$$\bar{\bar{x}}=\dfrac{\sum \bar{x}_i}{k}=\dfrac{7+6+8}{3}=\dfrac{21}{3}=7$$

مبانی رگرسیون و آزمون فرض‌های رگرسیون و همبستگی

۲۳ . گزینه ۲ درست است.

$$\text{شیب خط}: b_1=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sum x^2-\frac{\left(\sum x\right)^2}{n}}=\frac{43-\frac{8\times 16}{4}}{30-\frac{(8)^2}{4}}=\frac{43-32}{30-16}=\frac{11}{14}$$

با توجه به این‌که خط رگرسیون برآورد شده همیشه از نقطه $(\bar{x},\bar{y})$ عبور می‌کند، مقدار عرض از مبدأ به‌دست می‌آید:

$$\bar{y}=b_0+b_1\bar{x}\quad\rightarrow\quad \text{عرض از مبدأ}: b_0=\bar{y}-b_1\bar{x}=\frac{16}{4}-\frac{11}{14}\times\frac{8}{4}=\frac{34}{14}=\frac{17}{7}$$

۲۴ . گزینه ۳ درست است.

$$(1)\begin{cases}H_0:\beta=0\\ H_1:\beta\neq 0\end{cases}$$

$$(2)\begin{cases}t_{n-2}=\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}=\dfrac{0.8}{\sqrt{\dfrac{1-0.64}{18-2}}}=5.33\\ n=18\ ,\ R^2=0.64\ \rightarrow\ |r|=0.8\ \rightarrow\ r=+0.8\ \text{(با توجه به مثبت بودن شیب خط رگرسیون)}\end{cases}$$

۲۵ . گزینه ۴ درست است.

ضرایب معادله خط رگرسیون به دو روش حداقل مربعات و حداکثر درستنمایی برآورد می‌شوند که نتیجه هر دو تخمین با هم برابر و عبارت است از:

$$\begin{cases}\text{شیب خط}: b=\dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\dfrac{\sum x_i(y_i-\bar{y})-\bar{x}\overbrace{\cancel{\sum(y_i-\bar{y})}}^{0}}{\sum(x_i-\bar{x})^2}=\dfrac{\sum x_i(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\\ \text{عرض از مبدأ}: a=\bar{y}-b\bar{x}\end{cases}$$

یادآوری:

۱) خاصیت مهم میانگین: مجموع انحرافات داده‌ها از میانگین همواره صفر است $\sum(y_i-\bar{y})=0$ دقت کنید که رابطه بالا می‌توانست به صورت زیر هم باشد:

$$b=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{\sum y_i(x_i-\bar{x})-\bar{y}\overbrace{\cancel{\sum(x_i-\bar{x})}}^{0}}{\sum(x_i-\bar{x})^2}=\frac{\sum y_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

۲) درصورتی‌که در صورت سؤال ذکر می‌شد $E(Y|X=x)=\beta x$ یعنی خط از مبدأ مختصات می‌گذرد (عرض از مبدأ صفر $(\alpha=0)$) آنگاه برآورد شیب خط بصورت زیر بود و گزینه ۱ درست بود.

$$b=\frac{\sum y_i x_i}{\sum x_i^2}$$

۲۶ . گزینه ۳ درست است.

ابتدا باید معنی‌دار بودن شیب خط رگرسیون آزمون شود تا ارتباط خطی بین L و Q رد یا اثبات شود. آن‌گاه در صورت اثبات می‌توان درباره تأثیر یک واحد افزایش یا کاهش در L صحبت کرد.

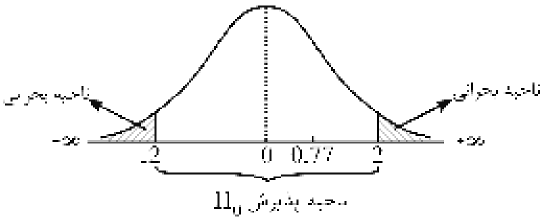
یادآوری: با توجه به داده‌های مسئله از آزمون t–استیودنت زیر برای معنی‌دار بودن شیب خط رگرسیون استفاده می‌کنیم:

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

$$t_{(n-2)} = \frac{b}{S_b} \rightarrow t_{20} = \frac{0.7}{0.9} = \frac{7}{9} = 0.77$$

در سطح 95% مقدار t در جدول t–استیودنت برای درجات آزادی مختلف حدود 2 است. $\left(t_{20,0.975} = 2.09\right)$

پس با توجه به شکل و مقدار آماره آزمون، فرض H_0 پذیرفته می‌شود؛ یعنی ثابت می‌شود که ارتباط خطی بین L و Q وجود ندارد و یا به عبارت دیگر با اطمینان 95% شیب خط رگرسیون برآوردشده معنی‌دار نیست، بنابراین کل خط رگرسیون نیز معنادار نیست.



دقت کنید، درصورتی‌که شیب خط رگرسیون معنی‌دار بود یعنی فرض H_0 رد می‌شد، جمله زیر در مورد تفسیر خط رگرسیون برآورد شده درست بود:

به ازای یک واحد افزایش در متغیر مستقل L مقدار متغیر وابسته Q به میزان 0.7 واحد افزایش می یابد (گزینه۲ درست بود).

۲۷. گزینه ۳ درست است.

آزمون هم‌خطی یا خودهمبستگی در رگرسیون چندمتغیره به این معناست که آیا بین چند متغیر مستقل موجود در خط رگرسیون $(x_1, x_2, \dots)$ همبستگی وجود دارد یا خیر؟

۲۸. گزینه ۱ درست است.

دقت کنید که در خط رگرسیون داده شده $(y_i = a + \varepsilon_i)$ شیب خط رگرسیون صفر بوده است زیرا جزء bx در خط رگرسیون دیده نمی‌شود $\left(Y = a + \underbrace{\cancel{bx}}_{0} + \varepsilon\right)$، بنابراین با توجه به صفر بودن شیب خط رگرسیون و اینکه خط رگرسیون همواره از نقطه $(\bar{x}, \bar{y})$ می‌گذرد برای عرض از مبدأ داریم:

$$\bar{Y} = a + b\bar{x} \rightarrow \text{عرض از مبدا} : a = \bar{y} - b\bar{x} \xrightarrow{b=0} a = \bar{y}$$

۲۹. گزینه ۱ درست است.

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{2n - \frac{0 \times 3n}{n}}{n - \frac{(0)^2}{n}} = \frac{2n}{n} = 2$$

$$\sum xy = 2n \;,\; \sum y = 3n \;,\; \sum x^2 = n \;,\; \bar{X} = 0 \rightarrow \sum x = 0$$

۳۰. گزینه ۲ درست است.

با توجه به اینکه $\bar{x}$ و $\bar{y}$ در جدول مورد نظر اعداد صحیحی هستند:

$$\begin{cases} \bar{x} = \frac{\sum x_i}{n} = \frac{12}{3} = 4 \\ \bar{y} = \frac{\sum y_i}{n} = \frac{21}{3} = 7 \end{cases}$$

به‌کارگیری رابطه زیر از نظر سادگی در محاسبه مناسب‌تر است:

x	y	$(x-\bar{x})$	$(Y-\bar{Y})$
5	2	1	−5
0	15	−4	8
7	4	3	−3

$\longrightarrow$

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
−5	1
−32	16
−9	9
$\sum=-46$	$\sum=26$

$$\text{شیب خط رگرسیون} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{-46}{26} = \frac{-23}{13}$$

۳۱. گزینه ۱ درست است.

$$\text{(خطا) پراکندگی مشاهدات حول خط رگرسیون} = SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$\hat{Y}_i$: برآورد مقدار y از خط رگرسیون

خودآزمایی

تعاریف

۱. کدام گزینه درباره فرض‌های آماری نادرست است؟

۱) فرض H_1 فقط یک وجه دارد.
۲) فرض‌ها از نظر منطقی با هم مخالف‌اند.
۳) فرض H_0 بیانگر عدم وجود تفاوت است.
۴) فرض‌ها از لحاظ احتمالات به هم پیوسته‌اند.

۲. فرض‌های صفر $\left(H_0\right)$:

۱) بیان ادعاها هستند.
۲) بدون شواهد معتبر، رد می‌شوند.
۳) بیان نفی ادعا هستند.
۴) ادعاهایی هستند که مورد تحقیق قرار می‌گیرند.

۳. ادعا شده است که پراکندگی جامعه «الف» کمتر از جامعه «ب» است. فرض H_0 کدام است؟

۱) $H_0 : \sigma_A^2 < \sigma_B^2$
۲) $H_0 : \sigma_A^2 \leq \sigma_B^2$
۳) $H_0 : \sigma_A^2 \geq \sigma_B^2$
۴) $H_0 : \sigma_A^2 > \sigma_B^2$

خطاهای آماری

خطای نوع اول

۴. در آزمون‌های فرض، زمانی خطای نوع اول رخ می‌دهد که:

۱) فرض صفر رد نشود، درصورتی‌که فرض صفر درست است.
۲) فرض صفر رد نشود، درصورتی‌که فرض صفر غلط است.
۳) فرض صفر $\left(H_0\right)$ رد شود، درصورتی‌که فرض صفر درست است.
۴) فرض صفر رد شود، درصورتی‌که فرض صفر غلط است.

۵. اگر آزمون فرضیه‌ای که تست می‌کنید دودامنه باشد، کدام‌یک از گزینه‌های زیر می‌تواند توان تست را افزایش دهد؟

۱) افزایش α
۲) کاهش حجم نمونه
۳) افزایش β
۴) هیچ‌کدام

۶. **در سطح $\alpha = 0.05$ یک آزمون فرضیه مورد بررسی است، این بدان معنی است که:**

۱) پنج درصد شانس وجود دارد که فرضیه H_0 درست باشد.

۲) حداکثر پنج درصد شانس وجود دارد که H_0 رد شود.

۳) پنج درصد شانس وجود دارد که فرضیه مقابل درست باشد.

۴) پنج درصد شانس وجود دارد که خطای نوع دوم رخ داده باشد.

خطای نوع دوم

۷. **در آزمون فرض‌ها، «خطای نوع دوم» کدام است؟**

۱) رد کردن H_0 وقتی H_1 درست است.

۲) رد نکردن H_0 وقتی اعتبار H_1 نشان داده نشده است.

۳) رد کردن H_0 وقتی H_0 درست است.

۴) رد نکردن H_0 وقتی H_1 درست است.

توان آزمون

۸. **کدام‌یک از گزینه‌ها درست است؟**

۱) $\alpha + \beta = 1$

۲) $\alpha = P($ رد نکردن H_0 وقتی H_0 نادرست است $)$

۳) $\alpha = P($ رد نکردن H_0 وقتی H_0 درست است $)$

۴) $1-\beta = P($ رد کردن H_0 وقتی H_0 درست نیست $)$

آزمون فرض میانگین جامعه

۹. **درصورتی‌که $Z = 1.96$ جدول و $Z = -3$ محاسبه شده باشد، چه قضاوتی می‌توان درباره H_0 و H_1 داشت؟**

۱) H_0 در ناحیه بحرانی قرار نمی‌گیرد.

۲) H_0 قبول می‌شود.

۳) H_0 رد می‌شود.

۴) H_1 رد می‌شود.

۱۰. **در آزمون فرض یک‌دامنه با سطح اطمینان 95% و $Z = 1.96$، Z متناظر با داده‌های نمونه کدام باشد تا فرضیه H_1 رد نشود؟**

۱) $|Z| < +1.96$

۲) $Z < +1.96$

۳) $Z > 1.96$

۴) $|Z| > 1.96$

۱۱. **تعبیر $P\left(-25.1 \le \mu_1 - \mu_2 \le -6.7\right) = 0.95$ این است که در سطح خطای 5 درصد می‌توان ادعا کرد:**

۱) $\mu_1 < \mu_2$

۲) $\mu_1 = \mu_2$

۳) $\mu_1 > \mu_2$

۴) اطلاعات برای اظهار نظر کافی نیست.

۱۲. **آماره آزمون، نرمال صفر و یک است و $H_0 : \mu_X = 40$ تعریف شده است. مقدار آماره آزمون 4.5 است. در سطح اطمینان 95 درصد کدام گزینه درست است؟ (مقدار جدول 1.96 فرض شود.)**

۱) H_0 تأیید می‌شود.

۲) H_1 تأیید می‌شود.

۳) H_1 تأیید نمی‌شود.

۴) اطلاعات بیشتری برای تصمیم‌گیری مورد نیاز است.

۱۳. فرض کنید $X_1,...,X_n$ یک نمونه تصادفی به حجم $n=36$ از توزیع پواسون با پارامتر λ باشد. برای آزمون فرض $H_0:\lambda=3$ در مقابل $H_1:\lambda=4$ در سطح $\alpha=\%5$ اگر $\sum x_i=144$ باشد، در آن صورت آماره آزمون برابر است با:

۱) 3 ۲) $\sqrt{12}$ ۳) 2 ۴) $\sqrt{10}$

۱۴. یک شرکت تعمیراتی ادعا می‌کند تعمیر یک دستگاه رایانه توسط این شرکت به‌طور متوسط حداکثر 60 دقیقه طول می‌کشد. در یک نمونه به حجم 25 دیده شده که متوسط زمان کارکرد برای تعمیر، 64 با انحراف معیار 10 دقیقه است. اگر $\alpha=\%5$ باشد آماره آزمون کدام است و از چه توزیعی استفاده می‌شود؟

۱) 0.4 و از توزیع نرمال استاندارد استفاده می‌شود.

۲) 2 و از توزیع نرمال استفاده می‌شود.

۳) 0.4 و از توزیع t–استیودنت با 24 درجه آزادی استفاده می‌شود.

۴) 2 و از توزیع t–استیودنت با 24 درجه آزادی استفاده می‌شود.

آزمون مقایسه میانگین دو جامعه

۱۵. برای آزمون تفاوت میانگین‌های دو جامعه مستقل با واریانس‌های برابر، زمانی که $n_1=20$ و $n_2=20$ باشد، درجه آزادی برابر است با:

۱) 39 ۲) 38 ۳) 19 ۴) 18

آزمون نسبت جامعه

۱۶. زمانی که آزمون $H_0:p\le 0.7$ و $H_1:p>0.7$ را در سطح α تست می‌کنیم کدام‌یک از گزینه‌های زیر فرضیه H_0 را رد می‌کند؟

۱) $Z<Z_\alpha$ ۲) $Z<-Z_\alpha$ ۳) $Z>Z_\alpha$ ۴) $Z<0$

آزمون واریانس جامعه

۱۷. اطلاعات $\bar{x}=60$، $S_X=15$ و $n=10$ از یک جامعه نرمال به‌دست آمده است. مقدار آماره آزمون برای فرضیه $H_0:\sigma_X^2=100$ کدام است؟

۱) 20.25 ۲) 6 ۳) 15 ۴) 1.35

آزمون مقایسه واریانس دو جامعه

۱۸. حد بالا و پایین تخمین فاصله‌ای $\dfrac{\sigma_1^2}{\sigma_2^2}$ در سطح اطمینان 99 درصد کوچک‌تر از 1 است. کدام گزینه درست است؟

۱) σ_1^2 بزرگ‌تر از σ_2^2 است.

۲) σ_1^2 کوچک‌تر از σ_2^2 است.

۳) σ_2^2 کوچک‌تر از σ_1^2 است.

۴) σ_1^2 و σ_2^2 با هم اختلاف معناداری ندارند.

آزمون نیکویی برازش (χ^2 ساده)

۱۹. مقادیر F_{o_i} و F_{e_i} به این صورت تعریف شده است:

F_{o_i}	5	10	8	7
F_{e_i}	5	8	9	8

مقدار آماره آزمون کدام است؟

۱) 1.254 ۲) 2.753 ۳) 0.251 ۴) 0.736

۲۰. در سؤال قبل اگر H_0 نشان‌دهنده توزیع یکنواخت باشد، تعداد درجات آزادی چقدر است؟

۱) 4 ۲) 2 ۳) 3 ۴) 1

۲۱. ادعا شده که توزیع تصادفات در سطح شهر دارای توزیع پواسون است. در آزمون این ادعا چند پارامتر برآورد خواهد شد؟

۱) 3 ۲) 2 ۳) 1 ۴) صفر

۲۲. برای آزمون فرضیه $H_0 : p_1 = p_2$، مقدار آماره آزمون برابر است با $Z = -2.70$. مقدار χ^2_α در آزمون همگونی کای ـ مربع چقدر است ؟

۱) 2.70 ۲) -2.70 ۳) 7.29 ۴) 10.14

آزمون استقلال (χ^2 مضاعف)

۲۳. یک جدول توافقی با 3 ردیف و 4 ستون وجود دارد. تعداد درجات آزادی آن کدام است؟

۱) 2 ۲) 3 ۳) 6 ۴) 12

۲۴. مقدار آماره‌ی χ^2 در جدول پیشایندی زیر تحت فرض استقلال برابر است با:

A \ B	1	2	
1	20	30	
2	40	10	
			100

۱) 17 ۲) 14.5 ۳) $\frac{100}{6}$ ۴) $\frac{100}{14}$

۲۵. مقدار آماره آزمون 8.35 است. اگر اعداد زیر نشان‌دهنده مقادیر بحرانی در سطح αهای مختلف باشند، در چه صورتی H_0 رد می‌شود؟

۱) 7.779 ۲) 9.488 ۳) 11.143 ۴) 13.277

آنالیز واریانس (آزمون برابری میانگین چند جامعه)

۲۶. در تحلیل واریانس یک‌عامله، اگر تعداد تیمارها 3 و $SST = 50$ و $SSE = 18$ باشد، $SS(Tr)$ کدام است؟

۱) 68 ۲) 36 ۳) 32 ۴) اطلاعات کافی نیست.

۲۷. در سؤال قبل، مقدار $Ms(Tr)$ کدام است؟

۱) $\frac{50}{3}$ ۲) 18 ۳) $\frac{32}{3}$ ۴) 16

۲۸. در دو سؤال قبل، اگر از هر تیمار 5 نمونه گرفته باشیم، آن‌گاه مقدار F با کدام گزینه برابر است؟

۱) 10.67 ۲) 5.44 ۳) 8.30 ۴) 12.00

۲۹. جدول ناقص آنالیز واریانس زیر ارائه شده است:

منبع تغییرات	مجموع مربعات	درجه آزادی	میانگین مربعات	F
رفتارها	140	5		
خطا			12	
جمع		29		

مجموع مربعات خطا کدام است؟

۱) 288 ۲) 180 ۳) 300 ۴) 240

۳۰. متوسط میزان محصول تولیدی چهار ماشین صنعتی در سه نوبت مختلف به شرح زیر است:

T_1	T_2	T_3	T_4
13	15	8	11
8	11	12	15
9	13	7	10

می‌خواهیم برابری متوسط مقدار تولیدی ماشین‌ها را بررسی کنیم. مقدار مجموع مربعات انحراف مربوط به تیمارها (SST) کدام است؟

۱) 30 ۲) 40 ۳) 50 ۴) 80

۳۱. در سؤال قبل درجه آزادی SSE کدام است؟

۱) 9 ۲) 3 ۳) 8 ۴) 12

۳۲. در روش تحلیلی واریانس که درآن $H_0 : \mu_1 = \mu_2 = \mu_3$ را در مقابل فرضیه H_a قرار می‌دهیم، از آماره:

۱) F استفاده می‌کنیم. ۲) χ^2 استفاده می‌کنیم. ۳) Z استفاده می‌کنیم. ۴) t استفاده می‌کنیم

مبانی رگرسیون و آزمون فرض‌های رگرسیون و همبستگی

۳۳. اگر $\sum(y-\bar{y})^2 = \sum(y-\hat{y})^2$ باشد، آن‌گاه ضریب همبستگی کدام است؟

۱) 1 یا 0 ۲) -1 یا 0 ۳) 0 ۴) 1

۳۴. پراکندگی مشاهدات حول خط رگرسیون با کدام‌یک از این عبارت‌ها بهتر توصیف می‌شود؟

۱) $\sum(y+\hat{y})^2$ ۲) $\sum(y-\hat{y})^2$ ۳) $\sum(y-\bar{y})^2$ ۴) $\sum(y+\bar{y})^2$

۳۵. فرض کنید می‌خواهید فرض $\beta = 0$ را در مقابل فرض $\beta \neq 0$ آزمون کنید. کدام‌یک از این موارد را باید قبل از بقیه محاسبه کرد؟

۱) S_b ۲) S_e ۳) آماره t ۴) فرقی نمی‌کند.

