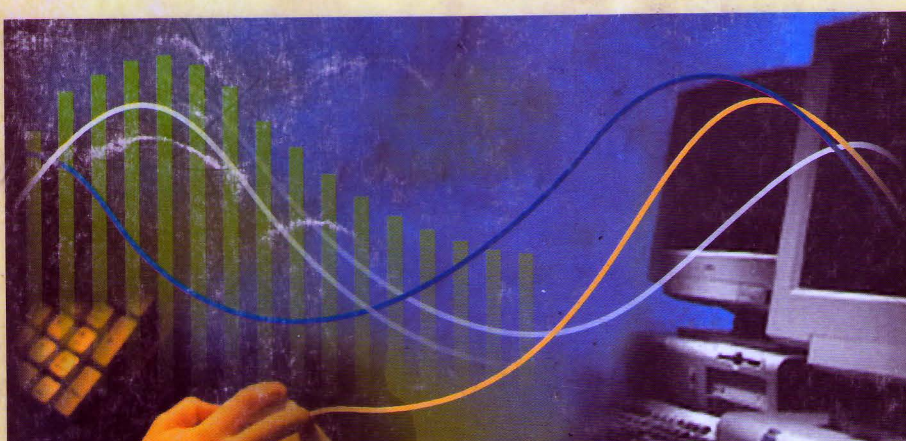
دانشگاه پیام نور

آمار و احتمالات مهندسی

(رشتهٔ کامپیوتر)

دکتر پرویز نصیری



درسنامه

دانشگاه پیام نور

آمار و احتمالات مهندسی

(رشتهٔ کامپیوتر)

دکتر پرویز نصیری

گروه آمار | دانشگاه پیام نور
(۳۰/ د) | ۱۱۵٤

نصیری، پرویز
آمار و احتمالات مهندسی (رشتهٔ کامپیوتر)/ تألیف پرویز نصیری.ــ تهران: دانشگاه پیام نور، ۱۳۸٤.
سیزده، ۲۹٦ص.ــ(دانشگاه پیام‌نور؛ ۱۱۵٤. گروه آمار ؛۳۰ / د)
فهرستنویسی بر اساس اطلاعات فیپا.
واژه‌نامه.
کتابنامه: ص. ۲۹۰ ـ ۲۹۱.

۱.آموزش از راه دور -- ایران. ۲. آمار-- آموزش برنامه‌ای.۳. احتمالات ــ آموزش برنامه‌ای. الف. دانشگاه پیام‌نور. ب. عنوان.
LC ۵۸۰۸ / ۹الف۱۳ن٦ | ۳۷۸ /۱۷۵۰۹۵۵
کتابخانه ملی ایران | ۱٦۵۱۸ ـ ۸٤ م

دانشگاه پیام نور

آمار و احتمالات مهندسی (رشتهٔ کامپیوتر)

دکتر پرویز نصیری

ویراستار علمی: محمدهادی خسروی‌راد

حروفچینی و نمونه‌خوانی: مدیریت تدوین

طراح جلد: آتنا عسگری‌فرد

لیتوگرافی: انتشارات دانشگاه‌پیام نور

چاپ و صحافی: چاپ راوی

شمارگان: ۸۰۰۰ نسخه

نوبت و تاریخ چاپ: چاپ اول مهر ۱۳۸٤، چاپ چهارم آذر ۱۳۸۵

شابک ۰ - ۱۷۷ - ۳۸۷ - ۹٦٤

ISBN 964 - 387 - 177 - 0

قیمت: ۱۷٦۰۰ ریال

بسم الله الرحمن الرحیم

پیشگفتار ناشر

کتابهای دانشگاه پیام نور حسب مورد و با توجه به شرایط مختلف به صورت درسنامه، آزمایشی، قطعی، متون آزمایشگاهی، فرادرسی، و کمک‌درسی چاپ می‌شود. کتاب **درسنامه** (د) نخستین ثمرهٔ کوششهای علمی صاحب اثر است که براساس نیازهای درسی دانشجویان و سرفصلهای مصوب تهیه می‌شود و پس از داوری علمی در گروههای آموزشی، بدون طراحی آموزشی و ویرایش چاپ می‌شود. با تجدیدنظر صاحب اثر و دریافت بازخوردها و اصلاح نارساییها، درسنامه با طراحی آموزشی، ویرایش، و طراحی فنی ـ هنری به صورت **آزمایشی** (آ) چاپ می‌شود. با دریافت نظرهای اصلاحی، صاحب اثر در کتاب تجدید نظر می‌کند و کتاب به صورت **قطعی** (ق) چاپ می‌شود. در صورت ضرورت، در کتابهای چاپ قطعی نیز تجدید نظرهای اساسی به عمل می‌آید یا متناسب با پیشرفت علوم و فناوری بازنویسی می‌شوند. **متون آزمایشگاهی** (م) متونی است که دانشجویان با استفاده از آن و راهنمایی مربیان کارهای عملی آزمایشگاهی را انجام می‌دهند. کتابهای **فرادرسی** (ف) و **کمک‌درسی** (ک) به منظور غنی‌تر کردن منابع درسی دانشگاهی تهیه می‌شوند. کتابهای فرادرسی با تأیید معاونت پژوهشی و کتابهای کمک‌درسی با تأیید شورای انتشارات تهیه می‌شوند.

مدیریت تدوین

فهرست

فصل ۴-توزیع احتمال‌های خاص

تقدیم به:

دکتر م . فهیمی نیری و دکتر س. جلیلی

پیشگفتار

برای نیل به خود کفایی و تحکیم استقلال کشورمان ، تربیت افراد متخصص در تمام زمینه‌ها امری لازم و اجتناب ناپذیر است ، یکی از عوامل مهم وموفق در امر آموزش وجود مراجع مفید دررشته‌های علوم و فنون است.کاربرد آمار و احتمالات در رشته‌های فنی و مهندسی و تجربه تدریس درس آمار و احتمالات مهندسی به کلیه رشته‌های فنی و مهندسی، بنده را بر آن داشته که کتاب حاضر را جهت استفاده دانشجویان تالیف نمایم. در تدوین مطالب کتاب سعی شده است که مطالب هر فصل پیش نیاز فصلهای بعدی باشد . همه فصلها مشتمل بر مثالها و مسائل متنوع است و با اطلاعات دبیرستانی مطالب اساسی آن قابل فهم است .

بر خود لازم می‌دانم از کلیه کسانی که در تهیه این کتاب همکاری داشته‌اند صمیمانه تشکر و قدردانی نمایم. به ویژه از اعضای شورای گروه آمار دانشگاه پیام نور ، مدیریت تدوین و همچنین از زحمات دکتر عباسی، آقای خسروی راد ، آقای نیک دست ، آقای خسروبیگی ، آقای مهندس قاضی دیزجی و دانشجویان رشته آمار خانم ندا میرزابیگی و خانم وحیده محمدی سپاسگزاری می‌کنم .

بدیهی است در تدوین هر اثر علمی لغزش و خطا اجتناب ناپذیر است . از این رو انتظار می‌رود دانش‌پژوهان و دانشجویان ارجمند نظرات اصلاحی خود را تذکر دهند .

پرویز نصیری

بهمن ماه۸۳

فصل ۱

آمار توصیفی

مقدمه

علم آمار با روشهای مورد استفاده در جمع‌آوری، ارائه، تجزیه وتحلیل و تفسیر داده‌ها سرو کار دارد. روشهایی که برای تجزیه و تحلیل مجموعه‌ای از داده‌ها بکار می‌روند تا حدود زیادی به روشی که برای جمع‌آوری اطلاعات بکار رفته است، بستگی دارند.
هرنوع عمل کردن روی داده‌ها که به پیش‌بینی‌ها یا استنباطهایی درباره گروه بزرگتری از داده‌ها منجر گردد، آمار استنباطی و مجموعه روشها و قوانینی که نتایج را ساده‌تر کند و به تفسیر آسانتری از جمع‌آوری و طبقه‌بندی از داده‌ها منجر گردد، آمار توصیفی شمرده می‌شود. تاریخ آمار توصیفی به تاریخ پیدایش نوع انسان برمی‌گردد.
در اوایل دوران مسیحیت، آمار برای کسب اطلاعات درباره مالیاتها، جنگلها، محصولات کشاورزی و حتی فعالیتهای ورزشی بکار می‌رفت.
برای اینکه تفاوت آمار توصیفی و آمار استنباطی کاملاً روشن شود، مجموعه‌ای از تعداد رایانه‌هایی را که در ۱۰ سال گذشته وارد بازار شده است درنظر بگیرید. هر مشخصه‌ای که داده‌ها را توصیف کند مثل میانگین، میانه، نما و دامنه رایانه‌ها مشخصه‌ای از رشته آمار توصیفی می‌باشد. اگر براساس داده‌های ۱۰ سال گذشته میانگین رایانه‌ها را برای سال یازدهم پیش‌بینی کنیم یا حدس بزنیم این کار درقلمرو آمار استنباطی قرار می‌گیرد.

۱-۱ مفاهیم اساسی

در این بخش به تعریف برخی از مفاهیم و اصطلاحات اساسی که در آمار همواره با آن روبرو هستیم می‌پردازیم.

۱-۱-۱ جامعه

جامعه از لحاظ آمار به مجموعه‌ای از افراد یا اشیایی که لااقل دارای یک صفت مشخص کننده باشند اطلاق می‌شود. جامعه آماری ممکن است متناهی (محدود) یا نامتناهی باشد. از بین جامعه‌های آماری که نامتناهی هستند، جامعه‌ای آماری را درنظر می‌گیریم که شمارا باشد مثلاً مجموعه اعداد طبیعی. فرض کنید جامعه آماری مورد مطالعه متناهی باشد. در این صورت حجم یا اندازه جامعه آماری را با N نشان می‌دهیم و اعضای آن را به صورت زیر نمایش می‌دهیم.

$$X_۱, X_۲, \dots, X_i, \dots, X_N$$

که X_i عضو iام جامعه است برای $i=۱,۲,\dots,N$.

مثال ۱-۱-۲ فرض کنید جامعه آماری مجموعه‌ای از انواع رایانه‌ها باشد. مثلاً پنتیوم I، پنتیوم II، پنتیوم IIIو پنتیوم IV. این جامعه آماری متناهی و $N=۴$ است.

۱-۱-۳ نمونه

هر بخشی از جامعه آماری را نمونه گویند. فرض کنید جامعه آماری یا جمعیت مورد بررسی مهندسین فارغ التحصیل رشته نرم افزار کامپیوتر تا سال ۱۳۸۳ در ایران باشد. فارغ التحصیلان با معدل بالاتر از ۱۵ یک نمونه محسوب می شود که جزئی از جامعه است. اگر نمونه انتخاب شده دارای حجم یا اندازه n باشد. اعضای نمونه را به صورت زیر نمایش می دهیم.

$$X_۱, X_۲, \dots, X_i, \dots, X_n$$

که X_i عضو iام نمونه است برای $i=۱,۲,\dots,n$.

۱-۱-۴ انواع داده های آماری

انواع داده‌های آماری به دو گروه، داده‌های دست اول (خام) و داده‌های دست دوم تقسیم بندی می شوند.
داده‌های خام، آن دسته از داده‌هایی هستند که روی آنها هیچگونه سازماندهی یا پردازش صورت نگرفته است مانند داده های سرشماری قبل از هرگونه دسته بندی یا پردازش.
داده‌های دست دوم، داده‌هایی خام هستند که روی آنها حداقل یک پردازش صورت گرفته باشد مانند مرتب کردن داده‌ها یا طبقه بندی داده‌ها.

۱-۱-۵ متغیر

متغیر علامتی است برای معرفی اعضای جامعه آماری و به ویژگی اطلاق می‌شود و تغییرات را از فردی به فردی یا از شئ به شئ دیگر نشان می دهد. به عنوان مثال مرتب کردن داده‌ها بوسیله چند دستگاه رایانه با مدلهای مختلف، متغیر تلقی می شود چون زمان لازم برای مرتب کردن در بین رایانه ها یکسان نیست بلکه از رایانه ای به رایانه‌ای دیگر فرق می کند. متغیر ممکن است کمی یا کیفی باشد.
متغیرکمی به متغیری گفته می شود که قابل اندازه‌گیری باشد یا برای اندازه‌گیری آن مقیاس وجود داشته باشد مانند طول قد، وزن، سرعت رایانه و ... متغیرکمی به نوبه خود ممکن است گسسته یا پیوسته باشد. متغیرکمی گسسته آنهایی هستند که قابل شمارش هستند. مانند تعداد برنامه‌های رایانه، سن، تعداد باسوادان، تعداد دانشجویان رشته نرم افزار کامپیوتر، تعداد روزهای بارانی در فصل بارندگی.
متغیرکمی پیوسته آنهایی هستند که بین هردو مقدار دلخواه بتوانند مقداری انتخاب کنند. مانند وزن یک شخص، قد یک شخص، سرعت باد و ... همگی متغیرکمی پیوسته هستند چون مقدار هر کدام از آنها می تواند به صورت فاصله بیان شود.

۱-۱-۶ عملگرهای جمع و ضرب

عملگر جمع به صورت زیر تعریف می شود

$$\sum_{i=1}^{n} X_i = X_۱ + X_۲ + \ldots + X_n$$

که علامت Σ یک حرف یونانی است و سیگما نامیده می شود. تعدادی از خواص آن عبارتند از:

وقتی که k ثابت است $\sum_{i=1}^{n} k = k + k + \ldots + k = nk$

$$\sum_{i=1}^{n}(X_i + Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$$

$$\sum_{i=1}^{n}(A + BX_i) = nA + B\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n}(A - BX_i)^2 = \sum_{i=1}^{n}(A^2 + B^2X_i^2 - 2ABX_i) = nA^2 + B^2\sum_{i=1}^{n} X_i^2 - 2AB\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} = \sum_{i=1}^{n}(X_{i1} + X_{i2} + \ldots + X_{im}) = \sum_{i=1}^{n} X_{i1} + \sum_{i=1}^{n} X_{i2} + \ldots + \sum_{i=1}^{n} X_{im}$$
$$= X_{11} + X_{21} + \ldots + X_{n1} + X_{12} + X_{22} + \ldots + X_{n2} + \ldots + X_{1m} + X_{2m} + \ldots + X_{nm}$$

عملگر ضرب به صورت زیر تعریف می شود

$$\prod_{i=1}^{n} X_i = X_۱ . X_۲ . \ldots . X_n$$

که Π (پی) نیز از حروف یونانی و علامت ضرب است. تعدادی خواص آن عبارتند از:

$$\prod_{i=1}^{n} kX_i = k^n \prod_{i=1}^{n} X_i$$

$$\prod_{i=1}^{n}(X_i + Y_i) = (X_1 + Y_1).(X_2 + Y_2). \ldots .(X_n + Y_n)$$

$$\prod_{i=1}^{n}(kX_i+Y_i)^2=\prod_{i=1}^{n}(k^2X_i^2+Y_i^2+2kX_iY_i)$$
$$=(k^2X_1^2+Y_1^2-2kX_1Y_1).(k^2X_2^2+Y_2^2-2kX_2Y_2)\dots(k^2X_n^2+Y_n^2-2kX_nY_n)$$
$$\prod_{i=1}^{n}\left(\frac{X_i}{k}\right)=\left(\frac{1}{k}\right)^n\prod_{i=1}^{n}X_i$$

۱-۲ شاخص‌های گرایش مرکزی

برای مقایسه جامعه‌های مختلف آماری وهمچنین برای افزایش دقت، نیاز به شاخص‌های عددی برای خصوصیات هر جامعه است. شاخص‌های توصیفی که از داده‌های یک جامعه محاسبه می شود پارامتر نامیده می شود شاخص‌های توصیفی که ازداده‌های یک نمونه محاسبه می شود شاخص‌های آماری نامیده می شود. در این بخش و بخش بعدی شاخص‌های گرایش مرکزی و شاخص‌های پراکندگی را مورد بحث قرار می‌دهیم. مهمترین شاخص‌های گرایش مرکزی که مورد بحث قرار خواهند گرفت عبارتند از میانگین، میانه، نما و چارکها.

۱-۲-۱ میانگین

میانگین یکی از متداول‌ترین و با اهمیت‌ترین شاخص‌های گرایش مرکزی است که در اینجا انواع آن در جامعه و نمونه مورد بحث قرار می گیرد.

۱-۲-۲ میانگین حسابی

فرض کنید جامعه مورد بررسی دارای N عضو $X_۱$، $X_۲$، ...، X_N باشد. میانگین جامعه از رابطه زیر بدست می آید.

$$\mu=\overline{X}=\frac{1}{N}\sum_{i=1}^{N}X_i=\frac{1}{N}(X_1+X_2+\dots+X_N)$$

که μ یک حرف یونانی است. برای محاسبه μ از تمام اعضای جامعه استفاده کردیم لذا μ یک پارامتر است.

فرض کنید x_1، x_2،...، x_n یک نمونه از جامعه مورد بررسی باشد. میانگین نمونه از رابطه زیر حساب می شود.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

۱-۲-۳ ویژگیهای میانگین نمونه

الف- متغیر است، چون از یک نمونه به نمونه دیگر تغییر می کند.

ب- در محاسبه آن از تمام داده‌های نمونه استفاده می شود.

ج- اگر به تک تک داده‌ها مقدار ثابت a اضافه یا کم شود، میانگین جدید از اضافه یا کم کردن a از میانگین قدیم بدست می آید.

د- اگر تک تک داده‌های نمونه را در عدد ثابت k ضرب یا بر آن تقسیم کنیم میانگین جدید از ضرب کردن میانگین قدیم در k یا تقسیم کردن بر k بدست می آید.

مثال ۱-۲-۴ فرض کنید کانون مهندسین نرم افزار کامپیوتر دارای ۷ عضو است که حقوق سالانه آنها عبارتند از:

۱۷۵۰	۱۴۰۰	۱۳۰۰	۲۰۰۰	۱۹۰۰	۱۷۰۰	۱۵۰۰

الف- میانگین جامعه را حساب کنید.

ب- اگر x_1،x_2،x_3 به ترتیب با مقادیر ۱۷۰۰، ۲۰۰۰ و ۱۴۰۰ یک نمونه باشند. میانگین نمونه را حساب کنید.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{7}\sum_{i=1}^{7} X_i = \frac{1}{7}(X_1 + X_2 + \ldots + X_7)$$

$$= \frac{1}{7}(1500 + 1700 + 1900 + 2000 + 1300 + 1400 + 1750) = 1650$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{1}{3}(x_1 + x_2 + x_3) = \frac{1}{3}(1700 + 2000 + 1400) = 1700$$

می‌توان برای محاسبه μ و $\bar{x}$، ابتدا داده‌ها را بر عدد ۱۰ تقسیم کرد و μ_1 و $\bar{x}_1$ را برای داده‌های جدید محاسبه کرد و با ضرب عدد ۱۰ در μ_1 و $\bar{x}_1$، μ و $\bar{x}$ را بدست آورد. با این فرض داریم،

۱۷۵	۱۴۰	۱۳۰	۲۰۰	۱۹۰	۱۷۰	۱۵۰

$$\mu_1 = \frac{1}{7}\sum_{i=1}^{7} x_i = \frac{1}{7}(150+170+190+200+130+140+175) = 165$$
$$\mu = 10\mu_1 = 10 \times 165 = 1650$$
$$\bar{x}_1 = \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{1}{3}(170+200+140) = 170$$
$$\bar{x} = 10\bar{x}_1 = 1700$$

۱-۲-۵ میانگین هندسی

اگر x_1، x_2، ...، x_n یک نمونه به حجم n از جامعه مورد بررسی باشد میانگین هندسی از رابطه زیر بدست می آید و با علامت G نمایش داده می شود.

$$G = \sqrt[n]{x_1, x_2, \ldots, x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

مثال ۱-۲-۶ جدول زیر جمعیت ایران را در ۵ سرشماری متوالی نشان می‌دهد.

سال سرشماری	۱۳۳۵	۱۳۴۵	۱۳۵۵	۱۳۶۵	۱۳۷۵
جمعیت	۱۸۹۵۴۷۰۴	۲۵۷۸۸۷۲۲	۳۳۷۰۸۷۴۴	۴۹۴۴۵۰۱۰	۶۲۰۱۳۱۷۱

میانگین هندسی را برای جمعیت حساب کنید.

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[5]{\prod_{i=1}^{n} x_i} = \sqrt[5]{x_1 . x_2 . x_3 . x_4 . x_5}$$
$$G = \sqrt{5.0\ 52377664 \times 10^{37}} = 34729550.42$$

یا به طور تقریبی

$$G = 34729550$$

گاهی اوقات، میانگین هندسی برای محاسبه رشد متوسط افزایش فاصله بین دو زمان مانند درصد متوسط افزایش تولید ناخالص ملی بین دو تاریخ یا رشد افزایش جمعیت، مصرف برق، مصرف گاز یا نرخ افزایش وجوه حاصل از بهره مرکب مورد استفاده قرار می گیرد. اگر بخواهیم رشد جمعیت را در مثال بالا حساب کنیم باید مراحل زیر را انجام دهیم.

سال	۱۳۳۵	۱۳۴۵	۱۳۵۵	۱۳۶۵	۱۳۷۵
درصدجمعیت نسبت به ده سال‌قبل	-	۱۳۶/۱	۱۳۰/۷	۱۴۶/۷	۱۲۵/۴

$$G=\sqrt[4]{\prod_{i=1}^{4} x_i}=\sqrt[4]{(136.1)(130.7)(146.7)(125\ .4)}=134.497$$

یعنی رشد متوسط برابر با ۱۳۴/۵ است.

می توان نشان داد که رشد متوسط اخیر از رابطه زیر نیز قابل محاسبه است.

$$G=\sqrt[n-1]{\frac{x_n}{x_1}}\times 100=\sqrt[4]{\frac{x_5}{x_1}}\times 100=134.497$$

یادآوری این نکته ضروری است که اگر حجم جامعه مورد بررسی متناهی باشد میانگین هندسی جامعه از فرمول زیر محاسبه می‌شود.

$$\sqrt[N]{\prod_{i=1}^{N} X_i}$$

۱-۲-۷ میانگین هارمونیک

اگر x_1، x_2، ...، x_n یک نمونه به حجم n از جامعه مورد بررسی باشد میانگین هارمونیک از رابطه زیر بدست می آید و با علامت H نمایش داده می شود.

$$\frac{1}{H}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} \qquad \text{یا} \qquad H=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

مثال ۱-۲-۸ در یک کارگاه تراشکاری یک قطعه خاص به وسیله سه رایانه در زمانهای $\frac{1}{4}$، $\frac{1}{3}$ و $\frac{1}{2}$ ساعت تراش داده می شود. میانگین هارمونیک را محاسبه کنید.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{3}{(4+3+2)} = \frac{1}{3}$$

۱-۲-۹ میانگین پیراسته

میانگین‌پیراسته حالت خاصی از میانگین‌حسابی است به طوری که تعدادی از مشاهدات به علت ناهماهنگ بودن، از داده‌ها حذف می‌شود و میانگین‌حسابی برای داده‌های باقی مانده محاسبه می‌شود. اگر k تا از مشاهدات حذف شده باشند میانگین‌پیراسته از رابطه زیر بدست می آید $(k < n)$.

$$\bar{x}_p = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i$$

مثال ۱-۲-۱۰ حقوق ماهانه ۱۰ نفر از کارمندان موسسه آموزشی در جدول زیر آمده است. میانگین حسابی و میانگین پیراسته را محاسبه نمایید.

کارمند	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
حقوق	۲۵	۳۵۰	۴۵۰	۵۰۰	۳۰۰	۶۰۰	۲۵۰	۳۲۰	۴۰۰	۲۵۰۰

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{10} \sum_{i=1}^{10} x_i = 569.5$$

متوسط حقوق این ده کارمند برابر با ۵۶۹/۵ است.

چون حقوق کارمند اول و کارمند دهم در مقایسه با بقیه غیرقابل تصور است لذا آنها را از داده‌ها حذف می‌کنیم و میانگین‌پیراسته را حساب می‌کنیم.

کارمند	۲	۳	۴	۵	۶	۷	۸	۹
حقوق	۳۵۰	۴۵۰	۵۰۰	۳۰۰	۶۰۰	۲۵۰	۳۲۰	۴۰۰

$$n = ۱۰ \quad , \quad k = ۲ \quad , \quad \bar{x}_p = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i = ۳۹۶/۲۵$$

۱-۲-۱۱ میانه

میانه یک مجموعه مورد بررسی، عبارتست از مقداری که مجموعه را به دو بخش مساوی تقسیم می‌کند به طوری که تعداد مقادیر مساوی یا بزرگتر از آن با تعداد مقادیر کوچکتر یا مساوی آن برابر باشد. برای محاسبه میانه ابتدا مشاهدات را به ترتیب صعودی مرتب می کنیم و دو حالت زیر را درنظر می‌گیریم:

الف- اگر حجم مشاهدات فرد باشد عدد میانی، میانه مشاهدات خواهد بود.

ب- اگر حجم مشاهدات زوج باشد میانه، میانگین دو عدد میانی خواهد بود.

۱-۲-۱۲ ویژگیهای میانه

الف- میانه مشاهدات را به دو بخش مساوی تقسیم می کند.

ب- منحصر به فرد است.

ج- تحت تأثیر داده های پرت قرار نمی‌گیرد.

د- محاسبه آن ساده است.

مثال ۱-۲-۱۳ میانه داده‌های زیر را حساب کنید.

۶/۸	۳/۸	۸/۳	۹/۲	۳/۳	۱۸/۰	۱۰/۷	۱۱/۳	۱۱/۶

ابتدا داده‌ها را مرتب می کنیم.

۱۸/۰	۱۱/۶	۱۱/۳	۱۰/۷	۹/۲	۸/۳	۶/۸	۳/۸	۳/۳

$n = ۹$ و ۹/۲ = میانه

مثال ۱-۲-۱۴ حقوق ۸ کارمند عبارتست از:

۱۵۰۰	۳۰۰	۲۹۰	۲۷۵	۲۶۰	۲۴۰	۲۲۵	۱۵۰

چون تعداد داده‌ها زوج است. میانه برابر است با

$$\frac{260+275}{2}=267.5$$

علیرغم اینکه ۱۵۰۰ یک مشاهده پرت است میانه تحت تأثیر آن قرار نگرفته است. در صورتی که میانگین داده‌ها برابر با ۴۰۵ است که هشت نفر حقوقشان از آن کمتر است.

۱-۲-۱۵ نما

نمای یک مجموعه عددی است که در آن مجموعه بیش از بقیه تکرار شده باشد. یک مجموعه‌ای از مشاهدات ممکن است یک یا چند نما داشته باشد و ممکن است فاقد نما باشد. در سه سری داده‌های زیر، سری اول فاقد نما، سری دوم دارای یک نما ۱۵ و سری سوم دارای دو نمای ۶۵ و ۶۷ است.

-۲	۳	۴	۷	۹	۱۱	۱۳	۱۷		
۱۲	۱۵	۱۸	۱۶	۱۵	۱۴	۱۶	۱۷	۱۵	
۵۶	۶۲	۶۳	۶۵	۶۵	۶۵	۶۷	۶۷	۶۷	۷۰

۱-۲-۱۶ چارکها

چارکهای یک مجموعه مورد بررسی عبارتست از کمیت‌ها یا مقادیری که مجموعه را به چهار قسمت مساوی تقسیم می‌کنند. محاسبه چارکها همانند میانه می‌باشد. اگر x_1، x_2، ...، x_n مجموعه مشاهدات باشند برای محاسبه چارکها به طریق زیر اقدام می‌کنیم.

ابتدا مشاهدات را مرتب می‌کنیم و میانه یا چارک دوم (Q_2) را محاسبه می‌کنیم. از آنجا که چارک دوم مشاهدات را به دو بخش مساوی تقسیم می‌کند. میانه را برای بخش سمت چپ و بخش سمت راست چارک دوم محاسبه می کنیم و اعداد محاسبه شده را به ترتیب چارک‌اول (Q_1) و چارک‌سوم (Q_3) می نامیم. بنابراین Q_1 کمیتی است که ٪۲۵ از مشاهدات کمتر یا مساوی آن است و Q_2 کمیتی است که ٪۵۰ از مشاهدات کمتر یا مساوی آن است و Q_3 کمیتی است که ٪۷۵ از مشاهدات کمتر یا مساوی آن است. محاسبه چارکها به این طریق وقتی از کارایی بالایی برخوردار است که حجم مشاهدات بیشتر از ۲۵ باشد.

مثال ۱-۲-۱۷ زمان عمل جراحی در یک بیمارستان برای ۱۱ مریض برحسب دقیقه به صورت زیر است. میانه و چارکها را حساب کنید.

۳۱	۴۰	۴۵	۳۶	۳۰	۲۹	۴۱	۳۹	۳۳	۳۰	۳۵

مرتب شده مشاهدات.

۴۵	۴۱	۴۰	۳۹	۳۶	۳۵	۳۳	۳۱	۳۰	۳۰	۲۹

$Q_2 =$ میانه $= ۳۵$ و $n = ۱۱$

میانه یا چارک دوم داده ها را به دو بخش زیر تقسیم کرده .

بخش سمت چپ میانه :	۳۳	۳۱	۳۰	۳۰	۲۹
بخش سمت راست میانه:	۴۵	۴۱	۴۰	۳۹	۳۶

لذا، $Q_1 = ۳۰$، $Q_3 = ۴۰$. یادآوری می‌شود که گاهی ممکن است میانه را به دو بخش بدست آمده اضافه کنیم و سپس چارک اول و سوم را محاسبه کنیم. مقادیر بدست آمده با مقادیر قبلی فرق خواهد کرد. برای مثال اخیر، با رعایت این نکته، $Q_1 = ۳۰/۵$ و $Q_3 = ۳۹/۵$ است.

۱-۳ شاخص‌های پراکندگی

پراکندگی یک مجموعه از مشاهدات به تنوعی که مقادیر مشاهدات از خود نمایش می‌دهند بر می‌گردد. اگر $x_1, x_2, \ldots, x_n$ مشاهدات به حجم n باشند معیارهای پراکندگی دامنه، واریانس، ضریب تغییر، متغیرهای استاندارد، انحراف چارکی، گشتاورها، چولگی و برجستگی را در این بخش مورد بحث قرار می دهیم.

۱-۳-۱ دامنه

برای محاسبه دامنه ابتدا داده‌ها را مرتب می کنیم و کوچکترین مشاهده را به $\underline{x}$ و

بزرگترین مشاهده را به $x_{\max}$ نمایش می دهیم و دامنه را از رابطه زیر بدست می‌آوریم.

$$R = \text{دامنه} = x_{\max} - x_{\min}$$

در بعضی از منابع دامنه را به صورت زیر نیز تعریف می کنند.

$$R = \text{دامنه} = x_{\max} - x_{\min} + 1$$

دامنه، فقط تغییرات پراکندگی بین بزرگترین و کوچکترین را تعیین می‌کند و محاسبه آن ساده است. سری داده‌هایی که دارای دامنه کوچکتر هستند از همگنی بیشتری برخوردار هستند. دامنه مشاهدات زیر برابر با ۹ = ۳۶ − ۴۵ است.

۳۶ ۳۹ ۴۰ ۴۱ ۴۵

۱-۳-۲ واریانس (پراش)

اگر $x_1, x_2, \ldots, x_n$ یک نمونه به حجم n و دارای میانگین برابر با $\bar{x}$ باشند واریانس از رابطه زیر محاسبه می‌شود و آن را در نمونه با S^2 نمایش می‌دهند.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

مثال ۱-۳-۳ واریانس مشاهدات زیر را بدست آورید.

۷۲ ۶۸ ۶۵ ۶۳ ۶۴ ۶۹ ۶۷ ۶۶ ۶۹

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{9}\sum_{i=1}^{9} x_i = 67$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{8}\sum_{i=1}^{9}(x_i - \bar{x})^2$$

$$= \frac{1}{8}\Big[(69-67)^2 + (66-67)^2 + (67-67)^2 + (69-67)^2 + (64-67)^2$$
$$+ (63-67)^2 + (65-67)^2 + (68-67)^2 + (72-67)^2\Big]$$
$$= \frac{1}{8}(4+1+0+4+9+16+4+1+25) = 8$$

در بعضی از منابع S^2 را به صورت $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ تعریف می‌کنند که در عمل اگر n بزرگ باشد دو تعریف واریانس یکسان است. اما برای اثبات نااریبی تحت شرایطی از تعریف اول استفاده می کنند. در ادامه کتاب از $S^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$ استفاده خواهیم کرد.

۱-۳-۴ ویژگیهای واریانس نمونه

۱- واریانس عدد ثابت c برابر با صفر است.

۲- اگرمقدار ثابت a رابه مشاهدات اضافه یا ازآنها کم کنیم واریانس تغییر نمی‌کند.

۳- اگر مشاهدات در مقدار ثابت k ضرب یا برآن تقسیم شود واریانس جدید از ضرب یا تقسیم واریانس قدیم در k^2 بدست می‌آید.

مثال ۱-۳-۵ مقدار فروش چهار فروشنده به صورت زیر داده شده است. میانگین و واریانس فروش را بدست آورید.

فروشنده	۱	۲	۳	۴
مقدار فروش (ریال)	۷/۵	۱۲/۵	۱۳	۱۷

$$n=4\ , \qquad \bar{x}=\frac{1}{4}\sum_{i=1}^{4}x_i \quad , \qquad S^2=\frac{1}{4}\sum_{i=1}^{4}(x_i-\bar{x})^2$$

فروشنده	فروش	$x_i-\bar{x}$	$(x_i-x)^2$
۱	۷/۵	−۵	۲۵
۲	۱۲/۵	۰	۰
۳	۱۳	۰/۵	۰/۲۵
۴	۱۷	۴/۵	۲۰/۲۵
جمع	۵۰	۰	۴۵/۵

باتوجه به جمع ستون دوم و چهارم

$$\bar{x} = \frac{1}{4}(50) = 12.5$$

$$S^2 = \frac{1}{4}(45/5) = 11.375$$

۱-۳-۶ انحراف معیار

انحراف معیار در نمونه جذر واریانس یا پراش می‌باشد.

$$\text{انحراف معیار} = \sqrt{S^2} = S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

انحراف معیار مثال اخیر برابر با

$$S = \sqrt{11.375} = 3.373$$

یادآوری ۱-۳-۷ میانگین و واریانس جامعه را به ترتیب به μ و σ^2 نمایش می‌دهند که σ^2 سیگما دو خوانده می‌شود و مقدارِی است ثابت (پارامتر).
اگر جامعه آماری مورد بررسی متناهی و دارای Nعضو باشد. σ^2 از از رابطه زیر بدست می‌آید.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

، که $\bar{X}$ یا μ میانگین جامعه است، و جذر σ^2 انحراف معیار جامعه خوانده می‌شود.

۱-۳-۸ متغیرهای استاندارد

اگر $x_1, x_2, \dots, x_n$ یک نمونه به حجم n از جامعه مورد بررسی باشند باتوجه به فرمولهای میانگین و واریانس نمونه، میانگین و واریانس نمونه قابل محاسبه است. با توجه به طبیعت جامعه، نمونه های $x_1, x_2, \dots, x_n$ واحد اندازه‌گیری خودشان را حفظ می‌کنند. مثلاً اگر هدف بررسی میزان شنوایی در جامعه آماری باشد هر مشاهده x_i واحد میزان شنوایی خود را به همراه دارد. برای حذف واحد اندازه‌گیری از متغیرهای x_i متغیرهای جدیدی به صورت زیر تعریف می شود.

$$z_i = \frac{x_i - \bar{x}}{S} \quad , \qquad i = ۱,۲,\ldots,n$$

که Z_iها فاقد واحد اندازه‌گیری‌اند و متغیرهای استاندارد خوانده می‌شوند.

۱-۳-۹ ویژگیهای متغیرهای استاندارد

۱- میانگین متغیرهای استاندارد برابر صفر است.

$$z_i = \frac{x_i - \bar{x}}{S} \quad , \qquad i = ۱,۲,\ldots,n$$

با جمع بستن روی طرفین رابطه اخیر داریم:

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S} = \frac{1}{S}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{1}{S}(\sum x_i - n\bar{x}) = \frac{1}{S}(n\bar{x} - n\bar{x}) = 0$$

۲- واریانس متغیرهای استاندارد برابر با ۱ است.

۳- متغیرهای استاندارد فاقد واحد اندازه‌گیری هستند.

۴- مقدار Z_i می تواند، منفی، صفر یا مثبت باشد.

مثال ۱-۳-۱۰ داده‌های زیر را به داده‌های استاندارد تبدیل کنید.

۵۱ ۵۳ ۵۴ ۵۸

$$n = 4 \ , \ \bar{x} = \frac{1}{4}\sum_{i=1}^{4} x_i = 54 \quad , \quad S^2 = \frac{1}{4}\sum (x_i - 54)^2 = 6.5$$

$$z_i = \frac{x_i - \bar{x}}{S} = \frac{x_i - 54}{\sqrt{6.5}} \ , \ i = 1,\ 2,\ 3,\ 4$$

$Z_۱$	$Z_۲$	$Z_۳$	$Z_۴$
-۱/۱۷۶۶۹۶	-۰/۳۹۲۲۳۳	۰	۱/۵۶۸۹۲۹

مثال کاربردی ۱-۳-۱۱ نرخ سود سپرده ۱۰۰ حساب بانک ملت دارای میانگین ۳/۵ و انحراف معیار ۲/۵ است. سود سپرده دو حساب به ترتیب برابر با ۴/۵ و ۲/۹ است. کدام حساب دارای سود سپرده خوب بوده است؟

داریم، $\bar{x} = ۳/۵$ و $S = ۲/۵$ ، $x_1 = ۴/۵$ ، $x_2 = ۲/۹$

$$z_1 = \frac{4.5-3.5}{2.5} = 0.4 \quad , \quad z_2 = \frac{2.9-2.5}{2.5} = 0.16$$

چون $z_1 > z_2$ است. پس حساب با سود سپرده ۴/۵ بهتر است.

۱-۳-۱۲ ضریب‌تغییر یا ضریب تعیین

ضریب‌تغییر یا ضریب پراکندگی یک مجموعه مشاهدات برابر است با نسبت انحراف معیار مشاهدات بر میانگین مشاهدات و آنرا معمولاً با $C.V$ نشان می‌دهند که برابر است با:

$$C.V = \frac{S}{\bar{x}} = \frac{\sqrt{\frac{1}{n}\sum(x_i-\bar{x})^2}}{\frac{1}{n}\sum_{i=1}^{n}x_i} = \frac{\sqrt{n\sum(x_i-\bar{x})^2}}{\sum x_i}$$

۱-۳-۱۳ ویژگیهای ضریب تغییر

۱- به واحد اندازه‌گیری بستگی ندارد.

۲- برای مقایسه دو صفت از یک جامعه با واحدهای اندازه‌گیری متفاوت مورد استفاده قرار می گیرد.

۳- مجموعه مشاهداتی که دارای $C.V$ کمتری است از سازگاری و همگنی بیشتری برخوردار هستند.

یادآوری: اگر σ^2 واریانس و μ میانگین جامعه باشند ضریب تغییر جامعه از رابطه زیر حساب می شود.

$$C.V = \frac{\sigma}{\mu}$$

مثال ۱-۳-۱۴ داده‌های زیر هزینه غذایی ده خانوار در مناطق مختلف می‌باشد. ضریب تغییر را حساب کنید.

۱۱۳۰ ۱۲۱۰ ۱۳۱۰ ۱۰۰۰ ۱۰۸۰ ۱۱۷۰ ۱۲۰۰ ۱۲۶۰ ۱۲۷۰ ۱۰۹۰

$n = ۱۰$ ، $\bar{x} = ۱۱۷۲$ ، $S = ۹۲/۲۸۲$ ، $C.V = ۰/۰۷۸۷$

معمولاً $C.V$ را به صورت درصد بیان می کنند. پس

$\%C.V = ۷/۸۷$

مثال کاربردی ۱-۳-۱۵ دو شخص در ۵ رقابت تیراندازی شرکت می‌کنند و تعداد شلیکهایی که در ۱۵ شلیک به هدف برخورد می‌کند درجدول زیر آورده‌شده‌است. کدام تیرانداز از مهارت بیشتری برخوردار است؟

شخص اول	۶	۱۲	۱۲	۱۰	۷	۴۷
شخص دوم	۱۲	۱۵	۷	۷	۴	۴۵

برای شخص اول $\bar{x}_۱ = ۹/۴$ ، $S_۱ = ۲/۴۹۷۹$ ، $C.V_۱ = ۲۶/۵۷۳$

برای شخص دوم $\bar{x}_۲ = ۹$ ، $S_۲ = ۳/۹۴۹۷$ ، $C.V_۲ = ۴۳/۸۸۶$

چون $C.V_۱ < C.V_۲$ است. شخص اول دارای مهارتی بیشتر است.

۱-۳-۱۶ انحراف چارکی

انحراف چارکی برای یک مجموعه مشاهدات به صورت زیر تعریف می‌شود.

$$Q.D = \frac{Q_3 - Q_1}{2}$$

که $Q_۱$ و $Q_۳$ به ترتیب چارک اول و سوم می‌باشند.

۱-۳-۱۷ ویژگیهای انحراف چارکی

۱- این شاخص چون میزان پراکندگی در اطراف مرکز توزیع را نشان می‌دهد از شاخص دامنه با ثبات‌تر است.

۲- این شاخص چون شامل ۲۵٪ از مشاهدات کوچک و بزرگ نیست تحت تأثیر داده‌های پرت قرار نمی‌گیرد.

۳- این شاخص برای داده‌های کلاس بندی نیز قابل محاسبه است.

مثال ۱-۳-۱۸ انحراف چارکی را برای داده‌های زیر بدست آورید.

۷/۵ ۱۲/۵ ۱۳ ۱۷

$$Q_1 = \frac{7.5+12.5}{2} = 10 \quad , \quad Q_3 = \frac{13+17}{2} = 15$$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{15-10}{2} = 2.5$$

۱-۳-۱۹ گشتاورها

اگر $x_1, x_2, \dots, x_n$ یک نمونه به حجم n از جامعه مورد بررسی و a یک عدد ثابت باشد گشتاورهای مرتبه rام حول نقطه a به صورت زیر تعریف می‌شود.

$$m'_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^r \quad , \quad r = 1, 2, \dots$$

که m'_r را گشتاورهای غیرمرکزی مرتبه rام نیز می‌نامند.

۱- برای $a = 0$، $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$

۲- برای $a = \bar{x}$، $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$ را گشتاور مرکزی حول میانگین نمونه می‌گویند و آن را با علامت m_r نمایش می‌دهند.

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r \quad , \quad r = 1, 2, \dots$$

۱-۳-۲۰ ویژگیهای گشتاورهای مرکزی

۱- برای $r=1$، $m_1 = 0$

۲- برای $r=2$، $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S^2$

۳- تغییر در مبدأ یا اضافه و کم کردن مقدار ثابت به مشاهدات تغییری در m_r ندارد.

۴- باتغییر در مقیاس یا ضرب و تقسیم کردن مقدار ثابت در مشاهدات، m_r در توان rام مقدار ثابت ضرب یا تقسیم می‌شود.

۵- $m_2 = m_2' - (\bar{x}-a)^2$

یادآوری ۱-۳-۲۱ گشتاورهای‌مرکزی وغیرمرکزی برای جامعه در فصل سوم تعریف می‌شود.

مثال ۱-۳-۲۲ دستمزدهای ۵ اپراتور رایانه عبارتند از:

۳۵۰ ۴۹۰ ۶۰۰ ۷۸۰ ۸۰۰

الف- گشتاورهای اول، دوم و سوم را حول نقطه $a=$ ۶۰۰ بدست آورید.