۳۶. فرض کنید همبستگی $n = 3$ داده از متغیرهای تصادفی -0.80 باشد، برای آزمون فرض استقلال دو متغیر ($\rho = 0$ در مقابل $\rho \neq 0$) ناچار به محاسبه آماره t هستیم. مقدار آماره مورد نظر کدام است؟

۱) 1.33 ۲) -2.22 ۳) 2.22 ۴) -1.33

پاسخنامه

سؤال	پاسخ	سؤال	پاسخ
۱	۱	۲۱	۳
۲	۲	۲۲	۳
۳	۳	۲۳	۳
۴	۳	۲۴	۳
۵	۱	۲۵	۱
۶	۲	۲۶	۳
۷	۴	۲۷	۴
۸	۴	۲۸	۱
۹	۳	۲۹	۱
۱۰	۳	۳۰	۱
۱۱	۱	۳۱	۳
۱۲	۲	۳۲	۱
۱۳	۲	۳۳	۳
۱۴	۴	۳۴	۲
۱۵	۲	۳۵	۲
۱۶	۳	۳۶	۴
۱۷	۱	۳۷	
۱۸	۲	۳۸	
۱۹	۴	۳۹	
۲۰	۳	۴۰	

برای دریافت پاسخ تشریحی سؤالات به سایت www.Tourani.ir مراجعه کنید.

سؤالات آزمون سراسری سال ۸۹

مدیریت و حسابداری

۱. سود شرکتی در 12 ماه سال به ترتیب $0,9,5,8,12,7,4,5,6,2,4,10$ می‌باشد. واریانس داده‌های نامطلوب کدام است؟

۱) 10.33 ۲) 10.82 ۳) 11.63 ۴) 13.20

۲. در 75 داده آماری $\sum_{i=1}^{75}(x_i-15)=0$ و $\sum_{i=1}^{75}(x_i-15)^2=432$ می‌باشد. اگر ضریب پراکندگی داده‌های $y_i=\frac{1}{2}x_i+a$ برابر 0.2 باشد، a کدام است؟

۱) 1.5 ۲) -1.5 ۳) 2 ۴) -2

۳. متغیرهای پیوسته در جدول زیر گروه‌بندی شده‌اند. متغیر 80 درصدی داده‌ها کدام است؟

حدود دسته	$12-15$	$15-18$	$18-21$	$21-24$	$24-27$	$27-30$
فراوانی	5	10	9	11	8	7

۱) 25.875 ۲) 25.625 ۳) 26.125 ۴) 26.225

۴. در 40 داده آماری مقدار انحراف معیار برابر 2.5 و $\sum(x_i-\bar{x})^3$ برابر 60 می‌باشد. نوع چولگی آن کدام است؟

۱) چوله به راست ـ تقریباً نرمال
۲) چوله به راست ـ تفاوت زیاد با نرمال
۳) چوله به چپ ـ تقریباً نرمال
۴) چوله به چپ ـ تفاوت زیاد با نرمال

۵. با حروف کلمه "SUCCESS" چند رمز عبور چهارحرفی می‌توان ساخت؟

۱) 114 ۲) 124 ۳) 141 ۴) 142

۶. هر یک از ارقام $1,2,3,4,5,6$ را بر روی شش کارت باریک نوشته به طور تصادفی در کنار هم قرار می‌دهیم تا عدد شش رقمی حاصل شود. با کدام احتمال عدد حاصل، مضرب 6 یا مضرب 5 می‌باشد؟

۱) $\frac{1}{2}$ ۲) $\frac{2}{3}$ ۳) $\frac{1}{3}$ ۴) $\frac{3}{4}$

۷. تابع احتمال توأم دو متغیر X و Y به صورت زیر است. مقدار کوواریانس کدام است؟

Y \ X	-1	2	5
2	0.2	0.3	0
4	0.1	0.15	0.25

۱) 1.05
۲) 1.25
۳) 1.45
۴) 1.65

۸. تیمی 5 مسابقه دارد، احتمال برد و باخت و مساوی در هر بازی به ترتیب 0.4، 0.25 و 0.35 است. با کدام احتمال این تیم 3 برد و یک باخت و یک مساوی ممکن است داشته باشد؟

۱) 0.108 ۲) 0.112 ۳) 0.124 ۴) 0.128

۹. سه کارگر A، B و C به ترتیب 40 درصد، 36 درصد و 24 درصد ظروف سرامیک فروشگاهی را تولید می‌کننـد. درصد صنایع دستی معیوب این کارگران به ترتیب 3، 2 و 1 می‌باشد. اگر یک ظرف تولیدی معیوب باشد، با کـدام احتمال این ظرف معیوب را کارگر C تولید کرده است؟

۱) $\frac{1}{6}$ ۲) $\frac{1}{8}$ ۳) $\frac{1}{9}$ ۴) $\frac{2}{9}$

۱۰. هشت مورد از حساب‌های شرکتی دارای اشتباه است. احتمال اینکه حسابرسی داخلی متوجه هر اشـتباه باشـد 0.6 است. با کدام احتمال، ممکن است پنجمین حساب اشتباه دومین حسابی باشد که وی متوجه شده است؟

۱) 0.09216 ۲) 0.0512 ۳) 0.0532 ۴) 0.01152

۱۱. در توزیع پواسون، انحراف معیار برابر 2 می‌باشد. در این توزیع احتمال اگر $P(X=0)=0.018$ باشـد، آن‌گـاه $P(X=3)$، کدام است؟

۱) 0.108 ۲) 0.186 ۳) 0.192 ۴) 0.204

۱۲. واریانس متغیر تصادفی X با تابع چگالی یکنواخت $f(x)=\begin{cases}\frac{1}{6} & ;-1<x<5\\ 0 & ; \text{جای دیگر}\end{cases}$ کدام است؟

۱) 2 ۲) 2.5 ۳) 3 ۴) 3.6

۱۳. از یک جامعه 97 عضوی با واریانس 6، نمونه‌های 16تایی انتخاب می‌کنیم. انحراف معیار توزیع میانگین نمونـه، کدام است؟

۱) 0.5625 ۲) 0.5675 ۳) 0.6525 ۴) 0.6575

۱۴. ضریب همبستگی بین دو صفت x و y در جدول زیر کدام است؟

x	2	3	4	5	6
y	4	2	5	1	3

۱) 0.25
۲) -0.25
۳) 0.3
۴) -0.3

۱۵. در یک کارگاه تولیدی از 40 نفر کارگر و 10 نفر کارمند در مورد وضع بهداشت محل کار به طور تصادفی سـؤال شده است. با استفاده از فراوانی مورد انتظار، آماره «کای دو» کدام است؟

	کارمند	کارگر
راضی	7	18
ناراضی	3	22

۱) 2
۲) 2.25
۳) 2.75
۴) 3

علوم اقتصادی

۱. مقدار r^2 از کدام فرمول به دست می‌آید؟ ($\hat{\beta}_2$ شیب خط رگرسیون و داده‌ها برحسب انحراف از میانگین می‌باشند.)

۱) $r^2=\hat{\beta}_2^2\frac{\sum x_t y_t}{\sum y_t^2}$ ۲) $r^2=\hat{\beta}_2^2\frac{\sum x_t^2}{\sum y_t^2}$ ۳) $r^2=\frac{\sum x_t^2 y_t^2}{\sum x_t^2 \sum y_t^2}$ ۴) $r^2=\hat{\beta}_2\left(\frac{S_X^2}{S_Y^2}\right)$

۲. منظور از میانگین مجذور خطا $\left[MSE\left(\hat{\beta}\right)\right]$ چیست؟

۱) $MSE\left(\hat{\beta}\right)=E\left(\hat{\beta}-\beta\right)^2$ ۲) $\left[MSE\left(\hat{\beta}\right)\right]=Var\hat{\beta}+(\text{تورش } \hat{\beta})$

۳) $MSE\left(\hat{\beta}\right)=Var\left(\hat{\beta}\right)-(\text{اریبی } \hat{\beta})^2$ ۴) $MSE\left(\hat{\beta}\right)=E\left[\beta-E\left(\hat{\beta}\right)\right]^2$

۳. اگر واریانس داده‌های آماری $x_1, x_2, \dots, x_n$ برابر 15 باشد، واریانس داده‌های آماری زیر کدام است؟
$3x_1+2, 3x_2+2, \dots, 3x_n+2$

۱) 135 ۲) 220 ۳) 155 ۴) 225

۴. به منظور مقایسه هزینه مواد غذایی خانوارها در 4 منطقه، از هر یک از این مناطق نمونه‌ای تصادفی به حجم $n=5$ خانوار انتخاب شده است. بر اساس نتایج مشاهدات، مجموع مربعات بین‌گروهی (تیمارها) برابر 12 و مجموع مربعات درون گروهی (خطاها) برابر با 32 به دست آمده است. مقدار عددی آماره آزمون عبارت است از:

۱) $F_{16,3}=\frac{3}{10}$ ۲) $F_{16,3}=0.5$ ۳) $F_{3,16}=\frac{10}{3}$ ۴) $F_{3,16}=2$

۵. میانگین و انحراف معیار حقوق در یک سازمان به‌ترتیب برابر 50 و 20 هزار تومان است. اگر حقوق‌ها در این سازمان به‌اندازه 25 درصد افزایش یابد، در آن صورت ضریب تغییرات حقوق:

۱) بیش از 25 درصد افزایش خواهد داشت. ۲) 25 درصد افزایش می‌یابد.

۳) کمتر از 25 درصد افزایش می‌یابد. ۴) هیچ تغییری نخواهد کرد.

۶. اگر آماره $T(X)=\sum_{i=1}^{n}a_i x_i$ یک تابع خطی از نتایج مشاهدات و $\sum_{i=1}^{n}a_i=1$ برقرار باشد، واریانس با کدام‌یک از شرایط ذیل حداقل می‌شود؟

۱) $a_1<a_2<a_3<\dots<a_n$ ۲) $a_1>a_2>a_3>\dots>a_n$

۳) $a_1=a_2=\dots=a_n=\frac{1}{n}$ ۴) $a_1=a_2=\dots=a_n=\frac{1}{\sqrt{n}}$

۷. اگر سه قطعه زمین مربعی شکل A، B و C به ابعاد 7، 5 و 1 متر با سه قطعه زمین مربعی یکسان تعویض شود، ابعاد زمین‌های مساوی چقدر است؟

۱) 2 ۲) 4 ۳) 2.5 ۴) 5

۸. اگر کمیت تصادفی X دارای تابع چگالی احتمال به صورت $f(x)=5e^{-5x}\ ;\ x>0$ باشد، دهک هشتم توزیع کدام است؟

۱) $5e^{-5}$ ۲) $\frac{1}{8}\ln(8)$ ۳) $\frac{1}{5}\ln(5)$ ۴) $8e^{-8}$

۹. در صورتی که متغیر تصادفی X دارای توزیع یکنواخت به صورت ذیل باشد، مقدار α و $E(X)$ به‌ترتیب برابراست با:

$$f(x)=\begin{cases}0 & x<0\\ \alpha & 0\le x\le 7\\ 0 & x>7\end{cases}$$

۱) $E(X)=3.5 \quad , \quad \alpha=\frac{1}{7}$

۲) $E(X)=171.5 \quad , \quad \alpha=7$

۳) $E(X)=2.5 \quad , \quad \alpha=7$

۴) $E(X)=2.5 \quad , \quad \alpha=\frac{1}{7}$

۱۰. برای هر سطح معنی‌دار و اندازه نمونه‌ای، مقدار بحرانی توزیع t:

۱) برابر صفر است.

۲) همیشه بزرگ‌تر از مقدار بحرانی توزیع Z می‌باشد.

۳) برابر با مقدار بحرانی توزیع Z می‌باشد.

۴) همیشه کوچک‌تر از مقدار بحرانی توزیع Z می‌باشد.

۱۱. با تغییر مدیریت در یک کارخانه، فروش در سال اول 2 برابر سال قبل، در سال دوم، سه برابر سال قبل و در سال سوم، چهار برابر سال قبل شده است. به طور متوسط، مقدار فروش از شروع مدیریت جدید چند برابر شده است؟

۱) 2.5 ۲) $\frac{36}{13}$ ۳) $2\sqrt[3]{3}$ ۴) 3

۱۲. سه کتاب ریاضی و چهار کتاب اقتصاد را در یک ردیف کنار هم قرار می‌دهیم. احتمال اینکه کتاب‌های ریاضی کنار هم و کتاب‌های اقتصاد نیز کنار هم قرار بگیرند، برابر است با:

۱) $\frac{2}{35}$ ۲) $\frac{1}{2}$ ۳) $\frac{12}{7!}$ ۴) $\frac{1}{6!}$

۱۳. برای دو مقدار x_1 و x_2، کمترین مقدار عبارت $\sum_{i=1}^{2}(x_i-a)^2$ برابر است با:

۱) $x_1^2+x_2^2$ ۲) صفر   ۳) $\frac{1}{2}(x_1+x_2)^2$ ۴) $\frac{1}{2}(x_1-x_2)^2$

۱۴. اگر تخمین نقطه‌ای واریانس قیمت نفت در 100 روز گذشته $S^2=10$ باشد، فاصله اعتماد 95 درصدی برای واریانس جامعه کدام است؟ (مقادیر بحرانی $X^2_{0.025,d.f}=128.422$, $X^2_{0.975,d.f}=73.361$ می‌باشند.)

۱) $\left(\frac{990}{128.422},\frac{990}{73.361}\right)$

۲) $\left(\frac{73.361}{990},\frac{128.422}{990}\right)$

۳) $\left(\frac{99}{1284.22},\frac{99}{733.61}\right)$

۴) $\left(\frac{733.61}{99},\frac{1284.22}{99}\right)$

۱۵. اگر اختلاف بین مقدار μ_0 و μ_1 که به ترتیب در فرضیه صفر و یک بیان شده‌اند $(H_0:\mu=\mu_0 \quad , \quad H_1:\mu=\mu_1)$ با $\Delta\mu$ بیان شود، توان آزمون:

۱) با افزایش $\Delta\mu$ کاهش می‌یابد.

۲) مستقل از $\Delta\mu$ است.

۳) با افزایش $\Delta\mu$ افزایش می‌یابد.

۴) با $\sqrt{\Delta\mu}$ متناسب است.

۱۶. اگر تخمین زننده $\tilde{X}=\frac{X_{min}+X_{max}}{2}$ و میانگین نمونه‌ای $\overline{X}$ در نمونه‌ای به حجم n با توزیع نرمال با امید ریاضی μ و واریانس σ^2 تعریف شده باشد، نسبت کارایی $\left(Var\left(\overline{X}\right)/Var\left(\tilde{X}\right)\right)$ عبارت است از:

۱) $\frac{n}{2}$ ۲) $\frac{2}{n}$ ۳) $\frac{\sqrt{n}}{2}$ ۴) $\frac{n}{\sqrt{2}}$

۱۷. تخمین حداکثر درستنمایی واریانس $\left(\sigma^2\right)$ در جامعه‌ای با توزیع نرمال مبتنی بر نمونه‌ای به حجم n برابر است با:

۱) $\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}$ ۲) $\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}$ ۳) $\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-2}$ ۴) $\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n+1}$

۱۸. در یک توزیع پواسون داریم $P(X=0)=\frac{1}{2}P(X=2)$ مقدار $P(X>0)$ برابر است با:

۱) $1-e^{-1}$ ۲) $\frac{1}{2}$ ۳) $1-e^{-2}$ ۴) $1-e^{-4}$

۱۹. محصولات کارخانه‌ای به تساوی توسط دو خط تولید A و B تولید می‌شوند. ده درصد محصولات خط A و 30 درصد محصولات خط B معیوب هستند. اگر کالایی به طور تصادفی انتخاب شود و سالم باشد، احتمال اینکه این محصول از خط تولید B باشد چیست؟

۱) $\frac{3}{4}$ ۲) $\frac{7}{16}$ ۳) $\frac{7}{8}$ ۴) $\frac{1}{3}$

۲۰. متغیر تصادفی X دارای توزیع کای ــ دو $\left(\chi^2\right)$ با امید ریاضی 15 است. ضریب تغییرات این متغیر کدام است؟

۱) $\frac{1}{\sqrt{2}}$ ۲) $\frac{1}{2}$ ۳) 2 ۴) $\sqrt{2}$

برنامه‌ریزی شهری و منطقه‌ای

۱. **کدام مقیاس دارای صفر قراردادی است؟**

۱) اسمی ۲) فاصله‌ای ۳) رتبه‌ای ۴) نسبی

۲. **اختلاف میانگین هندسی و میانگین حسابی داده‌های مقابل کدام است؟**

x	9	12	16
f	2	3	2

۱) 0.24 ۲) 0.27
۳) 0.25 ۴) 0.28

۳. **توزیع نمرات مسئولیت‌پذیری کارکنان یک مجموعه در جدول زیر تنظیم شده است. انحراف چارکی کدام است؟**

نمرات	<10	$10-12$	$12-14$	$14-16$	$16-18$	≥ 18
فراوانی	5	10	7	18	8	4

۱) 2.12 ۲) 2.16 ۳) 2.14 ۴) 2.18

۴. **در 40 داده آماری مجموع تمام داده‌ها 168 و مجموع مجذورات این داده‌ها 808 می‌باشد. ضریب پراکندگی این داده‌ها کدام است؟**

۱) 0.38 ۲) 0.25 ۳) 0.35 ۴) 0.42

۵. **نمرات دانشجویان دارای توزیع نرمال با میانگین 63.6 می‌باشد. 20 درصد این دانشجویان نمرات بیشتر از یا مساوی 72 دارند، اگر $P(Z \leq 0.84) = 0.8$، انحراف معیار کدام است؟**

۱) 8 ۲) 10 ۳) 9 ۴) 12

۶. **با حروف کلمه STANDARD چند رمز عبور چهارحرفی می‌توان ساخت؟**

۱) 576 ۲) 606 ۳) 596 ۴) 612

۷. **اگر $Cov(X,Y) = 0$ باشد، آن‌گاه رابطه بین X و Y چگونه است؟**

۱) رابطه غیرخطی یا مستقل از هم ۲) مستقل از هم
۳) رابطه غیرخطی ۴) رابطه‌ای وجود ندارد.

۸. **از محصولات تولیدی یک کارخانه 45 درصد تاریخ مصرف ندارند، 30 درصد برچسب قیمت ندارند و 25 درصد نه تاریخ مصرف دارند و نه برچسب قیمت، 4 درصد کل محصول معیوب است. اگر کالایی از بین آن‌ها انتخاب شود، با کدام احتمال این کالا تاریخ مصرف و برچسب قیمت دارد و سالم است؟**

۱) 0.45 ۲) 0.50 ۳) 0.48 ۴) 0.54

۹. **در یک توزیع آماری چارک اول 112.4 و چارک سوم 134.4 و میانه 121 می‌باشد. ضریب چولگی کدام است؟**

۱) 0.206 ۲) 0.224 ۳) 0.218 ۴) 0.226

۱۰. **در یک توزیع آماری میانگین برابر 36 و میانه 42 و این توزیع از چولگی معقولی برخوردار است، مد کدام است؟**

۱) 40 ۲) 48 ۳) 45 ۴) 54

۱۱. **هفت مجسمه متمایز را به چند طریق می‌توان در سه پارک مورد نظر به تعداد 2 و 2 و 3 مجسمه قرار داد؟**

۱) 84 ۲) 168 ۳) 105 ۴) 210

۱۲. به ازای کدام مقدار a تابع $x=0,1,2,3\ ;\ P(X=x)=\dfrac{\binom{4}{x}\binom{8}{3-x}}{a}$ **یک تابع احتمال است؟**

۱) 156 ۲) 220 ۳) 216 ۴) 224

۱۳. امید ریاضی متغیر تصادفی X برابر 3.5 و امید ریاضی X^2 برابر 14.5 محاسبه شده است، $V(-2X+5)$ کدام است؟

۱) 9 ۲) 7.5 ۳) 8.5 ۴) 6

۱۴. احتمال موفقیت در یک آزمایش برنولی 60 درصد است، اگر 96 بار این آزمایش تکرار شود انحراف معیار تعداد موفقیت‌ها در این توزیع کدام است؟

۱) 4.8 ۲) 3.2 ۳) 3.6 ۴) 2.4

۱۵. جهت بررسی اجرای یک پروژه بسیار وسیع شهری از برخی مناطق به تصادف نمونه‌گیری می‌کنیم. این نوع نمونه‌گیری کدام است؟

۱) تصادفی ساده ۲) خوشه‌ای ۳) منظم ۴) طبقه‌ای

۱۶. تابع چگالی احتمال به صورت $f(x)=\begin{cases}k(16x-x^3) & 0\le x\le 4\\ 0 & \text{سایر جای}\end{cases}$ **داده شده است. عدد k کدام است؟**

۱) $\frac{1}{32}$ ۲) $\frac{1}{8}$ ۳) $\frac{1}{16}$ ۴) $\frac{1}{64}$

۱۷. از جامعه‌ای با میانگین 72 و واریانس 12، به طور تصادفی نمونه 64 عضوی، انتخاب می‌کنیم. انحراف معیار میانگین کدام است؟

۱) 1.5 ۲) 1.25 ۳) 1.75 ۴) 1.8

۱۸. روابط عمومی شهرداری ادعا می‌کند که 80 درصد مردم از عملکرد کارکنان شهرداری راضی‌اند. یک نمونه 400 تایی از مردم انتخاب شده که 75 درصد آنان از عملکرد شهرداری راضی‌اند. آماره آزمون برای بررسی صحت این ادعا کدام است؟

۱) -2.5 ۲) -1.5 ۳) 1.25 ۴) 1.75

۱۹. ضریب همبستگی بین دو صفت x و y کدام است؟

۱) 0.4 ۲) 0.2
۳) 0.5 ۴) 0.6

x	5	9	11	7
y	8	12	10	6

سنجش از دور و سیستم اطلاعات جغرافیایی (GIS)

۱. در یک توزیع متقارن، داده‌های کمتر از چارک اول و داده‌های بیشتر از چارک سوم را حذف کرده‌ایم و میانگین پیراسته 72 می‌باشد. میانگین حسابی داده‌های اصلی کدام است؟

۱) 72 ۲) کمتر از 72 ۳) بیشتر از 72 ۴) قابل پیش‌بینی نیست.