ب- گشتاورهای مرکزی اول، دوم و سوم را بدست آورید.

i	x_i	$(x_i - 600)$	$(x_i - 600)^2$	$(x_i - 600)^3$
۱	۳۵۰	-۲۵۰	۶۲۵۰۰	-۱۵۶۲۵۰۰۰
۲	۴۹۰	-۱۱۰	۱۲۱۰۰	-۱۳۳۱۰۰۰
۳	۶۰۰	۰	۰	۰
۴	۷۸۰	۱۸۰	۳۲۴۰۰	۵۸۳۲۰۰۰
۵	۸۰۰	۲۰۰	۴۰۰۰۰	۸۰۰۰۰۰۰
جمع	۳۰۲۰	۲۰	۱۴۷۰۰۰	-۳۱۲۴۰۰۰

$m_1' = \frac{20}{5} = 4 \quad , \quad m_2' = 29400 \quad , \quad m_3' = -624800$

ب- حول $a = \bar{x} =$ ۶۰۴

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
۱	۳۵۰	-۲۵۴	۶۴۵۱۶	-۱۶۳۸۷۰۶۴
۲	۴۹۰	-۱۱۴	۱۲۹۹۶	-۱۴۸۱۵۴۴
۳	۶۰۰	-۴	۱۶	-۶۴
۴	۷۸۰	۱۷۶	۳۰۹۷۶	۵۴۵۱۷۷۶
۵	۸۰۰	۱۹۶	۳۸۴۱۶	۷۵۲۹۵۳۶
جمع	۳۰۲۰	۰	۱۴۶۹۲۰	-۴۸۸۷۳۶۰

$m_1=0$, $m_2=$۲۹۳۸۴ , $m_3=-$۹۷۷۴۷۲

اگر توجه داشته باشید m_r ها نظیر به نظیر از m'_r کوچکتر هستند.

۴-۱ جدول توزیع فراوانی

معمولاً داده‌های خام جمع‌آوری شده توسط تحلیل‌گر برای بررسی یک فرایند به طور تصادفی شکل متقارن ندارد. برای تجزیه و تحلیل این نوع داده‌ها و نتیجه‌گیری مناسب در مورد یک جامعه یا جمعیت مورد مطالعه ضرورت دارد که داده‌های جمع‌آوری شده به چند زیرگروه (کلاس) تقسیم‌بندی شود و تجزیه و تحلیل صورت گیرد. گروه‌بندی یا توزیع فراوانی، جدولی است که مجموعه‌ای از داده‌های خام را به تعداد مناسب کلاس تقسیم می‌کند و همچنین تعداد اقلامی از داده‌ها را که در هر کلاس قرارمی‌گیرند نشان می‌دهد. تشکیل چنین جدولی برای داده‌ها باعث از بین رفتن بعضی از اطلاعات می‌شود.

گروه‌بندی یا کلاس‌بندی، اهمیت داده‌ها را روشن می‌سازد و یک گویایی به آنها می‌افزاید که معمولاً بیشتر از زیانهایی است که داده‌ها در برابر جدول‌بندی متحمل شده‌اند. در کلاس‌بندی داده‌ها، قدم اول در ساختن یک توزیع فراوانی تصمیم‌گیری روی تعداد کلاسها و تعیین حدود هر کلاس است، یعنی یک کلاس از چه مقداری شروع و به چه مقداری ختم شود. به طور کلی تعداد کلاسها بستگی به تعداد مشاهدات و دامنه داده‌ها دارد ولی به ندرت اتفاق می‌افتد که توزیعی با کمتر از پنج کلاس یا بیشتر از ۱۵ کلاس نیز سودبخش باشد.

مثال ۱-۴-۱ داده‌ها زیر سن ازدواج ۵۲ نفر را در ۸ منطقه نشان می دهد. توزیع فراوانی سن ازدواج را تشکیل دهید.

۲۶	۲۳	۲۴	۲۵	۲۴	۲۵	۲۴	۲۳	۲۲	۲۶	۲۷	۲۵	۲۴
۲۸	۲۷	۲۲	۲۵	۲۵	۲۴	۲۵	۲۵	۲۴	۲۳	۲۵	۲۴	۲۸
۲۴	۲۷	۲۴	۲۵	۲۷	۲۵	۲۶	۲۸	۲۵	۲۴	۲۵	۲۶	۲۷
۲۶	۲۸	۲۵	۲۶	۲۷	۲۵	۲۷	۲۶	۲۴	۲۵	۲۴	۲۵	۲۵

چون اعداد تکراری ۲۲، ۲۳، ۲۴، ۲۵، ۲۶، ۲۷ و ۲۸ است. تعداد کلاس برابر با ۷ درنظر گرفته می‌شود.

درصد فراوانی	فراوانی	سن در هنگام ازدواج	شماره کلاس
۳/۸۵	۲	۲۲	۱
۵/۷۷	۳	۲۳	۲
۲۳/۰۸	۱۲	۲۴	۳
۳۲/۶۹	۱۷	۲۵	۴
۱۳/۴۶	۷	۲۶	۵
۱۳/۴۶	۷	۲۷	۶
۷/۶۹	۴	۲۸	۷
۱۰۰	۵۲	-	جمع

ستون چهارم جدول اخیر درصد فراوانی نامیده می‌شود و مقدار آن برای هر کلاس از تقسیم فراوانی آن کلاس بر مجموع فراوانی ضربدر عدد ۱۰۰ بدست می‌آید. بعنوان نمونه، کلاس چهارم شامل ۱۷ نفر است که در سن ۲۵ سالگی ازدواج کرده‌اند و درصد نسبت آنها برابر با $۳۲/۶۹ = ۱۰۰ \times \frac{17}{52}$ است.

جمع ستون سوم و ستون چهارم همیشه ثابت است که جمع ستون سوم همان حجم نمونه (n) و جمع ستون چهارم همیشه برابر با ۱۰۰ است. چون جمع این ستونها ثابت است اگر یکی از آنها مجهول باشد از تفاضل بقیه از جمع کل بدست می‌آید.

اغلب تحلیل‌گران با داده‌های خام (گسسته یا پیوسته) مواجه می‌باشند لذا ضرورت دارد صورت کلی جدول توزیع فراوانی را به صورت زیر شرح دهیم.

شماره کلاس	کلاس	حد متوسط	فراوانی	درصد فراوانی	فراوانی تجمعی	درصد فراوانی تجمعی
۱	a_1-b_1	$\frac{a_1+b_1}{2}$	f_1	$\frac{f_1}{n}\times 100$	F_1	$\frac{F_1}{n}\times 100$
۲	a_2-b_2	$\frac{a_2+b_2}{2}$	f_2	$\frac{f_2}{n}\times 100$	F_2	$\frac{F_2}{n}\times 100$
$\vdots$						
i	a_i-b_i	$\frac{a_i+b_i}{2}$	f_i	$\frac{f_i}{n}\times 100$	F_i	$\frac{F_i}{n}\times 100$
$\vdots$						
k	a_k-b_k	$\frac{a_k+b_k}{2}$	f_k	$\frac{f_k}{n}\times 100$	$F_k=n$	100
جمع	-	-	n	۱۰۰	-	-

در این جدول توزیع فراوانی که k کلاس است نمادها به صورت زیر تعریف می‌شوند.

a_1: حد پایین کلاس اول ، a_i : حد پایین کلاس iام

b_1: حد بالای کلاس اول ، b_i : حد بالای کلاس iام

$\frac{a_i+b_i}{2}$: حد متوسط یا نماینده کلاس iام برای $i=۱,۲,\ldots,k$

f_i: فراوانی یا فراوانی مطلق کلاس iام برای $i=۱,۲,\ldots,k$

$\frac{f_i}{n}\times 100$: درصد فراوانی کلاس iام برای $i=۱,۲,\ldots,k$

$F_1=f_1$: فراوانی تجمعی کلاس اول

$F_2=f_1+f_2$: فراوانی تجمعی کلاس دوم

$F_i=f_1+f_2+\ldots+f_i$: فراوانی تجمعی کلاس iام

در حقیقت فراوانی تجمعی کلاس iام برابر با فراوانی آن کلاس به علاوه فراوانی کلاسهای ماقبل است. برای کلاس kام یا آخر $F_k=\sum_{i=1}^{k} f_i=n$ و $\frac{F_i}{n}\times 100$ درصد فراوانی تجمعی کلاس iام تعریف می‌شود.

برای دقت زیاد، معمولاً، کلاس‌های طبقه‌بندی شده را به صورت کرانه‌دار می‌نویسند. مثلاً $a_1^*-b_1^*$ ، که a_1^* کران پایین کلاس اول و b_1^* را کران بالای کلاس اول و در

حالت کلی a_i^* و b_i^* کران پایین و بالای کلاس iام می‌باشد. کران پایین و بالای iام از رابطه زیر بدست می‌آید.

$$a_i^* = \frac{a_i + b_{i-1}}{2} \qquad , \qquad b_i^* = \frac{a_{i+1} + b_i}{2}$$

به عنوان نمونه، کران پایین و بالای کلاس سوم عبارتند از:

$$a_3^* = \frac{\text{حد بالا کلاس دوم} + \text{حد پایین کلاس سوم}}{۲} = \frac{a_3 + b_2}{2}$$

$$b_3^* = \frac{\text{حد بالا کلاس سوم} + \text{حد پایین کلاس چهارم}}{۲} = \frac{a_4 + b_3}{2}$$

۱-۴-۲ طول کلاس

طول کلاس از تفاضل حد پایین از حد بالا بدست می‌آید. آن را با l نمایش می‌دهند.

$$l_i = b_i - a_i \qquad i = ۱,۲,\ldots, k$$

در جدول توزیع فراوانی بین دامنه، طول کلاس و تعداد کلاس رابطه زیر همواره برقرار است.

طول کلاس × تعداد کلاس = دامنه

$$R = k.l$$

در تجزیه و تحلیل داده‌های خام همواره R قابل محاسبه است. اگر یکی از متغیرهای k و l معلوم باشد، دیگری از رابطه $R = k.l$ قابل محاسبه است.
اگر k و l هردو مجهول باشند. تعداد طبقات با استفاده از فرمول زیر که به فرمول استرژ معروف است بدست می‌آید.

$$k = ۱+۳/۳۲\log_{10}^{(n)}$$

که n حجم نمونه است. برای $n = ۱۳۰$ ، $k \approx ۸$ است.

مثال ۱-۴-۳ وزن ۳۰ کودک هنگام تولد برحسب کیلوگرم به صورت زیر ثبت شده است. جدول توزیع فراوانی را تشکیل دهید.

۲/۰ ۲/۱ ۲/۳ ۳/۰ ۳/۱ ۲/۷ ۲/۸ ۳/۵ ۳/۱ ۳/۷
۴/۰ ۲/۳ ۳/۵ ۴/۲ ۳/۷ ۳/۲ ۲/۷ ۲/۵ ۲/۷ ۳/۸
۳/۱ ۳/۰ ۲/۶ ۲/۸ ۲/۹ ۳/۵ ۴/۱ ۳/۹ ۲/۸ ۲/۲

ابتدا داده‌ها را به صورت زیر مرتب می‌کنیم.

۲/۰ ۲/۱ ۲/۲ ۲/۳ ۲/۳ ۲/۵ ۲/۶ ۲/۷ ۲/۷ ۲/۷
۲/۸ ۲/۸ ۲/۸ ۲/۹ ۳/۰ ۳/۰ ۳/۱ ۳/۱ ۳/۱ ۳/۲
۳/۵ ۳/۵ ۳/۵ ۳/۷ ۳/۷ ۳/۸ ۳/۹ ۴/۰ ۴/۱ ۴/۲

R = ۴/۲ − ۲/۰ = ۲/۲ و n = ۳۰

چون تعداد کلاسها و طول کلاسها داده نشده از رابطه استرژ استفاده می‌کنیم.

تعداد کلاس $= k =$ ۱ + ۳/۳۲ log (۳۰) = ۵/۹۰۴ ≈ ۶

فاصله کلاسها $= l = \frac{R}{k} = \frac{2.2}{6} = 0.37$

اگر مقدار l را به بالا گرد کنیم l = ۰/۴

کلاس کراندار	حد متوسط	درصد فراوانی تجمعی	فراوانی تجمعی	درصد فراوانی	فراوانی	کلاس	شماره کلاس
۲/۰-۲/۴	۲/۲	۱۷	۵	۱۷	۵	۲/۰-۲/۴	۱
۲/۴-۲/۸	۲/۶	۳۳	۱۰	۱۷	۵	۲/۴-۲/۸	۲
۲/۸-۳/۲	۳	۶۳	۱۹	۳۰	۹	۲/۸-۳/۲	۳
۳/۲-۳/۶	۳/۴	۷۷	۲۳	۱۳	۴	۳/۲-۳/۶	۴
۳/۶-۴/۰	۳/۸	۹۰	۲۷	۱۳	۴	۳/۶-۴/۰	۵
۴/۰-۴/۴	۴/۲	۱۰۰	۳۰	۱۰	۳	۴/۰-۴/۴	۶
-	-	-	-	۱۰۰	۳۰	-	جمع

مثال کاربردی ۱-۴-۴ در آزمایش هوش از ۵۰ دانش آموز که ۱۰۰ نمره داشته نتایج زیر بدست آمده است. داده‌های فوق را در ۸ کلاس رده بندی و جدول توزیع فراوانی را کامل کنید و به سئوالات زیر پاسخ دهید.

۸۱	۶۰	۳۲	۸۰	۹۴	۴۸	۷۲	۶۱	۶۶	۵۴	۳۹	۵۶	۶۸
۳۶	۴۹	۵۱	۷۶	۵۹	۶۰	۲۷	۵۴	۳۸	۷۰	۲۳	۲۹	۷۶
۴۳	۸۵	۵۵	۳۸	۶۶	۵۸	۴۹	۴۸	۴۳	۴۹	۶۴	۵۲	۵۷
	۴۴	۵۰	۵۷	۸۸	۷۲	۶۰	۶۵	۷۸	۷۱	۵۶	۳۹	

چون تعداد کلاسها داده شده با استفاده از رابطه $R = kl$ طول کلاسها برابر با

$$l = \frac{94-23}{8} = 8.875$$

اگر l را به بالا گرد کنیم، ۹ = l (یا۱+ ۲۳–۹۴= R درنظر بگیریم).

رده کراندار	حد متوسط	درصد فراوانی تجمعی	فراوانی تجمعی	درصد فراوانی	فراوانی	رده	شماره رده
۲۲/۵–۳۱/۵	۲۷	۶	۳	۶	۳	۲۳–۳۱	۱
۳۱/۵–۴۰/۵	۳۱	۱۸	۹	۱۲	۶	۳۲–۴۰	۲
۴۰/۵–۴۹/۵	۴۵	۳۴	۱۷	۱۶	۸	۴۱–۴۹	۳
۴۹/۵–۵۸/۵	۵۴	۵۶	۲۸	۲۲	۱۱	۵۰–۵۸	۴
۵۸/۵–۶۷/۵	۶۳	۷۴	۳۷	۱۸	۹	۵۹–۶۷	۵
۶۷/۵–۷۶/۵	۷۲	۸۸	۴۴	۱۴	۷	۶۸–۷۶	۶
۷۶/۵–۸۵/۵	۸۱	۹۶	۴۸	۸	۴	۷۷–۸۵	۷
۸۵/۵–۹۴/۵	۹۰	۱۰۰	۵۰	۴	۲	۸۶–۹۴	۸
–	–	–	–	۱۰۰	۵۰	–	جمع

الف- بیشترین و کمترین تعداد دانش‌آموزان از نظر هوشی در کدام رده قرار دارند؟

برای جواب این سئوال باتوجه به ستون چهارم جدول، بیشترین و کمترین فراوانی متعلق به رده چهارم و هشتم است.

ب- چند درصد از دانش‌آموزان نمره هوشی کمتر از ۵۹ دارند؟

برای جواب این سئوال از فراوانی تجمعی یا درصد فراوانی تجمعی استفاده می‌کنیم. باتوجه به ستونهای پنجم و ششم ۲۸ نفر یا ۵۶ درصد از دانش آموزان نمره‌هایشان کمتر از ۵۹ می باشد.

ج- چند درصد از دانش‌آموزان نمره هوشی بیشتر یا برابر ۵۹ دارند؟
چون ۵۶ درصد کمتر از ۵۹ است پس ۴۴ = ۵۶-۱۰۰ درصد نمره‌هایشان بیشتر یا برابر با ۵۹ است.

د- چند درصد دانش آموزان دارای نمره هوشی بیشتر یا مساوی ۵۰ و کمتر یا مساوی ۸۵ است؟

با توجه به ستون چهارم درصد فراوانی کلاسها چهارم، پنجم، ششم و هفتم را باهم جمع می‌کنیم.
۲۲+۱۸+۱۴+۸ = ۶۲

داده‌های دو مثال اخیر از نوع کمی پیوسته بودند. گاهی اوقات تحلیلگر با داده‌های گسسته یا کیفی سروکار دارد که با یک مثال نحوه تشکیل جدول توزیع فراوانی را توضیح می‌دهیم. در این نوع داده ها هر عدد یا گروه معرف همان صورتی است که نمایش داده شده است. حد پایین، حد بالا و حد متوسط کلاسها باهم برابرند.

مثال ۱-۴-۵ در آزمایش گروه خون که از یک کلاس ۳۰ نفری به عمل آمده و نتایج به صورت زیر ثبت شده است، جدول توزیع فراوانی را تشکیل دهید.

A	B	A	AB	O	A	A	B	O	AB	B	B	A	O	AB
A	A	B	A	O	A	A	AB	O	B	A	AB	O	A	B

شماره کلاس	کلاس	فراوانی	درصد فراوانی	فراوانی تجمعی	درصد فراوانی تجمعی
۱	A	۱۲	۴۰	۱۲	۴۰
۲	B	۷	۲۳	۱۹	۶۳
۳	AB	۵	۱۷	۲۴	۸۰
۴	O	۶	۲۰	۳۰	۱۰۰
جمع	-	۳۰	۱۰۰	-	-

مثال ۱-۴-۶ از ۲۰ نفر در مورد ارائه خدمات سازمانی به صورت عالی، خوب، متوسط و بد نظرخواهی شده و نتایج به صورت زیر ثبت شده است. جدول توزیع فراوانی را تشکیل دهید.

عالی، خوب، متوسط، بد، عالی، متوسط، خوب، عالی، بد، خوب، خوب، عالی، بد، خوب، خوب، بد، عالی، متوسط، خوب، عالی.

کلاس	عالی	خوب	متوسط	بد
فراوانی	۶	۷	۳	۴
درصد فراوانی	۳۰	۳۵	۱۵	۲۰

در بخشهای قبل فرمول محاسبه میانگین، واریانس، میانه، نما، چارکها و گشتاورها برای داده‌های بدون فراوانی ارائه و با چند مثال با نحوه محاسبه آنها آشنا شدیم. در این بخش، فرمول و نحوه محاسبه این شاخصها در جدول توزیع فراوانی ارائه می‌شود.
در حالت کلی فرض کنید $x_1, x_2, \ldots, x_k$ مشاهدات با فراوانیهای متناظر $f_1, f_2, \ldots, f_k$ باشند. یعنی مشاهده x_1 ، f_1 بار تکرار شده و در حالت کلی مشاهده x_i ، f_i بار تکرار شده است. حجم مشاهدات برابر با جمع فراوانی‌ها است. یعنی:

$$n=\sum_{i=1}^{k} f_i$$

۱-۴-۷ محاسبه میانگین و واریانس در جدول توزیع فراوانی

برای داده‌ها با فراوانی میانگین و واریانس به صورت زیر تعریف می‌شود.

میانگین حسابی
$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k} f_i x_i$$

میانگین هندسی
$$G=\sqrt[n]{x_1^{f_1}.x_2^{f_2}.....x_k^{f_k}}$$

میانگین هارمونیک
$$H=\frac{n}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

واریانس
$$S^2=\frac{1}{n}\sum_{i=1}^{k} f_i(x_i-\bar{x})^2$$

مثال ۱-۴-۸ دستمزد ماهیانه کارمندان یک واحد صنعتی در جدول زیر داده شده است. میانگین‌ها و واریانس را حساب کنید.

دستمزد	۳۵۰	۴۹۰	۶۰۰	۷۸۰	۸۰۰	۱۰۰۰
تعداد افراد	۴	۵	۷	۸	۴	۲

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k} f_i x_i=\frac{1}{30}\sum_{i=1}^{6} f_i x_i=\frac{19490}{30}=649.67$$

$$G=\sqrt[30]{350^4\times 490^5\times 600^7\times 780^8\times 800^4\times 1000^2}=622.46$$

$$H=\frac{30}{\sum_{i=1}^{6}\frac{f_i}{x_i}}=593.40$$

$$S^2=\frac{1}{30}\sum_{i=1}^{6} f_i(x_i-\bar{x})^2=32523.22$$

محاسبه میانگین و واریانس در جدول توزیع فراوانی، همانند محاسبه میانگین و واریانس برای داده‌های دارای فراوانی است که x_i در داده‌های دارای فراوانی همان حد متوسط در جدول توزیع فراوانی است.

مثال ۱-۴-۹ توزیع نمرات دانشجویان رشته مدیریت در درس آمار به صورت جدول زیر است. میانگین و واریانس آن را به دست آورید.

کلاس	۲۰-۳۰	۳۰-۴۰	۴۰-۵۰	۵۰-۶۰	۶۰-۷۰	۷۰-۸۰	۸۰-۱۰۰
فراوانی	۲	۵	۲۲	۳۴	۹	۳	۱

i	کلاس	فراوانی f_i	حد متوسط x_i	$f_i x_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
۱	۲۰-۳۰	۲	۲۵	۵۰	-۲۷/۴۳	۷۵۲/۴۰۴۹	۱۵۰۴/۸۰۹۸
۲	۳۰-۴۰	۵	۳۵	۱۷۵	-۱۷/۴۳	۳۰۳/۸۰۴۹	۱۵۱۹/۰۲۴۵
۳	۴۰-۵۰	۲۲	۴۵	۹۹۰	-۷/۴۳	۵۵/۲۰۴۹	۱۲۱۴/۵۰۷۸
۴	۵۰-۶۰	۳۴	۵۵	۱۸۷۰	۲/۵۷	۶/۶۰۴۹	۲۲۴/۵۶۶۶
۵	۶۰-۷۰	۹	۶۵	۵۸۵	۱۲/۵۷	۱۵۸/۰۰۴۹	۱۴۲۲/۰۴۴۱
۶	۷۰-۸۰	۳	۷۵	۲۲۵	۲۲/۵۷	۵۰۹/۴۰۴۹	۱۵۲۸/۲۱۴۷
۷	۸۰-۱۰۰	۱	۹۰	۹۰	۳۷/۵۷	۱۴۱۱/۵۰۴۹	۱۴۱۱/۵۰۴۹
کل	-	۷۶	-	۳۹۸۵	-	-	۸۸۲۴/۶۷۲۴

$$n=\sum_{i=1}^{7} f_i = 76 \quad , \quad \bar{x}=\frac{1}{n}\sum_{i=1}^{7} f_i x_i = \frac{1}{7}(3985)=52.43$$

$$S^2=\frac{1}{76}\sum_{i=1}^{7} f_i (x_i-\bar{x})^2 = \frac{8824.6724}{76}=116.11$$

به طور مشابه، گشتاورهای غیرمرکزی و مرکزی به صورت زیر تعریف می‌شوند:

$$m'_r=\frac{1}{n}\sum_{i=1}^{k} f_i (x_i-a)^r \quad , \quad r=۱,۲,\ldots$$

$$m_r=\frac{1}{n}\sum_{i=1}^{k} f_i (x_i-\bar{x})^r \quad , \quad r=۱,۲,\ldots$$

۱-۴-۱۰ محاسبه نما در جدول توزیع فراوانی

نمای یک مجموعه از مشاهدات مقدار یا مقادیری بود که بیشتر از بقیه تکرار شده بود. ولی در جدول توزیع فراوانی برای داده‌های پیوسته نما از فرمول زیر محاسبه می‌شود.

$$M=a_i+\frac{d_1}{d_1+d_2}\times l$$

که:

a_i = حد پایین کلاس که نما در آن قرار دارد.

l = طول کلاس که نما درآن قرار دارد.

اگر i شماره کلاسی که نما درآن قرار دارد، باشد؛ $d_۱$ و $d_۲$ به صورت زیر محاسبه می‌شود.

$$d_1=f_i-f_{i-1} \quad , \quad d_2=f_i-f_{i+1}$$

یافتن کلاسی که نما در آن قرار دارد نقش مهمی در محاسبه نما دارد. در یک جدول توزیع فراوانی نما در کلاسی است که دارای بیشترین فراوانی است.

مثال ۱-۴-۱۱ برای جدول توزیع فراوانی زیر نما را بدست آورید.

شماره کلاس	کلاس	فراوانی	فراوانی تجمعی
۱	۲/۰-۲/۴	۵	۵
۲	۲/۴-۲/۸	۵	۱۰
۳	۲/۸-۳/۲	۹	۱۹
۴	۳/۲-۳/۶	۴	۲۳
۵	۳/۶-۴/۰	۴	۲۷
۶	۴/۰-۴/۴	۳	۳۰

چون کلاس سوم بیشترین فراوانی را دارد. پس نما در این کلاس قرار دارد و i=۳ است و نما از رابطه زیر بدست می‌آید.

$$M = a_3 + \frac{d_1}{d_1 + d_2} \times l$$

$$l = 3.2 - 2.8 = 0.4 \quad , \quad d_1 = f_3 - f_2 = 4 \quad , \quad d_2 = f_3 - f_4 = 5$$

$$M = 2.8 + \frac{4}{4+5}(0.4) = 2.977$$

۱-۴-۱۲ محاسبه میانه در جدول توزیع فراوانی

در جدول توزیع فراوانی میانه از فرمول زیر محاسبه می‌شود.

$$m = a_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \times l$$

که f_i فراوانی کلاسی که میانه در آن قرار دارد و F_{i-1} فراوانی تجمعی کلاس ما قبل کلاسی که میانه در آن قرار دارد است. برای یافتن کلاسی که میانه در آن قرار دارد اقدامات زیر را انجام می‌دهیم.

۱- $\frac{n}{2}$ را محاسبه می کنیم.

۲- در ستون فراوانی تجمعی از بالا به پایین اعداد را با $\frac{n}{2}$ مقایسه می‌کنیم. اولین عددی که بزرگتر یا مساوی $\frac{n}{2}$ باشد کلاس متناظر کلاسی است که میانه درآن قرار دارد. برای مثال قبل داریم:

$$\frac{n}{2}=\frac{30}{2}=15 \qquad , \qquad 15 \le 19 \Rightarrow i=3$$

$$f_i=f_3=9 \qquad , \qquad F_{i-1}=F_2=10$$

$$m=2.8+\frac{15-10}{9}(0.4)=3.022$$

۱-۴-۱۳ محاسبه چارک‌ها در جدول توزیع فراوانی

چارک jام برای $j=1,2,3$ از فرمول زیر محاسبه می شود.

$$Q_j=a_i+\frac{\frac{j\times n}{4}-F_{i-1}}{f_i}\times l$$

که i شماره کلاسی که چارک jام در آن قرار دارد. یافتن کلاسی که چارک دوم در آن قرار دارد همانند میانه است. اما برای یافتن کلاسی که چارک اول و سوم در آن قرار دارد به ترتیب $\frac{n}{4}$ و $\frac{3n}{4}$ را با ستون فراوانی تجمعی مقایسه می‌کنیم و کلاس مورد نظر را پیدا می‌کنیم.

برای مثال ۱-۴-۱۱، چارک اول، دوم و سوم به شرح زیر محاسبه می‌شود.

چارک اول:

$$\frac{n}{4}=\frac{30}{4}=7.5$$

$$7.5 \le 10 \Rightarrow i=2$$

$$Q_1=2.4+\frac{7.5-5}{5}(0.4)=2.6$$

چارک دوم: همان میانه

چارک سوم:

$$\frac{3n}{4}=22.5 \quad , \qquad i=4$$

$$Q_3=3.2+\frac{22.5-19}{4}(0.4)=3.55$$

مثال کاربردی ۱-۴-۱۴ جدول زیر ۷۹۹ نفر را که از نظر سن گروه‌بندی شده‌اند نشان می‌دهد. میانگین، نما، میانه و چارک اول و سوم را محاسبه کنید و به سئوالات زیر پاسخ دهید.

سن	۲۰-۲۵	۲۵-۳۰	۳۰-۳۵	۳۵-۴۰	۴۰-۴۵	۴۵-۵۰	۵۰-۵۵	۵۵-۶۰
نفر	۵۰	۷۰	۱۰۰	۱۸۰	۱۵۰	۱۲۰	۷۰	۵۹

شماره گروه	گروه سنی	فراوانی	فراوانی تجمعی	حد متوسط	طول کلاس
۱	۲۰-۲۵	۵۰	۵۰	۲۲/۵	۵
۲	۲۵-۳۰	۷۰	۱۲۰	۲۷/۵	۵
۳	۳۰-۳۵	۱۰۰	۲۲۰	۳۲/۵	۵
۴	۳۵-۴۰	۱۸۰	۴۰۰	۳۷/۵	۵
۵	۴۰-۴۵	۱۵۰	۵۵۰	۴۲/۵	۵
۶	۴۵-۵۰	۱۲۰	۶۷۰	۴۷/۵	۵
۷	۵۰-۵۵	۷۰	۷۴۰	۵۲/۵	۵
۸	۵۵-۶۰	۵۹	۷۹۹	۵۷/۵	۵
		۷۹۹			

$$\bar{x}=\frac{1}{n}\sum f_i x_i=\frac{32192.5}{799}=40.291$$

کلاسی که نما در آن قرار دارد، کلاس چهارم است.

$$M=a_4+\frac{d_1}{d_1+d_2}(l)=35+\frac{180-100}{\cancel{180-150}}(5)=\cancel{48.333}\ 38,6\overline{3}$$

$80+30=110$

کلاسی که میانه در آن قرار دارد، کلاس چهارم است.

$$m = a_4 + \frac{\frac{n}{2} - F_3}{f_4}(l) = 35 + \frac{\frac{799}{2} - 220}{180}(5) = 39.986$$

$$Q_1 = a_3 + \frac{\frac{n}{4} - F_2}{f_3} \times l = 30 + \frac{\frac{799}{4} - 120}{100} \times 5 = 33.988$$

$$Q_3 = a_4 + \frac{\frac{3n}{4} - F_5}{f_6} \times l = 45 + \frac{\frac{799 \times 3}{4} - 550}{120} \times 5 = 47.052$$

الف- چند درصد از افراد سنشان کمتر یا مساوی ۳۳/۹۸۸ است؟
چون ۳۳/۹۸۸ چارک اول است. پس ۲۵٪ از افراد سنشان کمتر یا مساوی ۳۳/۹۸۸ است.
ب- چند درصد از افراد سنشان کمتر یا مساوی ۳۹/۹۸۶ است؟ ۵۰٪
ج- چند درصد از افراد سنشان بیشتر از ۴۷/۰۵۲ است؟
چون ۴۷/۰۵۲ چارک سوم است. ۷۵٪ از افراد سنشان کمتر یا مساوی ۴۷/۰۵۲ یا ۲۵٪ از افراد سنشان بیشتر از ۴۷/۰۵۲ است.

۱-۵ نمودارها

اطلاعات لازم و ضروری به وسیله داده‌ها جمع‌آوری و آماده می‌شوند تا مسائلی از قبیل مسائل اجتماعی، اقتصادی، کشاورزی، مهندسی و غیره را حل کنند. لذا داده‌های جمع‌آوری شده باید تجزیه و تحلیل شوند و نتایج خلاصه شده ارائه شود. یکی از روشهای مؤثر نمودارهای هندسی است که باتوجه به مزایای زیر همواره مورد توجه تحلیل‌گران آماری است.

۱- نمودارها و گرافها با یک نگاه گذرا اطلاعات مورد نیاز را به سهولت و قابل فهم در اختیار تحلیل‌گر و عموم مردم قرار می‌دهد.

۲- اعداد و ارقام به صورت خام، سرد و خشک می‌باشند که با تبدیل به نمودار یا گراف روح جذابی به خود می‌گیرند و رضایت خاطر تحلیل‌گر را تأمین می‌کنند و به همین دلیل اکثر مجلات و نشریات ترجیحاً برای جلب نظر خواننده خود از نمودارها و گراف استفاده می‌کنند.

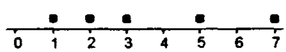
۳- گراف‌ها و نمودارها در مقابل توزیع فراوانی مدت زمان زیـادی در خـاطره و ذهن باقی می‌ماند.
۴- نمودارها و گرافها امکان سریع مقایسه داده‌ها را در زمانها و مکانهای مختلـف به تحلیل‌گر می‌دهند.

در این قسمت نمودارهای نقطه‌ای، دایره‌ای، میله‌ای، مـستطیلی، چندضـلعی فراوانـی و چندضلعی فراوانی تجمعی را مورد بحث قرار می‌دهیم.

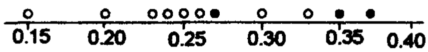
۱-۵-۱ نمودار نقطه‌ای

اگر x_1، x_2، ...، x_n مشاهدات به حجم n باشند برای رسم نمودار نقطه‌ای محـور xهـا را رسم می‌کنیم و مشاهدات را روی محور xها با نقطـه مـشخص مـی‌کنـیم. معمـولاً نمودار نقطه‌ای وقتی استفاده می‌شود که حجم نمونه کم باشد.

مثال ۱-۵-۲ سرعت ۵ نوع رایانه عبارتنـد از ۲، ۳، ۱، ۵، ۷ نمـودار نقطـه‌ای را رسـم کنید.



مثال کاربردی ۱-۵-۳ مقادیر مس در اجزای جوش خورده ماشین دریک دمای معـین دریک نمونه سه تایی عبارتست از ۰/۳۷، ۰/۳۵، ۰/۲۷ و در دمای دیگر در یـک نمونـه ۷تایی عبارتست از ۰/۲۳، ۰/۱۵، ۰/۲۵، ۰/۲۴، ۰/۳۰، ۰/۳۳ و ۰/۲۶ یک نمودار نقطـه‌ای رسم کنید که تفاوتهای ممکن در دو دمای مختلف را نشان دهد.



نمونه اول را با دایره‌های توپر و نمونه دوم را با دایره‌های توخالی نشان دادیم. یکسان بودن دو محصول غیرمتحمل به نظر می‌رسد.

۱-۵-۴ نمودار دایره‌ای

نمودار دایره‌ای عبارتست از تقسیم یک دایره به قسمتهای مختلف متناسب با گروه‌بندی توسط تحلیل‌گر آماری. در جدول توزیع فراوانی ناحیه هر گروه از رابطه زیر بدست می‌آید.

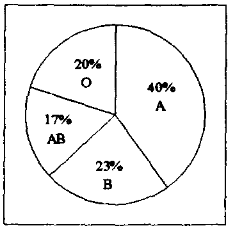
$$S_i = \frac{360}{\sum f_i} \times f_i \quad , \quad \sum_{i=1}^{k} S_i = 360$$

نمودار دایره‌ای برای انواع داده‌ها اعم از کیفی و کمی قابل استفاده است. اگر در فرمول ذکر شده به جای ۳۶۰، عدد ۱۰۰ قرار دهیم درصد فراوانی بدست می‌آید.

$$S_i = \frac{100}{\sum f_i} \times f_i$$

مثال ۱-۵-۵ در جدول توزیع فراوانی گروه خونی برای ۳۰ نفر داده شده است زاویه هرگروه و درصد از کل را حساب کنید.

گروه خونی	A	B	AB	O	کل
فراوانی	۱۲	۷	۵	۶	۳۰
درصدفراوانی	۴۰	۲۳	۱۷	۲۰	۱۰۰
s	۱۴۴	۸۴	۶۰	۷۲	۳۶۰

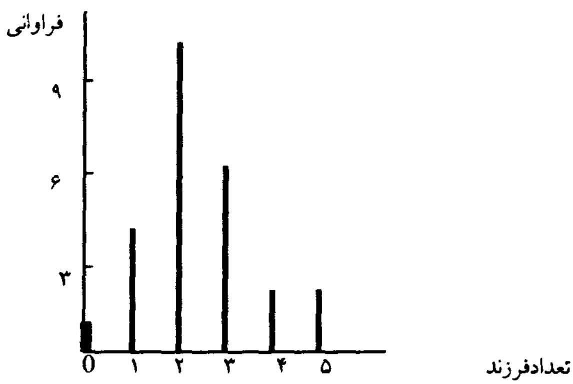


۱-۵-۶ نمودار میله‌ای

اگر مشاهدات یا داده‌ها کیفی یا کمی گسسته باشند برای نمایش هندسی آنها از نمودار میله‌ای استفاده می‌کنیم. مانند تعداد فرزندان خانواده‌ها، تعداد حوادث رانندگی، گروه خون، میزان علاقه و غیره. در جدول توزیع فراوانی با معرفی نماینده کلاسها می‌توان از نمودار میله‌ای استفاده کرد.

مثال ۱-۵-۷ تحقیقی که در مورد تعداد فرزندان ۲۵ خانواده به عمل آمده دارای توزیع فراوانی زیر است. نمودار میله‌ای را رسم کنید.

تعداد فرزندان	۰	۱	۲	۳	۴	۵
فراوانی	۱	۴	۱۰	۶	۲	۲



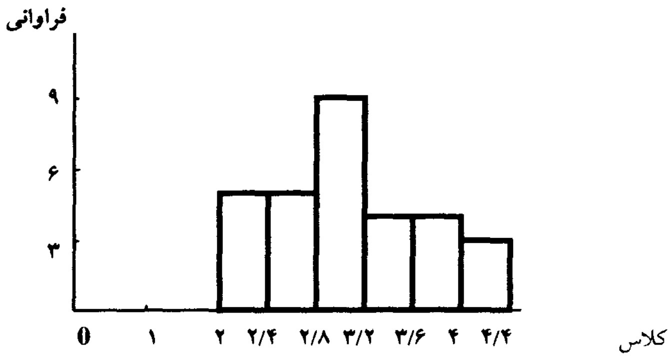
لازم به توضیح است که ما نمودار تعداد فرزند را در مقابل فراوانی رسم کردیم می‌توان تعداد فرزند در مقابل درصد فراوانی را نیز رسم کرد که نتیجه یکی خواهد بود.

۱-۵-۸ نمودار مستطیلی

عمومی‌ترین نمایش ترسیمی یک توزیع فراوانی نمودارمستطیلی است‌که از مستطیل‌های مجاور هم تشکیل گردیده است. ارتفاع آنها فراوانی کلاسها و قاعده آنها حدود کلاسها می‌باشد.

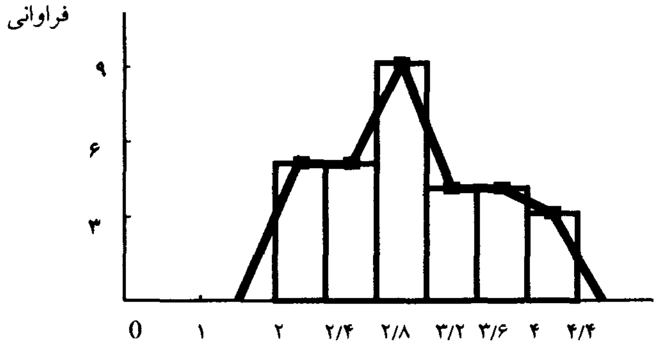
مثال ۱-۵-۹ برای جدول زیر نمودار مستطیلی را رسم کنید.

کلاس	۲/۰-۲/۴	۲/۴-۲/۸	۲/۸-۳/۲	۳/۲-۳/۶	۳/۶-۴/۰	۴/۰-۴/۴
فراوانی	۵	۵	۹	۴	۴	۳



۱-۵-۱۰ نمودار چندضلعی فراوانی

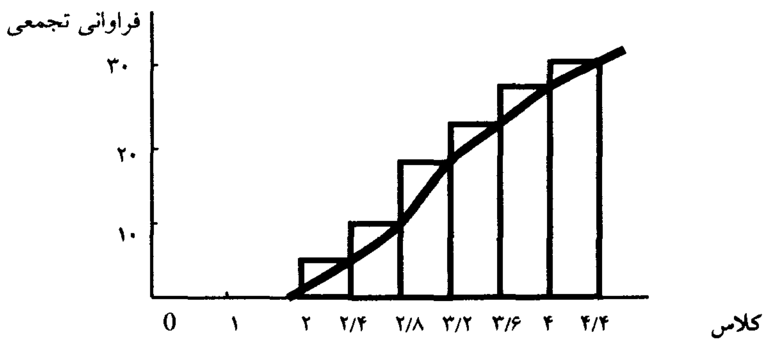
اگر نقطه‌های وسط قاعده‌های بالایی نمودار مستطیلی و نقطه‌های وسط کلاسهایی که بلافاصله در دو انتهای نمودار که فراوانی صفر هستند به هم وصل کنیم یک خط شکسته بدست می‌آید که آن را چندضلعی فراوانی گویند. دو کلاس انتهایی را که دارای فراوانی صفر هستند به این خاطر اضافه می‌کنیم که شروع و خاتمه چندضلعی به محور افقی ختم شود. برای مثال قبلی خط شکسته نمودار چندضلعی فراوانی عبارت است.



۱-۵-۱۱ نمودار چندضلعی فراوانی تجمعی

برای رسم نمودار چندضلعی فراوانی تجمعی از فراوانی تجمعی و حد بالای کلاسها استفاده می‌کنیم. به طوری که حد بالای کلاس‌ها روی محور xها و فراوانی تجمعی متناظر با آن را روی محور yها مشخص کرده و از وصل نقاط مذکور به هم نمودار چندضلعی فراوانی تجمعی به دست می‌آید. برای مثال ۱-۵-۹ نمودار فراوانی تجمعی ذیلاً رسم می‌شود.

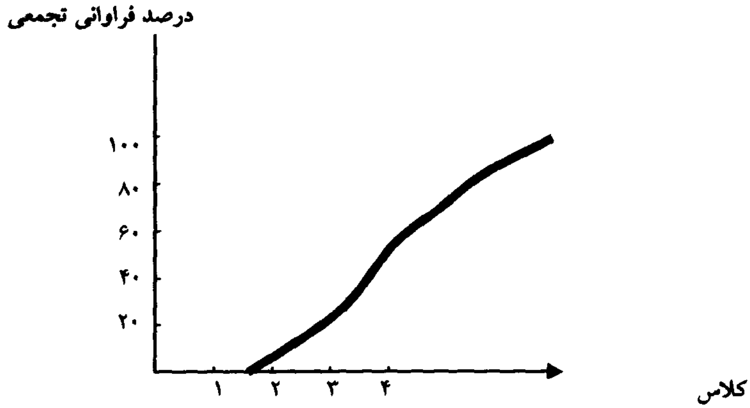
کلاس	۲/۰-۲/۴	۲/۴-۲/۸	۲/۸-۳/۲	۳/۲-۳/۶	۳/۶-۴/۰	۴/۰-۴/۴
فراوانی	۵	۵	۹	۴	۴	۳
فراوانی تجمعی	۵	۱۰	۱۹	۲۳	۲۷	۳۰



اگر در رسم نمودار چندضلعی فراوانی تجمعی به جای فراوانی تجمعی از درصد فراوانی تجمعی استفاده کنیم، منحنی حاصل را منحنی درصد فراوانی یا نمودار اجایو

گویند. برای مثال ۱-۵-۹ نمودار اجایو ذیلاً رسم می‌شود.

کلاس	۲/۰-۲/۴	۲/۴-۲/۸	۲/۸-۳/۲	۳/۲-۳/۶	۳/۶-۴/۰	۴/۰-۴/۴
فراوانی	۵	۵	۹	۴	۴	۳
فراوانی تجمعی	۵	۱۰	۱۹	۲۳	۲۷	۳۰
درصد فراوانی تجمعی	۱۶/۶۷	۳۳/۳۳	۶۳/۳۳	۷۶/۶۷	۹۰	۱۰۰



مثال ۱-۵-۱۲ برای جدول توزیع فراوانی زیر میانگین، میانه، نما و چارک اول و سوم را بدست آورید و نمودار چندضلعی آن را رسم کنید.

شماره کلاس	کلاس	فراوانی f_i	فراوانی تجمعی	حد متوسط x_i	$f_i x_i$
۱	۲/۵-۵/۵	۵	۵	۴	۲۰
۲	۵/۵-۸/۵	۷	۱۲	۷	۴۹
۳	۸/۵-۱۱/۵	۸	۲۰	۱۰	۸۰
۴	۱۱/۵-۱۴/۵	۱۰	۳۰	۱۳	۱۳۰
۵	۱۴/۵-۱۷/۵	۸	۳۸	۱۶	۱۲۸
۶	۱۷/۵-۲۰/۵	۷	۴۵	۱۹	۱۳۳
۷	۲۰/۵-۲۳/۵	۵	۵۰	۲۲	۱۱۰
		۵۰	-	-	۶۵۰

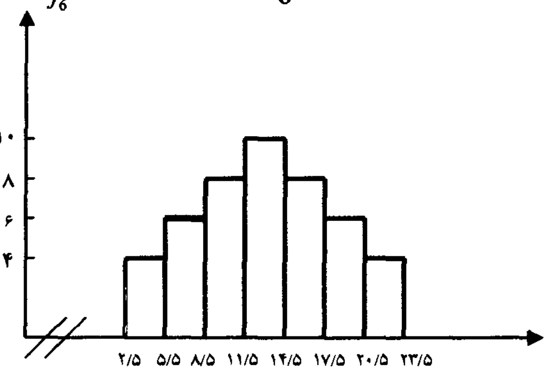
$$\bar{x} = \frac{1}{50}\sum_{i=1}^{7} f_i x_i = \frac{650}{50} = 13$$

$$M = a_4 + \frac{d_1}{d_1 + d_2} \times l = 11.5 + \frac{2}{4}(3) = 13$$

$$m = a_4 + \frac{\frac{n}{2} - F_3}{f_4} \times l = 11.5 + \frac{25-20}{10}(3) = 13$$

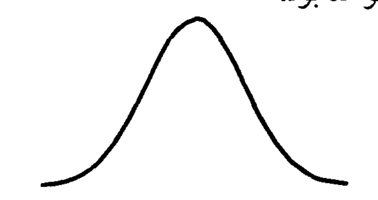
$$Q_1 = a_3 + \frac{\frac{n}{4} - F_2}{f_3} \times l = 8.5 + \frac{12/5-12}{8}(3) = 8.6875$$

$$Q_3 = a_4 + \frac{\frac{3n}{4} - F_5}{f_6} \times l = 14\ .5 + \frac{37.5-30}{8}(3) = 17.3125$$



علامت /// روی محور xها بیان این مطلب است که فاصله صفر تا ۲/۵ در روی محور انتخاب نشده است.

نمودار فراوانی چندضلعی باتوجه به اینکه میانه، میانگین و نما هر سه مساوی‌اند به صورت زیر متقارن خواهد بود.



این نمودار یا منحنی، منحنی نرمال نامیده می‌شود که دارای ویژگیهای زیادی است. این منحنی را تحت عنوان منحنی نرمال یا نرمال استاندارد در فصل چهار کاملاً توضیح می‌دهیم. آنچه در آمار توصیفی مهم است این است که اگر داده‌ها از یک همگنی یا تجانس برخوردار باشند منحنی چندضلعی فراوانی آنها منحنی نرمال خواهد بود. اگر

منحنی چندضلعی فراوانی منطبق بر منحنی نرمال نباشد ممکن است دارای چولگی یا برجستگی باشد.

۱-۶ چولگی و برجستگی

۱-۶-۱ چولگی

اگر منحنی چندضلعی فراوانی داده‌ها در مقایسه با منحنی نرمال، منطبق بر آن نباشد دارای چولگی است. معیارهای محاسبه میزان چولگی عبارتند از:

۱- ضریب چولگی پیرسن

$$SK_p = \frac{\bar{x} - M}{S}$$

۲- ضریب چولگی بر اساس گشتاور مرکزی مرتبه سوم

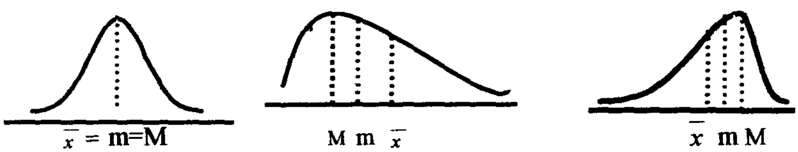
$$b = \frac{m_3}{S^3}$$

۱-۶-۲ ویژگیهای معیارهای چولگی

۱- معیارهای چولگی فاقد واحد اندازه‌گیری هستند.

۲- SK_p و b می توانند مقادیر منفی، صفر و مثبت باشند. در حالتی که منفی هستند چولگی وجود دارد و چولگی داده‌ها به طرف چپ است. در حالتی که SK_p و b مقدار صفر را اختیار کنند. داده‌ها فاقد چولگی است و مقادیر مثبت آنها دال بر چولگی به سمت راست است.

باتوجه به مقادیر SK_p، b و با درنظر گرفتن مقادیر میانگین، میانه و نما، سه شکل زیر موقعیت میانگین، میانه و نما و منحنی توزیع فراوانی داده ها را نشان می‌دهد.



فاقد چولگی — چولگی به راست — چولگی به چپ

مثال ۱-۶-۳ چولگی را برای مثال ۱-۵-۱۲ حساب کنید.

فاقد چولگی
$$SK_p = \frac{\bar{x}-M}{S} = \frac{13-13}{S} = \frac{0}{S} = 0$$

معیارهای میزان چولگی در جامعه را در فصل سوم بحث خواهیم کرد.

مثال ۱-۶-۴ ضرایب چولگی را برای جدول توزیع فراوانی زیر محاسبه نمایید.

x_i	f_i	F	$f_i x_i$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$
۲	۱	۱	۲	۴	۸-
۳	۳	۴	۹	۱	۱-
۴	۷	۱۱	۲۸	۰	۰
۵	۳	۱۴	۱۵	۱	۱
۶	۱	۱۵	۶	۴	۸
	۱۵		۶۰	۱۰	۰

$$\bar{x} = \frac{1}{15}\sum f_i x_i = \frac{60}{15} = 4 \quad , \quad M = 4$$

$$S^2 = \frac{1}{15}\sum (x_i-\bar{x})^2 = \frac{10}{15} = 0.67 \quad , \quad S = 0.82$$

$$m_3 = \frac{1}{15}\sum (x_i-\bar{x})^3 = \frac{0}{15} = 0$$

$$SK_p = \frac{\bar{x}-M}{S} = \frac{4-4}{0.82} = 0$$

$$b_3 = \frac{m_3}{S^3} = \frac{0}{(0.82)^3} = 0$$

۱-۶-۵ برجستگی

میزان کشیدگی یا پخی منحنی چندضلعی فراوانی داده‌ها در مقایسه با منحنی نرمال را برجستگی گویند و میزان برجستگی از رابطه زیر محاسبه می‌شود.

$$K=\frac{m_4}{S^4}-3=\frac{\frac{1}{n}\sum(x_i-\bar{x})^4}{\left[\frac{1}{n}\sum(x_i-\bar{x})^2\right]^2}-3$$

۱-۶-۶ ویژگیهای K

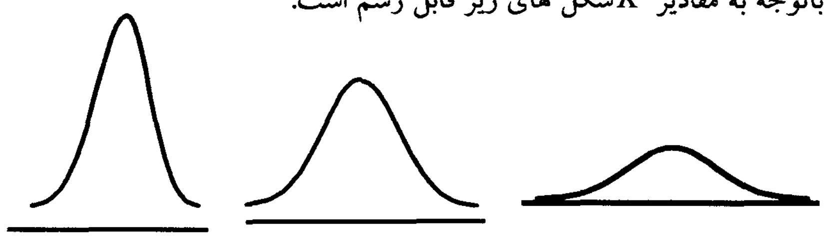
۱- مستقل از واحد اندازه‌گیری است.

۲- اگر $K=0$ باشد. میزان برجستگی صفر است و منحنی چندضلعی فراوانی بر منحنی نرمال منطبق است.

۳- اگر $K>0$ باشد. منحنی چندضلعی فراوانی در مقایسه با منحنی نرمال دارای برجستگی است.

۴- اگر $K<0$ منحنی چندضلعی فراوانی در مقایسه با منحنی نرمال دارای پخی است.

۵- باتوجه به مقادیر K شکل های زیر قابل رسم است.



$K>0$ $K=0$ $K<0$

مثال ۱-۶-۷ دستمزد هفتگی ۱۰۰ کارگر در یک کارخانه به صورت جدول زیر ثبت شده است. میزان برجستگی دستمزد کارگران را محاسبه کنید.

دستمزد هفتگی	فراوانی f	حد متوسط x	$f_i x$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
۱۲/۵–۱۷/۵	۱۲	۱۵	۱۸۰	−۱۳/۷۵	۱۸۹/۰۶۲۵
۱۷/۵–۲۲/۵	۱۶	۲۰	۳۲۰	−۸/۷۵	۷۶/۵۶۲۵
۲۲/۵–۲۷/۵	۲۵	۲۵	۶۲۵	−۳/۷۵	۱۴/۰۶۲۵
۲۷/۵–۳۲/۵	۱۴	۳۰	۴۲۰	۱/۲۵	۱/۵۶۲۵
۳۲/۵–۳۷/۵	۱۳	۳۵	۴۵۵	۶/۲۵	۳۹/۰۶۲۵
۳۷/۵–۴۲/۵	۱۰	۴۰	۴۰۰	۱۱/۲۵	۱۲۶/۵۶۲۵
۴۲/۵–۴۷/۵	۶	۴۵	۲۷۰	۱۶/۲۵	۲۶۴/۰۶۲۵
۴۷/۵–۵۲/۵	۳	۵۰	۱۵۰	۲۱/۲۵	۴۵۱/۵۶۲۵
۵۲/۵–۵۷/۵	۱	۵۵	۵۵	۲۶/۲۵	۶۸۹/۰۶۲۵
جمع	۱۰۰	–	۲۸۷۵	–	–

$$\bar{x}=\frac{1}{100}\sum f_i x_i=\frac{2875}{100}=28.75$$

$$S^2=\frac{1}{100}\sum f_i(x_i-\bar{x})^2=92.6875$$

$$m_4=\frac{1}{100}\sum f_i(x_i-\bar{x})^4=21822.72156$$

$$K=\frac{m_4}{S^4}-3=\frac{21822.72156}{8590.972656}-3=-0.459$$

۱-۷ کدگذاری

کدگذاری مجموعه‌ای داده‌ها عبارت از عملیاتی است که طی آن از هر مشاهده عدد ثابتی را کم (اضافه) کرده و نتیجه را بر عدد ثابتی تقسیم (ضرب) می‌نمایند. این عمل به منظور ساده کردن محاسبه معیارهای توصیفی انجام می‌شود. ولی در حال حاضر، استفاده وسیع از رایانه این هدف را برآورده ساخته است. اگر $x_۱$، $x_۲$،...، x_n مشاهدات به حجم n باشند و عدد a برای تغییر مبدأ اندازه‌گیری و عدد b برای تغییر واحد اندازه‌گیری باشند متغیرهای جدید $u_۱$، $u_۲$،...، u_n به صورت زیر تعریف می‌شوند.

$$u_i=\frac{x_i-a}{b}\qquad,\qquad i=۱,۲,\dots,n$$

اگر $\bar{x}$، S_x^2 و $\bar{u}$ ، S_u^2 به ترتیب میانگین و واریانس متغیرهای قدیم و جدید باشند روابط زیر در بین آنها برقرار است.

$$\bar{u}=\frac{1}{b}(\bar{x}-a)\qquad,\qquad S_u^2=\frac{1}{b^2}\,S_x^2$$

برای مثال S_u^2 به صورت زیر بدست می‌آید.

$$S_u^2=\frac{1}{n}\sum_{i=1}^{n}\left[u_i-\bar{u}\right]^2=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i-a}{b}-\frac{\bar{x}-a}{b}\right]^2=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i-\bar{x}}{b}\right]^2=\frac{S_x^2}{b^2}$$

باتوجه به تعریف u_i رابطه بین گشتاورهای مرکزی قدیم و جدید عبارتست از:

$$m_r^u = \frac{1}{b^r} m_r^x \qquad , \qquad m_r^x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$$

مثال ۱-۷-۱ برای جدول فراوانی زیر با استفاده از کدگذاری میانگین و واریانس را حساب کنید.

x_i	f_i	$u_i = \frac{x_i - 27.5}{5}$	$f_i u_i$	$u_i - \bar{u}$	$(u_i - \bar{u})^2$
۱۲/۵	۲۸	−۳	−۸۴	−۳/۵۹۲	۱۲/۹۰۲۴۴۶
۱۷/۵	۴۲	−۲	−۸۴	−۲/۵۹۲	۶/۷۱۸۴۶۴
۲۲/۵	۵۴	−۱	−۵۴	−۱/۵۹۲	۲/۵۳۴۴۶۴
۲۷/۵	۱۰۸	۰	۰	−۰/۵۹۲	۰/۳۵۰۴۶۴
۳۲/۵	۱۲۹	۱	۱۲۹	۰/۴۰۸	۰/۱۶۶۴۶۴
۳۷/۵	۶۱	۲	۱۲۲	۱/۴۰۸	۱/۹۸۲۴۶۴
۴۲/۵	۴۵	۳	۱۳۵	۲/۴۰۸	۵/۷۹۸۴۶۴
۴۷/۵	۳۳	۴	۱۳۲	۳/۴۰۸	۱۱/۶۱۴۴۶۴
جمع	۵۰۰	−	۲۹۶	−	−

$$\bar{u} = \frac{1}{500}\sum f_i u_i = \frac{296}{500} = 0.592$$

$$S_u^2 = \frac{1}{500}\sum_{i=1}^{n} f_i (u_i - \bar{u})^2 = 3.2095$$

در مثال فوق $a =$ ۲۷/۵ و $b =$ ۵ است.

$$\bar{u} = \frac{1}{b}(\bar{x} - a) \Rightarrow \bar{x} = b\bar{u} + a = 5 \times 0.592 + 27.5 = 30.46$$

$$S_u^2 = \frac{1}{b^2} S_x^2 \Rightarrow S_x^2 = b^2 S_u^2 = 25 \times 3.2095 = 80.2375$$

مثال ۱-۷-۲ با روش کدگذاری میانگین و واریانس جدول توزیع فراوانی زیر را محاسبه نمائید.

رده بندی	۱۴/۵−۱۹/۵	۱۹/۵−۲۴/۵	۲۴/۵−۲۹/۵	۲۹/۵−۳۴/۵	۳۴/۵−۳۹/۵
فراوانی	۹	۳۷	۳۱	۱۳	۱۰

ابتدا حد متوسط کلاس‌ها را بدست می‌آوریم و با توجه به حد متوسط مقادیر a و b را تشخیص می‌دهیم.

رده بندی	f_i	حد متوسط	$u_i = \frac{x_i - 27}{5}$	$f_i u_i$	$u_i - \bar{u}$	$(u_i - \bar{u})^2$	$f_i(u_i - \bar{u})^2$
۱۴/۵–۱۹/۵	۹	۱۷	−۲	−۱۸	−۱/۷۸	۳/۱۶۸۴	۲۸/۵۱۵۶
۱۹/۵–۲۴/۵	۳۷	۲۲	−۱	−۳۷	−۰/۷۸	۰/۶۰۸۴	۲۲/۵۱۰۸
۲۴/۵–۲۹/۵	۳۱	۲۷	۰	۰	۰/۲۲	۰/۰۴۸۴	۱/۵۰۰۴
۲۹/۵–۳۴/۵	۱۳	۳۲	۱	۱۳	۱/۲۲	۱/۴۸۸۴	۱۹/۳۴۹۲
۳۴/۵–۳۹/۵	۱۰	۳۷	۲	۲۰	۲/۲۲	۴/۹۲۸۴	۴۹/۲۸۴
جمع	۱۰۰	−	−	−۲۲	−	−	۱۲۱/۱۶

$$\bar{u} = \frac{1}{100}\sum f_i u_i = \frac{-22}{100} = -0.22$$

$$S_u^2 = \frac{1}{100}\sum f_i (u_i - \bar{u})^2 = 1.2116$$

باتوجه به ستون حدود متوسط $a = ۲۷$ و $b = ۵$ درنظر گرفته شده است. لذا میانگین و واریانس داده‌های اصلی از روابط زیر بدست می‌آید.

$$\bar{x} = b\bar{u} + a = 5 \times (-0.22) + 27 = 25.9$$

$$S_x^2 = b^2 S_u^2 = 25 \times (1.2116) = 30.29$$

۱–۸ جامعه آماری دوبعدی

در بخش ۱–۱–۱ با مفهوم جامعه آماری که از مقادیر یک صفت متغیر تشکیل شده بود آشنا شدیم. گاهی اوقات تحلیل‌گر آماری علاقه‌مند به مطالعه همزمان دو یا چند متغیر آماری می‌باشد. به عنوان مثال علاقه‌مند است وزن با قد اشخاص، مدل رایانه با سرعت آن یا بهره هوشی با پیشرفت تحصیلی را مورد مطالعه قرار دهد.

فرض کنید دو صفت مورد مطالعه در جامعه آماری به ترتیب X و Y باشند. همانطور که می‌دانیم جامعه آماری ممکن است متناهی یا نامتناهی باشد. با فرض آنکه جامعه آماری متناهی است می توانیم صفات مورد بررسی را به صورت زوج اطلاعاتی نمایش دهیم. برای مثال دو صفت قد (X) و وزن (Y) را در جامعه‌ای می‌توان به صورت زیر نمایش داد.

$$(X_۱, Y_۱), (X_۲, Y_۲), \ldots, (X_N, Y_N)$$

که زوج (X_1, Y_1) قد و وزن شخص اولی که مورد مشاهده قرار گرفته می‌باشد. در حالت کلی زوج (X_i, Y_i) قد و وزن شخص iام در جامعه است. برای نمایش بهتر می‌توان مشاهدات زوجی را به صورت زیر نیز نمایش داد.

i	۱	۲	۳	....	N
x_i	X_1	X_2	X_3	....	X_N
Y_i	Y_1	Y_2	Y_3	....	Y_N

مثال ۱-۸-۱ درآمد و هزینه تمام خانوارها در سرشماری ۱۳۷۵ را می توان به صورت زیر نمایش داد.

iخانوار	۱	۲	۳	....	N
Xدرآمد	X_1	X_2	X_3	....	X_N
Yهزینه	Y_1	Y_2	Y_3	....	Y_N

مشخص است که N در این بررسی بسیار بزرگ می‌باشد و بررسی جامعه به دلایل متعددی از جمله وقت‌گیربودن، هزینه بالا، طولانی بودن زمان نتیجه‌گیری و ... تحلیل‌گر را به انتخاب نمونه از جامعه و بررسی آن سوق می‌دهد. لذا در ادامه بحث به جای $(X_1,Y_1),\dots,(X_N,Y_N)$، یک نمونه به حجم n یعنی $(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$ را مورد بررسی قرار می‌دهیم.

i	۱	۲		n
x_i	x_1	x_2	....	x_n
y_i	y_1	y_2	....	y_n

مثال ۱-۸-۲ نمره آمار و برنامه نویسی ۵ دانشجو در جدول زیر ثبت شده است میانگین و واریانس نمره آمار و برنامه نویسی را حساب کنید.

شماره دانشجو	۱	۲	۳	۴	۵
نمره آمار (x)	۴۵	۲۰	۴۰	۲۵	۴۵
نمره برنامه‌نویسی (y)	۴۸	۳۵	۱۷	۲۳	۴۷

در این مثال $n=۵$ ، $\bar{x}=۳۵$ ، $\bar{y}=۳۴$ ، $S_x^2=۱۱۰$ و $S_y^2=۱۵۵/۲$.

در مواقعی که تعداد مشاهدات زیاد است، داده‌ها را معمولاً به صورت توزیع فراوانی دوطرفه یا جدول توزیع فراوانی دو متغیره گروه‌بندی می کنند که در این گروه‌بندی پس از مرتب کردن متغیرهای Xو Y یکی از آنها را برای سطرها و دیگری را برای ستونها عنوان قرار می‌دهند که صورت کلی آن در جدول زیر نمایش داده شده است.

$X \backslash Y$	y_1	y_2	$\cdots$	y_c	$f_{i\circ}$
x_1	f_{11}	f_{12}		f_{1c}	$f_{1\circ}$
x_2	f_{21}	f_{22}		f_{2c}	$f_{2\circ}$
$\vdots$					
x_r	f_{r1}	f_{r2}		f_{rc}	$f_{r\circ}$
$f_{\circ j}$	$f_{\circ 1}$	$f_{\circ 2}$		$f_{\circ c}$	$f_{\circ\circ}$

جدول فوق دارای r سطر و c ستون می‌باشد و f_{ij} معرف فراوانی تعداد زوجهایی با مولفه اول x_i و مولفه دوم y_j است.

$$f_{i\circ} = \sum_{j=1}^{c} f_{ij} \qquad , \qquad i = 1,2,\dots r$$

$$f_{\circ j} = \sum_{i=1}^{r} f_{ij} \qquad , \qquad j = 1,2,\dots c$$

$$f_{\circ\circ} = \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij} = n$$

که n تعداد کل مشاهدات و $\frac{f_{ij}}{n}$ فراوانی نسبی می‌باشد.

مثال ۱-۸-۳ نتایج اندازه‌گیری برحسب حجم محصول تولید شده در روز (x) و قیمت آن محصول (y) برای ۴۰ موسسه تولید کننده یک نوع محصول در زیر آورده شده است. جدول فراوانی دو متغیره را بنویسید.

x :	۲۰۰	۲۰۰	۴۰۰	۱۰۰	۴۰۰	۳۰۰	۴۰۰	۳۰۰	۱۰۰	۲۰۰	۴۰۰
y :	۱۲	۱۲	۱۰	۱۴	۱۰	۱۲	۱۲	۱۲	۱۶	۱۴	۱۲

۳۰۰	۴۰۰	۲۰۰	۲۰۰	۵۰۰	۳۰۰	۴۰۰	۳۰۰	۲۰۰	۵۰۰	۳۰۰	۵۰۰
۱۰	۱۲	۱۲	۱۴	۱۲	۱۲	۱۰	۱۰	۱۴	۱۰	۱۰	۱۰

۱۰۰	۳۰۰	۵۰۰	۴۰۰	۲۰۰	۳۰۰	۴۰۰	۳۰۰	۴۰۰	۳۰۰	۳۰۰	۲۰۰
۱۶	۱۲	۱۰	۱۲	۱۴	۱۲	۱۰	۱۲	۱۰	۱۲	۱۲	۱۲

۴۰۰	۲۰۰	۳۰۰	۲۰۰	۱۰۰
۱۰	۱۴	۱۲	۱۲	۱۶

$X \backslash Y$	۱۰	۱۲	۱۴	۱۶	$f_{i\circ}$
۱۰۰	۰	۰	۱	۳	۴
۲۰۰	۰	۵	۵	۰	۱۰
۳۰۰	۳	۹	۰	۰	۱۲
۴۰۰	۶	۴	۰	۰	۱۰
۵۰۰	۳	۱	۰	۰	۴
$f_{\circ j}$	۱۲	۱۹	۶	۳	۴۰

در حالت پیوسته بودن صفت‌های xو y، اگر فاصله‌های مقادیر صفت x فاصله (a,b) و فاصله مقادیر yفاصله (c,d) باشند. آنگاه در ابتدا فاصله مقادیر صفت xرا به فاصله‌های جزئی

$(a=x_1,x_2),(x_2,x_3),\dots,(x_r,x_{r+1=b})$

و فاصله مقادیر صفت yرا به فاصله‌های جزئی

$(c=y_1,y_2),(y_2,y_3),\dots,(y_c,y_{c+1=d})$

تقسیم می‌کنند و سپس زوجها را برحسب فاصله‌های جزئی مربوط به xو y گروه‌بندی می‌کنند. هرگروه به توسط دو فاصله (x_i,x_{i+1}) و (y_i,y_{i+1}) مشخص می‌شود.

مثال ۱-۸-۴ نمره آمار و حسابداری ۲۴ دانشجو به صورت زیر ثبت شده است. جدول فراوانی دو متغیره را تشکیل دهید.

۶	۱۷	۴	۵	۱۸	۲	۱۶	۳	۱	۰	۱۵	**نمره آمار**
۶	۱۶	۱۷	۹	۱۲	۹	۸	۷	۲	۱	۱۳	**نمره حسابداری**

۷	۹	۱۸	۱۲	۱۱	۱۱	۱۳	۱۰	۱۳	۸	۹	۱۴	۱۸
۳	۱۵	۱۵	۱۸	۷	۱۴	۱۱	۱۰	۴	۵	۳	۱۱	۱۹

نمره حسابداری	نمره آمار					
	۰–۴	۴–۸	۸–۱۲	۱۲–۱۶	۱۶–۲۰	$f_{i\circ}$
۰–۴	۲	۱	۱	۰	۰	۴
۴–۸	۱	۱	۲	۱	۰	۵
۸–۱۲	۱	۱	۱	۲	۱	۶
۱۲–۱۶	۰	۰	۲	۱	۲	۵
۱۶–۲۰	۰	۱	۰	۱	۲	۴
$f_{\circ j}$	۴	۴	۶	۵	۵	۲۴

خودآزمایی

۱- جامعه و نمونه را تعریف کنید و مشخص کنید کدامیک از موارد زیر جامعه و کدامیک نمونه را توصیف می‌کنند.

الف- فارغ التحصیلان رشته مهندسی

ب- فارغ التحصیلان مهندسی نرم افزار در سال ۱۳۸۳

ج- دانش‌آموزان گروه سنی ۱۴ تا ۱۶ در سرشماری ۱۳۷۵

د- دانش‌آموزان ممتاز گروه سنی ۱۴ تا ۱۶ در سرشماری ۱۳۷۵

هـ- رایانه‌های از رده خارج شده تا سال ۱۳۸۳

۲- نشان دهید.

$$\sum_{i=1}^{3} i = 6 \quad , \quad \sum_{i=1}^{3} i^2 = 14 \quad , \quad \sum_{i=1}^{3} i^3 = 36$$

$$\sum_{i=1}^{7} (x_i + 2) = \sum_{i=1}^{7} x_i + 14 \quad , \quad \sum_{i=1}^{3} (2i - 2) = 4$$

$$\sum_{i=0}^{5} i = 15 \quad , \quad \sum_{i=-1}^{3} i = 5$$

$$\sum_{i=1}^{3} (i+1)^{i+1} = 76 \quad , \quad \sum_{i=1}^{n} \frac{i}{2} = \frac{n(n+1)}{4}$$

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

۳- دستمزد ۱۵ کارگر در ادارات مختلف عبارتند از:

۱۱/۳۰	۱۱/۴۹	۱۸/۲۳	۲۹/۲۳	۱۲/۱۴	۱۲/۵۴	۸/۲۲	۱۱/۶۳
	۱۲/۸۱	۱۹/۷۷	۸/۲۳	۲۷/۵۶	۸/۶۴	۹/۱۶	۱۷/۰۰

میانگین حسابی، هندسی و هارمونیک دستمزد کارگران را بدست آورید.

۴- جدول زیر مصرف گاز مایع و رشد سالانه آن را از سال ۱۳۶۳ تا سال ۱۳۶۷ در کشور نشان می‌دهد. رشد متوسط را طی این دوره محاسبه کنید.

سال	۱۳۶۳	۱۳۶۴	۱۳۶۵	۱۳۶۶	۱۳۶۷
مصرف	۸۹۵	۱۰۵۴	۱۰۸۳	۱۰۷۶	۱۱۹۷
رشد سالانه	-	۱/۱۸	۱/۰۳	۰/۹۹	۱/۱۱

(رشد سالانه از تقسیم مصرف، هرسال به سال قبل بدست می‌آید)

۵- اگر $\bar{x}$ میانگین حسابی، Gمیانگین هندسی و Hمیانگین هارمونیک برای یک نمونه به حجم nباشند. نشان دهید که رابطه زیر همواره برقرار است.

$H \le G \le \bar{x}$

در چه حالت $H = G = \bar{x}$ است.

۶- اگر در سفر چهارساعته از تهران به سمنان با سرعت ۵۰ km/h در ساعت اول، ۶۵ km/hدر ساعت دوم، ۸۰ km/h در ساعت سوم و ۵۵ km/hدر ساعت چهارم طی مسیر شده باشد میانگین هارمونیک را بدست آورید.

۷- میانگین حسابی و هندسی دو عدد به ترتیب ۱۳ و ۱۲ است. دو عدد را پیدا کنید و میانگین هارمونیک آنها را بدست آورید.

۸- میانه سری مشاهدات زیر را بدست آورید.

سری اول

۲/۳   ۳/۵   ۱۴/۱   ۱۶/۶   ۱۸/۷   ۲۱/۳   ۲۹/۷   ۳۶/۶   ۱۷/۲

سری دوم

۶۵   ۶۹   ۷۴   ۸۴   ۹۳   ۱۰۲   ۱۰۵   ۱۰۲   ۹۸   ۸۸   ۷۴   ۶۶

سری سوم

۱۰   ۲   ۱۷   ۶   ۸   ۳   ۱۰   ۲   ۹   ۱۰   ۱   ۱۳   ۹   ۵   ۱۵

۹- نما یا مد سری مشاهدات زیر را بدست آورید.

سری اول		۳	۹	۵	۴	۵	۸	۵	۳
سری دوم				۲۴	۲۷	۱۰	۲۴	۲۷	۱۰
سری سوم	۹	۱۱	۱۰	۷	۹	۵	۸	۷	۵

۱۰- چارک اول، دوم و سوم مشاهدات زیر را بدست اورید.

۱۷/۷۷ ۲۲/۸۷ ۳۰/۱۷ ۲۱/۳۶ ۴۶/۵۹ ۵۲/۲۲

۱۱- جمعیت ۱۸ ناحیه عبارتست از:

۹۵	۷۵	۸۰	۱۰۷	۵۷	۶۸	۸۳	۷۶	۷۷
۷۴	۴۶	۸۷	۸۳	۱۲۱	۱۲۱	۱۱۹	۱۱۳	۱۰۰

دامنه، میانگین و واریانس جمعیت نواحی را بدست آورید.

۱۲- توزیع سنی ۱۳۰ مرد در اولین ازدواج در جدول فراوانی زیر داده شده است. نما، میانگین، واریانس و انحراف معیار را محاسبه کنید.

سن	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳	۲۴	۲۵	۲۶	۲۷	۲۸	۲۹
تعداد	۲	۱	۴	۸	۱۰	۱۲	۱۷	۱۹	۱۸	۱۴	۱۳	۱۲

۱۳- قد ۹ سرباز برحسب اینچ عبارتست از:

۷۲ ۶۸ ۶۵ ۶۳ ۶۴ ۶۹ ۶۷ ۶۶ ۶۹

قد این ۹ سرباز را استاندارد نمائید.

۱۴- نمره دانش‌آموزان یک کلاس در امتحان ریاضی دارای میانگین ۷۲ و انحراف معیار ۱۵ و در امتحان فیزیک دارای میانگین ۵۰ و انحراف معیار ۲۰ می‌باشد. اگر نمره دانش‌آموزی در ریاضی ۶۰ و در فیزیک ۳۵ باشد این دانش‌آموز در کدام کلاس مستعدتر است؟

۱۵- جدول زیرمقدار تولید شیرخشک غیرچربی در ۱۲ ماه گذشته را نشان می‌دهد.

۵۳/۲ ۷۶/۲ ۹۸/۷ ۱۲۶/۵ ۱۳۱/۴ ۱۱۱/۵ ۹۵/۸ ۸۱/۶ ۸۳/۵
۶۷/۱ ۴۹/۳ ۵۰/۳

دامنه، میانگین، واریانس و ضریب تغییرات مقدار تولیدات را بدست آورید.

۱۶- نرخ سود سپرده ۱۰۰ حساب بانک ملت دارای میانگین ۳/۵ و انحراف معیار ۳/۴۸ است و در بانک ملی دارای میانگین ۳/۲۵ و انحراف معیار ۳/۵۲ است. ضریب تغییر هر بانک را محاسبه کنید. برای پس انداز، کدام بانک را ترجیح می‌دهید؟

۱۷- اگر x_iها را دو به دو مقایسه کنیم و معدل مجموع n^2پرانتز $(x_i - x_j)^2$ را درنظر بگیریم، معیاری به صورت $\sigma_0^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2$ برای پراکندگی بدست می‌آید. نشان دهید که $\sigma_0^2 = 2S^2$ که S^2 واریانس نمونه است.

۱۸- داده‌های آماری زیر که عمر ۸ رایانه را نشان می‌دهد.

۱۳ ۸ ۱۰ ۱۵ ۷ ۴ ۲ ۵

مطلوبست:

الف) گشتاورهای مرتبه اول تا چهارم حول نقطه صفر

ب) گشتاورهای مرتبه اول تا چهارم حول میانگین نمونه

ج) ضریب تغییر

۱۹- جدول توزیع فراوانی زیر را در نظر بگیرید.

کلاس	۰–۱	۱–۲	۲–۳	۳–۴	۴–۵	۵–۶	۶–۷	۷–۸	۸–۹	۹–۱۰
فراوانی	۳	۹	۲	۸	۱۱	۱۳	۱۲	۸	۹	۵

الف) کلاسها را به صورت کراندار بنویسید.

ب) حد متوسط یا نماینده کلاس‌ها را بدست آورید.

ج) طول هر کلاس چند است.

د) درصد فراوانی و درصد فراوانی تجمعی را بدست آورید.

۲۰- برای داده‌های زیر، جدول توزیع فراوانی را کامل کنید و به سئوالات زیر پاسخ دهید؟

۴۷/۵	۶۰/۹	۳۱/۳	۳۲/۳	۴۲/۳	۴۲/۲	۴۴/۵	۵۰/۰	۵۳/۹	۴۹/۲
۵۳/۴	۴۵/۷	۵۲/۳	۵۱/۷	۴۵/۳	۴۰/۲	۵۷/۶	۴۱/۱	۳۷/۹	۴۳/۵
۵۰/۳	۴۸/۶	۵۹/۶	۵۵/۱	۶۷/۲	۳۲/۵	۵۰/۰	۴۵/۹	۴۵/۷	۵۱/۰
۳۵/۸	۵۶/۳	۶۲/۹	۶۲/۲	۴۴/۰	۴۱/۵	۳۸/۳	۴۷/۴	۴۶/۸	۴۵/۱
۳۱/۸	۳۴/۳	۵۲/۱	۵۵/۲	۴۱/۳	۴۹/۶	۴۷/۴	۴۸/۵	۳۳/۵	۳۸/۳
۳۱/۶	۴۷/۹	۳۲/۸	۴۹/۲	۴۸/۴	۳۹/۸	۴۱/۲	۴۷/۰	۴۶/۰	۴۹/۳
۳۹/۱	۴۳/۴	۴۱/۵	۴۵/۱	۴۴/۴	۴۶/۲	۴۴/۵	۵۴/۱	۵۴/۵	۳۹/۱
۳۷/۴	۳۳/۸	۲۸/۷	۳۳/۶	۳۷/۲	۴۸/۸	۴۳/۷	۴۳/۱	۴۱/۶	۴۳/۵
۴۸/۸	۴۷/۳	۳۹/۰	۴۸/۵	۵۰/۶	۵۱/۹	۴۵/۱	۵۳/۰	۴۴/۲	۳۸/۲

طول کلاس	درصد فراوانی تجمعی	فراوانی تجمعی	درصد فراوانی	حد متوسط	فراوانی	کلاس

الف) چند درصد از داده‌ها کمتر یا مساوی ۴۸ هستند؟

ب) بزرگترین و کوچکترین کلاس را مشخص کنید.

ج) چند درصد از داده‌ها بیشتر یا مساوی ۶۰ هستند؟

۲۱- جدول توزیع فراوانی زیر میزان منواکسیدکربن خارج شده از ۷۹۴ اتومبیل را نشان می‌دهد که براساس میزان منواکسیدکربن ایجاد شده توسط هر اتومبیل به گروه‌های به طول ۲۴ گرم در هر مایل گروه بندی شده‌اند

.

فاصله	۰-۲۴	۲۴-۴۸	۴۸-۷۲	۷۲-۹۶	۹۶-۱۲۰	۱۲۰-۱۴۴	۱۴۴-۱۶۸
حد متوسط	۱۲	۳۶	۶۰	۸۴	۱۰۸	۱۳۲	۱۵۶
فراوانی	۱۳	۹۸	۱۶۱	۱۸۹	۱۴۸	۸۵	۴۵

۱۶۸–۱۹۲	۱۹۲–۲۱۶	۲۱۶–۲۴۰	۲۴۰–۲۶۴	۲۶۴–۲۸۸	۲۸۸–۳۱۲	۳۱۲–۳۳۶	۳۳۶–۳۶۰
۱۸۰	۲۰۴	۲۲۸	۲۵۲	۲۷۶	۳۰۰	۳۲۴	۳۴۸
۳۰	۱۰	۵	۵	۱	۲	۱	۱

الف) درصد فراوانی، فراوانی تجمعی و درصد فراوانی تجمعی را بدست آورید.
ب) میانگین، واریانس و ضریب تغییر را حساب کنید.
ج) میانه، نما، چارک اول و سوم را حساب کنید.

۲۲– اندازه گیریهای ۱۵ نقطه جوش از ترکیبات سیکیان برحسب درجه سلسیوس به صورت زیر داده شده است. نمودار نقطه ای را رسم کنید.

۱۶۰	۱۳۲	۱۴۸	۱۵۷	۱۸۳	۱۴۶	۱۵۵	۱۶۲	۱۷۰
۱۵۳	۱۳۶	۱۴۱	۱۶۶	۱۵۰	۱۵۷			

۲۳– جدول زیر مساحت های قاره‌های سیاسی جهان را برحسب میلیون کیلومترمربع نشان می‌دهد. نمودار دایره آن را رسم کنید.

قاره	آفریقا	آسیا	امریکای شمالی	اروپا	امریکای جنوبی	شوروی سابق	استرالیا
مساحت	۳۰/۳	۲۶/۹	۲۴/۳	۴/۹	۱۷/۹	۲۰/۵	۸/۵

۲۴– تعداد اتومبیلهای بیمه شده توسط یک شرکت بیمه در چهل روز عبارتست از:

۲	۴	۵	۶	۳	۸	۷	۸	۴	۱
۷	۹	۷	۸	۵	۶	۹	۳	۶	۳
۴	۱۰	۶	۷	۵	۲	۹	۶	۱۰	۸
۱۱	۹	۷	۶	۷	۵	۶	۸	۵	۴

نمودار میله‌ای را رسم کنید.

۲۵– نمودار مستطیلی و نمودار چندضلعی فراوانی، جدول توزیع فراوانی زیر را رسم کنید.

دستمزد هفتگی	۱۰–۱۵	۱۵–۲۰	۲۰–۲۵	۲۵–۳۰	۳۰–۴۰	۴۰–۶۰	۶۰–۸۰
تعداد کارگر	۷	۱۹	۲۷	۱۵	۱۲	۱۲	۸

راهنمایی: فاصله کلاس مساوی نیست لذا قبل از رسم نمودار باید فاصله ها را یکسان کرد. مثلاً کلاس ۴۰–۳۰ با فراوانی ۱۲ تبدیل می شود به ۳۵–۳۰ با فراوانی ۷ و ۴۰–۳۵ با فراوانی ۵.

۲۶– جدول زیر قیمت ۵۸۰ رایانه فروخته شده در یک فروشگاه را نشان می‌دهد.

قیمت	۲۰–۲۵	۲۵–۳۰	۳۰–۳۵	۳۵–۴۰	۴۰–۴۵	۴۵–۵۰	۵۰–۵۵
تعداد	۲۱	۲۹	۱۹	۳۹	۴۳	۹۴	۷۳

۵۵–۶۰	۶۰–۶۵	۶۵–۷۰	۷۰–۷۵	۷۵–۸۰	۸۰–۸۵	۸۵–۹۰	۹۰–۹۵
۶۸	۳۶	۴۵	۲۷	۴۸	۲۱	۱۲	۵

الف) فراوانی تجمعی و درصد فراوانی تجمعی را بدست آورید.

ب) نمودار چندضلعی فراوانی تجمعی را رسم کنید.

ج) نمودار چندضلعی درصد فراوانی تجمعی (اجایو) را رسم کنید.

د) تعداد رایانه‌هایی را که قیمت آنها بین ۵۷ تا ۷۷ هستند بدست آورید.

۲۷– نمودار چندضلعی فراوانی و جدول توزیع فراوانی زیر را رسم کنید و آن را با نمودار منحنی نرمال مقایسه کنید.

برق مصرفی شرکت رایانه‌ای	۵–۲۴	۲۵–۴۴	۴۵–۶۴	۶۵–۸۴	۸۵–۱۰۴	۱۰۵–۱۲۴	۱۲۵–۱۴۴	۱۴۵–۱۶۴
تعداد شرکت	۴	۶	۱۴	۲۲	۱۴	۷	۵	۳

۲۸– ضریب چولگی پیرسن را برای جدول فراوانی زیر بدست آورید.

x	۱۲/۵	۱۷/۵	۲۲/۵	۲۷/۵	۳۲/۵	۳۷/۵	۴۲/۵	۴۷/۵
f	۲۸	۴۲	۵۴	۱۰۸	۱۲۹	۶۱	۴۵	۳۳

۲۹– نمره ۶۰ دانشجو در درس فیزیک در جدول زیر داده شده است. ضریب چولگی و میزان برجستگی را حساب کنید.

نمره	۰–۱۰	۱۰–۲۰	۲۰–۳۰	۳۰–۴۰	۴۰–۵۰	۵۰–۶۰	۶۰–۷۰	۷۰–۸۰	۸۰–۹۰
فراوانی	۲	۳	۱۲	۸	۱۰	۱۷	۴	۳	۱

۳۰- در یک سری مشاهدات هرگاه میزان چولگی خفیف باشد بین میانگین، میانه و نما رابطه تقریبی زیر برقرار است.

$\bar{x} - M \approx 3(\bar{x} - m)$

برای یک سری مشاهدات با $\bar{x}$=۱۷/۲ ، m = ۱۵ و S = ۵ ضریب چولگی پیرسن را بدست آورید.

۳۱- هزینه زندگی ده خانوار ۶ نفره در جدول زیر آمده است.

خانوار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
هزینه	۱۰۹۰	۱۲۷۰	۱۲۶۰	۱۲۰۰	۱۱۷۰	۱۰۸۰	۱۰۰۰	۱۳۱۰	۱۲۱۰	۱۱۳۰

میانگین، واریانس و ضریب تغییر را با استفاده از کدگذاری بدست آورید.

۳۲- جدول توزیع فراوانی دستمزد کارمندان یک کارخانه را نشان می‌دهد. میانگین، واریانس و ضریب تغییر را با استفاده از کدگذاری بدست آورید.

دستمزد	۳۰۰-۴۰۰	۴۰۰-۵۰۰	۵۰۰-۶۰۰	۶۰۰-۷۰۰	۷۰۰-۸۰۰	۸۰۰-۹۰۰	۹۰۰-۱۰۰۰	۱۰۰۰-۱۱۰۰
تعداد کارمندان	۱۵	۲۲	۱۸	۱۴	۹	۷	۵	۴

۳۳- مقدار کود مصرفی (x) برحسب تن در هکتار و محصول بدست آمده از هر هکتار (y) در جدول زیر آمده است. جدول دو متغیره xو yرا تشکیل دهید.

x	۱	۲	۵	۳	۲	۳	۳	۵	۴
y	۱۴	۱۴	۱۸	۱۵	۱۵	۱۷	۱۶	۱۸	۱۷

فصل ۲

احتمال

مقدمه

نظریه‌های احتمال و آمار نظام‌های ریاضی می‌باشند که در حوزه‌های مختلف بسیاری از فعالیتهای انسان کاربردهای مهمی یافته‌اند و دامنه روش علمی را بسط داده‌اند به طوری که بتوان روش مزبور را در آزمایشهایی که نتایج آنها کاملاً توسط شرایط آزمایشی تعیین نمی‌گردند، بکار برد.