۲. در 45 داده آماری مجموع تمام داده‌ها 81 و مجموع مربعات این داده‌ها 405 می‌باشد. ضریب پراکندگی متغیرهای جدیدی که هر یک از $\frac{3}{2}$ داده‌های مفروض 4.5 واحد بیشتر باشد، کدام است؟

۱) 0.125 ۲) 0.225 ۳) 0.25 ۴) 0.5

۳. داده‌های آماری پیوسته یک پژوهش در جدول زیر گروه‌بندی شده‌اند. چند درصد داده‌ها کمتر از 36.5 است؟

حدود دسته	$20-25$	$25-30$	$30-35$	$35-40$	$40-45$	$45-50$
فراوانی	12	17	19	20	15	7

۱) 56 ۲) 60 ۳) 63 ۴) 64

۴. در 25 داده آماری، $\sum(x_i-\bar{x})^4=6075$ و انحراف معیار جامعه برابر 3 می‌باشد. ضریب کشیدگی کدام است؟

۱) صفر ۲) 0.263 ۳) 1 ۴) 3

۵. با حروف کلمه «SKEWNESS» چند رمز عبور چهارحرفی می‌توان ساخت؟

۱) 224 ۲) 242 ۳) 268 ۴) 286

۶. شخصی در چهار آزمایش متوالی شرکت می‌کند. احتمال قبولی در آزمایش اول $\frac{1}{2}$، احتمال قبولی در هر آزمایش بعدی با شرط قبولی در آزمایش قبل $\frac{1}{2}$ و در صورت رد در آزمایش قبل $\frac{1}{4}$ می‌باشد و اگر حداقل در سه آزمایش قبول شود، پذیرفته خواهد شد. احتمال پذیرش وی کدام است؟

۱) $\frac{5}{32}$ ۲) $\frac{7}{32}$ ۳) $\frac{9}{32}$ ۴) $\frac{11}{32}$

۷. در توزیع احتمال جدول روبه‌رو، مقدار $V(-2X+3)$ کدام است؟

x	3	5	6	8
$f(x)$	0.2	0.4	0.15	α

۱) 5.9 ۲) 9.6 ۳) 10.8 ۴) 11.8

۸. در یک انتخابات، طبق پیش‌بینی 50 درصد به نامزد A، 30 درصد به نامزد B و 20 درصد به نامزد C رأی می‌دهند. از 6 نفر آماده رأی دادن، با کدام احتمال 3 نفر به نامزد A، 1 نفر به نامزد B و 2 نفر به نامزد C رأی می‌دهند؟

۱) 0.018 ۲) 0.054 ۳) 0.06 ۴) 0.09

۹. در ظرفی 6 مهره قرمز و 9 مهره آبی موجود است. یک مهره از ظرف خارج کرده و به جای آن دو مهره به رنگ دیگر در ظرف می ریزیم سپس یک مهره خارج می‌کنیم. اگر هر دو مهره خارج شده همرنگ باشند، با کدام احتمال هر دو مهره قرمز است؟

۱) $\frac{2}{5}$ ۲) $\frac{3}{5}$ ۳) $\frac{5}{17}$ ۴) $\frac{12}{17}$

۱۰. چهار درصد از «پیام‌های کوتاه» نارساست. در ارسال 150 پیام کوتاه، انحراف معیار تعداد پیام‌های نارسا کدام است؟

۱) 1.8 ۲) 2.4 ۳) 3.2 ۴) 4.5

۱۱. از هر 100,000 واحد کالا به طور متوسط 125 عدد آن معیوب است. با کدام احتمال از بین 4000 واحد این کالا حداکثر 2 کالا معیوب است؟ $\left(e^{-5}=0.007\right)$

۱) 0.108 ۲) 0.112 ۳) 0.129 ۴) 0.135

۱۲. امید ریاضی متغیر تصادفی X با تابع چگالی احتمال $f(x)=\begin{cases}\frac{1}{6\sqrt[3]{x}} & ; \quad 0<x<8 \\ 0 & ; \quad \text{دیگر جای}\end{cases}$**، کدام است؟**

۱) 3.2 ۲) 3.6 ۳) 4.2 ۴) 4.6

۱۳. در یک جامعه آماری یک نمونه 100 عضوی انتخاب کرده‌ایم. پس از محاسبات لازم $\bar{X}=35.5$ **و** $S=8$ **حاصل شده است. با 92 درصد اطمینان، میانگین جامعه در کدام فاصله است؟** $\left(S_0^{1.75}=0.46\right)$

۱) $(34.9\,,\,36.1)$ ۲) $(34.1\,,\,36.9)$ ۳) $(33.9\,,\,37.1)$ ۴) $(33.4\,,\,37.6)$

۱۴. شیب خط رگرسیون برآوردی y نسبت به x در جدول روبه‌رو، کدام است؟

x	5	7	11	12	15
y	3	4	6	8	9

۱) 0.525 ۲) 0.575
۳) 0.625 ۴) 0.675

۱۵. در محاسبه آماره کای دو، جدول توافقی با کدام درجه آزادی، بهتر است دستور اصلاح‌شده یتس را به کار برد؟

۱) 1 ۲) 2 ۳) بیشتر از 3 ۴) بیشتر از 5

محیط زیست

۱. اگر در بررسی فرضیه‌ای از تمام مشاهدات جامعه استفاده شود، آن را چه می‌نامند؟

۱) توصیفی ۲) استنباطی ۳) پارامتریک ۴) همبستگی

۲. میانگین هندسی داده‌های 12 و 54 و 6.75 و 18 و 24 کدام است؟

۱) 14.4 ۲) 15.6 ۳) 16 ۴) 18

۳. در داده‌های آماری دسته‌بندی‌شده جدول زیر، میانگین حسابی کدام است؟

حدود دسته	$215-225$	$225-235$	$235-245$	$245-255$	$255-265$
فراوانی تجمعی	11	28	51	72	90

۱) 241 ۲) 242 ۳) 243 ۴) 244

۴. انحراف چارکی در داده‌های آماری طبقه‌بندی‌شده زیر کدام است؟

فاصله طبقات	<8	$8-12$	$12-16$	$16-20$	$20-24$	≥ 24
فراوانی مطلق	4	12	16	18	9	5

۱) 3.74 ۲) 3.77 ۳) 3.82 ۴) 3.84

۵. در سه جامعه با تعداد مشاهدات 90 و 110 و 200، میانگین‌ها متفاوت و واریانس‌ها به ترتیب 7، 10 و 9 محاسبه شده‌اند. کدام عدد برای واریانس ترکیب این سه جامعه مورد قبول است؟

۱) 8.775 ۲) 8.825 ۳) 8.95 ۴) 8.59

۶. در یک پژوهش با میانگین 52 و واریانس 64 بنا بر قضیه «چی‌بی‌شف» حداقل 84 درصد داده‌ها در کدام بازه است؟

۱) $(32,72)$ ۲) $(40,64)$ ۳) $(42,62)$ ۴) $(44,60)$

۷. در یک توزیع با ضریب چولگی مثبت، کدام نابرابری برقرار است؟

۱) میانه > میانگین > مد ۲) میانگین > میانه > مد

۳) مد > میانه > میانگین ۴) مد > میانگین > میانه

۸. برای 60 داده آماری با میانگین 15 داریم $\sum(x_i-15)^2=300$ و $\sum(x_i-15)^4=4500$. ضریب کشیدگی کدام است؟

۱) 0.5 ۲) 1 ۳) 3 ۴) صفر

۹. به چند طریق می‌توان 9 نفر کارمند جدید را در اتاق‌های 2 نفره و 3 نفره و 4 نفره جای داد؟

۱) 1080 ۲) 1170 ۳) 1260 ۴) 1350

۱۰. از ظرفی که دارای 4 مهره سفید و 5 مهره سیاه است، 3 مهره بیرون می‌آوریم. اگر فقط دو مهره از آن‌ها همرنگ باشند، با کدام احتمال 2 مهره سفید و 1 مهره سیاه خارج شده‌اند؟

۱) $\frac{2}{5}$ ۲) $\frac{3}{5}$ ۳) $\frac{3}{7}$ ۴) $\frac{4}{7}$

۱۱. اعداد 1 تا 10 را بر روی 10 مهره یکسان نوشته‌ایم. به تصادف دو مهره از بین آن‌ها بیرون می‌آوریم، با کدام احتمال اعداد این دو مهره متوالی هم‌اند؟

۱) $\frac{1}{4}$ ۲) $\frac{1}{5}$ ۳) $\frac{1}{6}$ ۴) $\frac{2}{9}$

۱۲. شخصی در سه امتحان پی‌درپی شرکت می‌کند. احتمال قبولی وی در امتحان اول 0.6 و احتمال قبولی وی در امتحانات بعدی به شرط قبولی یا رد در امتحان قبلی به ترتیب 0.6 یا 0.3 است. شرط پذیرش، قبولی حداقل در دو امتحان است؛ احتمال پذیرش وی کدام است؟

۱) 0.405 ۲) 0.496 ۳) 0.504 ۴) 0.512

۱۳. به ازای کدام مقدار a تابع $P(X=x)=\frac{\binom{5}{x}}{a}\ ;\ x=0,1,2,3,4,5$ یک تابع احتمال است؟

۱) 16 ۲) 32 ۳) 48 ۴) 64

۱۴. اگر $E(X)=3.5$ و $E(X^2)=14.5$ باشد، $V(-2X+3)$ کدام است؟

۱) 7.5 ۲) 8 ۳) 10 ۴) 9

۱۵. اگر $E(X)=3.2\ ,\ E(Y)=2.5\ ,\ E(XY)=9$ باشد، آن‌گاه $V(X)+V(Y)-V(X-Y)$ برابر با کدام است؟

۱) -2 ۲) -1 ۳) 2 ۴) 1

۱۶. بنا بر پیش‌بینی‌های موجود، 80 درصد نهال‌های کاشته‌شده سبز می‌شوند. با کدام احتمال از 5 نهال کاشته‌شده، حداقل 2 نهال سبز می‌شود؟

۱) 0.99328 ۲) 0.99216 ۳) 0.98736 ۴) 0.98924

۱۷. در کدام توزیع، امید ریاضی مربع انحراف معیار است؟

۱) پواسون ۲) برنولی ۳) کای‌ـ دو ۴) t استودنت

۱۸. نسبت تلفات در پرندگان مهاجر 1.2 در هزار است. از بین 2500 پرنده مهاجر با کدام احتمال 5 پرنده تلف می‌شوند؟ $\left(\frac{1}{e^3}=0.048\right)$

۱) 0.0972 ۲) 0.0986 ۳) 0.1007 ۴) 0.1005

۱۹. متغیر تصادفی X با تابع چگالی احتمال $P(x)=\begin{cases}\frac{1}{3}(2x+k) & ;\ 0\le x\le 1\\ 0 & \text{جاهای دیگر}\end{cases}$ داده شده است. با تعیین k مقدار $P\left(0\le X\le \frac{1}{2}\right)$ کدام است؟

۱) $\frac{1}{2}$ ۲) $\frac{1}{3}$ ۳) $\frac{7}{12}$ ۴) $\frac{5}{12}$

۲۰. نمرات مسئولیت‌پذیری کارکنان یک سازمان توزیع نرمال با میانگین 16 و انحراف معیار 5 می‌باشد. اگر $P(Z\le 0.44)=0.67$ باشد، چند درصد کارکنان نمراتی کمتر از 13.8 دارند؟

۱) 17 ۲) 33 ۳) 34 ۴) 67

۲۱. در یک توزیع نرمال اندازه دو مقدار 33 و 24 بر حسب متغیر استاندارد Z برابر 2 و -1 می‌باشد. میانگین این جامعه کدام است؟

۱) 25.5 ۲) 26 ۳) 26.5 ۴) 27

۲۲. در یک توزیع آماری پیوسته دارای چولگی شدید، مقدار نمونه حداقل چقدر باشد تا برای آماره $\bar{X}$ بتوان تقریب نرمال به کار برد؟

۱) 30 ۲) 50 ۳) 60 ۴) 100

۲۳. پنجاه درصد از کارکنان شرکتی تحصیلات دانشگاهی دارند. با کدام احتمال حداقل 55 درصد افراد یک نمونه تصادفی 100 نفره از آنان تحصیلات دانشگاهی دارند؟ $\left(S_{-\infty}^{-1} = 0.1587\right)$

۱) 0.1587 ۲) 0.3174 ۳) 0.6587 ۴) 0.8413

۲۴. شیب خط رگرسیون دو صفت x و y در جدول زیر کدام است؟

X	5	3	4	7	1
Y	7	9	3	2	4

۱) 0.2 ۲) 0.25 ۳) -0.3 ۴) -0.4

۲۵. وارون عدد $F_{0.9,5,7}$ برابر کدام است؟

۱) $F_{0.05,7,5}$ ۲) $F_{0.1,5,7}$ ۳) $F_{0.1,7,5}$ ۴) $F_{0.9,7,5}$

پاسخ‌های تشریحی

مدیریت و حسابداری

۱. گزینه ۱ درست است.

واریانس سودهای نامطلوب (نیمه‌واریانس) برای شرکت عبارت است از واریانس سودهایی که از میانگین سود کمتر باشند. بنابراین:

ابتدا میانگین داده‌ها را محاسبه می‌کنیم:

$$\mu = \frac{\sum x_i}{N} = \frac{0+9+5+8+12+7+4+5+6+2+4+10}{12} = \frac{72}{12} = 6$$

سپس داده‌های نامطلوب $(x_i < \mu)$ را مشخص می‌کنیم:

داده‌های نامطلوب ; $x_i < 6 \quad \rightarrow \quad 0 , 2 , 4 , 4 , 5 , 5$

در انتها واریانس داده‌های نامطلوب را محاسبه می‌کنیم:

$$\text{نیمه‌واریانس} : S.V = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k} = \frac{(0-6)^2 + (2-6)^2 + (4-6)^2 + (4-6)^2 + (5-6)^2 + (5-6)^2}{6}$$

$$= \frac{36+16+4+4+1+1}{6} = \frac{62}{6} = 10.33$$

۲. گزینه ۲ درست است.

یادآوری: μ (میانگین) تنها شاخص مرکزی است که دو خاصیت زیر را دارد:

$$\sum (x_i - \mu) = 0$$

$$\sum (x_i - \mu)^2 : \min$$

بنابراین:

$$\sum_{i=1}^{75} (x_i - 15) = 0 \quad \rightarrow \quad \mu = 15$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{432}{75} = 5.76 \quad \rightarrow \quad \sigma = \sqrt{5.76} = 2.4$$

حال ازآنجاکه ضریب تغییرات y_i برابر 0.2 است، داریم:

$$CV_{Y_i}=\frac{\sigma_{Y_i}}{\mu_{Y_i}}=\frac{\sigma_{\frac{1}{2}X_i+a}}{\mu_{\frac{1}{2}X_i+a}}=\frac{\frac{1}{2}\sigma_X}{\frac{1}{2}\mu_X+a}=\frac{\frac{1}{2}(2.4)}{\frac{1}{2}(15)+a}=0.2$$

$$\rightarrow\ \frac{1}{2}\times 2.4=\frac{15}{2}\times 0.2+0.2a\ \rightarrow\ 1.2=1.5+0.2a\ \rightarrow\ 0.2a=-0.3\ \rightarrow\ a=-1.5$$

۳. گزینه ۱ درست است.

منظور از متغیر 80 درصدی همان صدک 80ام است؛ بنابراین:

الف) F_{c_i} را در جدول محاسبه می‌کنیم:

$C-L$	$12-15$	$15-18$	$18-21$	$21-24$	$24-27$	$27-30$	
F_i	5	10	9	11	8	7	$N=\sum F_i=50$
F_{c_i}	5	15	24	35	43	50	

ب) صدک در اولین طبقه‌ای است که $40=\frac{80\times 50}{100}=\frac{aN}{100}\geq F_{c_i}$ باشد؛ بنابراین طبقه $(24-27)$ طبقه صدک‌دار است.

ج) طبقه پیوسته است.

د) مقدار صدک هشتادم برابر است با:

$$P_a=L_i+\frac{\frac{aN}{100}-F_{c_{i-1}}}{F_i}\times I\ \xrightarrow[N=50]{a=80}\ P_{80}=24+\frac{\frac{80\times 50}{100}-35}{8}\times 3=24+\frac{5}{8}\times 3=25.875$$

۴. گزینه ۱ درست است.

با توجه به داده‌های مسئله ابتدا مقدار چولگی را به کمک رابطه گشتاوری به دست می‌آوریم:

راه حل اول:

$$Sk=\frac{\frac{\sum(x_i-\mu)^3}{N}}{\sigma^3}=\frac{\frac{60}{40}}{(2.5)^3}=\frac{1.5}{15.625}=0.096$$

اولاً، ازآنجاکه $Sk>0$، توزیع چوله به راست است.

ثانیاً، ازآنجاکه $|Sk|\leq 0.1$، توزیع تقریباً نرمال است.

راه حل دوم (تستی):

اولاً، ازآنجاکه مقدار $\sum(x_i-\mu)^3$ مثبت است پس چولگی مثبت و به سمت راست است.

ثانیاً، با جایگذاری مقادیر در کسر چولگی گشتاوری مشخص است که صورت خیلی کوچک‌تر از مخرج است و مقدار چولگی نمی‌تواند بزرگ‌تر از 0.5 باشد، پس اختلاف زیادی با نرمال ندارد.

یادآوری:

$|Sk|\leq 0.1\ \rightarrow$ تقریباً نرمال

$0.1<|Sk|\leq 0.5\ \rightarrow$ تفاوت اندک با نرمال

$|Sk|>0.5\ \rightarrow$ تفاوت فاحش با نرمال

۵. گزینه ۱ درست است.

کلمه SUCCESS شامل حروف S، U، C و E است که حرف S، 3 بار و حرف C، 2 بار در آن تکرار شده است.

برای ساختن رمزهای چهارحرفی باید حالات زیر را بررسی کنیم:

ـ تمام کلمات چهارحرفی با 4 حرف غیر تکراری S، U، C و E:

$$\binom{4}{4}\times 4! = 24$$

ـ تمام کلمات چهارحرفی با 2 حرف C و انتخاب 2 حرف دیگر از حروف S، U و E:

$$\binom{3}{2}\times \frac{4!}{2!} = 36$$

ـ تمام کلمات چهارحرفی با 2 حرف S و انتخاب 2 حرف دیگر از حروف C، U و E:

$$\binom{3}{2}\times \frac{4!}{2!} = 36$$

ـ تمام کلمات چهارحرفی با 3 حرف S و انتخاب 1 حرف دیگر از حروف C، U و E:

$$\binom{3}{1}\times \frac{4!}{3!} = 12$$

ـ تمام کلمات چهارحرفی با 2 حرف C و 2 حرف S:

$$\frac{4!}{2!2!} = 6$$

بنابراین، تعداد کل رمزهای چهارحرفی برابر است با:

$$24+36+36+12+6=114$$

۶. گزینه ۲ درست است.

$$\textbf{کل حالات} = 6\times 5\times \cdots \times 1 = 6!$$

حالات مساعد:

اولاً، برای آنکه عدد مضرب 5 باشد رقم یکان باید 5 باشد:

تعداد اعداد مضرب 5 :

حالت 5	حالت 4	حالت 3	حالت 2	حالت 1	حالت 1

$= 5\times 4\times 3\times 2\times 1\times 1 = 5!\times 1$

(زیر پنج خانهٔ اول: اعداد 1، 2، 3، 4 و 6 — زیر خانهٔ آخر: عدد 5)

ثانیاً، برای آنکه عدد مضرب 6 باشد باید هم مضرب 3 باشد (مجموع ارقام بر 3 بخش‌پذیر باشد که هست زیرا $1+2+3+\cdots+6=21$) و هم مضرب 2 (رقم یکان زوج باشد یعنی یکی از اعداد 2، 4 و 6)؛ بنابراین:

تعداد اعداد مضرب 6 :

					یکان
حالت 5	حالت 4	حالت 3	حالت 2	حالت 1	حالت 3

$= 5\times 4\times 3\times 2\times 1\times 3 = 5!\times 3$

(زیر پنج خانهٔ اول: 5 عدد باقی‌مانده — زیر خانهٔ یکان: یکی از اعداد 2، 4 یا 6)

$$\text{حالات مساعد} = 5!\times 1 + 5!\times 3 = 5!\times 4$$

و احتمال مورد نظر برابر است با:

$$\text{احتمال} = \frac{\text{حالات مساعد}}{\text{حالات ممکن}} = \frac{5!\times 4}{6!} = \frac{\cancel{5!}\times 4}{6\times \cancel{5!}} = \frac{4}{6} = \frac{2}{3}$$

۷. گزینه ۱ درست است.

Y \ X	-1	2	5	$f(y)$
2	0.2	0.3	0	0.5
4	0.1	0.15	0.25	0.5
$f(x)$	0.3	0.45	0.25	

$$\begin{cases} Cov(X,Y) = E(XY) - E(X)E(Y) = 6.6 - 1.85\times 3 = 1.05 \\ E(XY) = \sum\sum xy\,f(x,y) = 2\times(-1)\times 0.2 + 2\times 2\times 0.3 + 2\times 5\times 0 + 4\times(-1)\times 0.1 + 4\times 2\times 0.15 + 4\times 5\times 0.25 = 6.6 \\ E(X) = \sum x\,f(x) = -1\times 0.3 + 2\times 0.45 + 5\times 0.25 = 1.85 \\ E(Y) = \sum y\,f(y) = 2\times 0.5 + 4\times 0.5 = 3 \end{cases}$$

۸. گزینه ۲ درست است.

نعداد برد، باخت و مساوی دارای توزیع چندجمله‌ای است؛ بنابراین:

$$P(x_1 = 3\,,\, x_2 = 1\,,\, x_3 = 1) = \frac{5!}{3!\,1!\,1!}\times(0.4)^3\times(0.25)^1\times(0.35)^1 = 0.112$$

برای سادگی محاسبات بهتر است موارد زیر را در نظر بگیرید:

$$\begin{cases} \frac{5!}{3!1!1!} = 20 \\ (0.4)^3 = \left(\frac{4}{10}\right)^3 = \left(\frac{2}{5}\right)^3 = \frac{8}{125} \\ 0.25 = \frac{25}{100} = \frac{1}{4} \\ 0.35 = \frac{35}{100} = \frac{7}{20} \end{cases} \longrightarrow 20\times\frac{8}{125}\times\frac{1}{4}\times\frac{7}{20} = \frac{1120}{10000} = 0.112$$

۹. گزینه ۳ درست است.

با بررسی مسئله و اطلاعات آن مشخص می‌شود که شرایط قضیه بیز برقرار است:

	تولید		معیوب
A:	0.40	$\rightarrow$	0.03
B:	0.36	$\rightarrow$	0.02
C:	0.24	$\rightarrow$	0.01

$$P(\text{کارگر } C \mid \text{کالا معیوب}) = \frac{P(\text{کالا معیوب و کارگر } C)}{P(\text{کالا معیوب})} = \frac{\frac{24}{100}\times\frac{1}{100}}{\frac{40}{100}\times\frac{3}{100} + \frac{36}{100}\times\frac{2}{100} + \frac{24}{100}\times\frac{1}{100}} = \frac{24}{120+72+24} = \frac{24}{216} = \frac{1}{9}$$

۱۰. گزینه ۱ درست است.

وقوع r امین موفقیت در x امین آزمایش برنولی با احتمال p در هر آزمایش دارای توزیع دوجمله‌ای منفی است که در آن:

$$P(x)=\binom{x-1}{r-1}p^r q^{x-r}$$

در این سؤال احتمال آنکه دومین اشتباهی که حسابرسی متوجه می‌شود $(r=2)$، پنجمین اشتباه باشد $(x=5)$، از رابطه زیر به دست می‌آید:

$$P(x)=\binom{5-1}{2-1}(0.6)^2(0.4)^3=4\times\frac{36}{100}\times\frac{64}{1000}=0.09216$$

۱۱. گزینه ۳ درست است.

در هر توزیع پواسون با پارامتر λ داریم:

$$\begin{cases}P(x)=\dfrac{e^{-\lambda}\,\lambda^x}{x!}\\ \mu=\lambda\\ \sigma^2=\lambda \;\rightarrow\; \sigma=\sqrt{\lambda}\end{cases}$$

حال با توجه به اطلاعات مسئله داریم:

$$\begin{cases}P(X=0)=\dfrac{e^{-\lambda}\,\lambda^0}{0!}=0.018 \;\rightarrow\; e^{-\lambda}=0.018\\ \sigma=\sqrt{\lambda}=2 \;\rightarrow\; \lambda=4\end{cases}$$

بنابراین:

$$\begin{cases}P(X=3)=\dfrac{e^{-\lambda}\,\lambda^3}{3!}=\dfrac{0.018\times4^3}{3!}=\dfrac{18}{1000}\times\dfrac{64}{6}=\dfrac{192}{1000}=0.192\\ \lambda=4\;;\;e^{-\lambda}=0.018\end{cases}$$

۱۲. گزینه ۳ درست است.

برای یک متغیر تصادفی $\alpha<x<\beta$ با توزیع یکنواخت داریم:

$$\begin{cases}f(x)=\dfrac{1}{\beta-\alpha}\\ E(X)=\dfrac{\alpha+\beta}{2}\;,\;\sigma_X^2=\dfrac{(\beta-\alpha)^2}{12}\end{cases}$$

بنابراین، واریانس متغیر $-1<x<5$ با توزیع یکنواخت برابر است با:

$$\sigma_X^2=\frac{(\beta-\alpha)^2}{12}=\frac{(5-(-1))^2}{12}=\frac{36}{12}=3$$

۱۳. گزینه ۱ درست است.

ازآنجاکه جامعه محدود و انتخاب بدون جایگذاری است و داریم $\frac{n}{N}=\frac{16}{97}\simeq0.15>0.05$، برای محاسبه انحراف معیار توزیع میانگین نمونه از ضریب تصحیح استفاده می‌کنیم؛ بنابراین:

$$\begin{cases}\sigma_{\bar{X}}=\sqrt{\dfrac{N-n}{N-1}}\times\dfrac{\sigma_X}{\sqrt{n}}=\sqrt{\dfrac{97-16}{97-1}}\times\dfrac{\sqrt{6}}{\sqrt{16}}=\sqrt{\dfrac{81\times6}{96\times16}}=\sqrt{\dfrac{81}{16\times16}}=\dfrac{9}{16}=0.5625\\ n=16\;;\;N=97\;;\;\sigma_X^2=6\end{cases}$$

۱۴ . گزینه ۴ درست است.

$$\begin{cases} \bar{x} = \dfrac{\sum x}{n} = \dfrac{20}{5} = 4 \\ \bar{y} = \dfrac{\sum y}{n} = \dfrac{15}{5} = 3 \end{cases}$$

x	y	$x-\bar{x}$	$y-\bar{y}$
2	4	−2	1
3	2	−1	−1
4	5	0	2
5	1	1	−2
6	3	2	0

$\longrightarrow$

$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
4	1	−2
1	1	1
0	4	0
1	4	−2
4	0	0
$\sum = 10$	$\sum = 10$	$\sum = -3$

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{-3}{\sqrt{10\times 10}} = -0.3$$

۱۵ . گزینه ۱ درست است.

$F_{o_{ij}}$

	کارمند	کارگر	
راضی	7	18	25
ناراضی	3	22	25
	10	40	N = 50

$\longrightarrow$

$F_{e_{ij}}$

	کارمند	کارگر
راضی	$\frac{25\times 10}{50} = 5$	$\frac{25\times 40}{50} = 20$
ناراضی	$\frac{25\times 10}{50} = 5$	$\frac{25\times 40}{50} = 20$

$$\chi^2 = \frac{\sum\sum \left(F_{o_{ij}} - F_{e_{ij}}\right)^2}{F_{e_{ij}}} = \frac{(7-5)^2}{5} + \frac{(18-20)^2}{20} + \frac{(3-5)^2}{5} + \frac{(22-20)^2}{20} = \frac{4}{5} + \frac{4}{20} + \frac{4}{5} + \frac{4}{20} = \frac{10}{5} = 2$$

علوم اقتصادی

۱. گزینه ۲ درست است.

ازآنجاکه داده‌های مسئله برحسب انحراف از میانگین هستند، داریم:

$$\begin{cases}\sum(x-\bar{x})^2=\sum x_t^2\\ \sum(y-\bar{y})^2=\sum y_t^2\\ \sum(x-\bar{x})(y-\bar{y})=\sum x_t y_t\end{cases}$$

با توجه به رابطه r (ضریب همبستگی نمونه) داریم:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}\rightarrow r^2=\frac{\left[\sum(x-\bar{x})(y-\bar{y})\right]^2}{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}=\frac{\left(\sum x_t y_t\right)^2}{\sum x_t^2\sum y_t^2}$$

که در گزینه‌ها نیست (در گزینه ۳ اشتباهاً به‌جای $\left(\sum x_t y_t\right)^2$ از $\sum x_t^2 y_t^2$ استفاده شده است).

اگر معادله خط رگرسیون به صورت $\hat{y}=\hat{\beta}_1+\hat{\beta}_2 x$ باشد، $\hat{\beta}_2$ شیب خط رگرسیون است و داریم:

$$\hat{\beta}_2=r\frac{S_Y}{S_X}\rightarrow r=\hat{\beta}_2\frac{S_X}{S_Y}\rightarrow r^2=\hat{\beta}_2^2\frac{S_X^2}{S_Y^2}=\hat{\beta}_2^2\frac{\frac{\sum(x-\bar{x})^2}{\cancel{n-1}}}{\frac{\sum(y-\bar{y})^2}{\cancel{n-1}}}=\hat{\beta}_2^2\frac{\sum(x-\bar{x})^2}{\sum(y-\bar{y})^2}=\hat{\beta}_2^2\frac{\sum x_t^2}{\sum y_t^2}$$

که رابطه آخر در گزینه ۲ وجود دارد و پاسخ سؤال است.

در گزینه ۴ نیز اشتباهاً به‌جای $\hat{\beta}^2$ از $\hat{\beta}$ استفاده شده است.

۲. گزینه ۱ درست است.

$$MSE(\hat{\beta})=E(\hat{\beta}-\beta)^2=Var(\hat{\beta})+\left(E(\hat{\beta})-\beta\right)^2=\hat{\beta}\ \text{واریانس}+(\hat{\beta}\ \text{اریبی یا تورش})^2$$

همان‌طورکه دیده می‌شود:

گزینه ۳ نادرست است، زیرا به‌جای «+» از «−» استفاده شده است.

گزینه ۲ نادرست است، زیرا به‌جای $(\hat{\beta}\ \text{تورش})^2$ از $(\hat{\beta}\ \text{تورش})$ استفاده شده است.

۳. گزینه ۱ درست است.

$$\begin{cases}\sigma^2_{(3X+2)}=9\sigma_X^2=9\times15=135\\ \sigma_X^2=15\end{cases}$$

۴. گزینه ۴ درست است.

برای مقایسه میانگین k جامعه $(k\geq 2)$ از آزمون تحلیل واریانس آماره فیشر به صورت زیر استفاده می‌شود:

$$\begin{cases}F_{k-1,k(n-1)}=\dfrac{\frac{SS(Tr)}{k-1}}{\frac{SSE}{k(n-1)}}\rightarrow F_{3,16}=\dfrac{\frac{12}{3}}{\frac{32}{16}}=\dfrac{4}{2}=2\\ \text{مربعات بین گروهی}: SS(Tr)=12\\ \text{مربعات درون گروهی}: SSE=32\\ n=5\ ,\ k=4\end{cases}$$

۵. گزینه ۴ درست است.

$$CV_{X+0.25X}=CV_{1.25X}=\frac{\sigma_{(1.25X)}}{\mu_{(1.25X)}}=\frac{1.25\,\sigma_X}{1.25\,\mu_X}=\frac{\sigma_X}{\mu_X}=CV_X$$

دقت کنید، وقتی ضریب تغییرات $\left(CV_X=\frac{\sigma_X}{\mu_X}\right)$ تغییر نمی‌کند دیگری نیازی به مقادیر $\mu=50$ و $\sigma=20$ نیست.

۶. گزینه ۳ درست است.

$$T=\sum_{i=1}^{n}a_i x_i \xrightarrow{a_1=a_2=\dots=a_n=\frac{1}{n}} T(X)=\frac{\sum_{i=1}^{n}x_i}{n}=\bar{X}\rightarrow \text{کمترین واریانس}$$

۷. گزینه ۴ درست است.

همان‌طورکه می‌دانیم مساحت هر مربع به ضلع a برابر با a^2 است، بنابراین مجموع مساحت سه قطعه A، B و C برابر است با:

$$\text{مجموع مساحت سه قطعه A، B و C}=7^2+5^2+1^2=75$$

حال اگر بخواهیم این مجموع را بین سه قطعه مربع با ضلع یکسان a تقسیم کنیم، باید مجموع آن‌ها $3a^2$ شود؛ پس:

$$\text{مجموع مساحت سه قطعه یکسان}=3a^2=75 \rightarrow a^2=25\rightarrow a=5$$

۸. گزینه ۳ درست است.

$$P(X<\alpha)=0.8 \rightarrow \int_0^{\alpha}5e^{-5x}dx=0.8 \rightarrow \left[-e^{-5x}\right]_0^{\alpha}=0.8 \rightarrow 1-e^{-5\alpha}=0.8$$

$$\rightarrow e^{-5\alpha}=0.2=\frac{1}{5} \xrightarrow{\ln} -5\alpha=\ln\frac{1}{5} \rightarrow -5\alpha=-\ln 5 \rightarrow \alpha=\frac{\ln 5}{5}$$

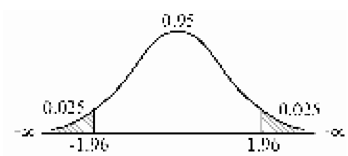
۹. گزینه ۱ درست است.

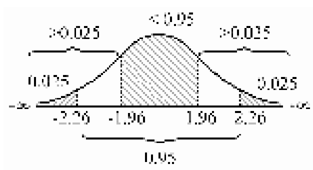
برای هر متغیر تصادفی $\alpha<x<\beta$ با توزیع یکنواخت داریم

$$\begin{cases} f(x)=\frac{1}{\beta-\alpha} \\ E(X)=\frac{\alpha+\beta}{2}\end{cases} \xrightarrow{\alpha=0,\beta=7} \begin{cases} f(x)=\frac{1}{7} \\ E(X)=\frac{7}{2}=3.5\end{cases}$$

۱۰. گزینه ۲ درست است.

ازآنجاکه توزیع t متقارن و از توزیع نرمال کوتاه‌تر است (پراکندگی بیشتر)، با توجه به شکل، مقدار ناحیه بحرانی برای سطوح برابر در آن از نرمال بیشتر است.





۱۱. گزینه ۳ درست است.

$$x_1 = \frac{\text{سال اول}}{\text{سال قبل}} = 2$$
$$x_2 = \frac{\text{سال دوم}}{\text{سال اول}} = 3 \quad \Rightarrow \quad \bar{x}_G = \sqrt[3]{2 \times 3 \times 4} = \sqrt[3]{24} = 2\sqrt[3]{3}$$
$$x_3 = \frac{\text{سال سوم}}{\text{سال دوم}} = 4$$

۱۲. گزینه ۱ درست است.

کل حالات: اگر کتاب‌های ریاضی را با R_1, R_2, R_3 و کتاب‌های اقتصاد را با E_1, E_2, E_3, E_4 نمایش دهیم، آن‌گاه:

$7! =$ حالات جایگشت کل کتاب‌ها

حالات مساعد: حال اگر بخواهیم کتاب‌های ریاضی کنار هم و کتاب‌های اقتصاد نیز کنار هم باشند، داریم:

$$\overset{R}{\boxed{R_1\ R_2\ R_3}} \quad \overset{E}{\boxed{E_1\ E_2\ E_3\ E_4}}$$

$2! \times 3! \times 4! =$ حالات مساعد

بنابراین احتمال مورد نظر برابر است با:

$$\text{احتمال} = \frac{\text{حالات مساعد}}{\text{حالات ممکن}} = \frac{2! \times 3! \times 4!}{7!} = \frac{2! \times \cancel{3!} \times \cancel{4!}}{7 \times \cancel{6} \times 5 \times \cancel{4!}} = \frac{2}{35}$$

۱۳. گزینه ۴ درست است.

کمترین مقدار $\sum_{i=1}^{2}(x_i - a)^2$ زمانی است که $a = \mu = \frac{x_1 + x_2}{2}$ باشد، بنابراین:

$$\min = \sum_{i=1}^{2}\left(x_i - \frac{x_1+x_2}{2}\right)^2 = \left(x_1 - \frac{x_1+x_2}{2}\right)^2 + \left(x_2 - \frac{x_1+x_2}{2}\right)^2 = \frac{(x_1-x_2)^2}{4} + \frac{(x_2-x_1)^2}{4} = 2\frac{(x_1-x_2)^2}{4} = \frac{(x_1-x_2)^2}{2}$$

یادآوری: همواره $(a-b)^2 = (b-a)^2$ است.

۱۴. گزینه ۱ درست است.

$$\begin{cases} \chi^2 = \dfrac{(n-1)S^2}{\sigma^2} \rightarrow \sigma^2 = \left[\dfrac{(n-1)S^2}{\chi^2_{0.025}}, \dfrac{(n-1)S^2}{\chi^2_{0.975}}\right] = \left[\dfrac{99 \times 10}{128.422}, \dfrac{99 \times 10}{73.361}\right] \\ n = 100 \ ; \ S^2 = 10 \end{cases}$$

یادآوری: ازآنجاکه در یک بازه همواره حد پایین کوچک‌تر است، عدد بزرگ‌تر در مخرج کسر حد پایین گذاشته می‌شود تا مقدار آن کوچک‌تر شود.

۱۵. گزینه ۳ درست است.

همان‌طورکه می‌دانیم، توان آزمون عبارت است از:

$P(H_0 \text{ نادرست} \mid H_0 \text{ رد}) =$ احتمال رد کردن فرض H_0 وقتی نادرست باشد

حال اگر اختلاف میان مقدار بیان شده در فرض صفر (H_0) و فرض مقابل (H_1) زیاد شود، در صورت نادرست بودن H_0، آن را با احتمال بیشتری رد می‌کنیم. برای مثال، در فرض‌های زیر:

$$(1)\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu = 12 \end{cases} \rightarrow \Delta\mu = 4 \qquad (2)\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu = 8.5 \end{cases} \rightarrow \Delta\mu = 0.5$$

فرض H_0 در حالت (1) با اطمینان بیشتری رد می‌شود؛ بنابراین، توان آزمون بیشتر خواهد بود.

۱۶. گزینه ۲ درست است.

ابتدا واریانس هر یک از تخمین‌زننده‌ها را محاسبه می‌کنیم:

$$\begin{cases} Var(\bar{X}) = \dfrac{\sigma^2}{n} \\ Var(\tilde{X}) = \dfrac{\cancelto{\sigma^2}{Var(x_{min})} + \cancelto{\sigma^2}{Var(x_{max})}}{4} = \dfrac{2\sigma^2}{4} = \dfrac{\sigma^2}{2} \end{cases}$$

حال نسبت کارایی را به دست می‌آوریم:

$$\text{کارایی } \tilde{X} \text{ به } \bar{X} = \frac{Var(\bar{X})}{Var(\tilde{X})} = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{2}} = \frac{2}{n}$$

۱۷. گزینه ۱ درست است.