توافقی که میان دانشمندان درباره اعتبار اکثر تئوریهای علمی وجود دارد به میزان قابل ملاحظه‌ای متکی بر این واقعیت است که آزمایشهایی که این تئوریها بر آنها متکی می‌باشند به هنگام تکرار، اساساً نتایج یکسانی را بدست می دهند. هنگامی که دانشمندی کشفی را اعلام می دارد سایر دانشمندان می‌توانند در نقاط دیگر صحت یا سقم یافته‌های وی را برای خود تعیین نمایند.

گاهی نتایج دو محقق به نظر متفاوت می‌آید، لکن این امر معمولاً بدین معنی می‌باشد که در این دو وضع، شرایط آزمایشی یکسان نیستند. اگر آزمایش پس از تکرار، تحت شرایط یکسان، نتایج یکسانی را به بار آورد می‌توان گفت نتایج توسط این شرایط تعیین می‌گردد یا آزمایش حتمیت دارد. این وضع حتمی‌گرایانه علم است که تئوری علمی را مجاز می سازد که آنچه رادر شرایط تعیین شده‌ای مشاهده می‌گردد پیش‌بینی کند. لکن، آزمایشهایی نیز وجود دارند که هر اندازه هم در ثابت نگهداشتن شرایط آزمایش بکوشیم، نتایج آنها را نمی‌توان پیش بینی کرد.

مثلاً پرتاب سکه، پرتاب تاس، نتیجه امتحان، پرتاب سکه و تاس آزمایشهایی هستند با نتایج غیرقابل پیش‌بینی. در حوزه‌های مختلف، نمونه‌های جالبتر و مهمتر بسیاری از این آزمایشها رخ می‌دهد. مثلاً دانه‌های بذر به ظاهر یکسان، گیاهانی با ارتفاع متفاوت پدید می‌آورند، توزینهای مکرر شئ واحدی با ترازویی شیمیایی، تغییرات کوچک را نمایش می‌دهند.

برای طبقه‌بندی آزمایش‌ها و ارائه مدل احتمال، مفهوم فضای نمونه، پیشامدها و قوانین شمارش را مورد بررسی قرار می‌دهیم.

۲-۱ فضای نمونه

مجموعه‌ای از همه برآمدهای ممکن یک تجربه تصادفی را فضای نمونه می‌گویند. و آن را با علامت S نمایش می‌دهند.

مثال ۲-۱-۱ در پرتاب یک تاس فضای نمونه را بنویسید.

$S = \{۱،۲،۳،۴،۵،۶\}$

شرکت در آزمون اداره راهنمایی و رانندگی یک تجربه تصادفی با فضای نمونه {قبول، مردود} = S می‌باشد. فضای نمونه، هر تجربه تصادفی که شامل تمام برآمدهای ممکن آن تجربه تصادفی است همواره قابل تصور است.

فضای نمونه مورد بررسی در مثالهای اخیر یک بعدی بودند ولی گاهی ممکن است فضای نمونه دو یا چند بعدی باشد، مثلاً در پرتاب یک سکه و تاس فضای نمونه دو بعدی به صورت زیر است.

$S = \{(H,۱),(H,۲),(H,۳),(H,۴),(H,۵),(H,۶),(T,۱),(T,۲),(T,۳),(T,۴),(T,۵),(T,۶)\}$

که مولفه اول هر زوج ظاهر شدن سکه و مولفه دوم ظاهر شدن تاس است و علامت H برای شیر بودن و T برای خط بودن سکه استفاده شده است.

به عنوان مثال دیگر، موسسه‌ای برای جایابی دو نوع تسهیلات تحقیقاتی جدید رایانه‌ای باید تصمیم بگیرد. برای یک هدف معین برایش مفید است که چند تا از آنها را در تبریز و چند تا از آنها را در تهران راه‌اندازی کند. می‌توان فضای نمونه را به صورت زیر نمایش داد.

$$S = \{(0,0),(۱,0),(0,۱),(۲,0),(0,۲),(۱,۱)\}$$

منظور از(۱،۱) این است که یکی را در تبریز و یکی در تهران راه‌اندازی نماید. عموماً فضاهای نمونه برحسب تعداد نقاطی که در بر دارند دسته‌بندی می‌شوند. که ممکن است فضای نمونه متناهی یا نامتناهی باشد. فضای نمونه مثالهای اخیر همگی متناهی بودند. اکنون مثالی را مورد بررسی قرار می‌دهیم که فضای نمونه آن نامتناهی است.
اگر شخصی مایل به بررسی نشر گاز اکسید نیتروژن ماشین‌ها باشد باید تعداد زیادی از ماشین‌ها را قبل از پیدا کردن اولین ماشین که دستور تنظیم را رعایت نکرده است، بررسی کند. ولی ممکن است ماشین مزبور در اولین بررسی یا دومین یا ... مشاهده شود.
واضح است که باید هزاران ماشین را قبل از رسیدن به ماشین مورد نظر بررسی کرده باشد و نتیجه اینکه نمی‌داند بررسی را برای چند ماشین باید انجام دهد. فضای نمونه حاصل از این بررسی درست مثل فضای نمونه‌ای است که برای انتخاب یک عضو از مجموعه اعداد طبیعی که مجموعه‌ای شمارا و نامتناهی است به کار می‌رود.

$$S = \{۱،۲،۳،...\}$$

اگر شخصی علاقه‌مند به بررسی نشر گاز اکسید نیتروژن یک ماشین معین برحسب گرم در هر مایل باشد فضای نمونه شامل همه نقاط روی یک مقیاس پیوسته است (یک فاصله معین روی خط اعداد حقیقی) از این رو یک پیوستار وجود دارد. به طور کلی فضای نمونه‌ای را که شامل تعدادی متناهی یا نامتناهی از عناصر شمارا باشد گسسته و فضای نمونه‌ای را که شامل همه نقاط روی یک خط یا روی یک صفحه باشد پیوسته گویند.

مثال ۲-۱-۲ یک سکه را آنقدر پرتاب می‌کنیم تا شیر ظاهر شود. فضای نمونه را بنویسید.

$$S = \{H, TH, TTH, ...\}$$

که S گسسته و نامتناهی شمارا است.

مثال ۲-۱-۳ فضای نمونه را برای عمر یک وسیله الکتریکی بنویسید.

$$S=\{t \mid t>0\}$$

که S فضای نمونه پیوسته است.

۲-۲ پیشامد

هر زیر مجموعه‌ای از فضای نمونه را یک پیشامد گویند. معمولاً پیشامدها را با حروف بزرگ A، B و C نمایش می‌دهند. در آمار گاهی اوقات فضای نمونه Sرا فضای پیشامدها نیز می‌گویند.

۲-۲-۱ رخداد یک پیشامد

در یک تجربه تصادفی پیشامد Aرخ می دهد، هرگاه تجربه تصادفی منجر به مشاهده یکی از اعضای A گردد.

۲-۲-۲ دو پیشامد ناسازگار

دو پیشامد A و B را پیشامدهای ناسازگار یا مجزا گویند. اگر اشتراک آنها تهی باشد. دو پیشامد مجزا از هم نمی‌توانند به طور همزمان رخ دهند.

۲-۲-۳ تفاضل پیشامد A از B

تفاضل پیشامد Aاز B، پیشامدی است که شامل تمام اعضای Bاست که در Aنباشد و آن را با $B-A$ یا $B\cap A^C$نشان می‌دهند.

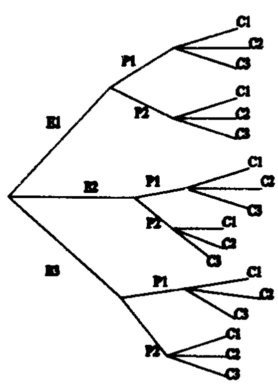
اعمال مجموعه‌ای روی پیشامدها همانند اعمال روی مجموعه‌ها است. لذا قوانین مجموعه‌ها که در جدول زیر آمده است، برای مجموعه از پیشامدها نیز صادق است.

$A \cup A = A$	$A \cap A = A$	قانون خودنمایی
$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$	قانون شرکت پذیری
$A \cup B = B \cup A$	$A \cap B = B \cap A$	قانون تعویض پذیری
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	قانون بخش پذیری
$A \cup \phi = A \ , \ A \cup S = S$	$A \cap S = A \ , \ A \cap \phi = \phi$	قانون همانی
$A \cup A^C = S \ , \ (A^C)^C = A$	$A \cap A^C = \phi \ , \ S^C = \phi$	قانون متمم ها
$(A \cup B)^C = A^C \cap B^C$	$(A \cap B)^C = A^C \cup B^C$	قانون دمورگان

۲-۳ شمارش

تعیین تعداد عناصر یک فضای نمونه متناهی به وسیله شمارش مستقیم، واقعاً مشکل یا لااقل خسته کننده است. برای درک بیشتر، فرض کنید یک مهندس رایانه درجه‌بندی سرویس کار رایانه‌ها را برحسب آنکه راه‌اندازی آنها آسان(E_1)، متوسط (E_2) و مشکل (E_3)، کم‌هزینه (P_1)، پرهزینه (P_2) و بالاخره تعمیر آن گران (C_1)، متوسط (C_2) و ارزان (C_3) باشد آزمون می‌کند. به وسیله این آزمون به چند طریق مختلف رایانه‌ها درجه‌بندی می‌شود به وضوح تعداد حالات زیادی وجود دارد. ممکن است یک رایانه با راه‌اندازی آسان، هزینه کم ولی با تعمیر گران نرخ گذاری شود. ممکن است با راه‌اندازی مشکل، پرهزینه ولی تعمیر ارزان داشته باشد، یا ممکن است برای حالتی که راه‌اندازی متوسط، کم هزینه ولی تعمیر با نرخ متوسط باشد درجه‌بندی شده باشد و غیره.

اگر به این روش ادامه دهیم همه ۱۸ حالت ممکن را می‌توانیم ثبت کنیم. برای حل دستی این نوع مسائل می‌توانیم از نمودار درختی استفاده کنیم که برای مسئله طرح شده نمودار درختی در زیر رسم شده است.



در این نمودار یک شاخه از سمت چپ به راست، یک دسته خاص بدست می‌دهد یعنی یک عنصر ویژه از فضای نمونه. ملاحظه می‌شود که روی هم رفته ۱۸ حالت ممکن وجود دارد. این نتیجه از مشاهده سه شاخه E که هرکدام از آنها به دو شاخه P و هر شاخه P نیز به سه شاخه C انشعاب یافته بدست آمده است. بنابراین ۱۸=۳×۲×۳ ترکیب شاخه یا مسیر وجود دارد. در تجربه تصادفی پرتاب دو تاس تعداد حالات ممکن یا فضای نمونه دارای ۳۶=۶×۶ عضو خواهد بود. اگر این مجموعه متناهی خیلی پر عضو باشد شمارش اعضای آن نه تنها با دست بلکه با رایانه نیز امکان‌پذیر نیست. با توجه به اندیشه‌های ریاضی می‌توان فرمولهایی ابداع کرد تا از دشواری کار شمارش رهایی پیدا کرد.

۲-۴ اصول شمارش

فرض کنید کار x با m طریق به نامهای $x_1, x_2, \ldots, x_m$ و کار y با n طریق به نامهای $y_1, y_2, \ldots, y_n$ قابل انجام باشند.

اصول شمارش عبارتند از:

۲-۴-۱ اصل اول شمارش

اگر انجام کار z منوط به انجام کار x یا y باشد آنگاه کار z را می‌توان به $m+n$ طریق با نامهای $x_1, x_2, \ldots, x_m$ ، $y_1, y_2, \ldots, y_n$ انجام داد.

مثال ۲-۴-۲ جعبه شماره ۱ شامل ۵ لامپ سفید و جعبه شماره ۲ شامل ۷ لامپ سبز است به چند طریق می توان یک لامپ سفید یا سبز انتخاب کرد.

در این مثال $m=5$ ، $n=7$ پس $m+n=12$ حالت.

۲-۴-۳ اصل دوم شمارش

اگر انجام کار z منوط به انجام پیاپی کار x و y باشد. آنگاه کار z را می‌توان به $m \times n$ طریق زیر انجام داد.

$$(x_1, y_1), (x_1, y_2), \ldots, (x_1, y_n)$$
$$(x_2, y_1), (x_2, y_2), \ldots, (x_2, y_n)$$
$$\vdots$$
$$(x_m, y_1), (x_m, y_2), \ldots, (x_m, y_n)$$

مثال ۲-۴-۴ تعداد طرق یا فضای نمونه را برای سه کتاب رایانه فارسی با برچسبهای I، IIو IIIو دو کتاب انگلیسی با برچسب های B و A را بنویسید.

با استفاده از اصل دوم شمارش،$m=۳$ ، $n=۲$ و تعداد طرق برابر با ۶=۲×۳ و با فضای نمونه زیر:

$$S = \{(I, A), (II, A), (III, A), (I, B), (II, B), (III, B)\}$$

مثال ۲-۴-۵ از ۴ دانشجوی رشته مهندسی رایانه و سه دانشجوی رشته آمار به چند طریق دو نماینده انتخاب کنیم که یکی دانشجوی رشته رایانه و دیگری دانشجوی رشته آمار باشد.

در این مثال،$m=۴$ ، $n=۳$ و در نتیجه$m \times n=۱۲$.

اصل دوم شمارش در حالت کلی انجام عملی مرکب از k مرحله است به طوریکه مرحله اول بتواند به $n_۱$ طریق انجام شود و برای هریک از این طرق، مرحله دوم بتواند به $n_۲$ طریق صورت گیرد و برای هریک از طرق این دو مرحله، مرحله دوم بتواند $n_۳$طریق صورت گیرد و الی آخر آنگاه کل عمل می تواند به $n_۱ \times n_۲ \times \ldots \times n_k$ طرق قابل انجام باشد.

مثال ۲-۴-۶ یک سکه دو ریالی، یک تاس و یک سکه ۵ ریالی هم زمان پرتاب می‌شوند. تعداد کل برآمدها را بنویسید.

در این مثال،$n_۱=۲$،$n_۲=۶$،$n_۳=۲$ پس ۲۴=۲×۶×۲ .

مثال ۲-۴-۷ چند عدد زوج سه رقمی از ارقام ۱، ۲، ۵، ۶ و ۹ می‌توان نوشت به طوریکه هر رقم فقط یک بار استفاده شود؟
از اینکه اعداد زوج باشد، برای رقم یکان فقط دو انتخاب وجود دارد پس کل طرق برابر است با ۲۴=۳×۴×۲ .

مثال ۲-۴-۸ از سه نفر دانشجو، ۴ نفر معلم و ۲ نفر استاد، می‌خواهیم یک کمیته علمی دو نفری تشکیل دهند که اعضای آن دارای تخصص متفاوت باشند. به چند طریق می‌توان این کمیته را تشکیل داد؟
طبق اصل دوم شمارش از ۳ نفر دانشجو و ۴ نفر معلم به ۱۲=۴×۳ طریق می‌توان یک کمیته دو نفری تشکیل داد که دارای تخصص متفاوت باشند و همانطور، دانشجو و استاد، استاد و معلم و طبق اصل اول شمارش:

۲۶=۲×۴+۲×۳+۴×۳

۲-۵ جایگشت

ترتیبی از مجموعه n شیء با آرایش معین جایگشت اشیاء خوانده می‌شود.

۲-۵-۱ جایگشت n شیء متمایز

جایگشت nشیء متمایز برابر با $n=n!$ $(n-۱)\times..\times.۳\times۲\times۱$ است. در حقیقت n شیء داریم و می‌خواهیم در nمکان قرار دهیم، چون اشیا متمایز هستند لذا، برای مکان اول n حالت وجود دارد. اگر مکان اول توسط یک شیء اشغال شود، برای مکان دوم $n-۱$ حالت وجود دارد. چون در مکان اول یک شیء قرار داده شده و تعداد اشیا از n به $n-۱$ تقلیل پیدا کرده است برای مکان سوم $n-۲$ حالت وجود دارد و الی آخر. طبق اصل دوم شمارش تعداد طرق برابر است با:

$$n(n-۱)(n-۲)\times...\times ۲\times ۱=n!$$

مثال ۲-۵-۲ سه کتاب A، B و Cرا به چند طریق در قفسه کنار هم قرار می‌گیرند؟

$ABC, ACB, BAC, BCA, CAB, CBA$ و $۳\times۲\times۱=۶=۳!$

مثال ۲-۵-۳ با ارقام ۷، ۴، ۳، ۱ بدون تکرار چند عدد چهاررقمی می‌توان نوشت؟
$۴\times۳\times۲\times۱=۲۴=۴!$

۲-۵-۴ جایگشت rتایی، nشیء متمایز

جایگشت rتایی، nشیء متمایز برابر با $n(n-۱)(n-۲)...(n-r+۱)$ است. در حقیقت nشیء داریم و می‌خواهیم در rمکان قرار دهیم. چون اشیا متمایز هستند. لذا برای مکان اول n حالت وجود دارد، برای مکان دوم $n-۱$ حالت و به همین ترتیب برای مکان rام، $n-r+۱$ حالت وجود دارد. طبق اصل دوم شمارش تعداد طرق برابر است با:

$$n(n-۱)(n-۲)...(n-r+۱)$$

اما

$$\frac{n(n-1)(n-2)...(n-r+1)(n-r)(n-r-1)...2\times1}{(n-r)(n-r-1)...2\times1}=\frac{n!}{(n-r)!}$$

در آمار جایگشت rتایی، n شیء متمایز را با نماد nPr نمایش می‌دهند.

$$nPr=\frac{n!}{(n-r)!}$$

مثال ۲-۵-۵ چهار کتاب A ، B ، C و D را به چند طریق می‌توان در دو قفسه جا داد؟

در این مثال $n=۴$ و $r=۲$ پس:

$$_{۴}P_{۲}=\frac{4!}{(4-2)!}=12$$

۲-۵-۶ جایگشت rتایی nشیء متمایز با تکرار اشیا

جایگشت rتایی، n شیء متمایز با تکرار اشیا برابر با $nP'r=n^r$

در حقیقت nشیء داریم و می‌خواهیم در rمکان قرار دهیم و بر خلاف حالت قبل هریک از اشیا می توانند تکرار شوند. یعنی در مکان اول یکی از n شیء می‌تواند قرار گیرد و همانطور برای مکان دوم، سوم و الی آخر، پس:

$$n P'r = n\times n\times n\times ...\times n = n^r$$

مثال ۲-۵-۷ با اعداد ۷، ۴، ۳، ۱ چند عدد سه رقمی با تکرار می توان نوشت؟ در این مثال، $n = ۴$ و $r = ۳$ پس:

$$۴P'۳=۴\times ۴\times ۴=۶۴$$

۲-۵-۸ جایگشت با اشیاء مکرر

تعداد جایگشت های n شیء که $n_۱$ شیء شبیه هم، $n_۲$ شیء شبیه هم و ... و n_k شیء شبیه هم هستند برابر است با:

$$\frac{n!}{n_1!n_2!...n_k!}=\binom{n}{n_1,n_2,...,n_k}$$

که درآن:

$$\sum_{i=1}^{k} n_i = n$$

در حقیقت n شیء داریم که $n_۱$ تا شبیه هم هستند، $n_۲$ تا شبیه هم هستند و الی آخر. اگر n شیء متمایز بودند از جابجایی $n_۱,n_۲,...,n_k$ شیء تبدیل‌های جدیدی به تعداد $n_۱!n_۲!...n_k!$ بدست می‌آمد ولی چون اشیا متمایز نیستند این تبدیل‌ها به یک صورت تلقی می‌شوند. لذا کل جایگشت مکرر برابر است با:

$$\frac{n!}{n_1!n_2!..n_k!}=\binom{n_1}{n_1,n_2,...,n_k}$$

مثال ۲-۵-۹ هشت پرچم به طور عمودی آویزان هستند. به طوریکه ۴ تای آنها قرمز، ۳ تای آنها آبی و یک پرچم سفید است. با این مجموعه پرچمها چه تعداد علامت مختلف می‌توان بوجود آورد؟

در این مثال، $n=۸$, $n_۱=۴, n_۲=۳, n_۳=۱$.و تکرار وجود دارد. لذا:

$$\binom{8}{4,3,1} = \frac{8!}{4!3!1!} = 280$$

مثال ۲-۵-۱۰ با اعداد ۹، ۷، ۷، ۴، ۴ چند عدد ۵ رقمی می‌توان نوشت؟

$\binom{5}{1,2,2} = \frac{5!}{1!\,2!\,2!} = 30$ ، $n=۵$, $n_1=۱$, $n_2=۲$, $n_3=۲$.

۲-۵-۱۱ جایگشت n شیء متمایز در محیط دایره

جایگشت n شیء متمایز روی محیط دایره برابر با $(n-۱)!$است.

مثال ۲-۵-۱۲ چند جایگشت دوری از ۵ ورزشکار که دور یک میز گرد نشسته‌اند وجود دارد؟ در این مثال $n=۵$ و $۲۴=(۵-۱)!$

۲-۶ ترکیب

هرگاه در جایگشت، آرایش و نظم اشیا کنار هم مورد توجه نباشد آن را ترکیب گویند. می‌دانیم تعداد جایگشت‌های ۲ حرفی A ، B، Cو D برابر با ۱۲ است و جایگشت‌ها عبارتند از:

$AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC$

باتوجه به فهرست بالا در می‌یابیم که چندین جفت از جایگشت‌ها بدون توجه به آرایش و نظم قرار گرفتنشان در کنار یکدیگر، به هم شبیه هستند. وقتی که جایگشت‌های مشابه را جور کنیم، خواهیم داشت:

$AB \quad AC \quad AD \quad BC \quad BD \quad CD$

$BA \quad CA \quad DA \quad CB \quad DB \quad DC$

اگر فرضاً بین جایگشت‌های ABو BAتمایزی قایل نشویم، تعداد جایگشت ۱۲ حالتی بالا به ۶ حالت تقلیل پیدا می‌کند.

$AB \quad AC \quad AD \quad BC \quad BD \quad CD$

۲-۶-۱ ترکیب rتایی، nشیء متمایز

ترکیب rتایی، nشیء متمایز برابر است با:

$$nC_r = \frac{n!}{r!(n-r)!}$$

در حقیقت از میان n شیء، r شیء انتخاب می‌کنیم که در جایگشت، این r شیء به $r!$ می‌توانند کنار هم قرار گیرند. از طرفی جایگشت‌های rتایی، nشیء برابر با $nP_r = \frac{n!}{(n-r)!}$ است. ولی در ترکیب هر $r!$ حالت، تنها به یک صورت نوشته می‌شود. زیرا جابجایی مطرح نیست. با یک تناسب ساده می‌توان نوشت.

تعداد ترکیب	تعداد جایگشت
۱	$r!$
x	nP_r

که نتیجه می‌شود:

$$x = \frac{nP_r}{r!} = \frac{n!}{r!(n-r)!} = nC_r$$

مثال ۲-۶-۲ به چند طریق می‌توان ازبین ۵ مهندس نرم‌افزار وسه مهندس سخت‌افزار یک کمیته سه نفری برای سایت رایانه تشکیل داد که شامل دو نفر مهندس نرم‌افزار و یک نفر مهندس سخت‌افزار باشد؟

تعداد ترکیبات ۲ از ۵ برابر با : $5C2 = \frac{5!}{2!3!} = 10$

تعداد ترکیبات ۱ از ۳ برابر با : $3C1 = \frac{3!}{1!\ 2!} = 3$

با استفاده از اصل دوم شمارش تعداد کل کمیته ها برابر با: ۱۰×۳=۳۰

۲-۶-۳ ترکیب rتایی، nشیء با تکرار اشیاء

فرض کنید سه حروف متمایز A ، B و C داریم و می‌خواهیم ترکیب ۲ تایی آنها را بدست آوریم، به طوری که هر حرف می‌تواند ۲ بار تکرار شود. ترکیب ۲ از ۳ برابر با

۳ و حالات برابر با AB ، AC و BC هستند اما هر حرفی می‌تواند ۲ بار تکرار شود مانند AA ، BB و CC که تعداد کل حالات برابر می‌شود با:

$$AB, AC, BC, AA, BB, CC$$

در حالت کلی ترکیبات rتایی، nشیء متمایز با تکرار اشیا به تعداد r از فرمول زیر بدست می‌آید.

$$n+r-1Cr = \frac{(n+r-1)!}{r!(n-1)!}$$

مثال ۲-۶-۴ تعداد ترکیبات دوتایی حروف A، B و C با تکرار مجاز ۲ برابر است با:

$$4C2 = \frac{4!}{2!2!} = 6, \qquad n=۳,\ r=۲,\ n+r-۱=۴$$

۲-۷ احتمال

در بخش‌های قبل فضای نمونه، پیشامد و شمارش اعضای فضای نمونه را مورد بررسی قرار دادیم. باتوجه به موارد ذکر شده برای ارائه مدل احتمال دو تعریف مفهوم کلاسیک و مفهوم فراوانی را برای احتمال ارائه می‌دهیم.

۲-۷-۱ مفهوم کلاسیک

فرض کنید فضای نمونه S دارای N برآمد هم شانس و پیشامد Aشامل n برآمد مورد نظر باشد. احتمال رخ داد پیشامد A به صورت زیر تعریف می‌شود و به علامت $P(A)$ نشان داده می‌شود.

$$P(A) = \frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات کل}} = \frac{n(A)}{n(S)} = \frac{n}{N}$$

که $n(A)$و $n(S)$ به ترتیب تعداد اعضای Aو S می‌باشند.

مثال ۲-۷-۲ کیسه‌ای شامل ۳ مهره سفید با برچسب های W_1، W_2و W_3 و ۴ مهره قرمز با برچسب‌های R_1، R_2 ، R_3 و R_4 است.

الف- فضای نمونه را بنویسید.

ب- اگر پیشامد A خارج شدن مهره سفید باشد اعضای A را بنویسید.

ج- اگر پیشامد B خارج شدن مهره قرمز باشد اعضای B را بنویسید.

د- احتمال رخ داد پیشامد A را حساب کنید.

هـ- احتمال رخ داد پیشامد B را حساب کنید.

$$S=\{W_1,W_2,W_3,R_1,R_2,R_3,R_4\}\ ,\ A=\{W_1,W_2,W_3\}\ ,\ B=\{R_1,R_2,R_3,R_4\}$$

$$P(A)=\frac{n(A)}{n(S)}=\frac{3}{7}\ ,\qquad P(B)=\frac{n(B)}{n(S)}=\frac{4}{7}$$

مثال ۲-۷-۳ در پرتاب تاس احتمال اینکه اعداد زوج ظاهر شوند چقدر است؟

$$S=\{۱,۲,۳,۴,۵\}\quad ,\quad A=\{۲,۴,۶\}$$

$$P(A)=\frac{n(A)}{n(S)}=\frac{3}{6}$$

عیب عمده مفهوم احتمال کلاسیک در محدود بودن قابلیت کاربردی آن است. زیرا وضعیت‌هایی وجود دارند که برای آنها حالاتی را که رخ می‌دهند نمی‌توان هم‌شانس دانست. به عنوان مثال، در یک منطقه در فصل بارندگی احتمال اینکه روزی به خصوصی بارانی باشد یا نباشد برابر با ۰/۵ نیست. یا در حسابهای بانکی شانس افراد برای برنده شدن یکسان نیست، بلکه متناسب با امتیازی که دارند می‌باشند.

۲-۷-۴ مفهوم فراوانی

احتمال یک پیشامد برابر با نسبت دفعاتی است که پیشامدهای از یک نوع در تکرار زیاد رخ خواهند داد، احتمال به مفهوم فراوانی تلقی می‌شود.

مثال ۲-۷-۵ سازمان هواشناسی اعلام می‌دارد که در یک منطقه در فصل بارندگی احتمال بارندگی در یک روز بخصوص برابر ۳۰ درصد است. عدد ۳۰ درصد براساس

تجربیات سالهای گذشته و شرایط جوی بدست آمده و از آن به عنوان یک الگو استفاده می‌شود.

مثال ۲-۷-۶ فرضاً کشوری اعلام می‌کند که در این کشور ۵۱ درصد از متولدین دختر هستند. عدد ۵۱ درصد براساس رکورد متولدین بیمارستانها بدست آمده است. چون همه می‌دانیم که شانس پسر بودن یا دختر بودن ۵۰ درصد است.

مثال ۲-۷-۷ در بازرسی یک دسته ابزار فنی فراوانی نسبی ابزارهای سالم ۰/۹۰ تعیین شده است. در صورتی که تعداد کل ابزارها ۲۰۰ باشند، تعداد ابزارهای سالم را بیابید.

اگر A پیشامد سالم بودن فرض شود، $P(A)=۰/۹۰$ ، $n(A)=n.P(A)=۱۸۰$

در مواقعی که فضای نمونه و پیشامد روی خط، صفحه یا فضا تعریف می‌شوند، احتمال را به صورت احتمال هندسی تعریف می‌کنند.

قطعه خط l را که قسمتی از قطعه خط L است درنظر بگیرید، نقطه ای به تصادف از قطعه خط انتخاب می‌کنیم. اگر احتمال قرار گرفتن نقطه روی قطعه خط l با طول آن متناسب باشد و به محل قرار گرفتن قطعه خط l روی خط L بستگی نداشته باشد در این صورت احتمال قرار گرفتن نقطه مزبور روی قطعه خط l برابر با:

$$P=\frac{\text{طول خط } l}{\text{طول خط } L}=\frac{l}{L}$$

به طور مشابه، در صفحه به صورت زیر تعریف می‌شود.

$$P=\frac{\text{مساحت } \mathrm{g}}{\text{مساحت } \mathrm{G}}$$

به همین ترتیب احتمال قرار گرفتن نقطه در داخل شکل فضای v که قسمتی از شکل فضای V است برابر با:

$$P=\frac{\text{حجم } \mathrm{v}}{\text{حجم } \mathrm{V}}$$

مثال ۲-۷-۸ روی قطعه خط به طول ۵ سانتی متر، قطعه خطی به طول ۲ سانتی متر علامت می‌زنیم. نقطه ای به تصادف از قطعه خط انتخاب می کنیم. احتمال اینکه نقطه انتخاب شده روی قطعه خط علامت‌دار باشد چقدر است؟

$$P = \frac{\text{طول قطعه خط علامت دار}}{\text{طول کل خط}} = \frac{2}{5} = 0.4$$

مثال ۲-۷-۹ دایره‌ای به شعاع r در داخل دایره به شعاع R قرار دارد. نقطه‌ای به تصادف انتخاب می‌شود. احتمال اینکه نقطه انتخابی در دایره کوچک باشد چقدر است؟

$$P = \frac{\text{مساحت دایره کوچک}}{\text{مساحت دایره بزرگ}} = \frac{\pi r^2}{\pi R^2} = \left(\frac{r}{R}\right)^2$$

۲-۸ تابع احتمال

فرض کنید S فضای نمونه و $A_\emptyset$، A_1، A_2 و ... . پیشامدهای دوبه دو ناسازگار از آن باشند.

تابعی را که به هر پیشامد عددی در بازه (۰،۱) نسبت دهد و در سه اصل زیر صدق کند تابع احتمال گویند.

اصل اول: احتمال هر پیشامد بزرگتر یا مساوی صفر است.

$P(A) \geq 0 \quad , \quad \forall A \subset S$

اصل دوم: احتمال فضای نمونه S برابر با ۱ می‌باشد.

$P(S) = 1$

اصل سوم:

$$P[A_1 \cup A_2 \cup \ldots] = \sum_{i=1} P(A_i) = P(A_1) + P(A_2) + \ldots$$

مثال ۲-۸-۱ در پرتاب یک سکه مجموعه توان را بنویسید و سه اصل تابع احتمال را مرور کنید. در پرتاب سکه $S = \{H, T\}$ و مجموعه توان برابر با:

$B = \{\phi, \{H\}, \{T\}, \{H, T\}\}$

۱- اگر $A=\phi$ یا $A=\{H\}$ یا $A=\{T\}$ یا $A=\{H,T\}$ باشد آنگاه

$$P(\phi)=\frac{1}{4},\ P(\{H\})=\frac{1}{4}\ ,\ P(\{T\})=\frac{1}{4}\quad ,\ P(\{H,T\})=\frac{1}{4}$$

که همگی بزرگتر یا مساوی صفر هستند.

۲- $P(B)=۱$

۳- اگر $A_۱=\{\phi\}$، $A_۲=\{H\}$ ، $A_۳=\{T\}$ و $A_۴=\{H,T\}$ باشند. آنگاه،

$$P(A_4)=\frac{1}{4},\ P(A_3)=\frac{1}{4}\ ,\ P(A_2)=\frac{1}{4}\quad ,\ P(A_1)=\frac{1}{4}$$

و

$$P(A_1\cup A_2)=P(A_1)+P(A_2)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$
$$P(A_1\cup A_2\cup A_3)=\sum_{i=1}^{3}P(A_i)=\frac{3}{4}$$

۲-۹ قوانین احتمال

قضیه ۲-۹-۱ اگر ϕ مجموعه تهی باشد آنگاه $P(\phi)=0$

برهان: می‌دانیم $S\cup\phi=S$ و S و ϕ دو مجموعه مجزا هستند. یعنی $S\cap\phi=\phi$ طبق اصل دوم و سوم.

$$P(S\cup\phi)=P(S)+P(\phi)$$
$$1=1+P(\phi)\Rightarrow P(\phi)=0$$

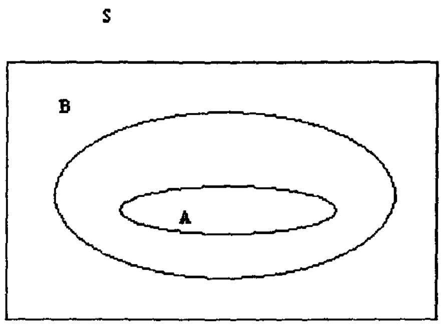
قضیه ۲-۹-۲ اگر A^C متمم پیشامد A باشد آنگاه $P(A^C)=۱-P(A)$

برهان: می دانیم $A\cap A^c=\phi$ و $A\cup A^c=S$ پس

$$P(A\cup A^c)=P(S)\ ,\quad P(A)+P(A^c)=1\ \Rightarrow P(A)\equiv 1-P(A^c)$$

قضیه ۲-۹-۳ اگر $A\subset B$ باشد آنگاه $P(A)\le P(B)$

برهان: اگر $A\subset B$ باشد B را می‌توان به صورت دو پیشامد مجزای A و $B\cap A^c$ نوشت.



$$B = A \cup (B \cap A^c)$$
$$P(B) = P(A) + P(B \cap A^c)$$

طبق اصل اول احتمال $P(B \cap A^c) \geq 0$ است. اگر آن را از طرف راست رابطه اخیر حذف کنیم نتیجه می‌شود $P(A) \leq P(B)$.

قضیه ۲-۹-۴ اگر A یک پیشامد باشد آنگاه $0 \leq P(A) \leq ۱$.

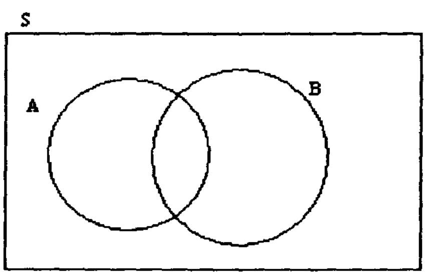
برهان: چون $\phi \subset A \subset S$ طبق قضیه ۲-۹-۳ داریم:

$$P(\phi) \leq P(A) \leq P(S) \qquad , 0 \leq P(A) \leq ۱$$

قضیه ۲-۹-۵ اگر A، B دو پیشامد دلخواه در S باشند آنگاه

$$P(A \cap B^c) = P(A) - P(A \cap B)$$

برهان: پیشامد A را می‌توان به دو پیشامد مجزای $A \cap B^c$ و $A \cap B$ تجزیه کرد.



$$A = (A \cap B^c) \cup (A \cap B)$$
$$P(A) = P(A \cap B^c) + P(A \cap B)$$
$$P(A \cap B^c) = P(A) - P(A \cap B)$$

قضیه ۲-۹-۶ اگر A و B دو پیشامد دلخواه در S باشند آنگاه

$$\boxed{P(A\cup B)=P(A)+P(B)-P(A\cap B)}$$

برهان: پیشامد $A\cup B$ را می‌توان به دو پیشامد مجزای $A\cap B^c$ و B تجزیه کرد.

$$A\cup B=B\cup(A\cap B^c)$$

$$P(A\cup B)=P(B)+P(A\cap B^c)$$

طبق قضیه ۲-۹-۵

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

مثال ۲-۹-۷ یک تاس دوبار پرتاب می‌شود. احتمال اینکه مجموع اعداد ظاهر شده بر ۲ یا ۳ بخش پذیر باشد چقدر است؟

فرض کنید A پیشامد اینکه مجموع اعداد بر ۲ بخش پذیر باشد:

A ={(۱،۱),(۳،۱),(۵،۱),(۲،۲),(۴،۲),(۶،۲),(۱،۳),(۳،۳),(۵،۳),(۲،۴),(۴،۴),(۶،۴),(۱،۵),
(۳،۵),(۵،۵),(۲،۶),(۴،۶),(۶،۶)}

و B پیشامد اینکه مجموع اعداد بر ۳ بخش پذیر باشد.

B ={(۲،۱),(۵،۱),(۱،۲),(۴،۲),(۳،۳),(۶،۳),(۲،۴),(۵،۴),(۱،۵),(۴،۵),(۳،۶),(۶،۶) }

$A\cap B$ ={(۵،۱),(۴،۲),(۳،۳),(۲،۴),(۱،۵),(۶،۶)}

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$
$$=\frac{18}{36}+\frac{12}{36}-\frac{6}{36}=\frac{24}{36}$$

قضیه ۲-۹-۸ اگر A ، B و C پیشامدهای دلخواه در S باشند آنگاه:

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$$
$$-P(B\cap C)+P(A\cap B\cap C)$$

برهان: به عهده خواننده

مثال ۲-۹-۹ در یک نظرخواهی برای بررسی سه شبکه تلویزیونی ۱، ۲ و ۳ از منطقه‌ای مسکونی که ۲۰۰۰ خانوار دارد اطلاعات زیر بدست آمده است:

۱- ۱۲۰۰ خانوار شبکه ۱ را تماشا می‌کنند.
۲- ۱۱۰۰ خانوار شبکه ۲ را تماشا می‌کنند.
۳- ۸۰۰ خانوار شبکه ۳ را تماشا می‌کنند.
۴- ۸۶۵ خانوار هم شبکه ۱ و هم شبکه ۲ را تماشا می‌کنند.
۵- ۴۵۰ خانوار هم شبکه ۱ و هم شبکه ۳ را تماشا می‌کنند.
۶- ۴۰۰ خانوار هم شبکه ۲ و هم شبکه ۳ را تماشا می‌کنند.
۷- ۱۰۰ خانوار هر سه شبکه را تماشا می‌کنند.

احتمال اینکه یک خانوار که به تصادف انتخاب می‌شود، حداقل یکی از شبکه را تماشا کند چقدر است؟

اگر تعریف کنیم، A پیشامد اینکه خانوارها شبکه ۱، B پیشامد اینکه خانوارها شبکه ۲ و C پیشامد اینکه خانوارها شبکه ۳ را تماشا می‌کنند، داریم:

$$P(A)=\frac{1200}{2000}\ ,\ P(B)=\frac{1100}{2000}\ ,\ \ P(C)=\frac{800}{2000}$$

$$P(A\cap B)=\frac{865}{2000}\ ,\ P(A\cap C)=\frac{450}{2000}\ ,\ \ P(B\cap C)=\frac{400}{2000}$$
$$P(A\cap B\cap C)=\frac{100}{2000}$$

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$$
$$-P(B\cap C)+P(A\cap B\cap C)$$

$$=\frac{1200}{2000}+\frac{1100}{2000}+\frac{800}{2000}-\frac{865}{2000}-\frac{450}{2000}-\frac{400}{2000}+\frac{100}{2000}=0.742$$

در بسیاری مواقع محاسبه احتمال وقوع پیشامد A مدنظر است، مشروط بر اینکه از وقوع پیشامد B آگاهی کافی داریم. برای محاسبه احتمال شرط A به شرط B احتمال شرطی را تعریف می‌کنیم.

۲-۱۰ احتمال شرطی

احتمال شرطی پیشامد A به شرط وقوع پیشامد B به صورت زیر تعریف می‌شود.

$$\boxed{P(A \mid B)=\frac{P(A\cap B)}{P(B)}} \quad , \qquad P(B)>0$$

در این تعریف، پیشامد B به عنوان فضای نمونه تحویل یافته است که فقط برآمدهایی را که در B است درنظر می‌گیریم. چون $P(A|B)$ برابر است با احتمال اینکه پیشامد A اتفاق بیفتد به شرطی که فضای نمونه جدید برابر با B باشد.

نکته ۲-۱۰-۱ اگر $P(A)\neq 0$ از $P(A \mid B)=\frac{P(A\cap B)}{P(B)}$ نتیجه می‌شود که:

$$P(A\cap B)=P(A \mid B)P(B)=P(B \mid A)P(A)$$

نکته ۲-۱۰-۲ اگر $P(B)=0$ باشد نتیجه می‌شود که $P(A\cap B)=0$ در این صورت $P(A|B)$ تعریف نشده است.

نکته ۲-۱۰-۳ اگر پیشامدهای A_1 ، A_2 ، ... ، A_k دوبه دو مجزا باشند. احتمال شرطی $\bigcup_{i=1}^{k} A_i$ به شرط B برابر با:

$$P\left[\bigcup_{i=1}^{k} A_i \mid B\right]=\frac{P\left(\bigcup_{i=1}^{k} A_i\cap B\right)}{P(B)}=\frac{\sum_{i=1}^{k} P(A_i\cap B)}{P(B)}$$

مثال ۲-۱۰-۴ در یک کلاس دانشگاهی که از ۶۰ دختر و ۴۰ پسر تشکیل شده است. مشاهده می‌شود که ۲۴ دختر و ۱۶ پسر عینکی هستند. احتمال اینکه یک دانشجو به تصادف انتخاب شود، عینکی باشد به شرط آنکه بدانیم پسر است چقدر است؟

B پیشامد اینکه پسر باشد $\qquad P(B)=\frac{40}{100}$

$A\cap B$ پیشامد اینکه شخصی انتخاب شده عینکی و پسر باشد $\qquad P(A\cap B)=\frac{16}{100}$

$$P(A \mid B)=\frac{P(A\cap B)}{P(B)}=\frac{16}{40}=0.4$$

مثال ۲-۱۰-۵ ظرفی شامل ۶ ژتون سفید و ۴ ژتون آبی است. دو ژتون پشت سرهم و بدون جایگذاری انتخاب می‌شود. احتمال اینکه اولین و دومین ژتون خارج شده هردو سفید باشند چقدر است؟

B: پیشامد اینکه اولین ژتون خارج شده سفید باشد. $P(B)=\frac{6}{10}$

$A|B$: پیشامد اینکه دومین ژتون خارج شده سفید باشد به شرطی که اولین ژتون خارج شده سفید بوده باشد. $P(A|B)=\frac{5}{9}$

$A\cap B$: پیشامد اینکه اولین و دومین ژتون هردو سفید باشد.

$$P(A\cap B)=P(A|B)P(B)=\frac{6}{10}\times\frac{5}{9}=\frac{1}{3}$$

۲-۱۱ دو پیشامد مستقل

دو پیشامد A و B را مستقل گوییم، اگر رخ داد یکی تأثیری در دیگری نداشته باشد. یعنی $P(A|B)=P(A)$ ، $P(B|A)=P(B)$ ، بنابراین A و B مستقل‌اند اگر :

$$P(A\cap B)=P(A|B)P(B)=P(B|A)P(A)=P(A).P(B)$$

مثال ۲-۱۱-۱ بسته ای حاوی ۱۰ رایانه است که چهار تا از آنها معیوب هستند. ۲ رایانه متوالیاً و با جایگذاری خارج می کنیم. احتمال اینکه هردو رایانه معیوب باشد چقدر است؟

A : پیشامد اینکه رایانه اول معیوب باشد، B پیشامد اینکه رایانه دوم معیوب باشد.

$$P(A)=\frac{4}{10}\ ,\qquad P(B)=\frac{4}{10}\ ,\qquad P(A\cap B)=P(A)P(B)=\frac{4}{10}\times\frac{4}{10}=0.16$$

مثال ۲-۱۱-۲ یک سکه دوبار متوالیاً پرتاب می‌شود. احتمال اینکه در ۲ پرتاب شیر ظاهر شود چقدر است؟

$$\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

قضیه ۲-۱۱-۳ اگر دو پیشامد A و B مستقل باشند آنگاه A^c و B نیز مستقل‌اند.