اگر X دارای توزیع نرمال با میانگین مجهول μ و واریانس مجهول σ^2 باشد، در این صورت برآورد MLE این پارامترها مانند روش گشتاوری عبارت است از:

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = \overline{X^2} - \bar{X}^2 \end{cases}$$

۱۸. گزینه ۳ درست است.

در توزیع پواسون با پارامتر λ (متوسط تعداد اتفاقات) داریم:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad ; \quad x = 0,1,2,\ldots$$

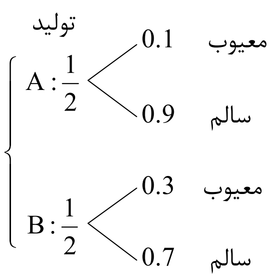
ازآنجاکه $P(X=0) = \frac{1}{2}P(X=2)$، داریم:

$$\frac{e^{-\lambda}\lambda^0}{0!} = \frac{1}{2} \times \frac{e^{-\lambda}\lambda^2}{2!} \rightarrow \lambda^2 = 4 \rightarrow \lambda = \pm 2 \xrightarrow{\lambda \ge 0} \lambda = 2$$

بنابراین $P(X>0)$ برابر است با:

$$\begin{cases} P(X>0) = 1 - P(X=0) = 1 - \dfrac{e^{-2}2^0}{0!} = 1 - e^{-2} \\ \lambda = 2 \end{cases}$$

۱۹. گزینه ۲ درست است.



بنا بر قضیه بیز داریم:

$$P(B \text{ خط} \mid \text{کالا سالم}) = \frac{P(B \text{ خط و کالا سالم})}{P(\text{کالا سالم})} = \frac{\frac{1}{2}\times 0.7}{\frac{1}{2}\times 0.9+\frac{1}{2}\times 0.7} = \frac{7}{16}$$

۲۰ . گزینه ۴ درست است.

اگر متغیر X دارای توزیع کای- دو با n درجه آزادی باشد، داریم:

$$\begin{cases} E(X)=\mu_X=n \\ \sigma_X^2=2n \\ \sigma_X=\sqrt{2n} \end{cases} \xrightarrow{n=15} \begin{cases} E(X)=15 \\ \sigma_X^2=30 \\ \sigma_X=\sqrt{30} \end{cases}$$

حال می‌توانیم ضریب تغییرات X را محاسبه کنیم:

$$CV_X=\frac{\sigma_X}{\mu_X}=\frac{\sqrt{30}}{15}=\frac{\sqrt{2}\times\sqrt{15}}{15}$$

احتمالاً طراح سؤال هنگام ساده کردن صورت و مخرج کسر آخر، به اشتباه $\sqrt{15}$ را با 15 ساده کرده است که جواب $\sqrt{2}$ یعنی گزینه ۴ در کلید سنجش انتخاب شده است.

برنامه‌ریزی شهری و منطقه‌ای

۱. گزینه ۲ درست است.

مقیاس فاصله‌ای دارای صفر قراردادی و مقیاس نسبی دارای صفر واقعی است.

۲. گزینه ۴ درست است.

$$\text{میانگین حسابی} : \bar{X} = \frac{\sum F_i X_i}{N} = \frac{2\times 9+3\times 12+2\times 16}{7} = \frac{86}{7} = 12.28$$

$$\text{میانگین هندسی} : \bar{X}_G = \sqrt[N]{X_1^{F_1}\times X_2^{F_2}\times .....} = \sqrt[7]{9^2\times 12^3\times 16^2} = \sqrt[7]{\left(3^2\right)^2\times(4\times 3)^3\times\left(4^2\right)^2}$$

$$= \sqrt[7]{3^4\times 4^3\times 3^3\times 4^4} = \sqrt[7]{3^7\times 4^7} = 12$$

$$\text{اختلاف میانگین‌ها} = \bar{X} - \bar{X}_G = 12.28 - 12 = 0.28$$

این سؤال دقیقاً در آزمون سراسری سال ۸۶ رشته‌های حسابداری و مدیریت نیز مطرح شده بود.

۳. گزینه ۳ درست است.

انحراف چارکی از رابطه $SIQR = \frac{Q_3 - Q_1}{2}$ محاسبه می‌شود؛ بنابراین، ابتدا باید چارک اول و سوم را محاسبه کنیم.

برای محاسبه چارک اول و سوم، F_{c_i} را در جدول به دست می‌آوریم:

$C-L$	<10	$10-12$	$12-14$	$14-16$	$16-18$	≥ 18	
F_i	5	10	7	18	8	4	$N=52$
F_{c_i}	5	15	22	40	48	52	

ـ چارک اول (Q_1) در اولین طبقه‌ای است که در آن $F_{c_i} \geq \frac{1N}{4} = 13$ باشد (طبقه $10-12$)؛ بنابراین:

$$Q_1 = 10 + \frac{13-5}{10}\times 2 = 11.6$$

ـ چارک سوم (Q_3) در اولین طبقه‌ای است که در آن $F_{c_i} \geq \frac{3N}{4} = 39$ باشد (طبقه $14-16$)؛ بنابراین:

$$Q_3 = 14 + \frac{39-22}{18}\times 2 = 15.88$$

و در نهایت داریم:

$$\text{انحراف چارکی} : SIQR = \frac{Q_3 - Q_1}{2} = \frac{15.88 - 11.6}{2} = 2.14$$

۴. گزینه ۱ درست است.

$$\begin{cases} CV = \frac{\sigma}{\mu} = \frac{1.6}{4.2} = \frac{16}{42} = 0.38 \\ \mu = \frac{\sum x_i}{N} = \frac{168}{40} = \frac{42}{10} \\ \sigma^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2 = \frac{808}{40} = \frac{202}{10} - \left(\frac{42}{10}\right)^2 = 2.56 \quad \rightarrow \quad \sigma = \sqrt{2.56} = 1.6 \end{cases}$$

۵ . گزینه ۲ درست است.

ابتدا با توجه به آنکه X (نمرات دانشجویان) دارای توزیع نرمال با $\mu = 63.6$ و نمره 20 درصد آن‌ها حداقل 72 است آن را به حالت استاندارد (Z) تبدیل می‌کنیم:

$$P(X \geq 72) = P\left(\frac{X-\mu}{\sigma} \geq \frac{72-63.6}{\sigma}\right) = P\left(Z \geq \frac{8.4}{\sigma}\right) = 0.2 \qquad (I)$$

حال ازآنجاکه $P(Z \leq 0.84) = 0.8$ است، داریم:

$$P(Z \leq 0.84) = 0.8 \quad \rightarrow \quad P(Z \geq 0.84) = 0.2 \qquad (II)$$

بنابراین:

$$(I),(II) \quad \rightarrow \quad \frac{8.4}{\sigma} = 0.84 \quad \rightarrow \quad \sigma = 10$$

۶ . گزینه ۲ درست است.

کلمه STANDARD شامل حروف S، T، A، N، D و R است که حرف A، 2 بار و حرف D نیز 2 بار در آن تکرار شده است.

برای ساختن رمزهای چهارحرفی باید حالات زیر را بررسی کنیم:

ـ کلمات چهارحرفی با 6 حرف غیرتکراری S، T، A، N، D و R:

$$\binom{6}{4} \times 4! = 15 \times 4! = 360$$

ـ کلمات چهارحرفی با 2 حرف تکراری D و 2 حرف دیگر از S، T، A، N و R:

$$\binom{5}{2} \times \frac{4!}{2!} = 10 \times 12 = 120$$

ـ کلمات چهارحرفی با 2 حرف تکراری A و 2 حرف دیگر از S، T، D، N و R:

$$\binom{5}{2} \times \frac{4!}{2!} = 10 \times 12 = 120$$

ـ کلمات چهارحرفی با 2 حرف تکراری D و 2 حرف تکراری A:

$$\frac{4!}{2!2!} = 6$$

بنابراین تعداد کل رمزهای چهارحرفی برابر است با:

$$360 + 120 + 120 + 6 = 606$$

۷ . گزینه ۱ درست است.

X, Y مستقل $\longrightarrow$ $\boxed{Cov(X,Y)=0}$

X, Y وابسته $\longrightarrow$ غیرخطی $\nearrow$ ، خطی $\searrow$ $\boxed{Cov(X,Y) \neq 0}$

۸ . گزینه ۳ درست است.

A : تاریخ ندارد $\quad \rightarrow \quad$ A′ : تاریخ دارد

B : برچسب ندارد $\quad \rightarrow \quad$ B′ : برچسب دارد

با توجه به اطلاعات مسئله داریم:

$$\begin{cases} P(A) = 0.45 \\ P(B) = 0.30 \\ P(A \cap B) = 0.25 \end{cases}$$

بنابراین درصد کالاهایی که هم تاریخ و هم برچسب دارند، برابر است با:

$$P(A'\cap B')=P(A\cup B)'=1-P(A\cup B)=1-\underbrace{(0.45+0.30-0.25)}_{0.5}=0.5$$

ازآنجاکه 0.04 (4 درصد) کل کالاها معیوب است پس 0.96 (96 درصد) کل کالاها از جمله کالاهایی که تاریخ و برچسب دارند نیز سالم است؛ بنابراین، احتمال آنکه کالای انتخاب‌شده تاریخ مصرف داشته باشد و سالم باشد برابر است با:

$$0.96\times 0.5=0.48$$

۹ . گزینه ۳ درست است.

با توجه به داده‌های مسئله، برای محاسبه ضریب چولگی از رابطه ضریب چولگی چندکی استفاده می‌کنیم:

$$\begin{cases} SK_Q=\dfrac{Q_3-2Q_2+Q_1}{Q_3-Q_1}=\dfrac{134.4-2\times121+112.4}{134.4-112.4}=\dfrac{4.8}{22}=0.218 \\ Q_1=112.4 \quad , \quad Q_2=Md=121 \quad , \quad Q_3=134.4 \end{cases}$$

۱۰ . گزینه ۴ درست است.

زمانی‌که توزیع آماری از چولگی معقولی برخوردار است روابط چولگی پیرسون با هم برابرند؛ یعنی:

$$Sk_1=Sk_2 \quad\rightarrow\quad \frac{\mu-Mo}{\sigma}=\frac{3(\mu-Md)}{\sigma} \quad\rightarrow\quad \mu-Mo=3(\mu-Md)$$

حال با توجه به آنکه $\mu=36$ و $Md=42$ است، داریم:

$$36-Mo=3(36-42) \quad\rightarrow\quad Mo=54$$

۱۱ . گزینه ۴ درست است.

تعداد حالات توزیع n شیء در k سلول به طوری که n_1 شیء در سلول اول و... و n_k شیء در سلول kام باشد، برابر است با:

$$\frac{n!}{n_1!n_2!\dots n_k!} \xrightarrow[n=7]{n_1=2\ ,\ n_2=2\ ,\ n_3=3} \text{تعداد حالات}=\frac{7!}{2!2!3!}=210$$

۱۲ . گزینه ۲ درست است.

راه حل اول: با توجه به تابع احتمال توزیع فوق هندسی داریم:

$$P(x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}=\frac{\binom{4}{x}\binom{8}{3-x}}{a} \quad\rightarrow\quad a=\binom{N}{n}=\binom{12}{3}=220$$

راه حل دوم: با توجه به تعریف تابع احتمال گسسته داریم:

$$\sum_{x=0}^{3}P(x)=1 \quad\rightarrow\quad \frac{\overbrace{\binom{4}{0}\binom{8}{3}+\binom{4}{1}\binom{8}{2}+\binom{4}{2}\binom{8}{1}+\binom{4}{3}\binom{8}{0}}^{\binom{12}{3}}}{a}=1 \quad\rightarrow\quad a=220$$

یادآوری:

$$\binom{N}{n}=\binom{k}{0}\binom{N-k}{n}+\binom{k}{1}\binom{N-k}{n-1}+\dots+\binom{k}{n}\binom{N-k}{0}$$

۱۳. گزینه ۱ درست است.

$$\begin{cases} V(-2X+5)=4V(X)=4\times 2.25=9 \\ V(X)=E(X^2)-(E(X))^2=14.5-(3.5)^2=2.25 \end{cases}$$

۱۴. گزینه ۱ درست است.

درصورتی‌که آزمایش برنولی با احتمال موفقیت $P=0.6$ را $n=96$ بار تکرار کنیم، توزیع حاصل، دوجمله‌ای است و داریم:

$$\sigma=\sqrt{npq}=\sqrt{96\times\frac{6}{10}\times\frac{4}{10}}=\sqrt{16\times 6\times\frac{6}{10}\times\frac{4}{10}}=\sqrt{\frac{64\times 36}{100}}=\sqrt{\frac{8^2\times 6^2}{10^2}}=\frac{8\times 6}{10}=4.8$$

۱۵. گزینه ۲ درست است.

وقتی جامعه وسیع باشد از نمونه‌گیری خوشه‌ای استفاده می‌کنیم.

۱۶. گزینه ۴ درست است.

با توجه به تعریف تابع احتمال داریم:

$$\int_{-\infty}^{+\infty} f(x)\,dx=1 \;\rightarrow\; \int_0^4 k\left(16x-x^3\right)dx=1 \;\rightarrow\; k\left[\frac{16x^2}{2}-\frac{x^4}{4}\right]_0^4=1 \;\rightarrow\; k[128-64]=1 \;\rightarrow\; k=\frac{1}{64}$$

۱۷. گزینه ۱ درست است.

انحراف معیار میانگین نمونه $(\sigma_{\bar{X}})$ از تقسیم انحراف معیار جامعه (σ_X) بر جذر تعداد نمونه $(\sqrt{n})$ به دست می‌آید؛ بنابراین:

$$\begin{cases} \sigma_{\bar{X}}=\dfrac{\sigma_X}{\sqrt{n}}=\dfrac{12}{\sqrt{64}}=\dfrac{12}{8}=1.5 \\ \sigma_X=12\;;\;n=64 \end{cases}$$

دقت کنید، طراح سؤال اشتباهاً به‌جای عبارت «انحراف معیار 12» از «واریانس 12» استفاده کرده است، زیرا اگر $\sigma_X^2=12$ را در نظر بگیریم، $\sigma_X=\sqrt{12}$ خواهد بود و خواهیم داشت:

$$\sigma_{\bar{X}}=\frac{\sigma_X}{\sqrt{n}}=\frac{\sqrt{12}}{\sqrt{64}}=\frac{2\sqrt{3}}{8}=\frac{\sqrt{3}}{4}$$

که در گزینه‌ها نیست.

۱۸. گزینه ۱ درست است.

با توجه به ادعای مطرح‌شده از طرف شهرداری درباره درصد رضایت مردم، باید از آزمون زیر استفاده کنیم:

$$p_0=0.8 \;\rightarrow\; \begin{cases} H_0: P=0.8 \\ H_1: P\neq 0.8 \end{cases}$$

$$\begin{cases} \text{آماره آزمون}: Z=\dfrac{\bar{p}-p_0}{\sqrt{\dfrac{p_0q_0}{n}}}=\dfrac{0.75-0.8}{\sqrt{\dfrac{0.8\times 0.2}{400}}}=\dfrac{-0.05}{\dfrac{0.4}{20}}=\dfrac{-1}{0.4}=-2.5 \\ \bar{p}=0.75\;,\;n=400 \end{cases}$$

۱۹ . گزینه ۴ درست است.

$$\begin{cases} \bar{x} = \dfrac{\sum x}{n} = \dfrac{5+9+11+7}{4} = \dfrac{32}{4} = 8 \\ \bar{y} = \dfrac{\sum y}{n} = \dfrac{8+12+10+6}{4} = \dfrac{36}{4} = 9 \end{cases}$$

x	y	$x-\bar{x}$	$y-\bar{y}$
5	8	−3	−1
9	12	1	3
11	10	3	1
7	6	−1	−3

$\longrightarrow$

$(x-\bar{x})^2$	$(y-\bar{y})^2$	$\sum(x-\bar{x})(y-\bar{y})$
9	1	3
1	9	3
9	1	3
1	9	3
$\sum = 20$	$\sum = 20$	$\sum = 12$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} = \frac{12}{\sqrt{20 \times 20}} = \frac{12}{20} = \frac{6}{10} = 0.6$$

سنجش از دور و سیستم اطلاعات جغرافیایی (GIS)

۱. گزینه ۱ درست است.

می‌دانیم در هر جامعه متقارن $\mu = Md = Mo$ است که در آن μ «میانگین حسابی» است. حال اگر به یک اندازه (مثلاً 25%) از ابتدا و انتهای مشاهدات حذف کنیم، باز هم جامعه متقارن و $\mu = Md = Mo$ است. اما این بار μ «میانگین پیراسته» است؛ پس در این سؤال ازآنجاکه میانگین پیراسته 72 است، میانگین حسابی نیز 72 خواهد بود.

۲. گزینه ۴ درست است.

اگر مشاهدات به صورت x_i $(i=1,2,\ldots,45)$ باشند، مشاهدات جدید به صورت $\frac{3}{2}x_i + 4.5$ هستند، بنابراین:

$$\begin{cases} CV_{\left(\frac{3}{2}X+4.5\right)} = \dfrac{\sigma_{\left(\frac{3}{2}X+4.5\right)}}{\mu_{\left(\frac{3}{2}X+4.5\right)}} = \dfrac{\frac{3}{2}\sigma_X}{\frac{3}{2}\mu_X + 4.5} = \dfrac{\frac{3}{2}\times\frac{12}{5}}{\frac{3}{2}\times\frac{9}{5}+4.5} = \dfrac{3.6}{7.2} = 0.5 \\ \mu_X = \dfrac{\sum x_i}{N} = \dfrac{81}{45} = \dfrac{9}{5} \\ \sigma_X^2 = \dfrac{\sum x_i^2}{N} - \left(\dfrac{\sum x_i}{N}\right)^2 = \dfrac{405}{45} - \left(\dfrac{9}{5}\right)^2 = \dfrac{144}{25} \rightarrow \sigma_X = \dfrac{12}{5} \end{cases}$$

۳. گزینه ۲ درست است.

ازآنجاکه درصد داده‌های کمتر از 36.5 مورد نظر است؛ 36.5 یک چندک است (میانه، چارک، دهک یا صدک) و چون صدک شامل تمام چندک‌هاست، 36.5 را صدک aام فرض می‌کنیم. درواقع، a درصد مشاهدات کمتر یا مساوی 36.5 هستند. برای به دست آوردن مقدار a، ابتدا F_{c_i} را در جدول محاسبه می‌کنیم:

$C-L$	$20-25$	$25-30$	$30-35$	$35-40$	$40-45$	$45-50$	
F_i	12	17	19	20	15	7	$N=\sum F_i = 90$
F_{c_i}	12	29	48	68	83	90	

با توجه به آنکه مقدار صدک برابر با 36.5 است، طبقه صدک‌دار طبقه $(35-40)$ خواهد بود؛ بنابراین:

$$P_a = L_i + \dfrac{\frac{aN}{100} - F_{c_{i-1}}}{F_i} \times I \rightarrow 36.5 = 35 + \dfrac{\frac{a\times 90}{100} - 48}{20} \times 5 \rightarrow 1.5 = \dfrac{0.9a - 48}{4} \rightarrow a = 60$$

۴. گزینه ۱ درست است.

$$\begin{cases} E = \dfrac{\frac{\sum (x_i - \bar{x})^4}{N}}{\sigma^4} - 3 = \dfrac{\frac{6075}{25}}{3^4} - 3 = \dfrac{243}{81} - 3 = 0 \\ \sum (x_i - \bar{X})^4 = 6075 \quad , \quad \sigma = 3 \ , \ N = 25 \end{cases}$$

۵ . گزینه ۴ درست است.

کلمه SKEWNESS شامل حروف S، K، E، W و N است که حرف S، 3 بار و حرف E، 2 بار در آن تکرار شده است.

برای ساختن رمزهای چهارحرفی باید حالات زیر را بررسی کنیم:

ـ رمزهای چهارحرفی با پنج حرف متمایز S، K، E، W و N:

$$\binom{5}{4}\times 4! = 120$$

ـ رمزهای چهارحرفی با 2 حرف E و 2 حرف دیگر از 4 حرف متمایز S، K، W و N:

$$\binom{4}{2}\times \frac{4!}{2!} = 72$$

ـ رمزهای چهارحرفی با 2 حرف S و 2 حرف دیگر از 4 حرف متمایز E، K، W و N:

$$\binom{4}{2}\times \frac{4!}{2!} = 72$$

ـ رمزهای چهارحرفی با 3 حرف S و 1 حرف دیگر از 4 حرف متمایز E، K، W و N:

$$\binom{4}{1}\times \frac{4!}{3!} = 16$$

ـ رمزهای چهارحرفی با 2 حرف S و 2 حرف E:

$$\frac{4!}{2!2!} = 6$$

بنابراین تعداد کل رمزهای چهارحرفی برابر است با:

$$120+72+72+16+6=286$$

۶ . گزینه ۲ درست است.