برهان: $A \cup A^c = S$, $B = (A \cap B) \cup (A^c \cap B)$

$$P(B) = P(A \cap B) + P(A^c \cap B)$$
$$= P(A)P(B) + P(A^c \cap B)$$

$$P(A^c \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^c)$$

قضیه ۲-۱۱-۴ پیشامدهای A_1 ، A_2 ، ...، A_k مستقل‌اند اگر و تنها اگر احتمال اشتراک هر ۲، ۳، ...، k تا از این پیشامدها مساوی حاصلضرب احتمالهای مربوطه به هر پیشامد باشد.

برای استقلال سه پیشامد A_1 ، A_2 و A_3 لازم است که :

1) $P(A_1 \cap B_2) = P(A_1)P(A_2)$, $P(A_1 \cap A_3) = P(A_1)P(A_3)$, $P(A_2 \cap A_3) = P(A_2)P(A_3)$

2) $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

قضیه ۲-۱۱-۵ اگر احتمال وقوع پیشامد A_1 برابر P_1 و احتمال وقوع پیشامد A_2 برابر P_2 و دو پیشامد A_1 و A_2 مستقل باشند آنگاه احتمال اینکه فقط یکی از آنها اتفاق بیفتد برابر است با: $\boxed{P_1(1-P_2)+(1-P_1)P_2}$

برهان: رخداد پیشامد A_1 برابر با رخ داد پیشامد A_1 اشتراکش با A_2^c و رخداد پیشامد A_2 برابر با رخداد پیشامد A_2 اشتراکش با A_1^c است. پس:

$A_1 = A_1 \cap A_2^c$, $A_2 = A_1^c \cap A_2$

چون A_1 و A_2 مستقل‌اند و مجزا هستند.

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = P(A_1 \cap A_2^c) + P(A_1^c \cap A_2)$$
$$= P(A_1)P(A_2^c) + P(A_1^c)P(A_2)$$
$$= P_1(1-P_2) + (1-P_1)P_2$$

مثال ۲-۱۱-۶ دو آژیر در ساختمانی برای آگاهی از خطر نصب شده است. این دو آژیر به طور مستقل کار می کنند و احتمال آنکه در مواقع خطر آژیر اول بکار افتد ۰/۹۵ و آژیر دوم کار کند ۰/۹ است. احتمال آنکه در موقع خطر فقط یکی از آنها کار کند چقدر است؟

A_i: پیشامد اینکه آژیر i ام به کار بیفتد. $P(A_1) = 0.95 = P_1 \quad P(A_2) = 0.90 = P_2$

$$P(A_1 \cup A_2) = P_1(1-P_2) + (1-P_1)P_2 = 0.95 \times 0.1 + 0.05 \times 0.90 = 0.14$$

قضیه ۲-۱۱-۷ (قانون جمع احتمالات) فرض کنید پیشامدهای A_1 ، A_2 ، ....، A_k پیشامدهای دو به دو مجزا از هم و اجتماع آنها S باشد و A یک پیشامد دلخواه از S باشد آنگاه:

$$\boxed{P(A) = \sum_{i=1}^{k} P(A \mid A_1)P(A_i)}$$

برهان:

$$A \cap S = A \quad , \quad \bigcup_{i=1}^{k} A_i = S$$

$$A = A \cap \left(\bigcup_{i=1}^{k} A_i\right) = A \cap (A_1 \cup A_2 \cup .... \cup A_k)$$
$$= (A \cap A_1) \cup (A \cap A_2) \cup .... \cup (A \cap A_k)$$

پیشامدهای طرف راست رابطه اخیر دوبه دو مجزا هستند. طبق اصل سوم

$$P(A) = P(A \cap A_1) + P(A \cap A_2) + ... + P(A \cap A_k)$$
$$= \sum_{i=1}^{k} P(A \cap A_i) = \sum_{i=1}^{k} P(A \mid A_i)P(A_i)$$

مثال ۲-۱۱-۸ توپ سفید رنگی را در داخل کیسه‌ای که حاوی دو توپ دیگر است می‌اندازیم و بعد به تصادف یک توپ خارج می‌کنیم. احتمال اینکه توپ خارج شده سفید باشد چقدر است؟

A : پیشامد اینکه توپ خارج شده سفید باشد.

A_1 : پیشامد اینکه هیچ یک از دو توپ قبلی سفید نباشد.

A_2 : پیشامد اینکه یکی از توپهای قبلی سفید باشد.

$A_۳$: پیشامد اینکه هردو توپ قبلی سفید باشد.

$$A_1 \cup A_2 \cup A_3 = S \ , \quad A \subset S \ , \quad P(A_1) = \frac{1}{3} \ , \quad P(A_2) = \frac{2}{3} \ , P(A_3) = \frac{3}{3} \ ,$$

$$P(A \mid A_1) = \frac{1}{3} \ , \quad P(A \mid A_2) = \frac{1}{3} \ , \quad P(A \mid A_3) = \frac{1}{3}$$

$$P(A) = \sum_{i=1}^{3} P(A \mid A_i) P(A_i) = \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} + \frac{1}{3} \times \frac{3}{3} = \frac{2}{3}$$

۲-۱۲ فرمول بیز

اگر پیشامدهای $A_۱$ ، $A_۲$ ، ...، A_k دو به دو مجزا و $\bigcup_{i=1}^{k} A_i = S$ باشد احتمال شرطی هریک از A_i ها به شرط اتفاق پیشامد A از S برابر با:

$$P[A_i \mid A] = \frac{P(A_i \cap A)}{P(A)}$$

با توجه به فرمول قضیه ۲-۱۱-۷

$$P[A_i \mid A] = \frac{P(A \mid A_i) P(A_i)}{\sum_{i=1}^{k} P(A \mid A_i) P(A_i)}$$

مثال ۲-۱۲-۱ دو برنامه‌نویس، برنامه‌ای را توسط دو دستگاه مختلف پانچ می‌کنند. احتمال اینکه برنامه‌نویس اول اشتباه کند ۰/۰۵ و این احتمال برای دومی ۰/۱ است. هنگامی که کارتهای پانچ شده را به ماشین می‌دهیم، متوجه می‌شویم که اشتباهی رخ داده است. احتمال اینکه این اشتباه از برنامه‌نویس اول باشد چقدر است؟ در صورتی که بدانیم هردو دستگاه پانچ درست کار می‌کنند.

A : پیشامد اینکه اشتباهی رخ داده باشد.

$A_۱$: پیشامد اینکه برنامه نویس اول اشتباه کرده باشد.

$A_۲$: پیشامد اینکه برنامه نویس دوم اشتباه کرده باشد.

$$P(A_1) = P(A_2) = \frac{1}{2} \quad , \quad P(A \mid A_1) = 0.05 \ , \quad P(A \mid A_2) = 0.1$$

$\frac{/۰۵}{/۱۵} = \frac{۱}{۳}$

$$P(A) = \sum_{i=1}^{2} P(A \mid A_i) P(A_i) = 0.05 \times \frac{1}{2} + 0.1 \times \frac{1}{2} = 0.075$$

$$P(A_1 \mid A) = \frac{P(A \mid A_1)P(A_1)}{P(A)} = \frac{\frac{1}{2} \times 0.05}{0.075} = \frac{1}{3}$$

مثال ۲-۱۲-۲ برنامه‌های نرم‌افزاری یک شرکت مهندسی توسط سه برنامه نویس نوشته می‌شود به طوری که سهم برنامه‌نویس اولی و دوم به ترتیب برابر با ۴۰٪ و ۲۰٪ می‌باشد. اگر اشتباهات این سه برنامه‌نویس به ترتیب ۲٪، ۵٪ و ۴٪ باشد. مطلوبست:

الف- احتمال آنکه یک برنامه انتخاب شده دارای اشتباه باشد.

ب- احتمال آنکه برنامه انتخاب شده دارای اشتباه باشد و توسط برنامه نویس سوم نوشته شده باشد.

A_i: پیشامد اینکه برنامه توسط برنامه نویس iام نوشته شده باشد.

$$P(A_1) = 0.4 \quad , \quad P(A_2) = 0.2 \quad , \quad P(A_3) = 1 - P(A_1) - P(A_2) = 0.4$$

A : پیشامد اینکه برنامه انتخاب شده دارای اشتباه باشد.

$$P(A \mid A_1) = 0.02 \, , \quad P(A \mid A_2) = 0.05 \quad , \quad P(A \mid A_3) = 0.04$$

$$P(A) = \sum_{i=1}^{3} P(A \mid A_i)P(A_i) = 0.02 \times 0.4 + 0.05 \times 0.2 + 0.04 \times 0.4 = 0.034$$

ب-

$$P(A_3 \mid A) = \frac{P(A_3 \cap A)}{P(A)} = \frac{P(A \mid A_3)P(A_3)}{P(A)} = \frac{0.04 \times 0.4}{0.034} = 0.471$$

خودآزمایی

۱- اگر $A=\{x|\ 1<x<9\}$ و $B=\{y|\ y<5\}$ باشد. مقادیر A^c، $A\cup B$، $A\cap B$، $(A\cup B)^c$، $A-B$ را پیدا کنید.

۲- یک سکه سه بار پرتاب می‌شود.
الف- فضای نمونه را بنویسید.
ب- پیشامد اینکه در سه پرتاب دو شیر ظاهر شود چیست؟
ج- پیشامد اینکه حداقل یک خط ظاهر شود چیست؟

۳- یک آزمایش شامل انداختن یک سکه و در صورت مشاهده شیر پرتاب مجدد آن می‌باشد و اگر در پرتاب اول خط ملاحظه شود، این بار یک تاس انداخته می شود.
الف- فضای نمونه را با رسم نمودار درختی بنویسید.
ب- فضای نمونه S مربوط به اتفاق A که در وجه فوقانی تاس عددی کمتر از ۴ ظاهر شده باشد را معین کنید.

۴- الف- به چند طریق ۵ نفر می‌توانند صف جهت سوار شدن به اتوبوس تشکیل دهند؟
ب- اگر ۲ نفر از آنها قبول نکنند که پشت سرهم قرارگیرند چند طریق وجود دارد؟

۵- ۶ رایانه را به چند طریق می‌توان دور یک میز دایره شکل چید؟

۶- از یک گروه ۵ نفر مهندس نرم افزار و ۳ نفر مهندس سخت افزار چند کمیته ۳تایی می‌توان تشکیل داد اگر:
الف- محدودیتی نباشد.
ب- با ۲ مهندس نرم افزار و یک مهندس سخت‌افزار
ج- با یک مهندس نرم‌افزار و دو مهندس سخت‌افزار در صورتی که یک مهندس سخت‌افزار مشخصی باید در کمیته شرکت کند.

۷- بسته‌ای شامل ۹ دستگاه رایانه است که ۳ تا از آنها معیوب هستند. به چند طریق می‌توان ۴ دستگاه رایانه از این بسته انتخاب کرد که:

الف- حداقل ۳ دستگاه سالم باشد.

ب- حداکثر ۲ دستگاه معیوب باشد.

ج- فقط ۱ دستگاه معیوب باشد.

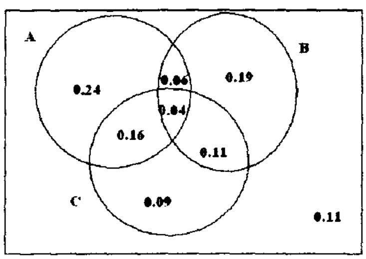
۸- پیشامد اینکه دقیقاً یکی از پیشامدهای A یا B رخ دهد را می‌توان به صورت $\left(A\cap B^c\right)\cup\left(A^c\cap B\right)$ بیان کرد، نشان دهید که:

$$P\left[(A\cap B^c)\cup(A^c\cap B)\right]=P(A)+P(B)-2P(A\cap B)$$

۹- الف- اگر A ، B و C سه پیشامد از فضای نمونه S باشند. نشان دهید که :

$$P\left[A\cup B\cup C\right]=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$

ب- فرمول بدست آمده در الف را برای نمودار وِن زیر امتحان کنید.



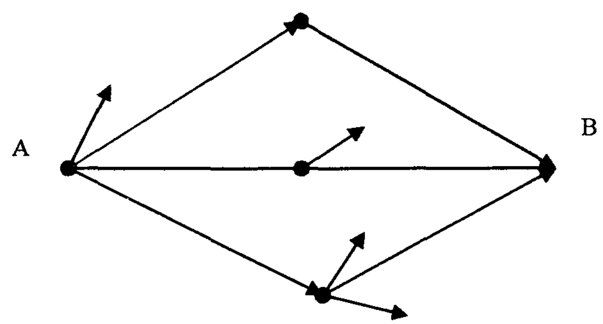
۱۰- فرض کنید $P(A)=P(B)=\frac{1}{3}$ و $P(A\cap B)=\frac{1}{10}$ باشد. احتمال‌های زیر را حساب کنید.

$P(A^c)$, $P(A\cup B^c)$, $P(B\cap A^c)$, $P(A^c\cup B^c)$

۱۱- احتمال اینکه یک ایستگاه تلویزیون بعد از نمایش یک برنامه مناظره ۰ ، ۱، ۲، ۳، ۴، ....، ۸ و یا لااقل ۹ شکایت دریافت کند به ترتیب ۰/۰۱، ۰/۰۳، ۰/۰۷، ۰/۱۵، ۰/۱۹، ۰/۱۸، ۰/۱۴، ۰/۱۲، ۰/۰۹، ۰/۰۲ است. احتمال های زیر را حساب کنید.

الف- بیشتر از ۴ شکایت دریافت کند.
ب- لااقل ۶ شکایت دریافت کند.
ج- از ۵ تا ۸ شکایت دریافت کند.
د- شکایتی دریافت نکند.

۱۲- شخصی می‌خواهد از نقطه A به نقطه B برود. ولی راه‌های مختلف اتصال این دو نقطه را نمی‌داند و بسته به راهی که انتخاب می‌کند ممکن است به نقطه B برسد یا نه. اگر راه‌های اتصال به شکل زیر باشد احتمال اینکه این شخص به نقطه B برسد چقدر است؟



۱۳- یک مؤسسه تحقیقاتی رایانه‌های مورد نیاز خود را با مدلهای ۱، ۲ و ۳ از چهار شرکت A ، B ، C و D خریداری می‌کند. آمار رایانه‌های خریداری شده در جدول زیر ثبت شده است.

مدل	A	B	C	D	جمع
۱	۲۰	۳۰	۱۵	۵	۷۰
۲	۱۰	۱۵	۱۰	۱۰	۴۵
۳	۱۵	۲۰	۲۵	۱۰	۷۰
جمع	۴۵	۶۵	۵۰	۲۵	۱۸۵

یک رایانه به تصادف از مؤسسه انتخاب می‌شود مطلوبست:
الف- احتمال اینکه رایانه مربوط به شرکت A باشد.
ب- احتمال اینکه رایانه مربوط به شرکت B یا C باشد.
ج- احتمال اینکه رایانه مربوط به شرکت D و دارای مدل ۳ باشد.

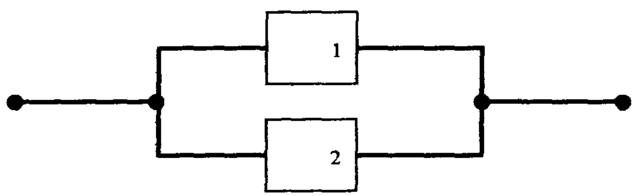
د- احتمال اینکه رایانه مربوط به شرکت D باشد به طوری می‌دانیم دارای مدل۲ است.

۱۴- در یک کلاس ۳۵ درصد دانشجویان مرد هستند و ۲۰ درصد از مردان و ۲۵ درصد زنان مردود شده‌اند. دانشجویی به تصادف انتخاب می شود. اگر این دانشجو مردود شده باشد. احتمال اینکه مرد باشد چقدر است؟

۱۵- یک سیستم مهندسی دارای دو عامل است که به طور مستقل از هم عمل می‌کنند. اگر ۰/۱ = (عامل اول خراب شود) P، ۰/۲ = (عامل دوم خراب شود) P احتمال اینکه سیستم در موقعیتهای زیر خراب نشود چقدر است؟

الف- دو عامل سری هستند.

ب- دو عامل موازی هستند.



۱۶- جعبه‌ای شامل ۲۰ الگوریتم است که ۵ تا از آنها ناقص است. اگر به تصادف ۳ الگوریتم متوالیاً و بدون جایگذاری از این جعبه انتخاب کنیم. احتمال اینکه هر ۳ الگوریتم دارای نقص باشد چقدر است؟

۱۷- احتمال اینکه در یک اندازه‌گیری کمیت فیزیکی خطا از حدی معین تجاوز کند ۰/۴ است. برای اندازه‌گیری این کمیت ۳ آزمایش مستقل انجام گرفته است. احتمال اینکه فقط در یکی از این آزمایس خطا از حد مجاز تجاوز کند چقدر است؟

۱۸- یک شرکت سازنده رایانه، رایانه‌های با مدلهای ۱، ۲، ۳ با نسبتهای ۲۰٪ ، ۳۰٪ و ۵۰٪ تولید می‌کند. به طوری که به ترتیب ۵٪، ۳٪ و ۲٪ از مدلهای تولید شده معیوب هستند. یک رایانه از تولیدات انتخاب می‌کنیم.

الف- احتمال اینکه معیوب باشد چقدر است؟

ب- می‌دانیم رایانه معیوب است. احتمال اینکه از مدل ۱ یا ۳ باشد چقدر است؟

۱۹- یک سایت رایانه‌ای دارای سه رایانه فعال می‌باشد که مستقل از یکدیگر عمل می‌کنند. احتمال اینکه یکی از آنها از کار بیافتد برابر ۰/۰۲۵ است. احتمال اینکه فقط یکی از آنها فعال باشد. چقدر است؟

۲۰- ظرف شماره ۱ شامل ۳ مهره قرمز و ۲ مهره سفید و ظرف شماره ۲ شامل ۲ مهره قرمز و ۵ مهره سفید است. یک تاس پرتاب می‌شود اگر تاس ۵ یا ۶ ظاهر شود یک مهره از ظرف ۱ انتخاب می شود و در غیر این صورت از ظرف شماره ۲ مطلوبست:

الف- احتمال اینکه یک مهره انتخاب شده سفید باشد.

ب- اگر مهره سفید انتخاب شود، احتمال اینکه از ظرف اول باشد چقدر است؟

فصل ۳

توزیع متغیرهای تصادفی

مقدمه

در این فصل ضمن تعریف متغیر تصادفی و برحسب اینکه متغیر تصادفی از نوع گسسته یا پیوسته باشد یک قالب احتمال ارائه می‌شود در حالی که متغیر تصادفی یک بعدی است تابع چگالی احتمال و تابع توزیع و در حالتی که متغیر تصادفی دوبعدی است، تابع چگالی احتمال توام، تابع توزیع توام، تابع چگالی حاشیه‌ای و تابع توزیع شرطی تعریف می‌شود و در پایان با تعریف گشتاورها و تابع مولد احتمال، میانگین، واریانس، چولگی و برجستگی در جامعه تعریف می‌شود.

۳-۱ متغیر تصادفی

در فصل‌های قبل با متغیر و فضای نمونه آشنا شدیم، با فرض اینکه هر تجربه تصادفی دارای فضای نمونه S باشد با تدوین یک قانون یا مجموعه‌ای از قوانین می‌توان اعضای فضای نمونه را به وسیله اعداد یا زوج اعداد (X_1,X_2) یا به طور کلی‌تر با n گانه مرتب اعداد $(X_1,X_2,...,X_n)$ افراز کرد.

مثال ۳-۱-۱ در پرتاب دو سکه فضای نمونه را بنویسید.

$S=\{HH,HT,TH,TT\}$

اگر متغیر X را تعداد شیرها روی فضای نمونه S برای پرتاب دو سکه و A را فضای نمونه وابسته به X تعریف کنیم داریم،

$X = 0$ هم ارز است با داشتن دو خط یا $X(TT) = 0$

$X =$ ۱ هم ارز است با داشتن یک شیر و یک خط یا ۱ $= X(HT) = X(TH)$

$X =$ ۲ هم ارز است با داشتن دو شیر یا ۲ $= X(HH)$

و $A = \{$ ۰،۱،۲ $\}$

چون مقادیر X از تجربه تصادفی به تجربه تصادفی دیگر تغییر می‌کند یک متغیر تصادفی است. هرتابع X ، که بر هرعضو s $(s \in S)$ یک و تنها یک عدد حقیقی $X(s) = x$ نسبت دهد متغیر تصادفی است و دارای فضای A است.

$$A = \{x \mid X(s) = x \quad , \quad s \in S\}$$

ممکن است عناصر مجموعه S خودشان اعداد حقیقی باشند در این صورت می‌توان نوشت $X(s) = s$ بنابراین $A = S$. مثال بارز این نکته پرتاب یک تاس است. چون $s = \{$۱،۲،۳،۴،۵،۶$\}$ و اگر X وجهی که ظاهر می‌شود تعریف شود داریم:

$$X(S_i) = i \qquad i = \text{۱،۲،۳،۴،۵،۶}$$

$$A = \{x | X(S) = x, s \in S\} = \{\text{۱،۲،۳،۴،۵،۶}\}$$

فرض کنید متغیر تصادفی X روی فضای نمونه S تعریف شده و A فضای نمونه X باشد. احتمال رخداد پیشامد $A \in A$ ، هم ارز است با احتمال رخداد پیشامد S' ، $S' \in S$ به طوری که

$$S' = \{s \mid s \in S \quad , \quad X(s) \in A\}$$

یعنی

$$P(X \in A) = P_x(A) = P(S')$$

که $P_x(A)$ تابع احتمال خوانده می‌شود. تابع احتمال $P_x(A)$ شرایط سه اصل ۲–۸ را دارا می‌باشد. یعنی:

۱– $\quad P_x(A) = P(S') \geq 0$

۲- $P_x(A) = P(S) = ۱$

۳- $P_x(A_1 \cup A_2 \cup \dots) = \sum_{i=1} P_x(A_i) \quad , \quad A_i \cap A_j = \phi \quad i \neq j$

در پرتاب دو سکه با فضای نمونه $S = \{HH, HT, TH, TT\}$ اگر هر عضو S شانس $\frac{1}{4}$ داشته باشد و

$$S_1 = \{TT\} \ , \quad S_2 = \{TH\} \ , \quad S_3 = \{HT\} \ , \quad S_4 = \{HH\}$$

$$P(S_1) = \frac{1}{4} \ , \quad P(S_2) = \frac{1}{4} \ , \quad P(S_3) = \frac{1}{4} \ , \quad P(S_4) = \frac{1}{4}$$

اکنون این احتمال‌ها را به زبان ساده‌تر که حاوی متغیر X است میتوان بیان کرد.

$$P(X=0) = P(S_1) = \frac{1}{4} \quad , \quad P(X=1) = P(S_2 \cup S_3) = \frac{2}{4} \quad ,$$

$$P(X=2) = P(S_4) = \frac{1}{4}$$

یا

X	۰	۱	۲
$P(X=x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

این جدول پخش احتمال روی عناصر A ، فضای متغیر تصادفی X را نشان می‌دهد.

مثال ۳-۱-۲ در پرتاب سه سکه اگر X تعداد شیرها تعریف شود پخش احتمال روی $A = \{0, ۱، ۲، ۳\}$ عبارتست از:

X	۰	۱	۲	۳
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

000 001 010 011 100 110 101 111
0 1 1 2 3
1 3/8 3/8 3/8

در مثالهای ذکر شده فضای نمونه S گسسته و شمارا بود. اکنون مثالی ارائه می‌دهیم که فضای نمونه یک فاصله است.

اگر از کارخانه‌ای که باطری با طول عمر استاندارد ۱ سال تولید می‌کند یک باطری به تصادف انتخاب شود فضای نمونه برای باطری انتخاب شده یک فاصله است و عبارتست از:

$$S=\{ s|0< s <۱\}$$

تابع چگالی احتمال در این حالت به صورت زیر تعریف می‌شود:

$$P(S') = \int_{S'} dz \qquad , \qquad S' \subset S$$

اگر $S' = \{s \mid \frac{1}{2} < S < \frac{3}{4}\}$ باشد آنگاه

$$P(S') = \int_{\frac{1}{2}}^{\frac{3}{4}} dz = z\Big|_{\frac{1}{2}}^{\frac{3}{4}} = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

اگر متغیر تصادفی X را برابر با $X = X(S)+۱$ تعریف کنیم داریم:

X فضای نمونه $= A=\{ x|۱< x <۲\}$

$$P_X(A) = P(A) = \int_A dz$$

که برای $A = \{ x|۱ < x <۱/۵ \}$ ،

$$P_X(A) = P(A) = \int_1^{1.5} dx = \int_{S'} dz = \int_0^{0.5} dz = \frac{1}{2}$$

در آمار معمولاً تابع احتمال متغیر تصادفی X ، $P_X(A)$ بیشتر از فضای نمونه Sو تابع احتمال $P(S')$مورد توجه است. بنابراین به جای $P(S')$ از $P_X(A)$ استفاده می‌شود.

مثال ۳-۱-۳ اگر $P(A)$ تابع چگالی احتمال متغیر تصادفی X به صورت زیر تعریف شود.

$$P(A)=\int_A f(x)\,dx$$

بـــه طـــوری کـــه $f(x)=\frac{3x^2}{8}$ ، A={ x|0< x <۲} و $A_۱$={ x| $0<x<\frac{1}{2}$ } ، $A_۲$={ x|۱< x <۲} دو مجموعه از A باشند آنگاه:

$$P_{A_1}(X)=P(X\in A_1)=\int_{A_1} f(x)\,dx=\int_0^{\frac{1}{2}}\frac{3x^2}{8}dx=\frac{1}{64}$$

$$P_{A_2}(X)=P(X\in A_2)=\int_{A_2} f(x)\,dx=\int_1^{2}\frac{3x^2}{8}dx=\frac{7}{8}$$

برای محاسبه $P(A_1\cup A_2)$ ، چون $A_1\cap A_2=\phi$ است.

$$P(A_1\cup A_2)=P(A_1)+P(A_2)=\frac{1}{64}+\frac{7}{8}=\frac{57}{64}$$

متغیرهای تصادفی معمولاً برحسب مقادیری که مـی تواننـد اختیـار کننـد، دسـته‌بنـدی می‌شوند. باتوجه به فضای نمونه متغیرهای تصادفی دونوع پخش یا توزیـع، گســسته و پیوسته را مورد توجه قرار می‌دهیم.

۳-۲ متغیر تصادفی گسسته

فرض کنید متغیر تصادفی X دارای فضای نمونه یک بعدی A باشد. به طـوری کـه A گسسته و شمارا باشد. هرگاه بتوان تابع احتمـال $P(A)$ ($A\subset A$) را برحـسب تابع $f(x)$ به شکل زیر تعریف کرد:

$$P(A)=P(X\in A)=\sum_A f(x)$$

به طوری که $f(x)$ در دو شرط زیر صدق کند.

۱- $f(x)\geq 0 \quad , \quad \forall\, x\in A$

۲- $\sum_A f(x)=۱$

X را متغیر تصادفی از نوع گسسته و $f(x)$ را تـابع احتمـال یـا پخـش گسـسته X گویند.

مثال ۳-۲-۱ فرض کنید X یک متغیر تصادفی از نوع گسسته با فضای نمونه $A=\{0,۱,۲,۳\}$ و $f(x)=\frac{1}{8}\binom{3}{x}$ باشد. برای :

$\frac{3!}{x!(3-x)!}$

$A_۱=\{0\}$, $A_۲=\{۱\}$, $A_۳=\{۲\}$, $A_۴=\{۳\}$

$$P(A_1)=P(X\in A_1)=\sum_{A_1} f(x)=\sum_{\{0\}} f(x)=f(0)=\frac{1}{8}$$

$$P(A_2)=P(X\in A_2)=\sum_{A_2} f(x)=\sum_{\{1\}} f(x)=f(1)=\frac{3}{8}$$

$$P(A_3)=P(X\in A_3)=\sum_{A_3} f(x)=\sum_{\{2\}} f(x)=f(2)=\frac{3}{8}$$

$$P(A_4)=P(X\in A_4)=\sum_{A_4} f(x)=\sum_{\{3\}} f(x)=f(3)=\frac{1}{8}$$

اگر توجه داشته باشید $f(x)$ یک قالب احتمال یا تابع احتمال برای پرتـاب ۳ سـکه و متغیر تصادفی X تعداد شیرها در سه پرتاب است.

X	۰	۱	۲	۳
$f(x)=P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

۳-۳ متغیر تصادفی پیوسته

فرض کنید متغیر تصادفی X دارای فضای یک بعدی A باشـد. بـه طـوری کـه A پیوسته یا یک بازه از اعداد حقیقی باشد. هرگاه بتوان تابع احتمـال $P(A)$ ($A\subset A$) را برحسب تابع $f(x)$ به شکل زیر تعریف کرد:

$$P(A)=P(X\in A)=\int_A f(x)dx$$

به طوری که $f(x)$ در دو شرط زیر صدق کند.

۱- $f(x) \geq 0 \quad , \quad \forall\, x \in A$

۲- $\int_A f(x)\,dx = ۱$

X را متغیر تصادفی از نوع پیوسته و $f(x)$ را تابع چگالی یا پخش پیوسته X گویند.

مثال ۳-۳-۱ فرض کنید X یک متغیر پیوسته با فضای $A=\{x | ۱<x<۲\}$ و $f(x)=۱$ باشد. برای $A_۱=\{x | ۱<x<۱/۵\}$ ، $P(A_1)$ برابر است با:

$$P(A_1) = \int_{A_1} f(x)\,dx = \int_1^{1.5} dx = \frac{1}{2}$$

اگر توجه داشته باشید $f(x)$ یک تابع چگالی برای عمر باطری است.

مثال ۳-۳-۲ اگر متغیر پیوسته X دارای فضای $A=\{x | 0<x<\infty\}$ باشد مقدار c را طوری پیدا کنید که $f(x)=c\,e^{-2x} \;,\; x>0$ یک تابع چگالی احتمال بشود. $f(x)$ باید در دو شرط $f(x)>0$ و $\int_A f(x)dx = ۱$ صدق کند لذا:

$$\int_A f(x)dx = \int_0^\infty c\, e^{-2x}dx = 1 \Rightarrow \frac{1}{c} = \int_0^\infty e^{-2x}dx = \frac{1}{2}$$

e^{-2x}

در نتیجه $c=۲$ و $f(x)=2e^{-2x} \;,\; x>0$

۳-۴ تابع توزیع $F(x)$

فرض کنید متغیر تصادفی X ، با فضای A ، دارای تابع چگالی احتمال $f(x)$ باشد و هدف محاسبه احتمال پیشامد $X \leq x$ یا به عبارت دیگر محاسبه احتمال اینکه متغیر تصادفی X مقدار کمتر یا مساوی x را اختیار کند، باشد احتمال پیشامد $X \leq x$ را به صورت زیر تعریف می‌کنیم:

$$F(x) = P[X \leq x]$$

$F(x)$ را تابع توزیع یا تابع توزیع تجمعی X گویند.

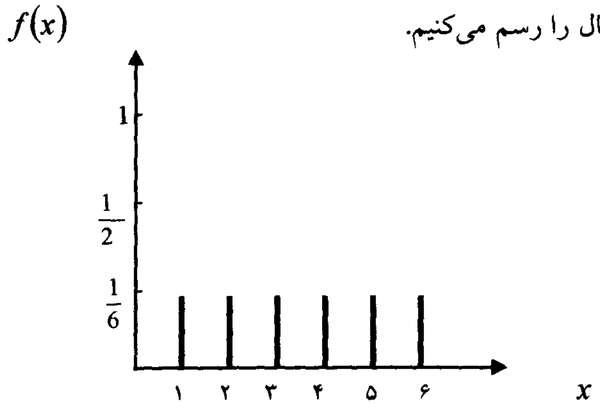
تابع توزیع متغیرهای تصادفی از نوع گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شوند.

$$\begin{cases} F(x)=P(X\le x)=\sum\limits_{t\le x} f(t) \\ \\ F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)\,dt \end{cases}$$

مثال ۳-۴-۱ فرض کنید متغیر تصادفی X از نوع گسسته و دارای تابع احتمال زیر باشد. تابع توزیع X را بدست آورید.

$$P(X=x)=f(x)=\frac{1}{6}\ , \qquad x=\text{۱،۲،۳،۴،۵،۶}$$

ابتدا نمودار تابع احتمال را رسم می‌کنیم.



$F(x)$ برای xهایی که کمتر از ۱ هستند به صورت زیر محاسبه می‌شود.

$$F(\text{۱})=P(X\le \text{۱})=\sum_{t\le 1} f(t)$$

اگر x کمتر از ۱ باشد یعنی $X<\text{۱}$، $f(x)$ برابر با صفر است چون $f(x)$ در این نقاط صفر است.

$$F(\text{۱})=P(X\le \text{۱}-\varepsilon)=\sum_{x<t} 0=0$$

که ε عدد کوچکی است.

$$F(\text{۱})=P(X\le \text{۱})=\sum_{t\le 1} f(t)=f(\text{۱})=\frac{1}{6}$$

$$F(۱)=P(X\leq ۱)=\sum_{t\leq 1} f(t)=f(۱)=\frac{1}{6}$$

$F(x)$ برای $۱\leq x<۲$

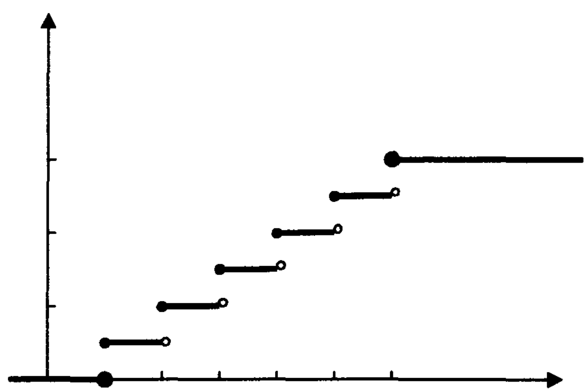
$$F(x)=P(X\leq x)=P(X=۱)+P(X=۲)=f(۱)+f(۲)=\frac{2}{6}$$

$F(x)$ برای $۲\leq x<۳$ ، در این فاصله x می‌تواند ۲ و کمتر از آن را اختیار کند.

$$F(۲)=P(X\leq ۲)=P(X=۱)+P(X=۲)=f(۱)+f(۲)=\frac{2}{6}$$

به همین ترتیب اگر حساب کنیم.

$$F(x)=\begin{cases} 0 & if \quad x<1 \\ \frac{1}{6} & if \quad 1\leq x<2 \\ \frac{2}{6} & if \quad 2\leq x<3 \\ \frac{3}{6} & if \quad 3\leq x<4 \\ \frac{4}{6} & if \quad 4\leq x<5 \\ \frac{5}{6} & if \quad 5\leq x<6 \\ 1 & if \quad x\geq 6 \end{cases}$$

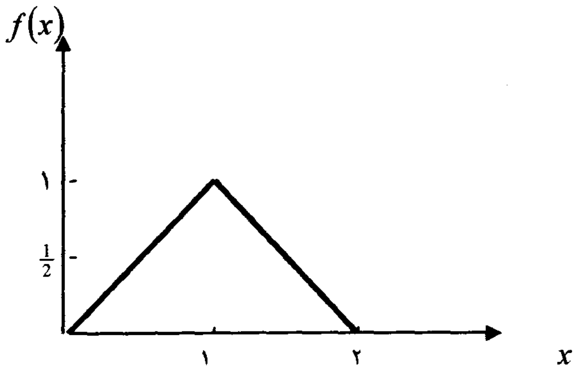


چنانچه در شکل می‌بینید $F(x)$ یک تابع پله‌ای است و در هر فاصله‌ای که شامل نقاط ۱، ۲، ۳، ۴، ۵، ۶ نباشد. مقداری است ثابت ولی در این نقاط دارای جهش‌های $\frac{1}{6}$ است. همچنین دیده می‌شود که $F(x)$ همه جا از راست پیوسته است.

مثال ۳-۴-۲ فرض کنید متغیر تصادفی X از نوع پیوسته دارای تابع چگالی زیر باشد.

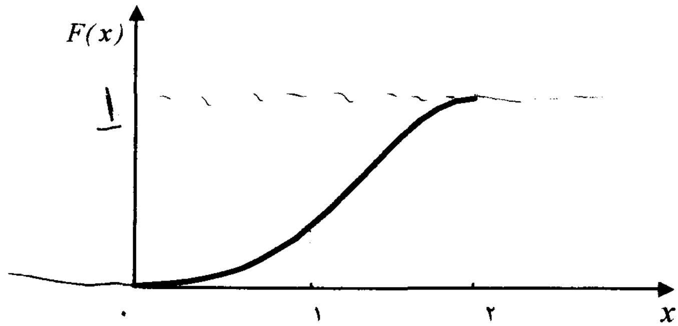
$$f(x)=\begin{cases} x & 0<x\le 1 \\ 2-x & 1\le x\le 2 \end{cases}$$

نمودار $f(x)$ را رسم کنید و تابع توزیع $F(x)$ را بدست آورید.



$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)\,dt$$

$$F(x)=\begin{cases} 0 & if & x\le 0 \\ \int_0^x t\,dt=\dfrac{x^2}{2} & if & 0<x\le 1 \\ \int_0^1 t\,dt+\int_0^x (2-t)\,dt=2x-\dfrac{x^2}{2}-1 & if & 1<x\le 2 \\ 1 & if & x\ge 2 \end{cases}$$



۳-۴-۳ خواص تابع توزیع (گسسته یا پیوسته)

۱- $0 \le P(X \le x) \le ۱$ یا $0 \le F(x) \le ۱$

۲- $F(x)$ یک تابع غیر نزولی است.

۳- $\lim_{x \to +\infty} F(x) = F(+\infty) = ۱$ و $\lim_{x \to -\infty} F(x) = F(-\infty) = 0$

۴- $F(x)$ در هر نقطه x از راست پیوسته است.

۵- $P(a \le X \le b) = F(b) - F(a)$

۶- $f(x) \approx P(X = x) = F(x) - F(x^-)$

که درآن $F(x^-)$ حد چپ $F(x)$ در نقطه x است.

۷- $F(x^-) \le F(x) \le F(x^+)$

۸- الف: در متغیر پیوسته $f(x) = \dfrac{dF(x)}{dx}$ یا $F(x) = \int_{-\infty}^{x} f(t)\,dt$

ب: در متغیر گسسته $f(x) = F(x) - F(x^-)$

که در آن

$$\boxed{F(x)=\sum_{x\le t} f(t)}$$

باتوجه به تعریف تابع توزیع $F(x)$، ذکر این نکته ضروری است که تابع چگالی احتمال، تابع توزیع را مشخص می‌کند و بالعکس.

مثال ۳-۴-۴ اگر متغیر تصادفی گسسته X دارای تابع توزیع $F(x)$ زیر باشد. تابع احتمال آن را بدست آورید.

$$F(x)=\begin{cases} 0 & x<0 \\ \frac{1}{16} & 0\le x<1 \\ \frac{5}{16} & 1\le x<2 \\ \frac{11}{16} & 2\le x<3 \\ \frac{15}{16} & 3\le x<4 \\ 1 & x\ge 4 \end{cases}$$

$f(x) = F(x) - F(x^-)$

متغیر X گسسته است با فضای $A=\{0,1,2,3,4\}$:

$$F(-\infty)=0=P(X\le -\infty)$$

باتوجه به خواص

$$f(x)=F(x)-F(x^-)$$

$$f(0)=F(0)-F(0^-)=\frac{1}{16}-0=\frac{1}{16}$$

$$f(1)=F(1)-F(0)=\frac{5}{16}-\frac{1}{16}=\frac{4}{16}$$

$$f(2)=F(2)-F(1)=\frac{11}{16}-\frac{5}{16}=\frac{6}{16}$$

$$f\,(۳)=F\,(۳)-F\,(۲)=\frac{15}{16}-\frac{11}{16}=\frac{4}{16}$$

$$f\,(۴)=F\,(۴)-F\,(۳)=1-\frac{15}{16}=\frac{1}{16}$$

x	۰	۱	۲	۳	۴
$f(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

مثال ۳-۴-۵ اگر متغیر تصادفی پیوسته X دارای تابع توزیع $F(x)$ باشد تابع چگالی احتمال آن را بدست آورید.

$$F(x)=\begin{cases} 0 & x<1 \\ \dfrac{(x-1)^2}{8} & 1\le x<3 \\ 1 & x\ge 3 \end{cases}$$

$f(x)=\dfrac{dF(x)}{dx}$

همانطور که ملاحظه می‌کنید $F(x)$ از راست پیوسته است چون:

$$F(1)=0$$

$$\lim_{x\to 1^+}F(x)=\lim_{x\to 1^+}\frac{(x-1)^2}{8}=\lim_{\varepsilon\to 0}\frac{(1+\varepsilon-1)^2}{8}=0$$

$$F(3)=1$$

$$\lim_{x\to 3^+}F(x)=\lim_{\varepsilon\to 0}\ F(3+\varepsilon)=1$$

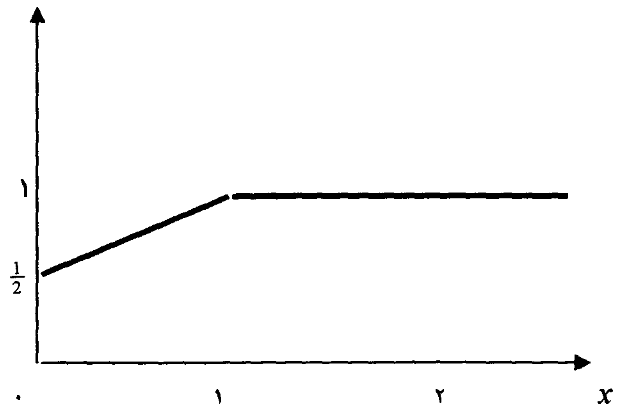
$$\lim_{x\to 3^-}F(x)=\lim_{\varepsilon\to 0}\ F(3-\varepsilon)=\lim_{\varepsilon\to 0}\frac{(3-\varepsilon-1)^2}{8}=\frac{1}{2}$$

$$f(x)=\frac{d\ F(x)}{dx}=\begin{cases} 0 & x<1 \\ \dfrac{x-1}{4} & 1\le x<3 \end{cases}$$

مثال ۳-۴-۶ اگر متغیر تصادفی X دارای تابع توزیع $F(x)$ زیر باشد. مطلوبست $P(X=0)$

$$F(x)=\begin{cases} 0 & x<0 \\ \dfrac{x+1}{2} & 0\le x<1 \\ 1 & x\ge 1 \end{cases}$$

$$P(X=0)=F(0)-F(0^-)=\frac{0+1}{2}-0=\frac{1}{2}$$



از نمودار $F(x)$ مشخص است که نه همواره پیوسته و نه یک تابع پله‌ای است. بنابراین $F(x)$ نه از نوع پیوسته است و نه از نوع گسسته بلکه ترکیبی از این دو توزیع است.