ازآنجاکه شرط پذیرش، قبولی در حداقل 3 آزمایش است. سپس باید در 4 آزمایش یا اصلاً شکست نخورد یا فقط 1 شکست بخورد.

	اول	دوم	سوم	چهارم	
1 شکست	q	p	p	p	$\rightarrow \frac{1}{2}\times\frac{1}{4}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{32}$
	p	q	p	p	$\rightarrow \frac{1}{2}\times\frac{1}{2}\times\frac{1}{4}\times\frac{1}{2}=\frac{1}{32}$
	p	p	q	p	$\rightarrow \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{4}=\frac{1}{32}$
	p	p	p	q	$\rightarrow \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{16}$
0 شکست	p	p	p	p	$\rightarrow \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{16}$

بنابراین احتمال پذیرش (حداقل 3 موفقیت یا حداکثر 1 شکست) برابر است با:

$$\frac{1}{32}+\frac{1}{32}+\frac{1}{32}+\frac{1}{16}+\frac{1}{16}=\frac{7}{32}$$

۷ . گزینه ۴ درست است.

با توجه به تعریف تابع احتمال داریم:

$$\sum f(x)=1 \;\rightarrow\; 0.2+0.4+0.15+\alpha=1 \;\rightarrow\; \alpha=0.25$$

بنابراین:

x	3	5	6	8
f(x)	0.2	0.4	0.15	0.25

$$\begin{cases} V(-2X+3)=4V(X)=4\times 2.95=11.8 \\ V(X)=E\left(X^2\right)-\left(E(X)\right)^2=33.2-(5.5)^2=33.2-30.25=2.95 \\ E\left(X^2\right)=\sum x^2 f(x)=\underbrace{3^2\times 0.2}_{1.8}+\underbrace{5^2\times 0.4}_{10}+\underbrace{6^2\times 0.15}_{5.4}+\underbrace{8^2\times 0.25}_{16} \\ E(X)=\sum x\,f(x)=\underbrace{3\times 0.2}_{0.6}+\underbrace{5\times 0.4}_{2}+\underbrace{6\times 0.15}_{0.9}+\underbrace{8\times 0.25}_{2} \end{cases}$$

۸ . گزینه ۴ درست است.

تعداد رأی‌ها دارای توزیع چندجمله‌ای است؛ بنابراین:

$$P(x_1=3\,,\,x_2=1\,,\,x_3=2)=\frac{6!}{3!\,1!\,2!}(0.5)^3(0.3)^1(0.2)^2=60\times\frac{1}{8}\times\frac{3}{10}\times\frac{4}{100}=\frac{9}{100}$$

۹ . گزینه ۳ درست است.

$$P(\text{هر دو قرمز} \mid \text{هر دو همرنگ})=\frac{P(\text{هردو همرنگ قرمز})}{P(\text{هر دو همرنگ})}=\frac{P(\text{هردو همرنگ قرمز})}{P(\text{هر دو همرنگ قرمز})+P(\text{هر دو همرنگ آبی})}$$

$$=\frac{\frac{6}{15}\times\frac{5}{16}}{\frac{6}{15}\times\frac{5}{16}+\frac{9}{15}\times\frac{8}{16}}=\frac{30}{30+72}=\frac{30}{102}=\frac{5}{17}$$

یادآوری: اگر هر دو بار مهره قرمز خارج کنیم:

6 قرمز ، 9 آبی		5 قرمز ، 11 آبی
$\frac{6}{15}=$ بار اول قرمز	$\xrightarrow[\text{2 آبی اضافه}]{\text{1 قرمز حذف}}$	$\frac{5}{16}=$ بار دوم قرمز

اگر هر دو بار مهره آبی خارج کنیم:

6 قرمز ، 9 آبی		8 قرمز ، 8 آبی
$\frac{9}{15}=$ بار اول آبی	$\xrightarrow[\text{2 قرمز اضافه}]{\text{1 آبی حذف}}$	$\frac{8}{16}=$ بار دوم آبی

۱۰ . گزینه ۲ درست است.

اگر آزمایش برنولی با احتمال موفقیت $p=0.04$ را $n=150$ بار تکرار کنیم، توزیع حاصل (تعداد پیام نارسا) دارای توزیع دوجمله‌ای است که در آن:

$$\sigma=\sqrt{npq}=\sqrt{150\times\frac{4}{100}\times\frac{96}{100}}=\sqrt{6\times\frac{16\times 6}{100}}=\frac{6\times 4}{10}=2.4$$

۱۱ . گزینه ۳ درست است.

معیوب بودن هر کالا دارای توزیع برنولی با احتمال $p=\frac{125}{100000}$ است. حال اگر نمونه‌ای با $n=4000$ کالا انتخاب کنیم، توزیع تعداد معیوب در نمونه دوجمله‌ای است، اما شرایط تقریب به توزیع پواسون برای آن فراهم شده است؛ زیرا:

$$n=4000\ge 100\ ,\ \ np=4000\times\frac{125}{100,000}=5\le 10$$

بنابراین، برای حل سؤال از توزیع پواسون با پارامتر $\lambda=np=5$ استفاده می‌کنیم:

$$P(X\le 2)=P(X=0)+P(X=1)+P(X=2)=\frac{e^{-5}5^0}{0!}+\frac{e^{-5}5^1}{1!}+\frac{e^{-5}5^2}{2!}$$
$$=e^{-5}+5e^{-5}+12.5e^{-5}=18.5e^{-5}=18.5\times 0.007=0.129$$

۱۲ . گزینه ۱ درست است.

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^8 x\times\frac{1}{6x^{\frac{1}{3}}}dx=\int_0^8\frac{x^{\frac{2}{3}}}{6}dx=\frac{1}{6}\left[\frac{3}{5}x^{\frac{5}{3}}\right]_0^8=\frac{3}{30}8^{\frac{5}{3}}=\frac{1}{10}\left(2^3\right)^{\frac{5}{3}}=\frac{2^5}{10}=3.2$$

۱۳ . گزینه ۲ درست است.

ازآنجاکه واریانس جامعه $\left(\sigma^2\right)$ نامعلوم و $n=100>30$ است، حتماً جامعه نرمال بوده و مورد (۲ - الف) برای تخمین میانگین جامعه برقرار می‌شود و داریم :

$$Z=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\rightarrow\mu\in\bar{x}\pm Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}=35.5\pm 1.75\times\frac{8}{\sqrt{100}}=35.5\pm 1.4\ \ \rightarrow\ \ (34.1,36.9)$$

۱۴ . گزینه ۳ درست است.

$$\begin{cases}\bar{x}=\frac{\sum x}{n}=\frac{50}{5}=10\\ \bar{y}=\frac{\sum y}{n}=\frac{30}{5}=6\end{cases}$$

x	y	$x-\bar{x}$	$y-\bar{y}$
5	3	−5	−3
7	4	−3	−2
11	6	1	0
12	8	2	2
15	9	5	3

$\longrightarrow$

$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
25	15
9	6
1	0
4	4
25	15
$\Sigma=64$	$\Sigma=40$

$$b=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}=\frac{40}{64}=\frac{5}{8}=0.625$$

۱۵ . گزینه ۱ درست است.

درصورتی‌که تعداد طبقات جدول $k=2$ باشد، درجه آزادی برابر با $k-1=1$ می‌شود و بهتر است از تصحیح یتس برای آزمون استفاده کنیم.

محیط زیست

۱. گزینه ۱ درست است.

در آمار توصیفی با سرشماری از تمام مشاهدات جامعه استفاده می‌شود.

۲. گزینه ۴ درست است.

میانگین هندسی : $\bar{x}_G = \sqrt[N]{x_1 \times x_2 \times \cdots \times x_N} = \sqrt[5]{12 \times 54 \times 6.75 \times 18 \times 24}$

برای محاسبه مقدار میانگین هندسی کافی است تمام اعداد زیر رادیکال را تجزیه کنیم تا بتوانند توانی از 5 یا مضرب 5 را بگیرند:

$$12 \times 54 \times \overset{\frac{27}{4}}{\boxed{6.75}} \times 18 \times 24 = \left(3 \times 2^2\right) \times \left(2 \times 3^3\right) \times \left(\frac{3^3}{2^2}\right) \times \left(2 \times 3^2\right) \times \left(2^3 \times 3\right) = 2^5 \times 3^{10}$$

بنابراین:

$$\bar{x}_G = \sqrt[5]{2^5 \times 3^{10}} = 2 \times 3^2 = 18$$

۳. گزینه ۲ درست است.

برای محاسبه میانگین حسابی در جدول فراوانی از رابطه زیر استفاده می‌کنیم:

$$\mu = \frac{\sum F_i x_i}{N}$$

که در آن x_i مرکز طبقه، F_i فراوانی مطلق طبقه و N تعداد مشاهدات است.

اولاً، ازآنجاکه در جدول فراوانی به جای فراوانی مطلق $\left(F_i\right)$، فراوانی تجمعی $\left(F_{c_i}\right)$ داده شده است. ابتدا فراوانی مطلق طبقات را از رابطه زیر به دست می‌آوریم:

فراوانی تجمعی طبقه ماقبل − فراوانی تجمعی طبقه = فراوانی مطلق طبقه

و با توجه به آنکه فراوانی تجمعی طبقه آخر همان تعداد مشاهدات است، $N = 90$ خواهد بود؛ بنابراین داریم:

$C-L$	$215-225$	$225-235$	$235-245$	$245-255$	$255-265$	
F_{c_i}	11	28	51	72	90	
F_i	11	17	23	21	18	$N = 90$

ثانیاً، ازآنجاکه حدود دسته‌های جدول اعداد بزرگی هستند، برای محاسبه میانگین به صورت زیر عمل می‌کنیم:

الف) طبقه $(232-248)$ دارای بیشترین فراوانی است؛ بنابراین جدول به صورت زیر خواهد بود:

x'	-2	-1	$\boxed{0}$	1	2
$C-L$	$215-225$	$225-235$	$\boxed{235-245}$	$245-255$	$255-265$
F_i	11	17	$\boxed{23}$	21	18

ب) میانگین را با مرکز طبقات جدید محاسبه می‌کنیم:

x'	-2	-1	0	1	2	
F_i	11	17	23	21	18	$N = \sum F_i = 90$
$F_i x'_i$	-22	-17	0	21	36	$\sum F_i x'_i = 18$

$$\longrightarrow \quad \mu_{X'} = \frac{\sum F_i x'_i}{N} = \frac{18}{90} = \frac{2}{10}$$

ج) میانگین داده‌های اصلی (μ_X) برابر است با:

$$\begin{cases} \mu_X = \mu_{X'} \times I + a \quad \to \quad \mu_X = \dfrac{2}{10} \times 10 + 240 = 242 \\ a = \dfrac{L_k + U_k}{2} = \dfrac{235+245}{2} = 240 \\ \mu_{X'} = \dfrac{2}{10} \;,\; I = 10 \end{cases}$$

۴ . گزینه ۲ درست است.

انحراف چارکی از رابطه $SIQR = \dfrac{Q_3 - Q_1}{2}$ محاسبه می‌شود؛ بنابراین، ابتدا باید چارک اول و سوم را محاسبه کنیم.

برای محاسبه چارک اول و سوم، F_{c_i} را در جدول به دست می‌آوریم:

$C-L$	<8	$8-12$	$12-16$	$16-20$	$20-24$	≥ 24	
F_i	4	12	16	18	9	5	$N=64$
F_{c_i}	4	16	32	50	59	64	

ـ چارک اول (Q_1) در اولین طبقه‌ای است که در آن $\dfrac{1N}{4} = 16$ $F_{c_i} \geq$ باشد (طبقه $8-12$)؛ بنابراین:

$$Q_1 = 8 + \frac{16-4}{12} \times 4 = 12$$

ـ چارک سوم (Q_3) در اولین طبقه‌ای است که در آن $\dfrac{3N}{4} = 48$ $F_{c_i} \geq$ باشد (طبقه $16-20$)؛ بنابراین:

$$Q_3 = 16 + \frac{48-32}{18} \times 4 = 19.55$$

و در نهایت داریم:

$$\text{انحراف چارکی}: ISQR = \frac{Q_3 - Q_1}{2} = \frac{19.55-12}{2} = 3.77$$

۵ . گزینه ۳ درست است.

ابتدا به اطلاعات مسئله و رابطه واریانس کل (ترکیبی) دقت می‌کنیم:

	اول	دوم	سوم
n_i	90	110	200
σ_i^2	7	10	9
μ_i	$(\mu_1 \neq \mu_2 \neq \mu_3) \neq \mu$		

$$\sigma^2 = \frac{\sum n_i \sigma_i^2}{\sum n_i} + \overbrace{\boxed{\frac{\sum n_i (\mu_i - \mu)^2}{\sum n_i}}}^{>0} > \frac{\sum n_i \sigma_i^2}{\sum n_i}$$

همان‌طور که مشاهده می‌شود، با توجه به متفاوت بودن میانگین‌های سه جامعه انتظار داریم واریانس کل (σ^2) از میانگین واریانس‌های جوامع $\left(\dfrac{\sum n_i \sigma_i^2}{\sum n_i}\right)$ بیشتر باشد؛ پس:

$$\frac{\sum n_i \sigma_i^2}{\sum n_i} = \frac{90 \times 7 + 110 \times 10 + 200 \times 9}{90+110+200} = 8.825$$

در نتیجه باید $\sigma^2 > 8.825$ باشد که تنها گزینه ۳ چنین است.

دقت کنید، اگر در سؤال مطرح می‌شد میانگین‌های سه جامعه برابرند $(\mu_1 = \mu_2 = \mu_3)$، آن‌گاه میانگین کل (μ) نیز با آن‌ها برابر بود و داشتیم:

$$\sigma^2 = \frac{\sum n_i\sigma_i^2}{\sum n_i} + \overset{=0}{\boxed{\frac{\sum n_i(\mu_i-\mu)^2}{\sum n_i}}} = \frac{\sum n_i\sigma_i^2}{\sum n_i} = 8.825$$

و پاسخ درست، گزینه ۲ بود.

۶. گزینه ۱ درست است.

بنا بر قضیه چی‌بی‌شف داریم:

$$P(\mu - k\sigma \le X \le \mu + k\sigma) \ge 1 - \frac{1}{k^2}$$

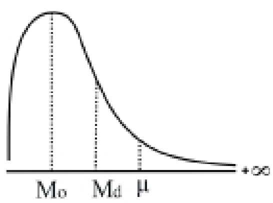
ازآنجاکه حداقل 84 درصد داده‌ها در بازه $(\mu - k\sigma\ ,\ \mu + k\sigma)$ وجود دارد، خواهیم داشت:

$$1 - \frac{1}{k^2} = \frac{84}{100} \to \frac{1}{k^2} = \frac{16}{100} \to k^2 = \frac{100}{16} \to k = \frac{10}{4} = \frac{5}{2}$$

حال با توجه به آنکه $\mu = 52$، $\sigma^2 = 64$، $\sigma = 8$ و $k = \frac{5}{2}$ است، می‌توان بازه مورد نظر را مشخص کرد:

$$(\mu - k\sigma\ ,\ \mu + k\sigma) \to \left(52 - \frac{5}{2}\times 8\ ,\ 52 + \frac{5}{2}\times 8\right) \to (32, 72)$$

۷. گزینه ۲ درست است.



$Sk > 0 \to$ چوله به راست، تمایل به چپ $\to$

۸. گزینه ۴ درست است.

$$\begin{cases} E = \dfrac{\dfrac{\sum(x_i-\mu)^4}{N}}{\sigma^4} - 3 = \dfrac{\dfrac{4500}{60}}{5^2} - 3 = 0 \\[2ex] \sigma^2 = \dfrac{\sum(x_i-\mu)^2}{N} = \dfrac{300}{60} = 5 \end{cases}$$

۹. گزینه ۳ درست است.

تعداد حالات توزیع n شیء در k سلول به طوری که n_1 شیء در سلول اول و … و n_k شیء در سلول kام باشد، برابر است با:

$$\frac{n!}{n_1!n_2!\cdots n_k!} \xrightarrow[n=9]{n_1=2\ ,\ n_2=3\ ,\ n_3=4} \text{تعداد حالات} = \frac{9!}{2!3!4!} = 1260$$

۱۰. گزینه ۳ درست است.

$$P(\text{مهره سیاه } 1 \text{ مهره سفید و } 2 \mid \text{مهره همرنگ } 2) = \frac{P(\text{مهره سیاه } 1 \text{ مهره همرنگ سفید و } 2)}{P(\text{مهره همرنگ } 2)}$$

$$= \frac{\dfrac{\binom{4}{2}\binom{5}{1}}{\cancel{\binom{9}{3}}}}{\dfrac{\binom{4}{2}\binom{5}{1}+\binom{5}{2}\binom{4}{1}}{\cancel{\binom{9}{3}}}} = \frac{\binom{4}{2}\binom{5}{1}}{\underbrace{\binom{4}{2}\binom{5}{1}}_{\text{سفید و 1 سیاه 2}}+\underbrace{\binom{5}{2}\binom{4}{1}}_{\text{سیاه و 1 سفید 2}}} = \frac{3}{7}$$

۱۱. گزینه ۲ درست است.

حالات ممکن : $\binom{10}{2} = 45$ حالت

حالات مساعد : $1,2,3,4,5,6,7,8,9,10 = 9$ حالت

$$\text{احتمال} = \frac{\text{حالات مساعد}}{\text{حالات ممکن}} = \frac{9}{45} = \frac{1}{5}$$

۱۲. گزینه ۳ درست است.

ازآنجاکه شرط پذیرش، قبولی در حداقل 2 امتحان است پس باید در سه امتحان یا اصلاً شکست نخورد یا فقط 1 شکست بخورد.

	اول	دوم	سوم	
1 شکست	q	p	p	$\rightarrow \frac{4}{10}\times\frac{3}{10}\times\frac{6}{10}=\frac{72}{1000}$
	p	q	p	$\rightarrow \frac{6}{10}\times\frac{4}{10}\times\frac{3}{10}=\frac{72}{1000}$
	p	p	q	$\rightarrow \frac{6}{10}\times\frac{6}{10}\times\frac{4}{10}=\frac{144}{1000}$
0 شکست	p	p	p	$\rightarrow \frac{6}{10}\times\frac{6}{10}\times\frac{6}{10}=\frac{216}{1000}$

بنابراین احتمال پذیرش (حداقل 2 موفقیت یا حداکثر 1 شکست) برابر است با:

$$\frac{72+72+144+216}{1000} = \frac{504}{1000}$$

۱۳. گزینه ۲ درست است.

راه حل اول:

حالت خاص از توزیع دوجمله‌ای $\binom{n}{x}p^x q^{n-x}$ زمانی که $p=\frac{1}{2}$ و $q=\frac{1}{2}$ باشد، به صورت زیر است:

$$\begin{cases} P(x)=\binom{n}{x}\left(\frac{1}{2}\right)^n \\ x=0,1,2,\cdots,n \end{cases} \xrightarrow{n=5} \begin{cases} P(x)=\binom{5}{x}\left(\frac{1}{2}\right)^5=\dfrac{\binom{5}{x}}{32} \\ x=0,1,2,\cdots,5 \end{cases}$$

بنابراین در این سؤال:

$$\frac{\binom{5}{x}}{a}=\frac{\binom{5}{x}}{32} \to a=32$$

راه حل دوم:

در هر تابع احتمال گسسته باید مجموع مقادیر تابع احتمال برابر با 1 شود، پس

$$\sum_{x=0}^{5}P(x)=1 \to \frac{\binom{5}{0}+\binom{5}{1}+\cdots+\binom{5}{5}}{a}=1 \to \frac{2^5}{a}=1 \to a=32$$

یادآوری:

$$\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$$

۱۴. گزینه ۴ درست است.

$$\begin{cases} V(-2X+3)=4V(X)=4\times 2.25=9 \\ V(X)=E(X^2)-(E(X))^2=14.5-(3.5)^2=2.25 \end{cases}$$

۱۵. گزینه ۳ درست است.

$$\begin{cases} V(X)+V(Y)-V(X-Y)=V(X)+V(Y)-\left[V(X)+V(Y)-2Cov(X,Y)\right]=2Cov(X,Y)=2 \\ Cov(X,Y)=\underbrace{E(XY)}_{9}-\underbrace{E(X)}_{3.2}\underbrace{E(Y)}_{2.5}=9-\underbrace{3.2\times 2.5}_{8}=1 \end{cases}$$

۱۶. گزینه ۱ درست است.