۳-۵ تابع احتمال و تابع توزیع توام دو متغیر تصادفی

در پرتاب سه سکه اگر X و Y به ترتیب تعداد شیرها و تعداد خط‌ها در سه پرتاب تعریف شود فضای نمونه این متغیرهای تصادفی مجموعه دوبعدی $\mathcal{A}$ است.

$\mathcal{A}=\{(x,y) \mid x=0,۱,۲,۳ \quad y=0,۱,۲,۳\}$

تابع احتمال توام برای X و Y برای هر زیر مجموعه A از $\mathcal{A}$ را می‌توان به صورت زیر تعریف کرد.

$$P(A)=P((X,Y)\in A) \quad , \qquad A\subset \mathcal{A}$$

باتوجه به اعضای مختلف A جدول توام احتمال X و Y برابر است با:

X \ Y	0	۱	۲	۳
0	**0**	**0**	**0**	$\frac{1}{8}$
۱	**0**	**0**	$\frac{3}{8}$	**0**
۲	**0**	$\frac{3}{8}$	**0**	**0**
۳	$\frac{1}{8}$	**0**	**0**	**0**

اگر احتمال‌ها را به وسیله تـابع تـوام $P[X=x,Y=y]$ و $f(x,y)$ بـرای هـر زوج از مقادیر (x,y) در برد متغیرهای X و Y بیان کنیم، داریم:

$$f(x,y)=\frac{\binom{3}{x}\binom{3}{x+y-3}}{n} \qquad \begin{array}{l} x=0,۱,۲,۳ \\ y=0,۱,۲,۳ \\ x+y=۳ \end{array}$$

فرض کنید متغیرهای تصادفی X و Y دارای فضای دوبعدی A باشد. به طوری که A گسسته و شمارا باشد. هرگاه بتوان تابع احتمال $P(A)$ را برحسب تـابع $f(x,y)$ به شکل زیر تعریف کرد

$$P(A)=P[(X,Y)\in A]=\sum_A\sum f(x,y)$$

به طوری که $f(x,y)$ در دو شرط زیر صدق کند.

۱- $f(x,y)\geq 0 \qquad \forall(x,y)\in A$

۲- $\sum_A\sum f(x,y)=۱$

(X,Y) را متغیرهای تصادفی توام از نوع گسسته و $f(x,y)$ را تابع چگالی احتمال یا پخش توام گسسته گویند.

مهم **مثال ۳-۵-۱** مقدار c را طوری پیدا کنید که تابع توام $f(x,y)$ یک تابع چگالی احتمال شود.

$$f(x,y)=c\,x.y \quad , \quad x=۱،۲،۳ \qquad y=۱،۲،۳$$

در این مثال

$$A=\{(۱،۱)،(۲،۱)،(۳،۱)،(۱،۲)،(۲،۲)\ (۳،۲)،(۱،۳)،(۲،۳)،(۳،۳)\}$$

باتوجه به شرط دوم

$$\sum_A\sum f(x,y)=\sum_{x=1}^{3}\sum_{y=1}^{3}c\ xy=۱$$

$$\sum_A\sum f(x,y)=\sum_{x=1}^{3}\sum_{y=1}^{3}cxy=۱$$

$$c=\frac{1}{36}$$

در نتیجه

$$f(x,y)=\frac{x.y}{36} \quad , \quad x=۱،۲، ۳ \qquad y=۱،۲،۳$$

برای حالتی که X و Y متغیرهای تصادفی از نوع پیوسته اند، می‌توان تابع چگالی احتمال $f(x,y)$ را روی همه صفحه تعریف کرد. تابع دومتغیره $f(x,y)$ را تابع چگالی احتمال توام متغیر X و Y گوییم اگر و تنها اگر برای هر ناحیه A از صفحه xy

$$P[(X,Y)\in A]=\int_A\int f(x,y)\,dy\,dx$$

و $f(x,y)$ همواره در دو شرط زیر صدق نماید.

$$\left\{\begin{array}{ll} f(x,y)\geq 0 & \forall A\in A \\ \int_A\int f(x,y)\,dy\,dx=۱ & \end{array}\right.$$

-۱

-۲

مثال ۳-۵-۲ به ازای چه مقداری از c، $f(x,y)$ یک تابع چگالی است و مقدار احتمال را برای $A=\{(x,y)\,|\,0<x<\frac{1}{2}\ ,\ 0<y<۱\}$ بدست آورید.

$$f(x,y)=c(xy+\frac{x^2}{2})\ ,\ 0<x<۱ \qquad 0<y<۲$$

واضح است که

$$\int_A\int f(x,y)\,dy\,dx=۱$$

$$c\int_0^1\int_0^2(xy+\frac{x^2}{2})\,dy\,dx=۱$$

$$\frac{1}{c}=\int_0^1\int_0^2(xy+\frac{x^2}{2})\,dy\,dx=\int_0^2\left(\frac{y}{2}+\frac{1}{6}\right)dy=\frac{4}{3}$$

نتیجه می‌شود $c=\frac{3}{4}$ و

$$f(x,y)=\frac{3}{4}(xy+\frac{x^2}{2}) \quad ,\ 0<x<۱ \qquad 0<y<۲$$

$$P[(X,Y)\in A]=\int_A\int\frac{3}{4}(xy+\frac{x^2}{2})\,dy\,dx=\int_0^{\frac{1}{2}}\int_0^1\frac{3}{4}(xy+\frac{x^2}{2})\,dy\,dx$$
$$=\int_0^{\frac{1}{2}}\ \frac{3}{4}(\frac{x}{2}+\frac{x^2}{2})\,dx=\frac{3}{4}\left(\frac{4}{48}\right)=\frac{1}{16}$$

۳-۶ تابع توزیع توام

اگر X و Y متغیرهای تصادفی با تابع چگالی احتمال توام $f(x,y)$ باشند تابع توزیع یا تابع توزیع تجمعی توام X و Y در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شود.

$$F(x,y)=P[X\le x\ ,\ Y\le y]=\sum_{S\le x}\sum_{t\le y} f(s,t)$$

$$F(x,y)=P[X\le x\ ,\ Y\le y]=\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dt\,ds$$

مثال ۳-۶-۱ اگر X و Y دارای تابع احتمال توام $f(x,y)$ باشند مطلوبست $F(x,y)$ ، $F(0,۲)$ و $F(۱,۱)$:

$$f(x,y)=\frac{1}{36}\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y} \qquad x=0,۱,۲$$

$$y=0,۱,۲$$

$$x+y\le ۲$$

$$F(x,y)=P[X\le x\ ,\ Y\le y]=\sum_{S=0}^{x}\sum_{t=0}^{y}f(s,t)$$

$$=\sum_{S=0}^{x}\sum_{t=0}^{y}\ \frac{1}{36}\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y}$$

$$F(0,2)=P[X\le 0\ ,\ Y\le 2]=\sum_{S=0}^{0}\sum_{t=0}^{2}f(s,t)$$

$$=f(0,0)+f(0,1)+f(0,2)=\frac{1}{6}+\frac{2}{9}+\frac{1}{36}=\frac{15}{36}$$

$$=P[X\le 1\ ,\ Y\le 1]=\sum_{S=0}^{1}\sum_{t=0}^{1}f(s,t)$$

$$=f(0,0)+f(0,1)+f(1,0)+f(1,1)=\frac{1}{6}+\frac{2}{9}+\frac{1}{3}+\frac{1}{6}=\frac{16}{18}$$

واضح است که $F(۲,۲)=۱$.

مثال ۳-۶-۲ اگر تابع چگالی احتمال توام X و Y به صورت زیر باشد تابع توزیع توام را بدست آورید و مقدار احتمال را روی A حساب کنید.

$$f(x,y)=e^{-(x+y)} \qquad , \qquad x>0 \qquad , \qquad y>0$$

$$A=\{(x,y)|0<x<۵ \qquad , \qquad ۲<y<۷\}$$

$$F(x,y)=P[X\le x\ ,\ Y\le y]=\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dt\,ds$$
$$=\int_{0}^{x}\int_{0}^{y} e^{-(s+t)}\,dt\,ds=\int_{0}^{x} e^{-s}(1-e^{-y})\,ds=(1-e^{-x})(1-e^{-y})$$

در نتیجه

$$F(x,y)=\begin{cases}(1-e^{-x})(1-e^{-y}) & x>0\ ,\ y>0\\ 0 & \text{سایرنقاط}\end{cases}$$

$$P[(X,Y)\in A]=\int_{A}\int f(s,t)\,dy\,dx=\int_{0}^{5}\int_{2}^{7} e^{-(x+y)}\,dy\,dx$$
$$=\int_{0}^{5} e^{-x}(e^{-2}-e^{-7})\,dx=(1-e^{-5})(e^{-2}-e^{-7})=0.13352$$

تبصره ۳-۶-۳ محاسبه احتمال (X,Y) روی A به طوری که

$A=\{(x,y)\mid a\le x<b\quad ,\quad c\le y<d\}$

تعریف شود از فرمول زیر قابل محاسبه است.

۸۵ $\boxed{P[a\le X<b\ ,\ c\le Y<d]=F(b,d)-F(a,d)-F(b,c)+F(a,c)}$

مثال ۳-۶-۴ برای مثال ۳-۶-۲ اگر A به صورت زیر باشد احتمال روی A را حساب کنید.

$A=\{(x,y)\mid 0<x<۵\quad ,\quad ۲<y<۷\}$

واضح است $a=0$ ، $b=۵$، $c=۲$ و $d=۷$.

$$P(0<X<۵\ ,\ ۲<Y<۷)=F(۵,۷)-F(0,۷)-F(۵,۲)+F(0,۲)$$
$$=(۱-e^{-5})(۱-e^{-7})-(۱-e^{0})(۱-e^{-7})-(۱-e^{-5})(۱-e^{-2})+(۱-e^{0})(۱-e^{-2})$$
$$==(۱-e^{-5})(e^{-2}-e^{-7})=۰/۱۳۳۵۲$$

تابع چگالی احتمال و تابع توزیع توام در حالت n متغیره $X_۱$، $X_۲$، ...، X_n قابل تعمیم است برای $(x_۱, x_۲, \ldots, x_n)$. تابع چگالی احتمال توام در دوحالت گسسته و پیوسته به ترتیب عبارتند از:

حالت گسسته:

$$f(x_1, x_2, ..., x_n) = P[X_1 = x_1 \ , X_2 = x_2, \ ..., \ X_n = x_n]$$

$$F(x_1, x_2, ..., x_n) = P[X_1 \leq x_1 \ , X_2 \leq x_2, \ ..., X_n \leq x_n]$$

حالت پیوسته:

$$F(x_1, x_2, ..., x_n) = P[X_1 \leq x_1 \ , X_2 \leq x_2, \ ..., \ X_n \leq x_n]$$
$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_n} f(s_1, s_2, ..., s_n) ds_1, ds_2, ..., ds_n$$

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, \ x_2, ..., x_n)}{\partial x_1, \partial x_2, \ ..., \partial \ x_n}$$

۷-۳ تابع چگالی احتمال و تابع توزیع حاشیه‌ای

فرض کنید دو متغیر تصادفی X و Y دارای تابع چگالی احتمال توام $f(x, y)$ باشند و بخواهیم احتمال پیشامد $-\infty < X \leq x$ را حساب کنیم. محاسبه احتمال پیشامد $-\infty < X < x$ برای دو متغیر تصادفی X و Y با تابع چگالی احتمال توام $f(x, y)$ هم ارز است با محاسبه احتمال پیشامد ($-\infty < X \leq x$ و $-\infty < Y < \infty$)، پس:

$$P[-\infty < X \leq x] = P[-\infty < Y < \infty \ , \ \ -\infty < X \leq x]$$

محاسبه احتمال رابطه اخیر در حالت گسسته و پیوسته به ترتیب برابرند با:

$$P[-\infty < X \leq x, \ \ -\infty < Y < \infty] = \sum_{S \leq x} \sum_{-\infty}^{\infty} f(s, y) = F(x, Y)$$

$$P[-\infty < X \leq x, \ \ -\infty < Y < \infty] = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(s, y) dy \, ds$$

اگر تعریف کنیم:

$$F_X(x) = P[X \leq x] = P[-\infty < X \leq x \ , \ -\infty < Y < \infty] = \sum_{s \leq x} \sum_{-\infty}^{\infty} f(s, y) = \sum_{s \leq x} f_x(s)$$

$$F_X(x)=P[X\le x]=P[-\infty<X\le\infty,\ -\infty<Y<\infty]=\int_{-\infty}^{x}\int_{-\infty}^{\infty}f(s,y)\,dy\,ds=\int_{-\infty}^{x}f_x(s)ds$$

تابع‌های $F_X(x)=\sum_{s\le x}f_x(s)$ و $F_X(x)=\int_{-\infty}^{x}f_x(s)ds$ را به ترتیب تابع توزیع حاشیه‌ای X در حالت گسسته و پیوسته گویند. با معلوم بودن تابع توزیع حاشیه‌ای، تابع چگالی احتمال حاشیه‌ای برای متغیرهای گسسته و پیوسته به ترتیب از رابطه‌های زیر بدست می‌آیند.

تابع حاشیه‌ای X در حالت گسسته $\quad f_x(x)=F_X(x)-F_X(x^-)$

تابع حاشیه‌ای Y در حالت گسسته $\quad f_y(y)=F_Y(y)-F_Y(y^-)$

تابع حاشیه‌ای X در حالت پیوسته $\quad f_x(x)=\dfrac{\partial F(x,y)}{\partial y}$

تابع حاشیه‌ای Y در حالت پیوسته $\quad f_y(y)=\dfrac{\partial F(x,y)}{\partial x}$

اگر تابع چگالی احتمال توام $f(x,y)$ معلوم باشد تابع چگالی احتمال حاشیه‌ای X و Y برای حالت گسسته و پیوسته به ترتیب از روابط زیر بدست می‌آیند.

$$f_X(x)=\sum_{y}f(x,y)\quad,\quad f_Y(y)=\sum_{x}f(x,y)$$

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy\quad,\quad f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,dx$$

یادآوری می‌شود که هریک از توابع چگالی حاشیه‌ای X و Y به نوبه خود تابع چگالی احتمال می‌باشند و در تمام شرایط تابع چگالی بودن صدق می‌کنند.

مثال ۳-۷-۱ اگر متغیرهای تصادفی X و Y دارای تابع احتمال توام $f(x,y)$ باشند تابع احتمال و تابع توزیع حاشیه‌ای X و Y را بدست آورید.

$$f(x,y)=\frac{x+y}{21}\qquad x=۱,۲,۳\qquad y=۱,۲$$

$$f(x)=\sum_{y}f(x,y)=\sum_{y=1}^{2}\frac{x+y}{21}=\frac{2x+3}{21}$$

$$f_x(x) = \frac{2x+3}{21} \qquad x = ۱,۲,۳$$

بنابراین

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{5}{21} & 1 \le x < 2 \\ \frac{12}{21} & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

$F_x(x) = \sum_{s \le x} f(s)$

$$f_y(y) = \sum_x f(x,y) = \sum_{x=1}^{3} \frac{x+y}{21} = \frac{6+3y}{21}$$

$$f_y(y) = \frac{6+3y}{21} \qquad y = ۱,۲$$

$$F_Y(y) = \begin{cases} 0 & y < 1 \\ \frac{9}{21} & 1 \le y < 2 \\ 1 & y \ge 2 \end{cases}$$

مثال ۳-۷-۲ اگر متغیرهای تصادفی پیوسته X و Y دارای تابع چگالی احتمال تـوام $f(x,y)$ باشند تابع چگالی احتمال و تابع توزیع حاشیه ای X و Y را بدست آورید.

$$f(x,y) = ۲ \qquad 0 < x < y < ۱$$

$$f_x(x) = \int_x^1 f(x,y)\,dy = \int_x^1 2dy = 2(1-x)$$

$$f_x(x) = ۲(۱- x) \qquad 0 < x < ۱$$

$$F_X(x)=\begin{cases}0 & x<0\\ 2x-x^2 & 0\le x<1\\ 1 & x\ge 1\end{cases}$$

$$f_y(y)=\int_x^y f(x,y)dx=\int_x^y ۲\,dx=۲\,y$$

$$f_y(y)=۲\,y \qquad 0<y<۱$$

$$F_Y(y)=\begin{cases}0 & y<0\\ y^2 & 0\le y<1\\ 1 & y\ge 1\end{cases}$$

۳-۸ تابع چگالی احتمال و تابع توزیع شرطی

فرض کنید دو متغیر تصادفی X و Y از نوع گسسته، دارای تابع احتمال توام $f(x,y)$، تابع‌های چگالی احتمال حاشیه‌ای $f_x(x)$، $f_y(y)$ و فضای نمونه A باشند. دو پیشامد $A_۱$ و $A_۲$ را به صورت زیر درنظر می‌گیریم.

$$A_۱\subset A=\{(x,y)|x=x_1,-\infty<y<+\infty\}$$
$$A_۲\subset A=\{(x,y)|-\infty<x<+\infty,y=y_1\}$$

می‌دانیم که

$$P(A_1)=P[(x,y)\in A_1]=\sum_x\sum_y f(x,y)=\sum_x f_x(x)=f(x)$$

$$P(A_2)=P[(x,y)\in A_2]=\sum_x\sum_y f(x,y)=\sum_y f_y(y)=f(y)$$

احتمال شرطی پیشامد $A_۱$ به شرط $A_۲$ برابر است با:

$$P(A_1\mid A_2)=\frac{P(A_1\cap A_2)}{P(A_2)}=\frac{f(x_1,y_1)}{f(y)}$$

اگر تعریف کنیم

$$f(x_1 \mid y_1)=\frac{f(x_1,y_1)}{f(y_1)}$$

آنگاه تابع احتمال شرطی x_1 به شرط y_1 برابر است با:

$$f(x_1 \mid y_1)=\frac{f(x_1,y_1)}{f(y_1)} \qquad , \qquad f(y_1)>0$$

برای سادگی تابع احتمال X به شرط Y را به صورت زیر تعریف می کنیم.

$$f(x \mid y)=\frac{f(x,y)}{f(y)} \qquad , \qquad f(y)>0$$

و تابع احتمال Y به شرط X را به صورت $f(y|x)$ تعریف می کنیم.

$$f(y \mid x)=\frac{f(x,y)}{f(x)} \qquad , \qquad f(x)>0$$

در حالتی که متغیرهای تصادفی X و Y پیوسته باشند از همین نماد استفاده می‌کنیم.

$$f(x \mid y)=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dx}=\frac{f(x,y)}{f(y)} \qquad , \qquad f(y)>0$$

$$f(y \mid x)=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dy}=\frac{f(x,y)}{f(x)} \qquad , \qquad f(x)>0$$

یادآوری می‌شود که تابع چگالی احتمال شرطی نیز بـه نوبـه خـود یـک تـابع چگـالی احتمال است و در تمام شرایط چگالی بودن صدق می‌کند.

مثال ۳-۸-۱ فرض کنید دو متغیر تصادفی گسسته X و Y دارای تابع احتمـال تـوام زیر باشند. توابع احتمال شرطی و توابع توزیع شرطی را بدست آورید.

$$f(x,y)=\frac{1}{28}\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y} \qquad x=0,۱,۲$$

$$y=0,۱,۲$$

$$x+y\leq ۲$$

$$f_x(x)=P[X=x]=\sum_{y=0}^{2}f(x,y)=\frac{1}{28}\sum_{y=0}^{2}\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}$$

$$=\frac{1}{28}\binom{3}{x}\left[\binom{3}{2-x}+2\binom{3}{1-x}+\binom{3}{-x}\right]$$

$$P(X=x)=f_x(x)=\begin{cases}\frac{10}{28} & x=0\\ \frac{15}{28} & x=1\\ \frac{3}{28} & x=2\end{cases}$$

$$f_y(y)=P[Y=y]=\sum_{x=0}^{2}\frac{1}{28}\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}$$

$$=\frac{1}{28}\binom{2}{y}\left[\binom{3}{2-y}+3\binom{3}{1-y}+3\binom{3}{-y}\right]$$

$$f_y(y)=P(Y=y)=\begin{cases}\frac{15}{28} & y=0\\ \frac{12}{28} & y=1\\ \frac{1}{28} & y=2\end{cases}$$

تابع‌های احتمال شرطی X به شرط Y و Y به شرط X به ترتیب برابرند با:

$$P[X=x\mid Y=y]=f(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{P[X=x,Y=y]}{P[Y=y]}$$

$$f(x\,|\,y)=\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{28\ f_y(y)} \qquad x=0,۱,۲$$

اگر y=0 باشد، $f(x\,|\,y=0)$ برابر است با:

$$f(x\,|\,0)=\frac{\binom{3}{x}\binom{3}{2-x}}{28\ f_y(0)} \qquad x=0,۱,۲$$

x	0	۱	۲	جمع
$f(x\,|\,0)$	$\frac{3}{15}$	$\frac{9}{15}$	$\frac{3}{15}$	۱

برای ۱= y، ۲= y تابع‌های احتمال شرطی به ترتیب برابرند با:

x	0	۱	۲	جمع
$f(x\,|\,1)$	$\frac{3}{6}$	$\frac{3}{6}$	0	۱

x	0	۱	۲	جمع
$f(x\,|\,2)$	۱	0	0	۱

تابع‌های احتمال شرطی X به شرط $Y=y$ را می‌توان تواماً در یک جدول به صورت زیر نمایش داد.

$x\,|\,Y=y$	0	۱	۲
0	$\frac{3}{15}$	$\frac{3}{6}$	1
۱	$\frac{9}{15}$	$\frac{3}{6}$	0
۲	$\frac{3}{5}$	0	0
جمع	۱	۱	۱

به طور متشابه، تابع‌های احتمالی شرطی Y به شرط $X=x$.

$y \mid X = x$	0	۱	۲	جمع
0	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	۱
۱	$\frac{9}{15}$	$\frac{6}{15}$	0	۱
۲	۱	0	0	۱

اگر $f(x|y)$ و $f(y|x)$ به ترتیب تابع احتمال شرطی X به شرط Y و Y به شرط X باشند. تابع‌های توزیع شرطی آنها به ترتیب برابرند با:

$$F_x(x \mid Y = y) = P[X \le x \mid Y = y] = \sum_{S \le x} f_x(s \mid Y = y)$$

$$F_y(y \mid X = x) = P[Y \le y \mid X = x] = \sum_{t \le y} f_y(t \mid x = x)$$

در مثال اخیر تابع توزیع شرطی X به شرطی $Y =$ ۱ برابر است با:

$$F_x(x \mid Y = 1) = \sum_{S \le x} f_x(x \mid y = 1) = \begin{cases} 0 & x < 0 \\ \frac{3}{6} & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

و توزیع شرطی Y به شرط $X = 0$ برابر است با:

$$F_y(y \mid X = 0) = \begin{cases} 0 & y < 0 \\ \frac{3}{10} & 0 \le y < 1 \\ \frac{9}{10} & 1 \le y < 2 \\ 1 & y \ge 2 \end{cases}$$

مثال ۳-۸-۲ فرض کنید دو متغیر تصادفی پیوسته X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند. توابع چگالی احتمال شرطی و توابع توزیع شرطی را بدست آورید.

$$f(x,y) = ۲ \qquad 0 < x < y < ۱$$

$$f_x(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_x^1 2\,dy=2(1-x) \qquad 0<x<1$$

$$f_y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_0^y 2\,dx=2y \qquad 0<y<1$$

$$f(x\mid y)=\frac{f(x,y)}{f(y)}=\frac{2}{2y}=\frac{1}{y} \qquad 0<x<y$$

$$f(y\mid x)=\frac{f(x,y)}{f(x)}=\frac{2}{2(1-x)}=\frac{1}{1-x} \qquad x<y<1$$

تابع توزیع شرطی

$$\mathrm{F}(x\mid Y=y)=\int_{-\infty}^{x} f(s\mid y)\,ds=\int_{-\infty}^{x}\frac{ds}{y}=\frac{x}{y}$$

$$F(x\mid y)=\frac{x}{y} \quad , \qquad 0<x<y$$

$$F(y\mid X=x)=\int_{\infty}^{y} f(t\mid x)\,dt=\int_{-\infty}^{y}\frac{dt}{1-x}=\int_{x}^{y}\frac{dt}{1-x}$$

$$F(y\mid x)=\frac{y-x}{1-x} \qquad x<y<۱$$

۳-۹ استقلال دو متغیر تصادفی

فرض کنید دو متغیر تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$، توابع چگالی حاشیه‌ای $f_x(x)$، $f_y(y)$ و توابع چگالی احتمال شـرطی $f(x|y)$ و $f(y|x)$ باشند. گوییم دو متغیر تصادفی X و Y به طور احتمالی مستقل‌اند اگر:

$$\boxed{f(x,y)=f_x(x)f_y(y)} \checkmark$$ شرط استقلال

یا تابع چگالی احتمال شرطی X به شرط Y مستقل از Y باشد یا تابع چگالی احتمال شرطی Y به شرط X مستقل از X باشد. یا:

$$f(x,y)=f(X\mid y)f_Y(y)=f(Y\mid x)f_X(x)$$

مثال ۳-۹-۱ اگر دو متغیر تصادفی گسسته X و Y دارای تابع احتمال توام $f(x,y)$ باشند نشان دهید که X و Y از هم مستقل‌اند.

$$f(x,y)=\frac{e^{-\lambda}\lambda^{x}}{x!}\binom{n}{y}P^{y}(1-P)^{n-y} \qquad x=0,۱,۲,\ldots \quad y=0,۱,۲,\ldots,n$$

$$f_x(x)=\sum_{y=0}^{n} f(x,y)=\frac{e^{-\lambda}\lambda^{x}}{x!}\sum_{y=0}^{n}\binom{n}{y}P^{y}(1-P)^{n-y}$$

$$f_x(x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x=0,۱,۲,\ldots$$

$$f_y(y)=\sum_{x=0}^{\infty} f(x,y)=\binom{n}{y}P^{y}(1-P)^{n-y}\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$f_y(y)=\binom{n}{y}P^{y}(1-P)^{n-y} \qquad y=0,۱,۲,\ldots,n$$

چون $f(x,y)=f_x(x)f_y(y)$ است پس X و Y از هم مستقل‌اند.

$$P[X=x\mid Y=y]=f(x\mid y)=f_x(x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x=0,۱,۲,\ldots$$

$$P[Y=y\mid X=x]=f(y\mid x)=f_y(y)=\binom{n}{y}P^{y}(1-P)^{n-y} \qquad y=0,۱,۲,\ldots,n$$

مثال ۳-۹-۲ اگر دو متغیر تصادفی پیوسته X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند نشان دهید که دو متغیر X و Y از هم مستقل نیستند.

$$f(x,y)=\frac{1+xy}{4} \qquad |x|<1 \quad , \quad |y|<1$$

$$f_x(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\frac{1}{4}\int_{-1}^{1}(1+xy)dy=\frac{1}{2} \qquad |x|<1$$

$$f_y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\frac{1}{4}\int_{-1}^{1}(1+xy)dx=\frac{1}{2} \qquad |y|<1$$

چون $f(x,y) \neq f_x(x)f_y(y)$ پس دو متغیر x و y به طور احتمالی مستقل نیستند.

مثال ۳-۹-۳ اگر X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند نشان دهید که X و Y مستقل‌اند.

$$f(x,y) = 6e^{-2x-3y} \qquad x>0 \quad , \quad y>0$$

$$f(x,y) = (2e^{-2x})(3e^{-3y}) = f_x(x)f_y(y)$$

به طوری که $f_x(x) = 2e^{-2x},\ x>0$ و $f_y(y) = 3e^{-3y},\ y>0$ پس X و Y مستقل‌اند.

اگر دو متغیر تصادفی X و Y دارای تابع توزیع توام $F(x,y)$ و توابع توزیع حاشیه‌ای $F_x(x)$ و $F_y(y)$ باشند X و Y را مستقل گوییم اگر:

$$\boxed{F(x,y) = F_x(x).F_y(y)}$$

۳-۱۰ امید ریاضی

در بخش آمار توصیفی، مقدار میانگین مشاهدات $x_1, x_2, \dots, x_n$ با فراوانیهای $f_1, f_2, \dots, f_k$ از ضرب $\frac{f_i}{\sum f_i}$ در x_i و جمع آنها تعریف شد. حال سئوال این است که در مسائلی که شامل متغیرهای تصادفی یا توزیع‌ها هستند، چگونه می‌توان مفهوم مقدار متوسط یا امید ریاضی را تعریف کرد. ساده ترین تعریف امید ریاضی برای متغیر تصادفی گسسته از ضرب مقدار متغیر تصادفی در شانس آن و جمع آنها تعریف می‌شود. برای مثال، در یک بخت آزمایی ۱۰۰ بلیط وجود دارد که ۸۰ تای آن پوچ، ۱۰ تای آن هریک برنده ۱۵ ریال، ۷ تای آن هر یک برنده ۳۰ ریال، ۲ تای آن هر یک برنده ۱۰۰ ریال و یکی از آنها برنده ۱۰۰۰ ریال است.

اگر متغیر X را مقدار حک شده روی بلیط‌ها تعریف کنیم، داریم:

x	0	۱۵	۳۰	۱۰۰	۱۰۰۰
فراوانی (f)	80	10	7	2	1
شانس	$\frac{80}{100}$	$\frac{10}{100}$	$\frac{7}{100}$	$\frac{2}{100}$	$\frac{1}{100}$

$$\text{امید ریاضی} = \left(0\times\frac{80}{100}\right)+\left(15\times\frac{10}{100}\right)+\left(30\times\frac{7}{100}\right)+\left(100\times\frac{2}{100}\right)+\left(1000\times\frac{1}{100}\right)$$
$$=0+1.5+2.1+2+10=15.6$$

مقدار ۱۵/۶ میانگین یا امید ریاضی توزیع است. در حقیقت اگر کسی تمام بلیط‌های این بخت‌آزمایی را از قرار هر بلیط ۱۵/۶ ریال بخرد نه بازنده است و نه برنده و حتی اگر دفعات زیادی در این بخت‌آزمایی شرکت کند و هر بار برای قیمت یک بلیط ۱۵/۶ ریال بپردازد بازهم نه برنده است و نه بازنده.

اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x)$ باشد. امید ریاضی در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شود.

$$\begin{cases} E(X)=\sum_{x} x\,.\,f(x) \\ E(X)=\int_{-\infty}^{\infty} x\,f(x)\,dx \end{cases}$$

در ادبیات آماری امید ریاضی را معمولاً با μ نمایش می‌دهند.

مثال ۳-۱۰-۱ محموله‌ای شامل ۱۵ دستگاه رایانه است که ۵ تای آنها معیوب است. اگر سه دستگاه رایانه برای خرید انتخاب شود به طور متوسط چند دستگاه معیوب را می‌توان انتظار داشت؟

اگر متغیر X تعداد معیوب‌ها باشد آنگاه:

$$f(x)=P(X=x)=\frac{\binom{5}{x}\binom{10}{3-x}}{\binom{15}{3}} \qquad x=0,\ ۱,\ ۲,۳$$

x	0	۱	۲	یا ۳
$f(x)$	$\frac{24}{91}$	$\frac{45}{91}$	$\frac{20}{91}$	$\frac{2}{91}$

$$E(X)=\sum_{x=0}^{3} x\,f(x)=0\times f(0)+1\times f(1)+2\times f(2)+3\times f(3)$$
$$=0\times\frac{24}{91}+1\times\frac{45}{91}+2\times\frac{20}{91}+3\times\frac{2}{91}=1$$

به طور متوسط، انتظار یک رایانه معیوب در خرید سه رایانه وجود دارد.

مثال ۳-۱۰-۲ نقطه‌ای به تصادف در داخل دایره‌ای به شعاع ۲ و مرکز صفر انتخاب می‌شود. اگر فاصله این نقطه تا مرکز با متغیر X نشان داده شود امید ریاضی X را حساب کنید.

تابع چگالی احتمال x برابر با

$$f(x)=\begin{cases}\dfrac{x}{2} & 0<x<2\\ 0 & \text{جاهای دیگر}\end{cases}$$

$$E(X)=\int_{-\infty}^{\infty} x\ f(x)dx=\int_0^2 x.\frac{x}{2}dx=\int_0^2 \frac{x^2}{2}dx=\frac{4}{3}$$

به طور متوسط فاصله نقطه انتخاب شده تا مرکز دایره $\frac{4}{3}$ است. یعنی اگر به تعداد زیاد مثلاً ۶۰ بار آزمایش را تکرار کنیم و ۶۰ مقدار $x_1, x_2, \ldots, x_{60}$ بدست آمده را جمع کنیم و تقسیم بر ۶۰ کنیم، انتظار می‌رود که معدل بدست آمده در همسایگی $\frac{4}{3}$ باشد.

باتوجه به تعریف امید ریاضی، می‌توان امید ریاضی تابعی از متغیر تصادفی را نیز تعریف کرد. اگر $h(x)$ تابعی از متغیر تصادفی X باشد امید ریاضی $h(x)$ در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شود.

$$E(h(X))=\sum_x h(x)f(x)$$

$$E(h(X))=\int_{-\infty}^{\infty} h(x)f(x)dx$$

اگر $h(X)=X^2+X+7$ باشد امید ریاضی $h(X)$ را برای مثالهای ۳-۱۰-۱ و ۳-۱۰-۲ بدست آورید.

$$E(h(X)) = \sum_{x=0}^{3} h(x)\, f(x) = \sum_{x=0}^{3} (x^2 + x + 7)\, f(x)$$

$$= \sum_{x=0}^{3} x^2 f(x) + \sum_{x=0}^{3} x\, f(x) + 7\sum_{x=0}^{3} f(x)$$

$$= \frac{1}{91}[0 \times 24 + 1 \times 45 + 4 \times 20 + 9 \times 2] + E(X) + 7$$
$$= \frac{143}{91} + 1 + 7 = 9.571$$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx = \int_0^2 (x^2 + x + 7)\, f(x)\, dx$$

$$= \int_0^2 x^2 f(x)\, dx + \int_0^2 x\, f(x)\, dx + 7\int_0^2 f(x)\, dx$$

$$= \int_0^2 \frac{x^3}{2} dx + E(X) + 7 = 2 + \frac{4}{3} + 7 = 10.333$$

امید ریاضی را می‌توان برای تمامی توابع چگالی احتمال توام و شرطی در حالت گسسته و پیوسته تعریف کرد. اگر $f(x,y)$ و $f(x|y)$ به ترتیب توابع چگالی احتمال توام و شرطی متغیرهای X و Y باشند و $h(X,Y)$ تابعی از X و Y باشد.

امید ریاضی $h(X,Y)$ در حالت گسسته و پیوسته به صورت زیر تعریف می‌شود.

$$E[h(X,Y)] = \sum_{x} \sum_{y} h(x,y)\, f(x,y)$$

$$E[h(X,Y)] = \iint h(x,y)\, f(x,y)\, dy\, dx$$

امید ریاضی شرطی X به شرط Y در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شود

$$E(X \mid Y = y) = \sum_{x} x f(x \mid Y = y)$$

$$E(X \mid Y=y)=\int_{-\infty}^{\infty} x\, f(x \mid Y=y)\, dx$$

مثال ۳-۱۰-۳ اگر متغیرهای تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند.

$$f(x,y)=\frac{3}{4}\left(xy+\frac{x^2}{2}\right) \qquad 0<x<۱ \quad \text{و} \quad 0<y<۲$$

الف- توزیع‌های حاشیه‌ای X و Y را بدست آورید.
ب- امید ریاضی $h(X,Y)=۲XY+X+Y$ را بدست آورید.
ج- امید ریاضی $E(X|Y=۱)$ را بدست آورید.

$$f_x(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^2 \frac{3}{4}(xy+\frac{x^2}{2})\,dy=\frac{3}{4}x^2+\frac{3}{2}x$$

$$f_x(x)=\frac{3}{4}x^2+\frac{3}{2}x \qquad 0<x<۱$$

$$f_y(y)=\frac{3}{16}y^2+\frac{1}{4} \qquad 0<y<۲$$

$$E(h(X,Y))=E(2\times Y+X+Y)=\int_0^1\int_0^2 (2xy+x+y).\frac{3}{4}(xy+\frac{x^2}{2})\,dy\,dx$$

$$=\frac{3}{2}\int_0^1\int_0^2 x^2y^2dy\,dx+\frac{3}{4}\int_0^1\int_0^2 x^3y\,dy\,dx+\frac{3}{4}\int_0^1\int_0^2 x^2y\,dy\,dx$$

$$+\frac{3}{8}\int_0^1\int_0^2 \frac{x^3}{2}dy\,dx+\frac{3}{4}\int_0^1\int_0^2 xy^2\,dy\,dx+\frac{3}{8}\int_0^1\int_0^2 x^2y\,dy\,dx=\frac{23}{6}$$

برای محاسبه $E(X|Y=۱)$ تابع چگالی شرطی X به شرط Y را بدست می‌آوریم.

$$f(x\mid y)=\frac{f(x,y)}{f_y(y)}=\frac{3xy+\frac{3}{2}x^2}{\frac{3}{4}y^2+1} \qquad 0<x<1 \;,\quad 0<y<2$$

$$E(X \mid Y=1)=\int_0^1 x\ f(x \mid y=1)\,dx=\int_0^1 x\left(\frac{3x+\frac{3}{2}x^2}{\frac{7}{4}}\right)dx$$

$$=\frac{4}{7}\int_0^1 (3x^2+\frac{3}{2}x^3)\,dx=\frac{11}{14}$$

۳-۱۰-۴ ویژگیهای امید ریاضی

۱- امید ریاضی مقدار ثابت برابر با خودش است. $E(c)=c$

۲- $E\left[\sum_{i=1}^{n} a_i X_i\right]=\sum_{i=1}^{n} a_i E(X_i)$

۳- $E\left[\sum a_i X_i+\sum b_i Y_i\right]=\sum a_i E(X_i)+\sum b_i (Y_i)$

۴- اگر X و Y مستقل از هم باشند. $E(XY)=E(X).E(Y)$

۳-۱۱ گشتاورها

اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x)$ و a یک عدد ثابت حقیقی باشد گشتاورهای مرتبه rام حول نقطه a در جامعه در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شود.

$$E(X-a)^r=\sum_x (x-a)^r f(x)$$

$$E(X-a)^r=\int_{-\infty}^{\infty} (x-a)^r f(x)\,dx$$

برای $r=1$ و $a=0$ می‌توان از فرمولهای بالا امید ریاضی در حالت گسسته و پیوسته را تعریف کرد. در ادبیات آماری، معمولاً گشتاورهای حول نقطه میانگین جامعه (μ) را گشتاورهای مرکزی می‌گویند و با نماد μ_r نمایش می‌دهند.

$$\mu_r = E(X-\mu)^r = \sum_x (x-\mu)^r f(x)$$

$$\mu_r = E(X-\mu)^r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\,dx$$

$\gamma = \frac{m_3}{S^3}$

$k = \frac{m_4}{S^4} - 3$

برای $r=1$ ، $\mu_1=0$ است. (نشان دهید که برای $r=3$ ، $\mu_3=0$ است.)

۳-۱۱-۱ واریانس

اگر در فرمول گشتاورهای مرکزی r برابر با ۲ درنظر گرفته شود واریانس جامعه بدست می‌آید و معمولاً آن را با نماد σ^2 یا $V(X)$ نمایش می‌دهند.

$$V(X) = \sigma^2 = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x)$$

$$V(X) = \sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

باتوجه به خواص عملگر E می‌توان واریانس X را به صورت زیر تعریف کرد.

$$V(X) = E(X-\mu)^2 = E[X^2+\mu^2-2\mu X] = E(X^2)+\mu^2-2\mu E(X)$$
$$= E(X^2)+\mu^2-2\mu^2 = E(X^2)-\mu^2$$

$V(x) = E(x^2) - \mu^2$

۳-۱۱-۲ ویژگیهای واریانس

۱- واریانس مقدار ثابت صفر است. $V(C)=0$

۲- $V(aX) = a^2 V(X)$

۳- $V(aX+a) = a^2 V(X) + 0$

۴- $V\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 V(X_i)$

۵- جذر واریانس را انحراف معیار گویند. $\sqrt{\sigma^2} = \sqrt{V(X)} = \sigma$

برای نمونه ویژگی ۳ را می توان به صورت زیر ثابت کرد.

$$\begin{aligned} V(aX+c) &= E[aX+c-E(aX+c)]^2 \\ &= E[aX+c-aE(X)-c]^2 = E[aX-aE(X)]^2 \\ &= E[a(X-E(X))]^2 = a^2\,E(X-\mu)^2 = a^2V(X) \end{aligned}$$

۳-۱۱-۳ کوواریانس دو متغیر تصادفی

اگر متغیرهای تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند کوواریانس آنها به صورت زیر تعریف می‌شود.

$$Cov(X,Y) = E(X-\mu_x)(Y-\mu_y)$$

که μ_x و μ_y به ترتیب امید ریاضی X و Y می‌باشند. رابطه بالا را می‌توان به صورت زیر نیز نوشت:

$$\begin{aligned} Cov(X,Y) &= E[XY-\mu_yX-\mu_xY+\mu_x\mu_y] \\ &= E(XY)-\mu_yE(X)-\mu_xE(Y)+\mu_x\mu_y \\ &= E(XY)-\mu_x\mu_y-\mu_x\mu_y+\mu_x\mu_y = E(XY)-\mu_x\mu_y \end{aligned}$$

قضیه ۳-۱۱-۴ اگر دو متغیر X و Y مستقل باشند آنگاه:

$$Cov(X,Y) = 0$$

قضیه ۳-۱۱-۵ اگر X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند آنگاه:

$$V(X+Y) = V(X)+V(Y)+۲\,Cov(X,Y)$$

۳-۱۲ ضریب همبستگی دو متغیر تصادفی

ضریب همبستگی دو متغیر تصادفی X و Y را در جامعه با ρ نمایش می‌دهند و به صورت زیر تعریف می‌شود.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X-\mu_x)^2 . E(Y-\mu_y)^2}}$$

۳-۱۲-۱ ویژگیهای ضریب همبستگی

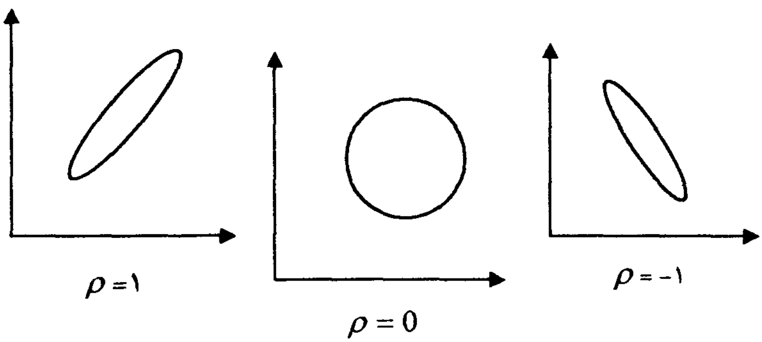
۱- همواره $-1 \le \rho \le 1$ و مستقل از واحد اندازه‌گیری است.

۲- هنگامی که $\rho = 1$ است همبستگی دو متغیر X و Y شدید و هم سو است. یعنی افزایش یکی باعث افزایش دیگری می‌شود.

۳- هنگامی که $\rho = -1$ است همبستگی دو متغیر X و Y شدید و خلاف هم است. یعنی افزایش یکی باعث کاهش دیگری می‌شود.

۴- هنگامی که ρ در همسایگی صفر است همبستگی دو متغیر ضعیف است.

۵- باتوجه به مقدار ρ ، نمودار پراکنش X و Y به صورت زیر دسته‌بندی می‌شود.



۶- $\rho_{aX+b,\, cY+d} = \rho_{X,Y}$

یعنی اگر متغیرهای X و Y را در مقدار ثابت ضرب کنیم و مقدار ثابت به آنها اضافه کنیم تغییری در همبستگی ایجاد نمی‌شود.(ثابت کنید.)

مثال ۳-۱۲-۲ اگر X و Y دارای تابع چگالی احتمال تـوام زیر باشند میانگین، واریانس، کوواریانس و ضریب همبستگی X و Y را محاسبه کنید.

$$f(x,y) = x + y \qquad 0 < x < 1 \qquad 0 < y < 1$$

$$f_x(x) = x + \frac{1}{2} \qquad 0 < x < 1$$

$$f_y(y) = y + \frac{1}{2} \qquad 0 < y < 1$$

$$E(X)=\int_0^1 x\,f_x(x)\,dx=\int_0^1 x\,(x+\frac{1}{2})\;dx=\frac{7}{12}$$

$$E(Y)=\int_0^1 y\,f_y(y)\,dy=\int_0^1 y(y+\frac{1}{2})\;dy=\frac{7}{12}$$

$$E(X^2)=\int_0^1 x^2\,f_x(x)\,dx=\int_0^1 x^2\,(x+\frac{1}{2})\;dx=\frac{5}{12}$$

$$E(Y^2)=\frac{5}{12}$$

$$V(X)=E(X^2)-(E(X))^2=\frac{5}{12}-\left(\frac{7}{12}\right)^2=\frac{11}{144}$$

$$V(y)=\frac{11}{144}$$

$$E(XY)=\int_0^1\int_0^1 xy\,f(x,y)\,dy\,dx=\int_0^1\int_0^1 xy\,(x+y)\,dy\,dx=\frac{1}{3}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{3}-\left(\frac{7}{12}\right)\left(\frac{7}{12}\right)=\frac{-1}{144}$$

$$\rho=\frac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}}=\frac{-\frac{1}{144}}{\sqrt{\frac{11}{144}\cdot\frac{11}{144}}}=\frac{-1}{11}$$

تعریف ۳-۱۲-۳ دو متغیر تصادفی X و Y را ناهمبسته گوییم اگر $\rho_{x,y}=0$ یا $Cov(X,Y)=0$ باشد.

اگر دو متغیر تصادفی X و Y مستقل باشند آنگاه $Cov(X,Y)=0$ و $\rho=0$ و نتیجه می‌شود که X و Y ناهمبسته هم هستند. اما ناهمبسته بودن ($\rho=0$) مستقل بودن را نتیجه نمی‌دهد.

۳-۱۳ چولگی و برجستگی در جامعه

در آمار توصیفی میزان چولگی و برجستگی را به ترتیب از فرمولهای زیر محاسبه کردیم.

$$b = \frac{m_3}{S^3}$$

$$k = \frac{m_4}{S^4} - 3$$

میزان چولگی و برجستگی در جامعه به ترتیب از فرمولهای زیر محاسبه می‌شود

چولگی $$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{E(X-\mu)^3}{\sigma^3}$$

برجستگی $$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{E(X-\mu)^4}{\sigma^4} - 3$$

مثال ۳-۱۳-۱ اگر متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد α_3 و α_4 را محاسبه کنید.

x	-۳	-۲	-۱	0	۱	۲	۳
$f(x)$	۰/۰۶	۰/۰۹	۰/۱	۰/۵	۰/۱	۰/۰۹	۰/۰۶

$$\mu = E(X) = \sum x\, f(x) = \sum_{x=-3}^{3} x\, f(x) = 0$$

$$\mu_3 = E(X-0)^3 = E(X^3) = \sum_{x=-3}^{3} x^3 f(x) = 0$$

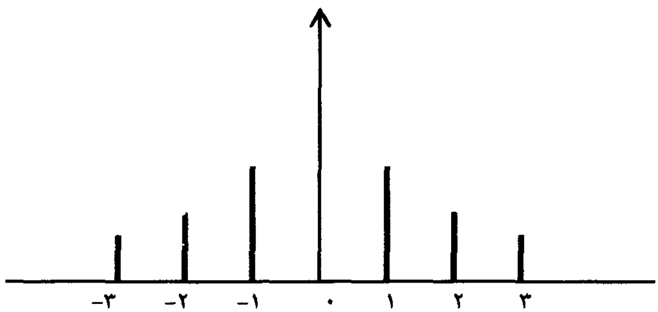
$$\mu_4 = E(X-0)^4 = E(X^4) = \sum_{x=-3}^{3} x^4 f(x) = 12.8$$

$$\sigma^2 = E(X-\mu)^2 = E(X-0)^2 = E(X^2) = \sum_{x=-3}^{3} x^2 f(x) = 2$$

$$\sigma = \sqrt{2} = 1.414$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{0}{2\sqrt{2}} = 0$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} - 3 = \frac{12.8}{2\times 2} = 0.2$$



باتوجه به نمودار، توزیع فاقد چولگی است. ولی دارای کشیدگی زیاد می‌باشد . مقدار آن برابر با $\alpha_4=0/2$ است.

۳-۱۴ تابع مولد گشتاورها

اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x)$ باشد تابع مولد گشتاورها در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می‌شود.

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$$

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

به طوری که $|t|<h$.

تابع مولد گشتاورها تعریف خاصی از امید ریاضی است. اگر $h(X) = e^{tX}$ تعریف شود $E[h(X)]$ همان تعریف تابع مولد گشتاورها است و در مواقعی که محاسبه $E(X^r)$ برای بعضی از توزیع‌ها وقت‌گیر است از $M_X(t)$ استفاده می‌شود.

$$M_X(t)=E[e^{tX}]=\int_{-\infty}^{\infty} e^{tx}f(x)dx$$

از $M_X(t)$ نسبت به t مشتق‌های متوالی می‌گیریم.

$$M'_X(t)=\frac{d}{dt}E[e^{tX}]=E(Xe^{tX})$$

$$M''_X(t)=\frac{d^2}{dt^2}E[e^{tX}]=E(X^2e^{tX})$$

پس از r بار مشتق‌گیری

$$M^r_X(t)=\frac{d^r}{dt^r}E[e^{tX}]=E(X^re^{tX})$$

برای $t=0$ در مشتق rام

$$E(X^r)=M^r_X(0)$$

۳-۱۴-۱ ویژگیهای تابع مولد گشتاورها

۱- $M_X(0)=1$

۲- $M'_X(0)=E(X)=\mu$

۳- $M''_X(0)-(M'_X(0))^2=V(X)$

۴- $M_{\frac{X+a}{b}}(t)=e^{\frac{at}{b}}M_X\left(t/b\right)$

مثال ۳-۱۴-۲ اگر متغیر تصادفی دارای تابع احتمال زیر باشد $M_X(t)$ را بدست آورید.

$$f(x)=\frac{e^{-\lambda}\lambda^x}{x!}\qquad x=0,1,2,\ldots\quad,\quad \lambda>0$$

$$M_X(t)=E[e^{tX}]=\sum_{x=0}^{\infty}e^{tx}\frac{e^{-\lambda}\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}=e^{-\lambda}\,e^{\lambda e^t}=e^{-\lambda(1-e^t)}$$

مثال ۳-۱۴-۳ اگر متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد $M_X(t)$ را حساب کنید و نشان دهید که ۱= $M_X(0)$ و $M'_X(0)=0$ است.

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \qquad -\infty<x<\infty$$

$$M_X(t)=\int_{-\infty}^{\infty}\frac{e^{tx}}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\frac{t^2}{2}}e^{-\frac{1}{2}(x-t)^2}dx$$

$$e^{\frac{t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t)^2}dx=e^{\frac{t^2}{2}}$$

$$M'_X(0)=0 \text{ ، } M'_X(t)=t\,e^{\frac{t^2}{2}} \text{ ، } M_X(0)=1 \text{ و } M_X(t)=e^{\frac{t^2}{2}}\ .$$

۳-۱۵ نامساوی مارکف و چبیشف

فرض کنید متغیر تصادفی X دارای فضای مفروض A باشد به طوری که اعضای A همه مثبت باشند و a یک عدد بزرگتر از صفر باشد. نامساوی مارکف را تحت قضیه زیر بیان می‌کنیم.

قضیه ۳-۱۵-۱ اگر متغیر تصادفی X دارای فضای A باشد و $E(X)$ موجود باشد آنگاه همواره:

$$P[X\geq a]\leq\frac{E(X)}{a}$$

برهان: تابع اشاره $I(X)$ را به صورت زیر تعریف می‌کنیم.

$$I(X)=\begin{cases}1 & \text{اگر} \quad X\geq a\\ 0 & \text{اگر} \quad X<a\end{cases}$$

چون $X\geq 0$ است پس:

$$I(X)\leq\frac{X}{a}$$

از طرفین رابطه اخیر امید ریاضی می‌گیریم:

$$E[I(X)] \le E\left[\frac{X}{a}\right]$$

$$1 \times P(X \ge a) \le \frac{E(X)}{a} \quad , \qquad P(X \ge a) \le \frac{E(X)}{a}$$

۳-۱۵-۲ نامساوی چبیشف

اگر متغیر تصادفی X دارای میانگین μ و واریانس σ^2 باشد آنگاه برای هر $k > 0$

$$P[|X - \mu| \ge k] \le \frac{\sigma^2}{k^2}$$

اثبات این نامساوی باتوجه به قضیه مارکف بسیار ساده است. اگرچه می‌توان آن را مستقیماً نیز ثابت کرد.

اگر k^2را برابر a و $(X-\mu)^2$ را X فرض کنیم شرایط مارکف تأمین می‌شود و

$$P[(X-\mu)^2 \ge k^2] \le \frac{E(X-\mu)^2}{k^2}$$

$$P[|X-\mu|^2 \ge k^2] \le \frac{\sigma^2}{k^2}$$

$$P[|X-\mu| \ge k] \le \frac{\sigma^2}{k^2}$$

اهمیت نامساوی مارکف و چبیشف در این است که ما را قادر می‌سازد با معلوم بودن میانگین و واریانس جامعه، کرانهای بالا و پایین را برای مقادیر مختلف احتمال بدست آوریم، گرچه فرم تابع چگالی احتمال معلوم نیست.

مثال ۳-۱۵-۳ فرض کنید تعداد محصولات تولید شده در یک کارخانه در طول هفته یک متغیر تصادفی با میانگین $\mu = ۵۰$ و واریانس $\sigma^2 = ۲۵$ باشد. مطلوبست:

الف- احتمال اینکه تولید محصول در یک هفته معین بیش از ۷۵ باشد.

ب- احتمال اینکه محصول یک هفته معین بین ۴۰ و ۶۰ باشد.

$$P[X \ge 75] \le \frac{E(X)}{75} \quad , \qquad P(X \ge 75) \le \frac{50}{75} = \frac{2}{3}$$

$$P[۴۰ \le X \le ۶۰] = P[-۱۰ \le X - ۵۰ \le ۱۰] = P[-۱۰ \le X - \mu \le ۱۰]$$
$$= P[|X - \mu| \le ۱۰] = ۱ - P[|X - \mu| > ۱۰]$$

اما

$$P[|X - \mu| > 10] \le \frac{E(X-\mu)^2}{100} = \frac{25}{100} = \frac{1}{4}$$

بنابراین

$$P[40 \le X \le 60] \ge 1 - \frac{1}{4}$$

$$P[40 \le X \le 60] \ge 0.75$$

خودآزمایی

۱- یک سکه نا اریب سه بار پرتاب می‌شود. اگر متغیر X تعداد شیرها در سه پرتاب باشد تابع چگالی احتمال X را بنویسید و احتمالات زیر را حساب کنید.

$P(X=0)\quad,\quad P(X=۲)\quad,\quad P(X\geq ۳)\quad,\quad P(X<۳)$

۲- سکه‌ای طوری تنظیم شده که شانس ظاهر شدن شیر درآن دو برابر خط است. اگر متغیر X تعداد شیرها در سه پرتاب باشد تابع چگالی احتمال X را بنویسید.

۳- یک سکه نااریب را آنقدر پرتاب می‌کنیم که اولین شیر ظاهر شود. اگر متغیر تصادفی X تعداد پرتابهای مورد نیاز برای ظاهر شدن اولین شیر باشد فضای نمونه X را بنویسید و یک قالب احتمال برای آن بنویسید.

۴- سایتی دارای ۶ رایانه است که دوتای آن معیوب هستند. اگر به تصادف دوتا از این رایانه‌ها را انتخاب کنیم و متغیر تصادفی X تعداد رایانه معیوب باشد تابع چگالی احتمال و تابع توزیع X را بدست آورید.

۵- جعبه‌ای شامل چهار مهره به شماره‌های ۳,۲,۱ و ۴ است. دو مهره به تصادف از جعبه استخراج می کنیم. اگر X عدد بزرگتر دو مهره انتخابی باشد تابع چگالی احتمال و تابع توزیع X را بدست آورید.

۶- از فضای نمونه $S=\{s|0<s<۱۰\}$ یک نقطه اختیار می‌کنیم. فرض کنید $S'\subset S$. اگر تابع مجموعه احتمال $P(S')$ به صورت زیر باشد

$$P(S')=\int_{S'}\frac{1}{10}dz$$

و $X=X(s)=۲s-۱۰$ تعریف کنیم تابع احتمال متغیر تصادفی X را پیدا کنید.

۷- تابع $f(x)$ به صورت زیر تعریف شده است

$$f(x)=c\left[\left(\frac{1}{2}\right)^{x}-\frac{1}{2}\right] \qquad x=0,\ ۱,۲,\ldots$$

الف- به ازای چه مقدار c تابع $f(x)$ یک تابع چگالی احتمال است.
ب- تابع توزیع آن را بدست آورید.

۸- متغیر تصادفی X دارای تابع چگالی احتمال به صورت زیر است

$$f(x)=c(۸-x) \qquad x=0,\ ۱,۲,۳,۴,۵$$

الف- مقدار c و $F(x)$ را پیدا کنید.
ب- احتمال $P(X>۲)$ را حساب کنید.

۹- اگر X دارای تابع توزیع $F(x)$ باشد.

$$F(x)=\begin{cases}0 & x<0\\ 1-\left(\frac{1}{2}\right)^{x+1} & x=0,1,2,\ldots\end{cases}$$

الف- تابع چگالی احتمال X را پیدا کنید.
ب- احتمال $P(۱۰<X\le ۲۰)$ را حساب کنید.
ج- مقدار P [X زوج باشد] را بیابید.

۱۰- یک متغیر تصادفی پیوسته X که مقادیرش بین $x=۲$ و $x=۵$ فرض شده دارای تابع چگالی احتمال $f(x)=\dfrac{2(1+x)}{27}$ است. تابع توزیع X را بدست آورید و احتمالات زیر را حساب کنید.

$$P(X<۴) \quad , \quad P(۳\le X<۴) \quad , \quad P(0<X<۳)$$

۱۱- به ازای چه مقدار c تابع زیر یک تابع چگالی است؟

$$f(x)=c \quad x^{-(c+1)} \qquad 0<x<\infty$$

۱۲- متغیر تصادفی X دارای تابع توزیعی است به گونه‌ای که:

$$F(x)=\begin{cases} \frac{x}{2} & 0<X\le {1}/{2} \\ x-\frac{1}{2} & 1<X<{3}/{2} \end{cases}$$

الف- نمودار $F(x)$ را رسم کنید.

ب- مطلوبست $P(X<\frac{1}{2})$.

۱۳- در یک فروشگاه رایانه تقاضای سالانه برای یک بسته‌بندی نرم‌افزار یک متغیر تصادفی گسسته است. صاحب فروشگاه چهار نسخه از بسته بندی را با قیمت ۸۰۰۰ تومان برای هر نسخه سفارش می‌دهد و هر نسخه را به ۲۸۰۰۰ تومان به مشتریها می‌فروشد. در پایان سال بسته‌بندی منسوخ می‌شود و فروشنده سرمایه‌اش را برای نسخه‌های فروش نرفته از دست می‌دهد. اگر تابع توزیع X به صورت زیر باشد تابع احتمال X را بدست آورید و احتمالات زیر را حساب کنید.

$P(X<۲)$, $P(۱<X<۳)$, $P(X=۴)$

$$F(x)=\begin{cases} 0 & x<0 \\ 0.1 & 0\le x<1 \\ 0.4 & 1\le x<2 \\ 0.7 & 2\le x<3 \\ 0.9 & 3\le x<4 \\ 1 & x\ge 1 \end{cases}$$

۱۴- متغیر تصادفی X دارای تابع چگالی احتمال زیر می‌باشد. عدد m را میانه توزیع گوییم اگر داشته باشیم $P(X\le m)=P(X\ge m)=\frac{1}{2}$. میانه را برای متغیر X پیدا کنید.

$$f(x)=\frac{2x}{9} \qquad 0<x<۳$$

۱۵- ضرایب معادله درجه دوم $x^2+bx+c=0$ را با ریختن دو تاس سالم تعیین می‌کنیم، به طوری که اولین برآمد b و دومین برآمد c درنظر گرفته شود. مطلوبست:
الف- فضای نمونه برای (b,c)
ب- احتمال اینکه معادله دارای جواب باشد چقدر است؟

۱۶- از بسته‌ای که شامل ۳ رایانه با مدل I، ۲ رایانه با مدل II و ۳ رایانه با مدل III است یک نمونه تصادفی ۴ تایی انتخاب می‌کنیم. اگر متغیرهای تصادفی X و Y به ترتیب تعداد رایانه‌های مدل I و II باشند.
الف- تابع احتمال توام (X,Y) را بدست آورید.
ب- احتمال $P[(X,Y)\in A]$ را به طوری که $A=\{(x,y)\,|\,x+y\le 2\}$ باشد حساب کنید.

۱۷- اگر X و Y دارای تابع احتمال توام $f(x,y)$ باشد مطلوبست:

$$f(x,y)=c\frac{2^{x+y}}{x!y!} \qquad x=0,1,2 \qquad y=0,1,2$$

الف- مقدار c
ب- تابع چگالی حاشیه‌ای X و Y
ج- آیا X و Y ازهم مستقل‌اند (راهنمایی $\sum_{x=0}^{\infty}\frac{2^x}{x!}=e^2$)

۱۸- فرض کنید X_1 و X_2 متغیرهای تصادفی گسسته با تابع توام به صورت جدول زیر باشند.

X_1 \ X_2	۱	۲	۳
۱	$\frac{1}{12}$	$\frac{1}{6}$	0
۲	0	$\frac{1}{9}$	$\frac{1}{5}$
۳	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$

الف- تابع چگالی حاشیه‌ای X_1 و X_2 را بیابید.
ب- آیا X_1 و X_2 مستقل اند؟
ج- $P[X_1 \leq X_2]$ را حساب کنید.

۱۹- تابع توزیع مشترک X و Y عبارتست از

$$F(x,y)=P(X\leq x,\ Y\leq y)=\begin{cases}0 & x\leq 0\ ,\ \ y\leq 0\\ 1-2^{-x}-2^{-y}+2^{-xy} & x>0\ ,\ \ y>0\end{cases}$$

مطلوبست احتمال اینکه X یکی از مقادیر خود را در فاصله (۲و۱) و Y در فاصله (۵ و ۳) اختیار کند.

۲۰- اگر

$$f_x(x)=\frac{1}{10} \qquad x=0,1,2,\ldots,9$$

$$f(y|x)=\frac{1}{10-x} \qquad y=x,x+1,\ldots,9$$

مطلوبست تابع احتمال توام $f(x,y)$ تابع چگالی حاشیه $f_Y(y)$.

۲۱- اگر متغیر تصادفی X دارای تابع احتمال زیر باشد میانگین، واریانس و انحراف توزیع را حساب کنید.

x	۱۳۱	۱۴۰	۱۶۰	۱۸۰
$f(x)$	۰/۰۵	۰/۱۰	۰/۲۵	۰/۶۰

۲۲- متغیر تصادفی گسسته X دو مقدار x_1 و x_2 را اختیار می‌کند ($x_2 > x_1$). احتمال اینکه x_1 را اختیار کند برابر با ۰/۶ است. تابع چگالی X را مشخص کنید و میانگین و واریانس را بدست آورید.

۲۳- اگر X دارای تابع چگالی احتمال زیر باشد مطلوبست $E(X)$ و $E(X^2+9)$.

$$f(x)=\begin{cases}\dfrac{x}{2} & 0<x\le 1\\ \dfrac{1}{2} & 1<x\le 2\\ \dfrac{3-x}{2} & 2<x<3\end{cases}$$

۲۴- بازرگانی یک قلم جنس را به ارزش ۱۶۰۰ تومان می‌خرد و به قیمت ۲۰۰۰ تومان می‌فروشد. احتمال خرابی برای ۴,۳,۲,۱,۰ یا۵ قلم بیشتر به ترتیب برابر با ۰/۰۵ ، ۰/۱۵، ۰/۳۰ ، ۰/۲۵ ، ۰/۱۵ و ۰/۱۰ می باشد. امید سود حاصل از انبار کردن ۴,۳,۲,۱,۰ یا۵ قلم یا بیشتر را محاسبه کنید.

۲۵- R شعاع اندازه‌گیری شده یک دایره دارای تابع چگالی احتمال زیر است. مطلوبست:

$f(r)=۶r(۱-r) \qquad 0<r<۱$

الف- میانگین شعاع

ب- میانگین محیط

ج- میانگین سطح

۲۶- اگر X و Y دو متغیر تصادفی باشند.

الف- نشان دهید که $Cov(aX+c,bY+d)=ab\,Cov(X,Y)$.

ب- اگر $E(X)=۷$، $E(Y)=۱$ و $E(XY)=۸$ باشد مقدار $Cov(۲X-۱,۳Y+۴)$ را محاسبه کنید.

۲۷- دو متغیر تصادفی X و Y دارای تابع احتمال توام زیر است.

(x,y)	(0,0)	(0,۱)	(۱,0)	(۱,۱)	(۲,0)	(۲,۱)
$f(x,y)$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{6}{18}$	$\frac{1}{18}$

مطلوبست، $E(X+Y)$ ، $E(X|Y=0)$ ، $Cov(X,Y)$ و ρ .

۲۸- ثابت کنید گشتاورهای مرکزی مرتبه دوم μ_2 از گشتاورهای مرتبه دوم نسبت به نقطه دلخواه c یعنی $\mu_2' = E(X-c)^2$ کوچکتر است.

۲۹- درستی تساوی زیر را نشان دهید.

$$\mu_3 = E(X^3) - 3E(X).E(X^2) + 2(E(X))^3$$

۳۰- اگر X و Y دو متغیر مستقل باشند و تعریف کنیم $Z = X+Y$ نشان دهید که گشتاورهای مرکزی مرتبه سوم Z برابر با مجموع گشتاورهای مرکزی مرتبه سوم X و Y است. یعنی

$$\mu_3^z = \mu_3^x + \mu_3^y$$

۳۱- فرض کنید X و Y دارای تابع چگالی احتمال توام زیر باشند.

$$f(x,y) = \frac{1}{20} \qquad\qquad (x,y) \in A$$

که $A = \{(x,y) | 0 < x < 10 \;,\; x-1 < y < x+1\}$. $f(x,y)$ را رسم کنید و ضریب همبستگی X و Y را محاسبه کنید.

۳۲- براساس اطلاعات گذشته احتمالهای نیروگاه اتمی مربوط به عمر مفید (سال) در جدول زیر ثبت شده است. چولگی و برجستگی را بررسی کنید.

عمر مفید (سال)	۱۰	۲۰	۳۰	۴۰
احتمالهای نیروگاه اتمی	۰/۰۵	۰/۲۵	۰/۵۰	۰/۲۰

۳۳- تابع مولد گشتاورها را برای توابع چگالی احتمال زیر محاسبه کنید.

۱- $f(x) = p^x(1-p)^{1-x} \qquad x = 0, 1$

۲- $f(x) = p(1-p)^x \qquad x = 0,1,2,...$

۳- $f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}} \qquad x>0$

۴- $f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ x^{\alpha-1}(1-x)^{\beta-1} \qquad x>0$

۳۴- اگر متغیر تصادفی X دارای تابع مولد گشتاورهای زیر باشد میانگین و واریـانس X را حساب کنید.

$$M_X(t)=e^{\mu t+\frac{1}{2}t^2\sigma^2}$$

۳۵- اگر متغیر تصادفی X دارای تابع احتمال زیر باشد. تابع مولد گشتاورهای X را بدست آورید.

x	$-\sqrt{3}$	0	$\sqrt{3}$
$f(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

۳۶- نـشان دهیـد کـه تـابع چگـالی احتمـال زیـر دارای تـابع مولـد گـشتاورهای $M_X(t)=\frac{1}{1-t^2}$ است.

$$f(x)=\frac{1}{2}e^{-|x|} \qquad -\infty<x<\infty$$

۳۷- اگر X دارای تابع چگالی احتمـال $f(x)=۱$ و $0<x<۱$ احتمـال زیـر را بـه طریق مستقیم و با استفاده از نامساوی چبیشف حساب کنید و آنها را مقایسه کنید.

$P[|X-\mu|>۴]$

که $\mu=E(X)$ است.

۳۸- در نامساوی چبیشف، اگر بخواهیم وقتی متغیر تـصادفی مقـداری بـین $\mu-k\sigma$ و $\mu+k\sigma$ اختیار می‌کند با احتمال حـداقل ۰/۹۵ همـراه باشـد، کمتـرین مقـدار k چـه خواهد بود؟ (μ و σ به ترتیب میانگین و انحراف معیار متغیر است.)

۳۹- متغیر تصادفی X دارای میانگین μ=۸ و واریانس $\sigma^۲$=۹ می باشد. مطلوبست $P[-۴<X<۲۰]$.

۴۰- اگر X دارای تابع چگالی احتمال زیر باشد مطلوبست $P[0<X<۲]$.

x	0	۱	۳
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

فصل ۴

توزیع احتمال‌های خاص

مقدمه

هدف این فصل معرفی چند توزیع احتمال خاص از نوع گسسته و پیوسته با ارائه الگو می‌باشد. با ارائه فرم تابع چگالی احتمال هر توزیع چند ویژگی توزیع نیز مورد بحث قرار می‌گیرد.

۴-۱ توابع احتمال خاص گسسته

در این بخش توابع احتمال یکنواخت، برنولی، دوجمله‌ای، دو جمله‌ای منفی، هندسی، فوق هندسی، پواسن، سری لگاریتمی و سری لگاریتمی مارکف با ارائه الگو معرفی می‌شود.

۴-۱-۱- تابع احتمال یکنواخت

متغیر تصادفی X دارای تابع احتمال یکنواخت با پارامتر k است اگر تابع احتمال آن به صورت زیر باشد:

$$f(x)=P(X=x)=\frac{1}{k} \qquad x=۱,۲,\ldots,\ k$$

الگو: جعبه‌ای شامل صفحه کلید با شماره‌های ۱ تا k است. اگر هم شانس بودن را برای همه شماره‌ها یکسان درنظر بگیریم و تعریف کنیم

X : شماره صفحه کلید خارج شده آنگاه X دارای تابع چگالی احتمال یکنواخت گسسته است. باتوجه به تعریف امید ریاضی و واریانس در فصل سوم، میانگین و واریانس این توزیع محاسبه می شود.

$$E(X)=\sum_{x=1}^{k} xf(x)=\sum_{x=1}^{k} x.\frac{1}{k}=\frac{1}{k}\sum_{x=1}^{k} x=\frac{k+1}{2}$$

برای محاسبه واریانس $E(X^{2})$ را محاسبه می کنیم.

$$E(X^2)=\sum_{x=1}^{k} x^2 f(x)=\frac{1}{k}\sum_{x=1}^{k} x^2=\frac{1}{k}\frac{k(k+1)(2k+1)}{6}=\frac{(k+1)(2k+1)}{6}$$

$$V(X)=E(X^2)-(E(X))^2=\frac{(k+1)(2k+1)}{6}-\frac{(k+1)^2}{4}=\frac{(k+1)(k-1)}{12}$$

مثال ۴-۱-۲- انتخاب یک ماه بخصوصی از ۱۲ ماه سال دارای توزیع یکنواخت با $k=۱۲$ است.

$$f(x)=\frac{1}{12} \qquad x=۱,۲,\dots,۱۲$$

۴-۱-۳ تابع احتمال برنولی

متغیر تصادفی X دارای تابع احتمال برنولی با پارامتر (شانس) p است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=p(X=x)=p^x(1-p)^{1-x} \qquad x=0,۱$$

الگو: جعبه ای شامل صفحه کلیدهایی از نوع دست دوم و نو با نسبت های p-۱ و p است. یک صفحه کلید به تصادف از جعبه خارج کنیم و اگر متغیر X را به صورت زیر تعریف کنیم:

$$X=\begin{cases}0 & \text{اگر صفحه کلید خارج شده دست دوم باشد}\\ 1 & \text{اگر صفحه کلید خارج شده نو باشد}\end{cases}$$

آنگاه X دارای تابع برنولی است.

۴-۱-۴ تابع احتمال دو جمله‌ای

متغیر تصادفی X دارای تابع احتمال دوجمله‌ای با پارامترها n و p است. اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, ۱, ۲, \ldots, n$$

nCx

الگو: جعبه‌ای شامل صفحه کلیدهای از نوع دست دوم و نو با نسبت‌های p و p-۱ است. از این جعبه در شرایط یکسان و به تصادف n صفحه کلید یکی یکی و با جایگذاری خارج می‌کنیم و اگر تعریف کنیم

X : تعداد صفحه کلیدهای نو خارج شده آنگاه X دارای توزیع دوجمله‌ای است.

مثال ۴-۱-۵ تعداد سئوالاتی که لازم است یک کاربر رایانه پاسخ گوید ۱۰۰ سئوال چهار جوابی است. اگر او به تصادف پاسخ گوید مطلوبست:

۱- احتمال اینکه دقیقاً به ۵۰ سئوال پاسخ صحیح دهد.

۲- احتمال اینکه به بیش از ۵۰ سئوال پاسخ صحیح دهد.

$$n = 100 \ , \ p = \frac{1}{4} \ , \quad f(x) = \binom{100}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{100-x} \qquad x = 0, ۱, \ldots, ۱۰۰$$

$$f(50) = \binom{100}{50} \left(\frac{1}{4}\right)^{50} \left(\frac{3}{4}\right)^{50}$$

$$P(X > 50) = 1 - P(X \le 50) = 1 - F(50) = 1 - \sum_{t=0}^{50} \binom{100}{t} \left(\frac{1}{4}\right)^t \left(\frac{3}{4}\right)^{100-t}$$

۴-۱-۶ ویژگیهای توزیع دو جمله‌ای

۱- $$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1$$

۲- دارای نمای منحصر به فرد است.

۳- برای $n=۱$ ، تابع چگالی احتمال دو جمله‌ای همان تابع احتمال برنولی است.

۴- دارای میانگین np و واریانس $np(۱-p)$ است.

۵- $$f(x)=\frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\left(\frac{p}{1-p}\right)^{x}f(0)$$

۶- برای مقادیر مختلف n و p می توان از جدول ضمیمه (۱) مقدار $F(x)$ را محاسبه کرد.

۴-۱-۷ تابع احتمال دو جمله‌ای منفی (پاسکال)

متغیر تصادفی X دارای تابع احتمال دو جمله‌ای منفی با پارامتر r و p است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x)=p(X=x)=\binom{x+r-1}{r-1}p^{r}(1-p)^{x} \qquad x=0,۱,۲,\ldots$$

الگو: جعبه‌ای شامل الگوریتم برنامه‌های درست و نادرست با نسبت‌های p و $q=۱-p$ است. از این جعبه برنامه‌ها را یک به یک و بدون جایگذاری و به تصادف خارج می‌کنیم. خارج کردن برنامه‌ها را آنقدر ادامه می‌دهیم که r امین برنامه خارج شده درست باشد. اگر: rامین برنامه درست باشد

X : تعداد برنامه‌های نادرست بیش از r امین برنامه درست باشد آنگاه X دارای توزیع دوجمله‌ای منفی یا پاسکال است. همانطور که می دانید تعداد کل برنامه‌های خارج شده برابر با $X+r$ است.

مثال ۴-۱-۸ فرض کنید خط تولیدی تا زمانی نسبت به تولید کالا اقدام می‌کند که اولین تولید معیوب باشد، اگر شانس تولید کالای معیوب ۰/۰۵ و تولید کالاها از هم مستقل باشند مطلوبست:

P ٤ ,۹۵

الف- احتمال اینکه اولین کالای معیوب تولید شده در پنجمین تولید باشد. ۲٤۴ ۹٤۱

ب- احتمال اینکه اولین کالای معیوب تولید شده در پنجمین تولید و قبل از آن باشد.

$q=0/05 \quad , \quad p=0/95 \quad , \quad r=1 \quad , \quad x+r=5$

$$p(X=1)=\binom{4}{0}(0.05)(0.95)^4=0.0407$$

$$p(X\le 5)=\sum_{t=0}^{4}\binom{t+1-1}{1-1}p^1(1-p)^t=\sum_{t=0}^{4}p(1-p)^t$$
$$=0.50+0.05\times0.95+0.05\times(0.95)^2+0.05\times(0.95)^3+0.05\times(0.95)^4$$
$$=0.2262$$

۴-۱-۹ ویژگیهای توزیع دو جمله‌ای منفی

۱- $\sum_{x=0}^{\infty}\binom{x+r-1}{r-1}p^r(1-p)^x=\sum_{x=0}^{\infty}\binom{x+r-1}{x}p^r(1-p)^x=p^r p^{-r}=1$

۲- دارای میانگین $\dfrac{r(1-p)}{p}$ و واریانس $\dfrac{r(1-p)}{p^2}$ است.

۴-۱-۱۰ تابع احتمال هندسی

متغیر تصادفی X دارای تابع احتمال هندسی با پارامتر p است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x)=p(X=x)=p(1-p)^x \qquad x=0,1,2,\dots$$

الگو: اگر در الگو دوجمله‌ای منفی r را برابر با یک درنظر بگیریم الگوی فوق یک الگو برای توزیع هندسی است. یا به عبارتی، توزیع هندسی حالت خاص توزیع دوجمله‌ای منفی است وقتی که $r=1$ است.

مثال ۴-۱-۱۱ اگر احتمال خرابی رایانه‌ای در طول ماه ۰/۰۰۵ باشد احتمال خرابی آن در ماه پنجم چقدر است؟

در این مثال $p=0/005$ و $x=5$ پس،

$$p[x=5]=(0/005)(1-0/005)^5=(0/005)(0/995)^5=0/0049$$

۴-۱-۱۲ ویژگیهای توزیع هندسی

۱- این توزیع فاقد حافظه است یعنی $p[X \geq s+t \mid X \geq t] = p[X \geq s]$

۲- دارای میانگین $\dfrac{1-p}{p}$ و واریانس $\dfrac{1-p}{p^2}$ است.

۳- $\displaystyle\sum_{x=0}^{\infty} p(1-p)^x = 1$

۴-۱-۱۳ تابع احتمال فوق هندسی

متغیر تصادفی X دارای تابع احتمال فوق هندسی با پارامترهای N، k و n است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p[X = x] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0, ۱, ۲, \ldots, n$$

الگو: جعبه‌ای شامل N مهره است که k تا از آنها سفید و بقیه سیاه هستند از این جعبه به تصادف n مهره بدون جایگذاری خارج می‌کنیم. طبق اصل دوم شمارش و قانون ترکیب تعداد حالات برابر با $\binom{N}{n}$ است. اگر X تعداد مهره های سفید در n مهره خارج شده باشد آنگاه X دارای توزیع فوق هندسی است. چون تعداد انتخاب x مهره سفید از k مهره سفید برابر با $\binom{k}{x}$ و تعداد $n-k$ مهره سیاه از $N-k$ برابر با $\binom{N-k}{n-k}$ است.

$k=50$ $N=200$

مثال ۴-۱-۱۴ جعبه‌ای شامل ۲۰۰ برنامه رایانه‌ای است که ۵۰ تا از آنها دارای خطا است. از این جعبه به تصادف ۳۰ برنامه خارج می‌کنیم مطلوبست: $n=30$

$x=0$ الف- احتمال اینکه نمونه انتخاب شده فاقد برنامه خطا باشد.

$x \leq 2$ ب- احتمال اینکه نمونه انتخاب شده حداکثر شامل ۲ برنامه دارای خطا باشد.

$$f(x)=P[X=x]=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}=\frac{\binom{50}{x}\binom{150}{30-x}}{\binom{200}{30}}$$

$$f(0)=\frac{\binom{50}{0}\binom{150}{30}}{\binom{200}{30}}$$

$$P(X\le 2)=P(X=0)+P(X=1)+P(X=2)=f(0)+f(1)+f(2)$$

۴-۱-۱۵ ویژگیهای توزیع فوق هندسی

۱- $\sum_{x=0}^{n}\binom{k}{x}\binom{N-k}{n-x}=\binom{N}{n}$

۲- N یک عدد صحیح مثبت، k یک عدد صحیح نامنفی $(k\le N)$ و n یک عدد نامنفی و حداکثر برابر با N است.

۳- دارای میانگین $\frac{nk}{N}$ و واریانس $\left(\frac{nk}{N}\right)\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)$ است.

۴-۱-۱۶ تابع احتمال پواسن

متغیر تصادفی X دارای تابع احتمال پواسن با پارامتر λ است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x)=p(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\qquad x=0,۱,۲,\dots$$

الگو: اگر در الگوی مربوط به توزیع دوجمله‌ای تعداد آزمایش ها n و شانس موفقیت هر آزمایش p به ترتیب از حد تصور بزرگتر و کوچکتر باشند به طوری که $np=\lambda$ باشد آنگاه می‌توان توزیع دوجمله‌ای را با استفاده از توزیع پواسن تقریب زد. اگر در دو جمله‌ای $p=\frac{\lambda}{n}$ قرار دهیم.

$$\lim_{n\to\infty} P(X=x)=\lim_{n\to\infty}\left[\frac{n!}{x!(n-x)}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}\right]$$

$$=\lim_{n\to\infty}\left[\frac{n(n-1)...(n-x+1)}{n^x}\left(\frac{\lambda^x}{x!}\right)\left(\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}\right)\right]$$

$$=\lim_{n\to\infty}\left[(1)\left(1-\frac{1}{n}\right)...\left(1-\frac{x-1}{n}\right)\right]\left(\frac{\lambda^x}{x!}\right)\left[\lim_{n\to\infty}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}\right]$$

$$=\frac{\lambda^x e^{-\lambda}}{x!}$$

مهم **مثال ۴-۱-۱۷** فرض کنید در سازمانی تعداد رایانه‌هایی که در طول یک سال از رده خارج می‌شوند دارای توزیع پواسن با $\lambda=۳$ است. مطلوبست :

الف- احتمال اینکه یک رایانه خارج شود.

ب- حداکثر دو رایانه خارج شود.

ج- در طول یک سال از ۲۰۰ رایانه چند رایانه از رده خارج می شود؟

$$f(x)=\frac{e^{-\lambda}\lambda^x}{x!}=\frac{e^{-3}3^x}{x!} \qquad x=0,۱,۲,...$$

$$f(1)=\frac{e^{-3}.3}{1}=0.149$$

$$P(X\le 2)=f(0)+f(1)+f(2)=0.05+0.149+0.224=0.423$$

ج- حدوداً ۳۰ رایانه، ۲۹/۸=۰/۱۴۹ × ۲۰۰

۴-۱-۱۸ ویژگیهای توزیع پواسن

۱- $\displaystyle\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=1$

۲- دارای نمای منحصر به فرد است.

۳- دارای میانگین λ و واریانس λ است.

۴- برای مقادیر مختلف λ می توان از جدول ضمیمه (۲) مقدار $F(x)$ را محاسبه کرد.

۴-۱-۱۹ تابع احتمال سری لگاریتمی

متغیر تصادفی X دارای تابع احتمال سری لگاریتمی با پارامتر α است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = P(X = x) = \frac{-1}{\ln(1-\alpha)}\ \frac{\alpha^x}{x} \qquad x = ۱,۲,\ldots$$

الگو: با استفاده از الگوی دو جمله‌ای منفی می‌توان الگویی برای این توزیع ارائه داد که از ذکر آن در اینجا خودداری می‌شود.

مثال ۴-۱-۲۰ تعداد مقالاتی که از استادان چاپ می‌شود دارای توزیع سری لگاریتمی با پارامتر $\alpha =$ ۰/۵۶ مطلوبست:

الف- احتمال اینکه استادی دارای یک مقاله چاپ شده باشد.

ب- احتمال اینکه حداقل دو مقاله چاپ شده داشته باشد.

ج- از بین ۱۵۴۰ استاد به طور متوسط چند نفر یک مقاله چاپ شده دارند؟

$$f(x) = P(X = x) = \frac{-1}{\ln(1-0.56)}\ \frac{(0.56)^x}{x} \qquad x = ۱,۲,\ldots$$

$$f(1) = \frac{-1}{\ln(0.44)}\ \frac{(0.56)^x}{1} = 0.682$$

$$f(x \geq 2) = 1 - P(X < 2) = 1 - P(X = 1) = 1 - 0.682 = 0.318$$

ج- $n\,P(X = 1) = 1540 \times 0.682 = 10\ 50$ نفر

توزیع سری لگاریتمی دارای میانگین $\frac{\alpha}{1-\alpha} \cdot \frac{-1}{\ln(1-\alpha)}$ است.

۴-۱-۲۱ تابع احتمال سری لگاریتمی مارکف

متغیر تصادفی X دارای تابع احتمال سری لگاریتمی مارکف با پارامترهای αو β است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = P(X=x) = \frac{-1}{\ln(1-\alpha)} \cdot \frac{(1-\beta)^x - \left[1-\frac{\beta}{1-\alpha}\right]^x}{x} \qquad x = ۱,۲,\ldots$$

۴-۱-۲۲ ویژگیهای سری لگاریتمی مارکف

۱- برای $\beta = ۱\text{-}\alpha$ توزیع سری لگاریتمی مارکف به توزیع سری لگاریتمی تبدیل می‌شود.

۲- دارای میانگین $\frac{-\alpha}{\beta \ln(1-\alpha)}$ است.

مثال ۴-۱-۲۳- طول نوبت بارندگی دارای توزیع سری لگاریتمی مارکف با $\alpha = ۰/۶۳$ و $\beta = ۰/۳$ است مطلوبست:

الف- احتمال اینکه طول نوبت بارندگی برابر با ۱ باشد.

ب- احتمال اینکه طول نوبت بارندگی حداکثر ۲ باشد.

$$f(x) = P(X=x) = \frac{-1}{\ln(1-0.63)} \cdot \frac{(1-0.3)^x - \left[1-\frac{0.3}{1-0.63}\right]^x}{x}$$

$$f(1) = 0.514$$

$$f(x \le 2) = f(1) + f(2) = 0.514 + 0.228 = 0.742$$

۴-۲ توابع چگالی احتمال خاص پیوسته

در این بخش توابع چگالی احتمال یکنواخت، نرمال، نرمال استاندارد، نمایی، گاما، کی‌دو، بتا، استودنت و فیشر ارائه می‌شود.

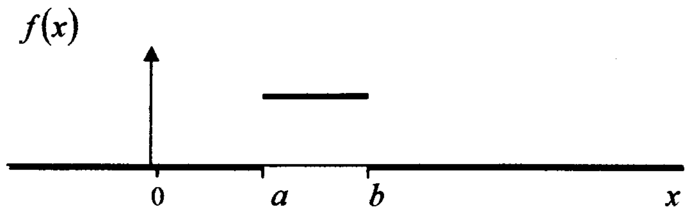
۴-۲-۱ تابع چگالی احتمال یکنواخت (مستطیلی)

متغیر تصادفی X دارای تابع چگالی احتمال یکنواخت با پارامترهای aو b است اگر تابع چگالی آن به صورت زیر باشد.

$$f(x)=\frac{1}{b-a} \qquad a<x<b$$

۴-۲-۲ ویژگیهای تابع چگالی احتمال یکنواخت

۱- نمودار $f(x)$ برای $-\infty<a<b<\infty$ به صورت زیر است.



۲- تابع توزیع $F(x)$ برابر است با:

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(t)dt=\begin{cases} 0 & x<a \\ \dfrac{x-a}{b-a} & a\leq x<b \\ 1 & x\geq b \end{cases}$$

۳- دارای میانگین $\frac{a+b}{2}$ و واریانس $\frac{(b-a)^2}{12}$ است.