ازآنجاکه درصد نهال‌های سبز دارای توزیع برنولی با $p=0.80$ است، پس تعداد نهال سبز در $n=5$ نهال کاشته‌شده دارای توزیع دوجمله‌ای خواهد بود، داریم:

$$P(X\geq 2)=1-P(X=0)-P(X=1)=1-\binom{5}{0}(0.8)^0(0.2)^5-\binom{5}{1}(0.8)^1(0.2)^4=0.99328$$

۱۷. گزینه ۱ درست است.

در هر توزیع پواسون با پارامتر λ (متوسط تعداد اتفاقات) داریم:

$$\begin{cases} E(X)=\lambda \\ \sigma_X^2=\lambda \\ \sigma_X=\sqrt{\lambda} \end{cases}$$

پس در توزیع پواسون امید و واریانس با هم برابرند و امید، مربع انحراف معیار است.

۱۸. گزینه ۱ درست است.

ازآنجاکه نسبت تلفات در پرندگان دارای توزیع برنولی با $p=\frac{1.2}{1000}$ است، پس تعداد تلفات در $n=2500$ پرنده مهاجر دارای توزیع دوجمله‌ای است؛ اما شرایط تقریب پواسون برقرار است، زیرا:

$$n=2500\geq 100 \;,\; np=2500\times\frac{1.2}{1000}=3\leq 10$$

بنابراین برای حل مسئله از توزیع پواسون با پارامتر $\lambda=np=3$ استفاده می‌کنیم:

$$P(x)=\frac{e^{-\lambda}\lambda^x}{x!}\xrightarrow{x=5,\lambda=3}\frac{e^{-3}3^5}{5!}=\frac{1}{e^3}\times\frac{243}{120}=\frac{48}{1000}\times\frac{243}{120}=0.0972$$

۱۹. گزینه ۴ درست است.

با توجه به تعریف تابع احتمال داریم:

$$\int_{-\infty}^{+\infty}f(x)dx=1 \to \int_0^1\frac{1}{3}(2x+k)dx=1 \to \frac{1}{3}\left[x^2+kx\right]_0^1=1 \to \frac{1}{3}[1+k]=1 \to k=2$$

بنابراین $P\left(0<X<\frac{1}{2}\right)$ برابر است با:

$$P\left(0<X<\frac{1}{2}\right)=\int_0^{\frac{1}{2}}\frac{1}{3}(2x+2)dx=\frac{1}{3}\left[x^2+2x\right]_0^{\frac{1}{2}}=\frac{1}{3}\left[\frac{1}{4}+1\right]=\frac{5}{12}$$

۲۰. گزینه ۲ درست است.

با توجه به آنکه متغیر X (نمرات مسئولیت‌پذیری) دارای توزیع نرمال با $\mu=16$ و $\sigma=5$ است، برای محاسبه درصد افرادی که نمرات کمتر از 13.8 گرفته‌اند، آن را به حالت استاندارد تبدیل می‌کنیم:

$$P(X<13.8)=P\left(\frac{X-\mu}{\sigma}<\frac{13.8-16}{5}\right)=P(Z<-0.44)$$

حال ازآنجاکه در نرمال استاندارد $P(Z<\alpha)=P(Z>-\alpha)$ است، داریم:

$$P(Z<-0.44)=P(Z>0.44)=1-\underbrace{P(Z\leq 0.44)}_{0.67}=0.33$$

۲۱. گزینه ۴ درست است.

$$Z_i=\frac{X_i-\mu}{\sigma}\xrightarrow[z_i=2,-1]{x_i=33,24}\begin{cases} 2=\frac{33-\mu}{\sigma} \to \sigma=\frac{33-\mu}{2} \\ -1=\frac{24-\mu}{\sigma} \to \sigma=\frac{24-\mu}{-1} \end{cases}$$

$$\frac{33-\mu}{2}=\frac{24-\mu}{-1} \to \mu-33=48-2\mu \to 3\mu=81 \to \mu=27$$

۲۲. گزینه ۱ درست است.

۲۳. گزینه ۱ درست است.

ازآنجاکه 50 درصد افراد جامعه $(p=0.5)$ تحصیلات دانشگاهی دارند، برای محاسبه احتمال اینکه حداقل 55 درصد یک نمونه 100تایی تحصیلات دانشگاهی داشته باشند. ابتدا به صورت زیر آن را به حالت استاندارد تبدیل می‌کنیم:

$$P(\bar{p}\geq 0.55)=P\left(\frac{\bar{p}-P}{\sqrt{\frac{pq}{n}}}\geq\frac{0.55-0.5}{\sqrt{\frac{0.5\times 0.5}{100}}}\right)=P\left(Z\geq\frac{0.05}{\frac{0.5}{10}}\right)=P(Z\geq 1)$$

حال با توجه به راهنمایی سؤال داریم:

$$S_{-\infty}^{-1}=0.1587 \;\rightarrow\; P(Z\leq -1)=0.1587 \xrightarrow{P(Z\leq\alpha)=P(Z\geq-\alpha)} P(Z\geq 1)=0.1587$$

۲۴. گزینه ۴ درست است.

$$\begin{cases}\bar{x}=\dfrac{\sum x}{n}=\dfrac{20}{5}=4\\[2ex] \bar{y}=\dfrac{\sum y}{n}=\dfrac{25}{5}=5\end{cases}$$

x	y	$x-\bar{x}$	$y-\bar{y}$
5	7	1	2
3	9	−1	4
4	3	0	−2
7	2	3	−3
1	4	−3	−1

→

$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
2	1
−4	1
0	0
−9	9
3	9
$\sum=-8$	$\sum=20$

$$b=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})}=\frac{-8}{20}=-0.4$$

۲۵. گزینه ۳ درست است.

$$\frac{1}{F_{\alpha,m,n}}=F_{1-\alpha,n,m} \;\rightarrow\; \frac{1}{F_{0.9,5,7}}=F_{0.1,7,5}$$

پیوست

سری‌های زمانی و مدل‌های پیش‌بینی

تعریف: سری زمانی مجموعه‌ای از مشاهدات است که برحسب زمان یا هر کمیت دیگری مرتب شده باشد و معمولاً آن را به صورت $x_1, x_2, ..., x_n$ نشان می‌دهند.

تجزیه و تحلیل سری‌های زمانی به مشاهداتی مربوط می‌شود که مستقل نیستند و به طور متوالی به هم وابسته‌اند. کاربرد اصلی تجزیه و تحلیل سری‌های زمانی «پیش‌بینی» است.

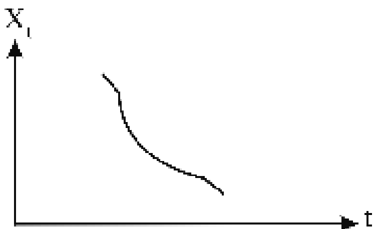
اجزای تشکیل‌دهنده سری زمانی

تغییرات سری زمانی می‌تواند به علت تغییرات بعضی از عوامل که تعدادی از آن‌ها طبیعی بوده و تعدادی دیگر ناشی از عوامل اقتصادی و اجتماعی هستند، به وجود آید. معمولاً برای تحلیل یک سری زمانی، تغییراتی که نتیجه چهار مؤلفه اصلی هستند، درنظر گرفته می‌شوند. این تغییرات عبارت‌اند از:

۱- روند

تعریف: روند عبارت است از تغییرات درازمدت در میانگین سری زمانی؛ به عبارت دیگر سیر طبیعی سری زمانی را در درازمدت روند گویند که در این حالت افت و خیزهای سری زمانی را نادیده گرفته و به نمای کلی آن توجه می‌کنند.

برای مثال، رشته‌کوه‌ها از نزدیک پستی و بلندی‌های زیادی دارند، ولی بر روی نقشه با یک منحنی ساده نشان داده می‌شوند.

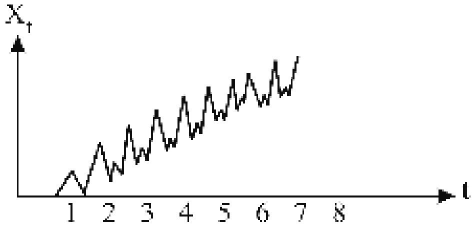


منحنی روند نزولی در سری زمانی

۲- تغییرات فصلی

تعریف: تغییرات فصلی، تغییراتی هستند که در دوره‌های تناوبی کوتاه پیش می‌آیند.

تغییرات فصلی مربوط به عواملی هستند که به طریقی منظم و چرخه‌ای روی یک دوره کمتر از یک سال عمل می‌کنند. اگر مشاهدات سری زمانی به صورت هر سه ماه، ماهانه، هفتگی و یا روزانه ثبت شوند، تغییرات فصلی در سری زمانی وجود دارد.

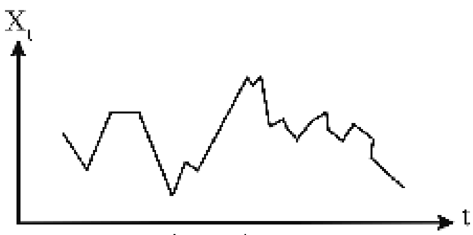


منحنی تغییرات فصلی

۳- تغییرات دوره‌ای

تعریف: حرکات نوسانی در یک سری زمانی با دوره نوسان بیشتر از یک سال را تغییرات دوره‌ای می‌نامند.

تغییرات دوره‌ای در سری‌های زمانی به دلیل افت و خیزهایی است که بعد از یک دوره بیشتر از یک سال رجعت می‌کنند. نوسانات دوره‌ای ممکن است به طور دقیق از طرح‌های مشابهی بعد از فواصل زمانی مساوی پیروی کنند، اما همیشه این طور نیست.
همان‌طورکه در شکل مشخص است، اگر خط روند، موازی با محور t تصور شود، می‌توان نوسانات دوره‌ای را بالا و پایین خط روند مشخص کرد؛ چنین نوساناتی را «تغییرات دوره‌ای» می‌نامند.

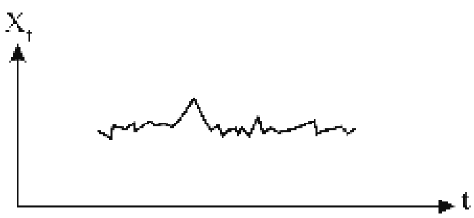


تغییرات دوره ای

۴- تغییرات نامنظم

تعریف: تغییراتی که کاملاً تصادفی و نتیجه نیرویی پیش‌بینی نشدنی است و به طریقی نامنظم عمل می‌کند را تغییرات نامنظم یا تصادفی می‌نامند.

این‌گونه تغییرات، طرح معینی را نشان نمی‌دهند و دوره زمان وقوع آن‌ها منظم نیست. برای مثال، این تغییرات به وسیله عواملی مانند سیل، زلزله، اعتصاب‌ها و غیره به وجود می‌آیند که رفتار آن‌ها نامنظم و پیش‌بینی نشدنی است و به طور معمول کوتاه‌مدت هستند، ولی گاهی اوقات اثر آن‌ها به اندازه‌ای زیاد است که باعث پیدایش تغییرات دوره‌ای و تغییرات دیگر می‌شود.



تغییرات نامنظم سری زمانی

مدل‌های سری زمانی

انواع مدل‌های سری زمانی:
- ۱- مدل جمعی
- ۲- مدل ضربی

مدل‌های پیش‌بینی

برای پیش‌بینی رفتار سری‌های زمانی و تعیین مدل پیش‌بینی، فنون مختلفی وجود دارد؛ این فنون را می‌توان به دو دسته روش‌های کمّی و روش‌های کیفی تقسیم کرد.

مدل‌های کمّی پیش‌بینی

اگر تحلیل‌گر بر اساس رفتار مشاهده‌شده از سری زمانی و تجزیه و تحلیل اجزای آن، بتواند مقادیر آینده را با استفاده از مبانی ریاضی پیش‌بینی کند، از مدل‌های کمّی برای پیش‌بینی در سری زمانی استفاده می‌شود. این مدل‌ها عبارت‌اند از:

۱- مدل ساده

برخی از مدل‌های کمّی پیش‌بینی بسیار ساده و ابتدایی‌اند و می‌توان از آن‌ها در هر نوع سری زمانی استفاده کرد. مدل‌های ساده معمولاً مبنایی برای ارزیابی صحت مدل‌های پیچیده یا نسبتاً پیچیده‌اند، به طوری که می‌توان با مقایسه خطای پیش‌بینی مدل‌های پیچیده با مدل‌های ساده، به درجه اعتبار آن‌ها پی برد. دو نوع از معروف‌ترین مدل‌های ساده عبارت‌اند از:

الف) مدل پیش‌بینی بدون تغییر: این مدل برای سری‌هایی به کار می‌رود که هیچ‌گونه رشد یا کاهش بلندمدتی از خود نشان نمی‌دهند.

ب) مدل پیش‌بینی با درصد تغییر: این مدل برای سری‌هایی به کار می‌رود که یک رفتار نمایی با روند صعودی یا نزولی از خود نشان می‌دهند.

۲- مدل میانگین متحرک

از مدل‌های میانگین متحرک برای پیش‌بینی سری‌هایی که تغییرات نامنظم دارند و فاقد روند و تغییرات فصلی هستند، استفاده می‌شود. به‌علاوه این نوع مدل‌ها برای پیش‌بینی کوتاه‌مدت به کار گرفته می‌شوند.

یکی از انواع میانگین‌های متحرک «میانگین متحرک مرکزی موزون» نامیده می‌شود. در این روش، بیشترین وزن به نقطه مرکزی داده می‌شود و سپس به تناسب دور شدن مشاهدات از مرکز، از وزن آن‌ها کاسته می‌شود.

۳- مدل نمو هموار ساده

در بسیاری از موارد، مدل نمو هموار ساده برای پیش‌بینی مقادیر آینده سری زمانی استفاده می‌شود. این روش برای آن دسته از سری‌های زمانی مفید است که تغییرات منظمی دارند و تغییرات فصلی و دوره‌ای در آن‌ها مورد نظر نیست.

۴- مدل هُلت ـ وینترز

مدل هُلت ـ وینترز برای پیش‌بینی در سری‌هایی که روند دارند به کار می‌رود. همچنین این مدل دارای یک حالت مجزا برای سری‌هایی با تغییرات غیرفصلی و یک حالت مجزا برای سری‌هایی با تغییرات فصلی است.

۵- مدل اتورگرسیو ـ میانگین متحرک تلفیقی

در این مدل‌ها که به مدل‌های ARIMA معروف‌اند، علاوه بر عامل روند به تغییرات فصلی و تصادفی نیز توجه می‌شود. از مدل‌های ARIMA می‌توان مدل‌های متعددی چون مدل رگرسیون ساده، رگرسیون چندگانه، میانگین متحرک و نمو هموار ساده را استخراج کرد.

۶- مدل اقتصادسنجی

مدل‌های اقتصادسنجی، مدل‌های احتمالی هستند و بر روابط احتمالی‌ای که بین یک متغیر وابسته و یک یا چند متغیر مستقل وجود دارد، تأکید دارند. «مدل رگرسیون خطی» یکی از مدل‌های معروف اقتصادسنجی است که در بسیاری از مواقع، مدل احتمالی مناسبی برای تعیین روند بلندمدت در سری‌های زمانی است.

مدل‌های کیفی پیش‌بینی

در بسیاری از موارد برای تصمیم‌گیری، اطلاعات کامل و دقیقی از گذشته در دست نیست و یا محیط به گونه‌ای آشفته است که نمی‌توان اطلاعات گذشته را ملاک پیش‌بینی آینده و تصمیم‌گیری قرار داد. در چنین شرایطی از روش‌های کیفی برای پیش‌بینی آینده استفاده می‌شود. در این روش‌ها مبنای پیش‌بینی، تجربه، تخصص، قضاوت و قدرت پیشگویی صاحب‌نظران و متخصصان آن حوزه تصمیم‌گیری است. مهم‌ترین روش‌های کیفی عبارت‌اند از:

۱- روش دلفی

این روش بر اساس نظر یک گروه از متخصصان و یک نفر هماهنگ‌کننده شکل می‌گیرد و هیچ عضوی از گروه از سایر اعضای آن خبر ندارد بلکه همه ارتباطات از طریق هماهنگ‌کننده اصلی انجام می‌شود. فرآیند روش دلفی به صورت زیر است:

۲- روش طوفان مغزی

یکی از روش‌های بسیار مهم و متداول در پیش‌بینی‌های کیفی، روش طوفان مغزی است. با استفاده از این روش می‌خواهیم تحرکی بزرگ در فکر و ذهن اعضای گروه به وجود آوریم و سیلی خروشان از عقاید و نظرهای بدیع به راه اندازیم تا پیش‌بینی نهایی به واقعیت نزدیک‌تر باشد.

روش طوفان مغزی بر اساس این فرضیه بنا شده است که جلسه‌ای تشکیل شود تا در آن هرکس بتواند آزادانه و بدون ترس از انتقاد، هر فکری که به نظرش می‌رسد، ابراز کند. در جلسات طوفان مغزی یافتن نظرهای جدید اولویت دارد و مطالعه امکان‌پذیری آن‌ها، به بعد موکول می‌شود.

۳- روش گروه اسمی

در روش طوفان مغزی، نظم و قاعده خاصی برای پیش‌بینی وجود ندارد و در واقع، با آزاد گذاشتن کامل افراد و با تأکید و تشویق بر خودجوشی می‌خواهیم به نظر و عقاید تازه‌ای برسیم و از این راه، تصمیم‌گیری گروهی را بهتر و غنی‌تر کنیم. درحالی‌که گروه اسمی، روشی است که ضمن تشویق و ترغیب افراد به نوآوری و خلاقیت و فراهم آوردن شرایط مناسب برای آن، به فرایند پیش‌بینی، نظم بیشتری می‌دهد.

مدل‌های تلفیقی

اغلب تحلیل‌گران برای انتخاب بهترین مدل پیش‌بینی، تلاش می‌کنند که مدل‌های متعدد را به طور مجزا بررسی و آزمون کنند. مدلی که پیش‌بینی‌های آن انحراف کمتری با واقعیت داشته باشد، به عنوان بهترین مدل انتخاب می‌شود. رویکرد انتخاب یک مدل خاص از بین مدل‌های مختلف پیش‌بینی بر این تفکر بنا شده است که بالاخره بهترین مدل برای تبیین رفتار یک سری زمانی وجود دارد و هیچ مدلی غیر از مدل انتخاب‌شده، پیش‌بینی درستی ندارد.

تحقیقات اخیر نشان داده است که ترکیب دو یا چند مدل خاص پیش‌بینی می‌تواند مقادیر پیش‌بینی درست‌تری ارائه دهد. «رویکرد تلفیقی» بر این فرض بنا شده است که برخی مدل‌های حذف‌شده، صرف نظر از مدل انتخاب‌شده اطلاعات مفیدی دارند که با قرار گرفتن در کنار مدل انتخابی و ترکیب آن‌ها با یکدیگر، واریانس خطای پیش‌بینی بسیار کوچک خواهد شد.

اگرچه مدل‌های پیشرفته‌ای چون ARIMA و اقتصاد سنجی، پیش‌بینی‌های خوبی در محیط‌های پویا و دوره‌های زمانی میان‌مدت و کوتاه‌مدت دارند، ولی تحقیقات بسیاری نشان می‌دهد که تلفیق مدل‌های پیش‌بینی به شدت خطای پیش‌بینی را کاهش می‌دهد.

سؤالات کنکور

۱. کدام‌یک از روش‌های زیر جزو روش‌های کمّی سری‌های زمانی است؟ **(مدیریت ـ ۸۰)**

۱) دلفی ۲) اسمی ۳) طوفان مغزی ۴) هلت ـ وینترز

حل: گزینه ۴ درست است.

۲. کدام‌یک از مدل‌های پیش‌بینی دارای متغیر مستقل هستند؟ **(مدیریت ـ ۸۱)**

۱) نمو هموار ساده ۲) هلت ـ وینترز ۳) میانگین متحرک وزنی ۴) اقتصادسنجی

حل: گزینه ۴ درست است.

۳. کدام‌یک از روش پیش‌بینی، برای محیط پویا مناسب‌تر است؟ **(مدیریت ـ ۸۲)**

۱) نمو هموار ۲) میانگین متحرک ۳) هلت ـ وینترز ۴) ARIMA

حل: گزینه ۴ درست است.

۴. از مدل نمو ساده برای پیش‌بینی در چه نوع سری‌های زمانی استفاده می‌شود؟ **(مدیریت ـ ۸۳)**

۱) سری‌های زمانی که تغییرات منظم دارند. ۲) سری‌های زمانی که تغییرات فصلی دارند.
۳) سری‌های زمانی که تغییرات دوره‌ای دارند. ۴) سری‌های زمانی که تغییرات فصلی و نامنظم دارند.

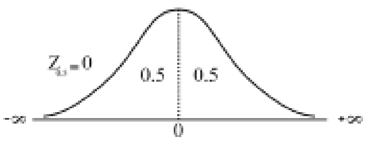
حل: گزینه ۱ درست است.

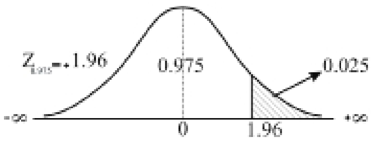
۵. در سری‌های زمانی چهار مؤلفه اصلی وجود دارد. اول روند کلی، دوم تغییرات فصلی و سوم تغییرات ادواری است، چهارمین آن کدام است؟ **(برنامه‌ریزی شهری ـ ۸۱)**

۱) تغییرات جزئی ۲) تغییرات کلی ۳) تغییرات اصلی ۴) تغییرات نامنظم

حل: گزینه ۴ درست است.

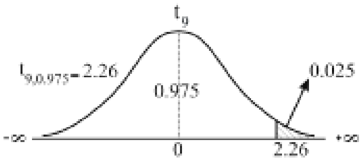
جدول توزیع نرمال استاندارد





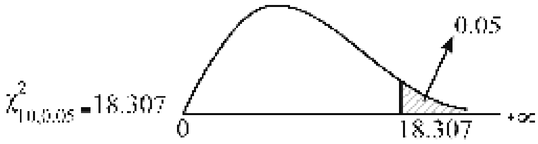
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	**0.5000**	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

جدول توزیع t ـ استیودنت



df	*t.90*	*t.95*	*t.975*	*t.99*	*t.995*
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.93
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	**2.26**	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.84
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.80
26	1.31	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
40	1.30	1.68	2.02	2.42	2.70
60	1.30	1.67	2.00	2.39	2.66
120	1.29	1.66	1.98	2.36	2.62
infinite=Z	**1.28**	**1.64**	**1.96**	**2.33**	**2.58**

جدول توزیع کای اسکور

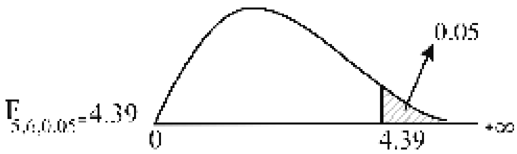


df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	**18.307**	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

	..	10	11	12	13	14	15	16	17	18	19
	0.1	9.39	9.40	9.41	9.41	9.42	9.42	9.43	9.43	9.44	9.44
	0.05	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44
2	0.025	39.40	39.41	39.41	39.42	39.43	39.43	39.44	39.44	39.44	39.45
	0.01	99.40	99.41	99.42	99.42	99.43	99.43	99.44	99.44	99.44	99.45
	0.001	999.40	999.41	999.42	999.42	999.43	999.43	999.44	999.44	999.44	999.45
	0.1	5.23	5.22	5.22	5.21	5.20	5.20	5.20	5.19	5.19	5.19
	0.05	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67
3	0.025	14.42	14.37	14.34	14.30	14.28	14.25	14.23	14.21	14.20	14.18
	0.01	27.23	27.13	27.05	26.98	26.92	26.87	26.83	26.79	26.75	26.72
	0.001	129.25	128.74	128.32	127.96	127.64	127.37	127.14	126.93	126.74	126.57
	0.1	3.92	3.91	3.90	3.89	3.88	3.87	3.86	3.86	3.85	3.85
	0.05	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81
4	0.025	8.84	8.79	8.75	8.71	8.68	8.66	8.63	8.61	8.59	8.58
	0.01	14.55	14.45	14.37	14.31	14.25	14.20	14.15	14.11	14.08	14.05
	0.001	48.05	47.70	47.41	47.16	46.95	46.76	46.60	46.45	46.32	46.21
	0.1	3.30	3.28	3.27	3.26	3.25	3.24	3.23	3.22	3.22	3.21
	0.05	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57
5	0.025	6.62	6.57	6.52	6.49	6.46	6.43	6.40	6.38	6.36	6.34
	0.01	10.05	9.96	9.89	9.82	9.77	9.72	9.68	9.64	9.61	9.58
	0.001	26.92	26.65	26.42	26.22	26.06	25.91	25.78	25.67	25.57	25.48
	0.1	2.94	2.92	2.90	2.89	2.88	2.87	2.86	2.85	2.85	2.84
	0.05	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88
6	0.025	5.46	5.41	5.37	5.33	5.30	5.27	5.24	5.22	5.20	5.18
	0.01	7.87	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45	7.42
	0.001	18.41	18.18	17.99	17.82	17.68	17.56	17.45	17.35	17.27	17.19
	0.1	2.70	2.68	2.67	2.65	2.64	2.63	2.62	2.61	2.61	2.60
	0.05	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46
7	0.025	4.76	4.71	4.67	4.63	4.60	4.57	4.54	4.52	4.50	4.48
	0.01	6.62	6.54	6.47	6.41	6.36	6.31	6.28	6.24	6.21	6.18
	0.001	14.08	13.88	13.71	13.56	13.43	13.32	13.23	13.14	13.06	12.99

جدول توزیع فیشر F



ν_2	p	ν_1	1	2	3	4	5	6	7	8	9
	0.1		8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	0.05		18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
2	0.025		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	0.01		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	0.001		998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
	0.1		5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	0.05		10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
3	0.025		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	0.01		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	0.001		167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
	0.1		4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	0.05		7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
4	0.025		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	0.01		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	0.001		74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
	0.1		4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	0.05		6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
5	0.025		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	0.01		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	0.001		47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
	0.1		3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	0.05		5.99	5.14	4.76	4.53	**4.39**	4.28	4.21	4.15	4.10
6	0.025		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	0.01		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	0.001		35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
	0.1		3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	0.05		5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
7	0.025		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	0.01		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	0.001		29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33

	..	10	11	12	13	14	15	16	17	18	19
	0.1	2.54	2.52	2.50	2.49	2.48	2.46	2.45	2.45	2.44	2.43
	0.05	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16
8	0.025	4.30	4.24	4.20	4.16	4.13	4.10	4.08	4.05	4.03	4.02
	0.01	5.81	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41	5.38
	0.001	11.54	11.35	11.19	11.06	10.94	10.84	10.75	10.67	10.60	10.54
	0.1	2.42	2.40	2.38	2.36	2.35	2.34	2.33	2.32	2.31	2.30
	0.05	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95
9	0.025	3.96	3.91	3.87	3.83	3.80	3.77	3.74	3.72	3.70	3.68
	0.01	5.26	5.18	5.11	5.05	5.01	4.96	4.92	4.89	4.86	4.83
	0.001	9.89	9.72	9.57	9.44	9.33	9.24	9.15	9.08	9.01	8.95
	0.1	2.32	2.30	2.28	2.27	2.26	2.24	2.23	2.22	2.22	2.21
	0.05	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79
10	0.025	3.72	3.66	3.62	3.58	3.55	3.52	3.50	3.47	3.45	3.44
	0.01	4.85	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46	4.43
	0.001	8.75	8.59	8.45	8.32	8.22	8.13	8.05	7.98	7.91	7.86
	0.1	2.25	2.23	2.21	2.19	2.18	2.17	2.16	2.15	2.14	2.13
	0.05	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66
11	0.025	3.53	3.47	3.43	3.39	3.36	3.33	3.30	3.28	3.26	3.24
	0.01	4.54	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15	4.12
	0.001	7.92	7.76	7.63	7.51	7.41	7.32	7.24	7.17	7.11	7.06
	0.1	2.19	2.17	2.15	2.13	2.12	2.10	2.09	2.08	2.08	2.07
	0.05	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56
12	0.025	3.37	3.32	3.28	3.24	3.21	3.18	3.15	3.13	3.11	3.09
	0.01	4.30	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91	3.88
	0.001	7.29	7.14	7.00	6.89	6.79	6.71	6.63	6.57	6.51	6.45
	0.1	2.14	2.12	2.10	2.08	2.07	2.05	2.04	2.03	2.02	2.01
	0.05	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47
13	0.025	3.25	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96
	0.01	4.10	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72	3.69
	0.001	6.80	6.65	6.52	6.41	6.31	6.23	6.16	6.09	6.03	5.98

	p	1	2	3	4	5	6	7	8	9
	0.1	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
8	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	0.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
	0.1	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	0.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
9	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	0.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
	0.1	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
10	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	0.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	0.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96
	0.1	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	0.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
11	0.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	0.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	0.001	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
	0.1	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	0.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
12	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	0.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
	0.1	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	0.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
13	0.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	0.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	0.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98

	..	10	11	12	13	14	15	16	17	18	19
	0.1	2.10	2.07	2.05	2.04	2.02	2.01	2.00	1.99	1.98	1.97
	0.05	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40
14	0.025	3.15	3.09	3.05	3.01	2.98	2.95	2.92	2.90	2.88	2.86
	0.01	3.94	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56	3.53
	0.001	6.40	6.26	6.13	6.02	5.93	5.85	5.78	5.71	5.66	5.60
	0.1	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.95	1.94	1.93
	0.05	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34
15	0.025	3.06	3.01	2.96	2.92	2.89	2.86	2.84	2.81	2.79	2.77
	0.01	3.80	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42	3.40
	0.001	6.08	5.94	5.81	5.71	5.62	5.54	5.46	5.40	5.35	5.29
	0.1	2.03	2.01	1.99	1.97	1.95	1.94	1.93	1.92	1.91	1.90
	0.05	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29
16	0.025	2.99	2.93	2.89	2.85	2.82	2.79	2.76	2.74	2.72	2.70
	0.01	3.69	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31	3.28
	0.001	5.81	5.67	5.55	5.44	5.35	5.27	5.20	5.14	5.09	5.04
	0.1	2.00	1.98	1.96	1.94	1.93	1.91	1.90	1.89	1.88	1.87
	0.05	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24
17	0.025	2.92	2.87	2.82	2.79	2.75	2.72	2.70	2.67	2.65	2.63
	0.01	3.59	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21	3.19
	0.001	5.58	5.44	5.32	5.22	5.13	5.05	4.99	4.92	4.87	4.82
	0.1	1.98	1.95	1.93	1.92	1.90	1.89	1.87	1.86	1.85	1.84
	0.05	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20
18	0.025	2.87	2.81	2.77	2.73	2.70	2.67	2.64	2.62	2.60	2.58
	0.01	3.51	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13	3.10
	0.001	5.39	5.25	5.13	5.03	4.94	4.87	4.80	4.74	4.68	4.63
	0.1	1.96	1.93	1.91	1.89	1.88	1.86	1.85	1.84	1.83	1.82
	0.05	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17
19	0.025	2.82	2.76	2.72	2.68	2.65	2.62	2.59	2.57	2.55	2.53
	0.01	3.43	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05	3.03
	0.001	5.22	5.08	4.97	4.87	4.78	4.70	4.64	4.58	4.52	4.47
	0.1	1.94	1.91	1.89	1.87	1.86	1.84	1.83	1.82	1.81	1.80
	0.05	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14
20	0.025	2.77	2.72	2.68	2.64	2.60	2.57	2.55	2.52	2.50	2.48
	0.01	3.37	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99	2.96
	0.001	5.08	4.94	4.82	4.72	4.64	4.56	4.49	4.44	4.38	4.33

	p	1	2	3	4	5	6	7	8	9
	0.1	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	0.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
14	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	0.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58
	0.1	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	0.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
15	0.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	0.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	0.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
	0.1	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	0.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
16	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	0.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98
	0.1	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	0.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
17	0.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	0.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
	0.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75
	0.1	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	0.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
18	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	0.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
	0.1	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	0.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
19	0.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	0.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	0.001	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39
	0.1	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
20	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24

	..	10	11	12	13	14	15	16	17	18	19
	0.1	1.92	1.90	1.87	1.86	1.84	1.83	1.81	1.80	1.79	1.78
	0.05	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.11
21	0.025	2.73	2.68	2.64	2.60	2.56	2.53	2.51	2.48	2.46	2.44
	0.01	3.31	3.24	3.17	3.12	3.07	3.03	2.99	2.96	2.93	2.90
	0.001	4.95	4.81	4.70	4.60	4.51	4.44	4.37	4.31	4.26	4.21
	0.1	1.90	1.88	1.86	1.84	1.83	1.81	1.80	1.79	1.78	1.77
	0.05	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08
22	0.025	2.70	2.65	2.60	2.56	2.53	2.50	2.47	2.45	2.43	2.41
	0.01	3.26	3.18	3.12	3.07	3.02	2.98	2.94	2.91	2.88	2.85
	0.001	4.83	4.70	4.58	4.49	4.40	4.33	4.26	4.20	4.15	4.10
	0.1	1.89	1.87	1.84	1.83	1.81	1.80	1.78	1.77	1.76	1.75
	0.05	2.27	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08	2.06
23	0.025	2.67	2.62	2.57	2.53	2.50	2.47	2.44	2.42	2.39	2.37
	0.01	3.21	3.14	3.07	3.02	2.97	2.93	2.89	2.86	2.83	2.80
	0.001	4.73	4.60	4.48	4.39	4.30	4.23	4.16	4.10	4.05	4.00
	0.1	1.88	1.85	1.83	1.81	1.80	1.78	1.77	1.76	1.75	1.74
	0.05	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04
24	0.025	2.64	2.59	2.54	2.50	2.47	2.44	2.41	2.39	2.36	2.35
	0.01	3.17	3.09	3.03	2.98	2.93	2.89	2.85	2.82	2.79	2.76
	0.001	4.64	4.51	4.39	4.30	4.21	4.14	4.07	4.02	3.96	3.92
	0.1	1.87	1.84	1.82	1.80	1.79	1.77	1.76	1.75	1.74	1.73
	0.05	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04	2.02
25	0.025	2.61	2.56	2.51	2.48	2.44	2.41	2.38	2.36	2.34	2.32
	0.01	3.13	3.06	2.99	2.94	2.89	2.85	2.81	2.78	2.75	2.72
	0.001	4.56	4.42	4.31	4.22	4.13	4.06	3.99	3.94	3.88	3.84
	0.1	1.86	1.83	1.81	1.79	1.77	1.76	1.75	1.73	1.72	1.71
	0.05	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00
26	0.025	2.59	2.54	2.49	2.45	2.42	2.39	2.36	2.34	2.31	2.29
	0.01	3.09	3.02	2.96	2.90	2.86	2.81	2.78	2.75	2.72	2.69
	0.001	4.48	4.35	4.24	4.14	4.06	3.99	3.92	3.86	3.81	3.77
	0.1	1.85	1.82	1.80	1.78	1.76	1.75	1.74	1.72	1.71	1.70
	0.05	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00	1.99
27	0.025	2.57	2.51	2.47	2.43	2.39	2.36	2.34	2.31	2.29	2.27
	0.01	3.06	2.99	2.93	2.87	2.82	2.78	2.75	2.71	2.68	2.66
	0.001	4.41	4.28	4.17	4.08	3.99	3.92	3.86	3.80	3.75	3.70

	p	1	2	3	4	5	6	7	8	9
	0.1	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	0.05	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
21	0.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	0.01	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	0.001	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
	0.1	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	0.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
22	0.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	0.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	0.001	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
	0.1	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	0.05	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
23	0.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	0.01	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	0.001	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89
	0.1	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	0.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
24	0.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	0.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	0.001	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80
	0.1	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	0.05	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
25	0.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	0.01	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	0.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71
	0.1	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	0.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
26	0.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	0.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	0.001	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
	0.1	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	0.05	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
27	0.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	0.01	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	0.001	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57

	..	10	11	12	13	14	15	16	17	18	19
	0.1	1.84	1.81	1.79	1.77	1.75	1.74	1.73	1.71	1.70	1.69
	0.05	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97
28	0.025	2.55	2.49	2.45	2.41	2.37	2.34	2.32	2.29	2.27	2.25
	0.01	3.03	2.96	2.90	2.84	2.79	2.75	2.72	2.68	2.65	2.63
	0.001	4.35	4.22	4.11	4.01	3.93	3.86	3.80	3.74	3.69	3.64
	0.1	1.83	1.80	1.78	1.76	1.75	1.73	1.72	1.71	1.69	1.68
	0.05	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97	1.96
29	0.025	2.53	2.48	2.43	2.39	2.36	2.32	2.30	2.27	2.25	2.23
	0.01	3.00	2.93	2.87	2.81	2.77	2.73	2.69	2.66	2.63	2.60
	0.001	4.29	4.16	4.05	3.96	3.88	3.80	3.74	3.68	3.63	3.59
	0.1	1.82	1.79	1.77	1.75	1.74	1.72	1.71	1.70	1.69	1.68
	0.05	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95
30	0.025	2.51	2.46	2.41	2.37	2.34	2.31	2.28	2.26	2.23	2.21
	0.01	2.98	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60	2.57
	0.001	4.24	4.11	4.00	3.91	3.82	3.75	3.69	3.63	3.58	3.53
	0.1	1.76	1.74	1.71	1.70	1.68	1.66	1.65	1.64	1.62	1.61
	0.05	2.08	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87	1.85
40	0.025	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.13	2.11	2.09
	0.01	2.80	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42	2.39
	0.001	3.87	3.75	3.64	3.55	3.47	3.40	3.34	3.28	3.23	3.19

	p	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
	0.1	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
	0.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
28	0.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
	0.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	0.001	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50
	0.1	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	0.05	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
29	0.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
	0.01	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	0.001	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
	0.1	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
30	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
	0.1	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	0.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
40	0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
	0.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	0.001	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02

مروری بر نماد مجموع ($\sum$)

قانون ۱:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

قانون ۲:

$$\begin{cases} \sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \cdots + x_n^2 \\ \left(\sum_{i=1}^{n} x_i\right)^2 = (x_1 + x_2 + \cdots + x_n)^2 \end{cases} \rightarrow \sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$$

قانون ۳:

$$\begin{cases} \sum x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\ \sum x_i \sum y_i = (x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n) \end{cases} \rightarrow \sum x_i y_i \neq \sum x_i \sum y_i$$

قانون ۴: اگر c, b, a اعداد ثابتی باشند،

$$\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} c = c + c + \cdots + c = nc$$

$$\sum_{i=1}^{n} (x_i \pm c) = \sum_{i=1}^{n} x_i \pm nc$$

$$\begin{aligned} \sum_{i=1}^{n} (x_i \pm y_i) &= (x_1 \pm y_1) + (x_2 \pm y_2) + \cdots + (x_n \pm y_n) \\ &= (x_1 + x_2 + \cdots + x_n) \pm (y_1 + y_2 + \cdots + y_n) \\ &= \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i \end{aligned}$$

$$\sum_{i=1}^{n} (a x_i \pm b y_i \pm c z_i) = a \sum_{i=1}^{n} x_i \pm b \sum_{i=1}^{n} y_i \pm c \sum_{i=1}^{n} z_i$$

قانون ۵: با توجه به آنکه $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ میانگین حسابی $x_1, x_2, \ldots, x_n$ است خواهیم داشت:

$$\begin{cases} \sum_{i=1}^{n} x_i = n\bar{x} \\ \sum_{i=1}^{n} \bar{x} = \underbrace{\bar{x} + \bar{x} + \cdots + \bar{x}}_{\text{n تا}} = n\bar{x} \end{cases}$$

حال می‌توانیم قانون زیر را مطرح کنیم:

$$\sum_{i=1}^{n}(x_i-\overline{x})=0$$

زیرا:

$$\sum_{i=1}^{n}(x_i-\overline{x})=\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\overline{x}=n\overline{x}-n\overline{x}=0$$

قانون ۶:

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}(x_i-\overline{x})y_i$$

اثبات:

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}\left(\underline{x_iy_i}-\underline{\underline{x_i\overline{y}}}-\underline{\overline{x}y_i}+\underline{\underline{\overline{x}\,\overline{y}}}\right)$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})y_i-\overline{y}\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})}_{0}$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})y_i$$

قانون ۷:

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}(y_i-\overline{y})x_i$$

اثبات:

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\sum_{i=1}^{n}\left(\underline{x_iy_i}-\underline{x_i\overline{y}}-\underline{\underline{\overline{x}y_i}}+\underline{\underline{\overline{x}\,\overline{y}}}\right)$$

$$=\sum_{i=1}^{n}(y_i-\overline{y})x_i-\overline{x}\underbrace{\sum_{i=1}^{n}(y_i-\overline{y})}_{0}$$

$$=\sum_{i=1}^{n}(y_i-\overline{y})x_i$$