۴- برای $a=0$ ، $b=۱$ تابع $f(x)$ را روی بازه (۰،۱) گویند و آن را بـه صـورت زیـر تعریف می کنند.

$$f(u)=۱ \qquad 0<u<۱$$

یا

$$f(u)=I_{(0,1)}^{(u)}$$

که $I_{(0,1)}^{(u)}$ را تابع نشانگر گویند.

ذکر این نکته ضروری است که تابع توزیع هر متغیر تصادفی همانند $0 < u < ۱$ و $f(u)$ عمل می کند چون $F(+\infty)=۱$ ، $F(-\infty)=0$ پس $u = F(x)$ است. از این خاصیت در آمار برای شبیه سازی متغیرهای تصادفی استفاده می‌کنند.

مثال ۴-۲-۳ اگر متغیر X دارای تابع چگالی احتمال $0 < x < ۵$، $f(x) = \frac{1}{5}$ باشد. مطلوبست:

الف- $P(X < ۱)$ ب- $P(0 < X < ۴)$

ج- $P(x = 0)$ د- $P(۴ < X < ۷)$

$$P(X < 1) = F(1) = \int_0^1 f(x)dx = \int_0^1 \frac{1}{5}dx = \frac{1}{5}$$

$$P(0 < X < 4) = F(4) - F(0) = \int_0^4 f(x)\,dx = \int_0^4 \frac{dx}{5} = \frac{4}{5}$$

$$P(X = 0) = \int_0^0 f(x) = 0$$

$$P(4 < X < 7) = \int_4^7 f(x)dx$$

$$= \int_4^5 f(x)dx + \int_5^7 f(x)dx = \int_4^5 \frac{1}{5}\,dx + \int_5^7 0\ dx = \int_4^5 \frac{1}{5}\ dx = \frac{1}{5}$$

۴-۲-۴ تابع چگالی احتمال نرمال

متغیر تصادفی نرمال یکی از توزیع‌های مهم آماری در حالت پیوسته است. متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\;e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$-\infty < x < \infty$$

$$-\infty < \mu < \infty$$

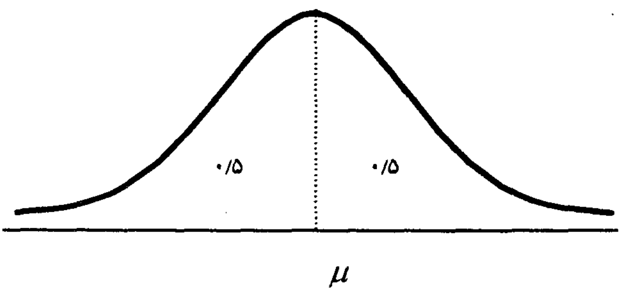
$$\sigma > 0$$

μ و $\sigma^۲$ پارامترهای توزیع نرمال هستند.

۴-۲-۵ ویژگیهای توزیع نرمال

۱- این توزیع نسبت به محور $y=\mu$ دارای تقارن است.

۲- $\int_{-\infty}^{\infty} f(x)\,dx = 1$



۳- $P[X \le \mu] = P[X \ge \mu] = ۰/۵$

۴- برای $\mu = 0$ و $\sigma^۲=۱$، توزیع نرمال را توزیع نرمال استاندارد گویند.

۴-۲-۶ توزیع نرمال استاندارد

متغیر تصادفی Z دارای توزیع نرمال استاندارد است اگر تـابع چگـالی احتمـال آن بـه صورت زیر باشد.

$$f(z)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2} \qquad -\infty < Z < \infty$$

همانطور که ملاحظه می‌کنید این توزیع فاقد پارامتر است و برای راحتی متغیر نرمال استاندارد را با Z نمایش می‌دهند. در حقیقت Z همان متغیر X است با میانگین صفر و واریانس یک.

$$f(x) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

مقادیر مختلف $F(z)$ را می توان با توجه به ویژگی Z از جدول ضمیمه (۳) بدست آورد که $F(z) = P(Z < z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$.

متغیر تصادفی نرمال استاندارد نسبت به محور $y = 0$ دارای تقارن است. یعنی:

$$\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \frac{1}{2}$$

مثال ۴-۲-۷ اگر متغیر Z دارای توزیع نرمال استاندارد باشد احتمالات زیر را حساب کنید.

$$P(Z > 0) \quad , \quad P(0 < Z < \infty) \quad , \quad P(۱ < Z < ۳)$$
$$P(-۲ < Z < 0) \quad , \quad P(-۱/۹۶ < Z < ۱/۹۶) \quad , \quad P(Z < ۱/۶۴)$$

باتوجه به جدول ضمیمه (۳)

$$P(Z > 0) = P(Z \leq 0) = 1 - 0.5 = 0.5$$

$$P(0 < Z < \infty) = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = F(+\infty) - F(0) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(1 < Z < 3) = P(Z < 3) - P(Z < 1) = F(3) - F(1) = 0.9987 - 0.8413 = 0.1574$$

$$\begin{aligned} P(-2 < Z < 0) &= P(Z < 0) - P(Z < -2) \\ &= P(Z < 0) - P(Z > 2) \\ &= P(Z < 0) - [1 - P(Z < 2)] \\ &= P(Z < 0) + P(Z < 2) - 1 \\ &= 0.5 + 0.9772 - 1 = 0.4772 \end{aligned}$$

$$P(-1.96 < Z < 1.96) = P(Z < 1.96) - P(Z < -1.96)$$
$$= P(Z < 1.96) - P(Z > 1.96) = P(Z < 1.96) - [1 - P(Z < 1.96)]$$
$$= 2P(Z < 1.96) - 1 = 2F(1.96) - 1 = 2(0.9750) - 1 = 0.95$$

$$P(Z > 1.64) = 1 - P(Z < 1.64) = 1 - F(1.64) = 1 - 0.9495 = 0.0505$$

در آمار توزیع نرمال را با نمـاد $N(\mu,\sigma^۲)$ و نرمـال اسـتاندارد را بـا $N(0,۱)$ نـشان می‌دهند و برای استفاده از جدول نرمال استاندارد برای متغیر تصادفی X بـا میـانگین μ و واریانس $\sigma^۲$ همواره تبدیل زیر را داریم.

$$Z = \frac{X-\mu}{\sigma}$$

باتوجه به تبدیل یا رابطه بین X و Z ، می توان احتمال مربوط به X را درهر فاصله با استفاده از جدول نرمال استاندارد محاسبه کرد. برای مثال:

$$P[a < X < b] = P\left[\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right]$$

$$= P\left[\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right]$$
$$= P\left(Z < \frac{b-\mu}{\sigma}\right) - P\left(Z < \frac{a-\mu}{\sigma}\right)$$
$$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

مثال ۴-۲-۸ اگر مقدار اشعه‌ای خاص که کاربر رایانه ممکن است در هر ساعت کاری دریافت کند دارای توزیـع نرمـال بـا میـانگین ۴/۳۵= μ و واریـانس ۰/۴۹= $\sigma^۲$ باشـد مطلوبست:

الف- مابین ۴ و ۵ واحد باشد.

ب- حداقل ۵/۷ واحد باشد.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \frac{1}{0.7\sqrt{2\pi}} e^{-\frac{1}{2\times 0.49}(x-4.35)^2}$$

$$P(4<X<5)=\int_4^5 \frac{1}{0.7\sqrt{2\pi}}\, e^{-\frac{1}{2\times 0.49}(x-4.35)^2}\, dx$$

محاسبه انتگرال وقت گیر است. اما با تبدیل X به Z داریم.

$$\begin{aligned}
P(4<x<5) &= P\left[\frac{4-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{5-\mu}{\sigma}\right]\\
&= P\left[\frac{4-4.35}{0.7}<\frac{X-\mu}{\sigma}<\frac{5-4.35}{0.7}\right]\\
&= P[-0.5<Z<0.93]=P(Z<0.93)-P(Z<-0.05)\\
&= P(Z<0.93)-P(Z>0.05)=P(Z<0.93)+P(Z<0.05)-1\\
&= 0.8238+0.5199-1=0.3437
\end{aligned}$$

$$P(X\geq 5.7)=P\left(\frac{X-\mu}{\sigma}>\frac{5.7-4.35}{0.7}\right)=P(Z>1.93)$$

$$=1-P(Z<1.93)=1-F(1.93)=1-0.9732=0.0268$$

۴-۲-۹ تابع چگالی احتمال نمایی

متغیر تصادفی X دارای تابع چگالی احتمال نمایی با پارامتر θ است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\frac{1}{\theta}\, e^{-x/\theta} \qquad x>0\ ,\quad \theta>0$$

توزیع نمایی کاربردهای مهمی دارد. از جمله در مدلهای صف بندی، می‌توان نشان داد که زمان انتظار مابین ورودی‌های متوالی از توزیع نمایی پیروی می‌کند.

۴-۲-۱۰ ویژگیهای توزیع نمایی

۱- $F(x)=P(X<x)=\int_{-\infty}^{x} f(t)\,dt=1-e^{-x/\theta}$

۲- فاقد حافظه است. $P[X>s+t\mid X>t]=P(X>s)$

۳- دارای میانگین θ و واریانس θ^2 است.

۴- اگر u دارای توزیع یکنواخت روی $(0,1)$ باشد آنگاه $-\ln(u)$ دارای توزیع نمایی با $\theta=1$ است.

مثال ۴-۲-۱۱ مدت زمانی که رایانه‌ای بدون نیاز به تعمیر کار کند، متغیری تصادفی نمایی با $\theta=4$ سال است. مطلوبست:

الف- احتمال اینکه رایانه‌ای در کمتر از ۳/۵ سال نیازی به تعمیر نداشته باشد.

ب- حداقل ۴/۵ سال نیاز به تعمیر نداشته باشد.

ج- بین ۲ الی ۴ سال نیازی به تعمیر نداشته باشد.

$$f(x)=\frac{1}{\theta}e^{-x/\theta}=\frac{1}{4}e^{-x/4} \qquad x>0$$

$$P(X<3.5)=\int_{-\infty}^{3.5}\frac{1}{4}e^{-x/4}dx=\int_{-\infty}^{3.5}\frac{1}{4}e^{-x/4}dx=F(3.5)$$
$$=1-e^{-3.5/4}=0.583$$

$$P(X>4.5)=1-P(X\le 4.5)=1-F(4.5)=e^{-4.5/4}=0.325$$

$$P(2<X<4)=P(X<4)-P(X<2)=F(4)-F(2)$$
$$=-e^{-1}+e^{-1/2}=0.2387$$

۴-۲-۱۲ تابع چگالی احتمال گاما

متغیر تصادفی X دارای تابع چگالی احتمال گاما با پارامترهای α و β است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} \qquad x>0\ ,\quad \alpha>0\ ,\quad \beta>0$$

حالت خاص: برای $\alpha=1$، $\beta=\theta$ توزیع گاما به توزیع نمایی تبدیل می‌شود. تابع چگالی احتمال گاما باتوجه به ویژگی تابع گاما تعریف می‌شود. چون:

$$\int_0^\infty f(x)dx = \int_0^\infty \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty (\beta u)^{\alpha-1}\, e^{-u}\beta\, du$$

$$= \frac{1}{\Gamma(\alpha)}\int_0^\infty e^{-u}u^{\alpha-1}\, du = 1$$

یا

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}du$$

۴-۲-۱۳ ویژگیهای توزیع گاما

۱- $F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}dx = 1 - \sum_{i=0}^{\alpha-1} \frac{1}{i!}\left(\frac{x}{\beta}\right)^i e^{-x/\beta}$

۲- دارای میانگین $\alpha\beta$ و واریانس $\alpha\beta^{2}$ است.

مثال ۴-۲-۱۴ در یک شهر مصرف برق روزانه دارای توزیع گاما با $\alpha=3$ و $\beta=2$ است. اگر ظرفیت روزانه ۱۲ میلیون کیلووات ساعت باشد. احتمال اینکه برق موجود برای یک روز کافی باشد چقدر است؟

$$P(X \le 12) = F(12) = 1 - \sum_{i=0}^{3} \frac{1}{i!}\left(\frac{12}{2}\right)^i e^{-12/2}$$

$$= 1 - e^{-6}[1+6+18+36] = 0.849$$

۴-۲-۱۵ تابع چگالی احتمال کی‌دو

متغیر تصادفی X دارای تابع چگالی احتمال کی‌دو با پارامتر r است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x) = \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})}\, x^{\frac{r}{2}-1}\, e^{-x/2} \qquad x>0\ , \quad r = ۱,۲,۳,\ldots$$

توزیع کی دو حالت خاص توزیع گاما است $\left(\alpha = \frac{r}{2}, \quad \beta = 2\right)$.

۴-۲-۱۶ ویژگیهای توزیع کی دو

۱- r را درجه آزادی توزیع گویند.

۲- دارای میانگین r و واریانس $2r$ است.

۳- $$F(x) = P(X \le x) = 1 - \sum_{i=0}^{r-1} \frac{1}{i!}\left(\frac{x}{2}\right)^i e^{-x/2}$$

۴- مقادیر مختلف $F(x)$ را می توان برای مقادیر مختلف r از جدول ضمیمه (۵) بدست آورد.

مثال ۴-۲-۱۷ باتوجه به جدول ضمیمه (۵) احتمالات زیر را حساب کنید.

الف- احتمال $P(X > ۵/۸۹۲)$ با $r = ۱۳$

ب- احتمال $P(X < ۳۴/۱۷۰)$ با $r = ۲۰$

ج- $P(۶/۲۶۲۲ < X < ۲۲۴/۹۹۶)$ با $r = ۱۵$

د- $P(X > x_0) = ۰/۰۲۵$ با $r = ۲۵$

باتوجه به درجه آزادی در ستون اول و عدد ۵/۸۹۲ ، $P(X > ۵/۸۹۲) = ۰/۹۵$

$P(X < ۳۴/۱۷۰) = ۱ - P(X > ۳۴/۱۷۰) = ۱ - ۰/۰۵ = ۰/۹۵$

$P(۲۲۴/۹۹۶ < X < ۶/۲۶۲۲) = P(X < ۲۴/۹۹۶) - P(X < ۶/۲۶۲)$

$= ۱ - P(X \ge ۲۴/۹۹۶) - ۱ + P(X \ge ۶/۲۶۲)$

$= P(X \ge ۶/۲۶۲) - P(X \ge ۲۴/۹۹۶) = ۰/۹۷۵ - ۰/۰۵ = ۰/۹۷$

با توجه به درجه آزادی و مقدار احتمال ۰/۰۲۵

$P(X > x_0) = ۰/۰۲۵ \Rightarrow x_0 = ۴۰/۶۴۶$

۴-۲-۱۸ تابع چگالی احتمال بتا

متغیر تصادفی X دارای تابع چگالی احتمال بتا با پارامترهای α و β است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,x^{\alpha-1}\,(1-x)^{\beta-1}\qquad 0<x<1\ ,\quad \alpha>0\ ,\quad \beta>0$$

۴-۲-۱۹ ویژگیهای توزیع بتا

۱- $\int_0^1 x^{\alpha-1}\,(1-x)^{\beta-1}dx=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$

۲- برای ۱= α و ۱= β توزیع بتا به توزیع یکنواخت پیوسته تبدیل می‌شود.

۳- دارای میانگین $\frac{\alpha}{\alpha+\beta}$ و واریانس $\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$ است.

مثال ۴-۲-۲۰ قسمت‌هایی از بزرگراهی که در طول یکسال احتیاج به تعمیر دارد، دارای توزیع بتا با ۳= α و ۲= β است. مطلوبست:

الف- احتمال اینکه نیمی از بزرگراه در یکسال احتیاج به تعمیر داشته باشد.

ب- احتمال اینکه ۷۵ درصد احتیاج به تعمیر داشته باشد.

$$P(X<0.5)=\int_0^{0.5}\frac{\Gamma(3+2)}{\Gamma(3)\Gamma(2)}x^2\,(1-x)\,dx=\frac{5}{16}$$

$$P(X<\frac{2}{3})=\int_0^{2/3}12\,x^2(1-x)\,dx=\frac{16}{27}$$

۴-۲-۲۱ تابع چگالی احتمال استودنت (توزیع t)

متغیر تصادفی X دارای توزیع t با پارامتر r است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\frac{\Gamma\left[\frac{(r+1)}{2}\right]}{\Gamma\left(\frac{r}{2}\right)}\cdot\frac{1}{\sqrt{r\pi}}\cdot\frac{1}{\left(1+\frac{x^2}{r}\right)^{\left(\frac{r+1}{2}\right)}} \qquad -\infty<x<\infty\ ,\quad r>0$$

که r را درجه آزادی توزیع t گویند.

۴-۲-۲۲ ویژگیهای توزیع t

۱- $\int_{-\infty}^{\infty} f(x)dx = ۱$

۲- برای $r>۱$ دارای میانگین صفر و برای $r>۲$ دارای واریانس $\frac{r}{r-2}$ است.

۳- در توزیع استودنت اگر درجه آزادی r از حد تصور بزرگتر باشد توزیع، بر توزیع نرمال استاندارد منطبق می‌شود.

۴- مقادیر مختلف $F(x)$ برای مقادیر مختلف درجه آزادی r از جدول ضمیمه (۴) قابل محاسبه است.

مثال ۴-۲-۲۳ اگر X دارای توزیع استودنت با $r=۷$ باشد. احتمالات زیر راحساب کنید.

$P(T>۲/۳۶۵)\ ,\quad P(T<۳/۴۹۹)\ ,\quad P(۱/۴۱۵<T<۱/۸۹۵)$

چون درجه آزادی برابر با ۷ است ردیف ۷ و ستون اول را درنظر می‌گیریم.

$P(T>۲/۳۶۵)=۰/۰۲۵$

$P(T<۳/۴۹۹)=۱-P(T\geq ۳/۴۹۹)=۱-۰/۰۰۵=۰/۹۹۵$

$P(۱/۴۱۵<T<۱/۸۹۵)=P(T<۱/۸۹۵)-P(T<۱/۴۱۵)$

$=P(T>۱/۴۱۵)-P(T>۱/۸۹۵)=۰/۱۰-۰/۰۵=۰/۰۵$

۴-۲-۲۴ تابع چگالی احتمال فیشر

متغیر تصادفی X دارای توزیع فیشر با پارامترهای r_1 و r_2 است، اگر تابع چگالی آن به صورت زیر باشد

$$f(x)=\frac{\Gamma\left[\frac{(r_1+r_2)}{2}\right]}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}\cdot\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}}\frac{x^{\frac{r_1}{2}-1}}{\left(1+\frac{r_1}{r_2}x\right)^{\frac{r_1+r_2}{2}}}\qquad x>0$$

که r_1 و r_2 به ترتیب درجه آزادی صورت و مخرج خوانده می‌شود برای مقادیر مختلف r_1 و r_2 مقادیر مختلف $F(x)$ از جدول ضمیمه (۶) قابل محاسبه است.

خودآزمایی

۱- توزیع یکنواخت را برای نمونه تصادفی سه تایی که از ۵ رایانه مختلف می‌توانند تشکیل شود بدست آورید.

۲- احتمال اینکه یک تیرانداز هدفی را بزند برابر با ۰/۷۵ است. یک قالب احتمال برای پیروزی و عدم پیروزی تیرانداز ارائه دهید و شانس عدم موفقیت ایشان را بدست آورید.

۳- احتمال اینکه در یک خط تولید کالایی معیوب باشد برابر با ۰/۰۰۱ است. در یک نمونه ۷ تایی مطلوبست احتمال اینکه:

الف- کالای معیوب پیدا نشود.

ب- بیش از ۵ کالا معیوب باشد.

ج- دقیقاً ۳ کالا معیوب باشد.

۴- یک محموله از پنجاه قطعه مکانیکی شامل چهل و دو قطعه سالم و هشت قطعه معیوب است. یک بازرس پنج قطعه را بدون جایگذاری انتخاب می‌کند.

الف- یک قالب احتمال ارائه دهید.

ب- احتمال اینکه درست سه قطعه سالم باشد چقدر است؟

ج- احتمال اینکه حداکثر سه قطعه سالم باشد چقدر است؟

۵- احتمال اینکه یک وسیله اندازه‌گیری یک جابه جایی مفرط را نشان دهد برابر با ۰/۰۵ است. احتمال اینکه ششمین وسیله‌ای که آزمون شده است اولین وسیله‌ای باشد که جابه جایی را نشان می دهد چقدر است؟

۶- برای تابع چگالی احتمال هندسی امید ریاضی و واریانس را بدست آورید.

$$f(x) = P(X = x) = P(1 - P)^x \qquad x = 0, ۱,۲,...$$

۷- در مسئله ۶ نشان دهید که

$$P[X \geq s+t \mid X \geq t] = P(X \geq s)$$

۸- در آبگیری ۲۰۰ ماهی قزل آلا وجود دارد. از این آبگیر ۵۰ ماهی را به تصادف صید کرده و پس از علامت گذاری آنها را به آبگیر بر می‌گردانیم. اگر مجدداً ۵۰ ماهی را به تصادف صید کنیم احتمال اینکه ۵ ماهی از آنها علامت‌گذاری شده باشند چقدر است؟ و احتمال اینکه حداکثر ۳ ماهی از ماهیها علامت‌گذاری شده باشد چقدر است؟

۹- یک چاپگر رایانه طوری طراحی شده که در هر ۱۵ ثانیه، ۲ صفحه از اطلاعات ذخیره شده خود را چاپ می‌کند. اگر این چاپگر به مدت ۳ دقیقه کار کند مطلوبست احتمال اینکه:

الف- صفحه‌ای چاپ نکند.

ب- حداقل چهار صفحه چاپ کند.

۱۰- تعداد تصادفاتی که در فواصل مساوی جاده اتفاق می‌افتد دارای توزیع پواسن با پارامتر $\lambda =$ ۰/۳ است. مطلوبست احتمال اینکه:

الف- تصادفی صورت نگیرد.

ب- حداقل یک تصادف اتفاق بیافتد.

ج- بین ۱ تا ۲ تصادف اتفاق بیافتد.

۱۱- در توزیع پواسن، تابع مولد گشتاورها را بدست آورید و از روی آن میانگین و واریانس توزیع را حساب کنید. برای محاسبه $E(X(X-1))$ چه راهی پیشنهاد می‌کنید؟

۱۲- اگر X دارای تابع احتمال سری لگاریتمی با پارامتر $\alpha =$ ۰/۶۳۷۹ باشد جدول زیر را کامل کنید.

x	۱	۲	۳	۴	۵	۶
$f(x)$						

۱۳- تجربه نشان داده که چاپگرهای شرکتی در هر ۱۰۰۰ صفحه یک اشتباه چاپ می‌کند. اگر ۱۰۰۰۰ صفحه به تصادف انتخاب شود احتمال اینکه ۶، ۷ یا ۸ صفحه اشتباه چاپ شود چقدر است؟

۱۴- سوابق در یک سایت رایانه نشان می‌دهد که احتمال اینکه رایانه‌ای از رده خارج شود برابر با ۰/۰۰۱۲ است. احتمال اینکه از بین ۱۰۰۰ رایانه حداقل ۲ رایانه از رده خارج شود چقدر است؟

۱۵- سختی یک آلیاژ معین (اندازه‌گیری شده با مقیاس شکنندگی) یک متغیر تصادفی یکنواخت با $a=۵۰$ و $b=۷۵$ است. تابع توزیع آن را بدست آورید و احتمال $P[۶۰<X<۷۰]$ را پیدا کنید.

۱۶- در آزمایشهای معینی، خطاهای حاصل از تعیین چگالی یک ماده، متغیر تصادفی است با تابع چگالی یکنواخت که درآن $a=-۰/۰۲۵$ و $b=۰/۰۲۵$ می باشد. مطلوبست احتمال خطا:

الف- مابین ۰/۰۱ و ۰/۱۵

ب- مابین ۰/۰۱۲- و ۰/۰۱۲

۱۷- اگر متغیر تصادفی Z دارای توزیع نرمال استاندارد باشد با استفاده از جدول ضمیمه (۳) احتمالات زیر را حساب کنید.

$P[Z<۱/۶۵]$, $P[|Z|<۱/۹۶]$, $P[Z<۰/۲]$

$P[|Z|<۱/۶۴]$, $P[-۰/۵<Z<۳/۲]$, $P[Z^{۲}\leq ۸]$

۱۸- اگر متغیر تصادفی X دارای توزیع نرمال با میانگین $\mu=۳$ و واریانس $\sigma^۲=۰/۱۶$ باشد احتمال‌های زیر را حساب کنید.

$$P[X>۳] \quad , \quad P[X>۳/۳] \quad , \quad P[|X|<۳/۳]$$
$$P[|X-۳|<۱/۵] \quad , \quad P[X<۴/۹۶] \quad , \quad P[۲/۸<X<۳/۱]$$

و مقدار c را به گونه‌ای بیابید که $P[۳-c<X<۳+c]=۰/۹۰۱$.

۱۹- تجربه نشان داده که توزیع نمرات دانشجویان در یک درس دارای توزیع نرمال با میانگین ۷۰ و انحراف معیار ۳ است. یک دانشجو به تصادف انتخاب می‌شود مطلوبست:

الف- احتمال اینکه نمره ایشان بین ۷۴ و ۹۶ باشد.

ب- احتمال اینکه نمره ایشان کمتر از ۷۴ باشد.

ج- چند درصد دانشجویان نمراتشان بیش از ۹۶ است؟

۲۰- استادی از ۵ دانشجو امتحان گرفته که نتایج آن عبارتست از:

۱۸	۱۲	۹/۲۵	۴/۲۵	۰/۷۵

میانگین و واریانس نمونه برابر با ۸/۸۵ و ۳۶/۱۲ . او علاقه مند است به دانشجویان متناسب با نمره‌ای که گرفته‌اند، نمره‌ای اضافه کند به طوری که میانگین نمرات جدید ۱۲ و واریانس آن ۱۶ بشود. چگونه می توان این کار را انجام داد؟ (راهنمایی: نمرات را به داده های استاندارد تبدیل کنید و سپس نمرات جدید را از رابطه $y=۴Z+۱۲$ بدست آورید.)

۲۱- فرض کنید متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ باشد به طوری که $P[X\le ۶۰]=۰/۱$ و $P[X\le ۹۰]=۰/۹۵$ مطلوبست μ و $\sigma^۲$.

۲۲- در فواصل زمانی مساوی مثلاً یک ساعت، تعداد تلفن به‌یک مرکز مخابره می‌شود. اگر فواصل زمانی از هم مستقل و دارای توزیع نمایی با $\theta=۲$ باشد احتمالات زیر را حساب کنید.

الف- $P=(X>۳)$

ب- $P[|X|<۳]$

ج- $P=(X>۷)$

۲۳- اگر X دارای توزیع نمایی با پارامتر θ باشد $M_X(t)$، $M'_X(0)$ و $M''_X(0)$ را بدست آورید.

۲۴- مقدار روزانه (برحسب اینچ) اندازه رسوب در مسیر رودخانه متغیر تصادفی گاما با پارامترهای $\alpha=۶$ و $\beta=/۲$ است. احتمال اینکه مقدار رسوب از یک حدی مثلاً دو اینچ تجاوز کند چقدر است؟

۲۵- تابع مولد گشتاورها را برای توزیع گاما بدست آورید و نشان دهید که تحت چه شرایط توزیع گاما به توزیع نمایی تبدیل می‌شود؟

۲۶- اگر X دارای توزیع کی‌دو با ۵ درجه آزادی باشد مطلوبست:

$E(X+۷)$ $\quad E(X^۲+۱)$ $\quad V(X+۴)$ $\quad V(۲X+۳)$

۲۷- اگر X دارای تابع چگالی بتا با پارامترهای α و β باشد نشان دهید که:

$$E(X)=\frac{\alpha}{\alpha+\beta} \quad , \quad V(X)=\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$$

۲۸- در یک استان، بخشهایی از بزرگراهی که در یک سال احتیاج به تعمیر دارد، دارای توزیع بتا با پارامترهای $\alpha=۳$ و $\beta=۲$ مطلوبست:

الف- به طور متوسط چند درصد از بخشهای بزرگراه در یک سال به تعمیر احتیاج دارد؟

ب- احتمال اینکه حداکثر نیمی از بخشهای بزرگراه در یک سال به تعمیر احتیاج داشته باشد چقدر است؟

۲۹- اگر X دارای استودنت با r درجه آزادی باشد نشان دهید که $E(X)=0$.

۳۰- فرض کنید X دارای توزیع فیشر با درجات آزادی $r_1=7$ و $r_2=25$ باشد. احتمالات زیر را با استفاده از جدول ضمیمه (۶) حساب کنید.

$P[X>3/46]$ $\quad$ $P[X<3/46]$ $\quad$ $P[|X|<3/46]$

فصل ۵

توزیع‌های نمونه‌گیری

مقدمه

در فصل چهارم توزیع‌های مختلف آماری را معرفی کردیم که بعضی از آنها دارای پارامتر یا پارامترهایی بودند. استنباط روی پارامتر یا پارامترها در اغلب موارد با سرشماری جامعه آماری امکان پذیر نیست. لذا در چنین مواقعی باید با استفاده از نمونه، استنباطی روی پارامتر یا پارامترها داشت. به عنوان مثال، فرض کنید برای برآورد کردن عمر متوسط (θ) مفید نوعی معین از ترانزیستور، مهندسی ۱۰ عدد از آنها را انتخاب می‌کند، برای مدت زمانی آنها را مورد آزمایش قرار می‌دهد و زمان از کار افتادن هریک از آنها را یادداشت می کند. اگر این زمانهای از کارافتادن، مقادیر متغیرهای تصادفی $X_1, X_2, \ldots, X_{10}$ باشند که دارای توزیع نمایی با پارامتر θ است، گوییم که این داده‌ها نمونه‌ای از جامعه‌نمایی را تشکیل می‌دهند.

تصورش آسان است که هر نمونه‌ای به طرزی معتبر قابل تعمیم در جامعه‌ای که از آن حاصل شده است نیست. در واقع اغلب روشهای استنباط که در اینجا مطرح می‌شود براساس یک نمونه تصادفی است.

در عمل، اغلب با نمونه‌های تصادفی از جامعه‌هایی سروکار داریم که متناهی، اما به قدر کافی بزرگ‌اند به طوری که گویی نامتناهی‌اند. در ادامه بحث هرگونه استنباط روی پارامتر یا پارامترها براساس نمونه‌هایی از جامعه‌های نامتناهی خواهد بود. برای هر

نمونه تصادفی $X_1, X_2, \ldots, X_n$ از جامعه‌ی نامتناهی برآوردگر، برآورد، توزیع مشترک و توزیع آماره را مورد بحث قرار می‌دهیم.

۵-۱ برآوردگر

هر تابعی از نمونه را که به پارامتر یا پارامترهای جامعه بستگی نداشته باشد برآوردگر یا آماره گویند. چون مقدار برآوردگر از نمونه‌ای به نمونه دیگر تغییر می‌کند متغیری است تصادفی. مقدار عددی آماره یا برآوردگر را برآورد گویند.

مثال ۵-۱-۱ اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از تابع چگالی احتمال پواسن با پارامتر λ باشند آماره‌ایی براساس نمونه تعریف کنید.

۱- $S(X_1, X_2, \ldots, X_n) = X_1$

۲- $S(X_1, X_2, \ldots, X_n) = X_n - X_1$

۳- $S(X_1, X_2, \ldots, X_n) = X_1 + X_2$

۴- $S(X_1, X_2, \ldots, X_n) = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

۵- $S(X_1, X_2, \ldots, X_n) =$ میانه X_i ها

۶- $S(X_1, X_2, \ldots, X_n) = \dfrac{1}{n}\sum_{i=1}^{n} X_i^r$ و $r = 1, 2, \ldots$

ملاحظه می‌شود که تعداد آماره که می‌توان تعریف کرد بیشمار است.

۵-۱-۲ ویژگیهای برآوردگر کارا

۱- ناریب باشد.

۲- دارای کمترین واریانس باشد.

۵-۲ توزیع مشترک

فرض کنید متغیر تصادفی گسسته X دارای تابع احتمال $f(x)=p(X=x)$ باشد و $X_1, X_2, ..., X_n$ نمونه‌های تصادفی مستقل از هم باشند. اگـر $x_1, x_2, ..., x_n$ مقـادیر متناظر مشاهده بـرای $X_1, X_2, ..., X_n$ باشـند. پیـشامدهای $X_1=x_1$ ، $X_2=x_2$ ، ...، $X_n=x_n$ به طور مجزا از هم مستقل‌اند و احتمال توام آنها برابر است با:

$$P[X_1=x_1, X_2=x_2, ..., X_n=x_n]=P(X_1=x_1)P(X_2=x_2)...P(X_n=x_n)$$

اگر از نماد $f(x_1, x_2, ..., x_n)$ به جای $P[X_1=x_1, X_2=x_2, ..., X_n=x_n]$ اسـتفاده کنیم.

$$f(x_1, x_2, ..., x_n)=f(x_1)f(x_2)...f(x_n)=\prod_{i=1}^{n} f(x_i)$$

$f(x_1, x_2, ..., x_n)$ را تابع چگالی احتمال توام یا تابع چگالی مشترک گویند.

برای $n=2$ ، $f(x_1, x_2)=f(x_1)f(x_2)$ تـابع چگـالی احتمـال تـوام X_1 و X_2 می‌باشد. در حالتی که متغیر تصادفی X از نوع پیوسته است تابع چگـالی مـشترک یـا توام را با $f(x_1, x_2, ..., x_n)$ نیز نمایش می‌دهند.

مثال ۵-۲-۱ اگر X_1, X_2, X_3 یک نمونه سه تایی از تابع احتمال دوجمله‌ای باشند توزیع مشترک آنها را بنویسید.

$$f(x)=\binom{n}{x}p^x(1-p)^{n-x} \qquad x=0, 1, 2, ..., n$$

تابع احتمال برای نمونه اول
$$f(x_1)=\binom{n}{x_1}p^{x_1}(1-p)^{n-x_1}$$

تابع احتمال برای نمونه دوم
$$f(x_2)=\binom{n}{x_2}p^{x_2}(1-p)^{n-x_2}$$

تابع احتمال برای نمونه سوم
$$f(x_3)=\binom{n}{x_3}p^{x_3}(1-p)^{n-x_3}$$

$$f(x_1,x_2,x_3)=f(x_1)f(x_2)f(x_3)$$
$$=\prod_{i=1}^{3}f(x_i)=p^{x_1+x_2+x_3}(1-p)^{n-x_1-x_2-x_3}\prod_{i=1}^{3}\binom{n}{x_i}$$

مثال ۵-۲-۲ اگر Z_1، Z_2، Z_3،...، Z_n یک نمونه n تایی از توزیع نرمال استاندارد باشند تابع چگالی احتمال توام یا مشترک را بنویسید.

$$f(z_1,z_2,\dots,z_n)=\prod_{i=1}^{n}f(z_i)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_i^2}=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}z_i}$$

مثال ۵-۲-۳ اگر X_1، X_2،...، X_n یک نمونه از $f(x)=I_{(0,1)}(x)$ باشند تابع چگالی احتمال توام آنها را بنویسید.

$$f(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}f(x_i)=\prod_{i=1}^{n}I_{(0,1)}(x_i)$$

۵-۳ توابع خطی از متغیرهای تصادفی مستقل

فرض کنید X_1، X_2،...، X_n نمونه تصادفی مستقل با توزیع مشترک $f(x_1,x_2,\dots,x_n)$ باشند. یک تابع خطی یا آماره را می‌توان در حالت کلی به صورت زیر تعریف کرد.

$$\sum_{i=1}^{n}a_iX_i=a_1X_1+a_2X_2+\dots+a_nX_n$$

که a_i ها مقادیر ثابت هستند.

برای سادگی فرض کنید متغیرهای تصادفی مستقل X_1 و X_2 دارای چگالی احتمال حاشیه‌ای $f_1(x_1)$، $f_2(x_2)$ و تابع چگالی احتمال توام $f(x_1,x_2)$ باشند. ترکیب خطی زیر را درنظر بگیرید.

$$y=a_1X_1+a_2X_2$$

اگر $X_۱$ و $X_۲$ به ترتیب دارای میانگین‌های $\mu_۱$ و $\mu_۲$ و واریانسهای $\sigma_۱^۲$، $\sigma_۲^۲$ باشند. میانگین و واریانس Y به طریق زیر محاسبه می‌شود:

$$\mu_y = E(a_1X_1 + a_2X_2) = \sum_{x_1}\sum_{x_2}(a_1x_1 + a_2x_2)f_1(x_1)f_2(x_2)$$

$$= a_1\left[\sum_{x_1} x_1 f_1(x_1)\right]\left[\sum_{x_2} f_2(x_2)\right] + a_2\left[\sum_{x_1} f_1(x_1)\right]\left[\sum_{x_2} x_2 f_2(x_2)\right]$$

$$= a_1E(X_1) + a_2E(X_2) = a_1\mu_1 + a_2\mu_2$$

$$\sigma_y^2 = V(Y) = E\left[Y - \mu_y^2\right] = E\left[(a_1X_1 + a_2X_2) - a_1\mu_1 - a_2\mu_2\right]^2$$

$$= E\left[(a_1(X_1 - \mu_1) + a_2(X_2 - \mu_2)\right]^2$$

$$= E\left[(a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\right]$$

$$= a_1^2V(X_1) + a_2V(X_2) + 0$$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2$$

مثال ۵-۳-۱ اگر $X_۱$ و $X_۲$ متغیرهای مستقل با میانگین های $\mu_۱=۶$ و $\mu_۲=۴$ و واریانس‌های $\sigma_۱^۲=۱۶$ و $\sigma_۲^۲=۹$ باشند میانگین و واریانس Y را محاسبه کنید.

$$Y = ۳X_۱ - ۲X_۲$$

$$\mu_y = E(Y) = ۳E(X_۱) - ۲E(X_۲)$$

$$= ۶ \times ۳ - ۴ \times ۲ = ۱۰$$

$$\sigma_y^۲ = V(Y) = V(۳X_۱ - ۲X_۲) = ۹V(X_۱) + ۴V(X_۲) = ۱۸۰$$

۵-۴ توزیع میانگین

در آمار توصیفی، میانگین نمونه تصادفی به صورت $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ تعریف شده بود. در این بخش، توزیع $\overline{X}$ را باتوجه به توزیع جامعه‌ای که نمونه از آن گرفته شده بدست می‌آوریم.

قضیه ۵-۴-۱ اگر $X_1, X_2, \ldots, X_n$ نمونه‌های مستقل و هم توزیع از جامعه‌ای با میانگین μ و واریانس σ^2 باشند آنگاه میانگین نمونه $\overline{X}$ دارای میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

برهان: چون $\overline{X}$ یک ترکیب خطی از X_i هاست، پس:

$$E(\overline{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$$

$$V(\overline{X}) = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\overline{X}$ تابعی از X_i هاست و به پارامترهای جامعه μ و σ^2 بستگی ندارد. یک آماره نااریب و دارای کمترین واریانس است.

مثال ۵-۴-۲ اگر X_1, X_2, X_3 یک نمونه ۳ تایی از جامعه ای با میانگین $\mu=10$ و واریانس $\sigma^2=5$ باشند کدام یک از آماره‌های زیر را ترجیح می‌دهید؟

$$\overline{X}_1 = \frac{1}{3}\sum_{i=1}^{3} X_i \quad , \quad \overline{X}_2 = \frac{1}{4}(X_1 + 2X_2 + X_3)$$

$$E(\overline{X}_1) = E\left(\frac{1}{3}\sum_{i=1}^{3} X_i\right) = \frac{1}{3}\sum_{i=1}^{3} E(X_i) = \frac{3\times 10}{3} = 10$$

$$E(\overline{X}_2) = E\left[\frac{1}{4}(X_1 + 2X_2 + X_3)\right] = \frac{4\times 10}{4} = 10$$

$$V(\overline{X}_1) = V\left(\frac{1}{3}\sum_{i=1}^{3} X_i\right) = \frac{1}{9}\sum_{i=1}^{3} V(X_i) = \frac{3\times 5}{9} = \frac{15}{9}$$

$$V(\overline{X}_2)=V\left[\frac{1}{4}(X_1+2X_2+X_3)\right]=\frac{1}{16}[V(X_1)+4V(X_2)+V(X_3)]$$
$$=\frac{1}{16}(5+20+5)=\frac{30}{16}=\frac{15}{8}$$

هردو آماره نااریب هستند. ولی چون $V(\overline{X}_1)$ کمتر از $V(\overline{X}_2)$ است پس $\overline{X}_1$ را نسبت به $\overline{X}_2$ ترجیح می‌دهیم.

اکنون توزیع $\overline{X}$ را وقتی که توزیع جامعه کاملاً مشخص است بدست می‌آوریم.

قضیه ۵-۴-۳ اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی n تایی از جامعه نرمال با میانگین μ و واریانس σ^2 باشند آنگاه $\overline{X}$ دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

برهان: چون $X \sim N(\mu,\sigma^2)$ است. تابع مولد گشتاورهای آن برابر است با:

$$M_X(t)=e^{\mu_t+\frac{1}{2}\sigma^2t^2}$$

$$M_{\overline{X}}(t)=E\left[e^{t\overline{X}}\right]=E\left[e^{t\cdot\frac{1}{n}\sum X_i}\right]=E\left[e^{\frac{t}{n}(X_1+X_2+\ldots+X_n)}\right]=E\left[e^{\frac{tX_1}{n}+\frac{tX_2}{n}+\ldots+\frac{tX_n}{n}}\right]$$
$$=\left(E\left[e^{\frac{tX}{n}}\right]\right)^n$$

چون X_i ها مستقل و هم توزیع‌اند.

$$=\left[M_x\left(\frac{t}{n}\right)\right]^n=\left[e^{\mu\frac{t}{n}+\frac{1}{2}\frac{\sigma^2t^2}{n^2}}\right]^n=e^{\mu t+\frac{\sigma^2t^2}{2n}}$$

$e^{\mu t+\frac{\sigma^2 t^2}{2n}}$ تابع مولد گشتاورهای متغیر تصادفی نرمال بـا میـانگین μ و واریـانس $\frac{\sigma^2}{n}$ است.

از اینکه $\overline{X}$ یک ترکیب خطی از X_i ها و X_i ها از هم مـستقل‌انـد، امیـد ریاضـی و واریانس مستقیماً به صورت زیر نیز محاسبه می‌شود.

$$E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}n\mu=\mu$$

$$V(\overline{X})=V\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

لم ۵-۴-۴ اگر شرایط قضیه ۵-۴-۳ برقـرار باشـد متغیـر $Z=\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیـع نرمال استاندارد است.

$$E(Z)=E\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)=\frac{\sqrt{n}}{\sigma}E(\overline{X}-\mu)=\frac{\sqrt{n}}{\sigma}(E\overline{X}-\mu)=0$$

$$V(\overline{X})=V\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)=\frac{n}{\sigma^2}V(\overline{X}-\mu)=\frac{n}{\sigma^2}.\frac{\sigma^2}{n}=1$$

لم اخیر در حقیقت استاندارد کردن متغیر تصادفی $\overline{X}$ است. با استفاده از این لم می‌توان احتمالهای مربوط به $\overline{X}$ را با استفاده از جدول نرمال استاندارد محاسبه کرد.

مثال ۵-۴-۵ یک نمونه ۲۵ تایی از جامعه ای از نرمـال بـا میـانگین μ=۷۵ و واریـانس σ^2=۱۰۰ انتخاب می‌کنیم، مطلوبست:

الف- $P=[۷۱< X <۷۹]$

ب- $P=[۷۱< \overline{X} <۷۹]$

X دارای توزیع نرمال با میانگین ۷۵ و واریانس ۱۰۰ و استاندارد آن برابر است با:

$$Z = \frac{X - 75}{100}$$

$$P[71 < X < 79] = P\left[\frac{71-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{79-\mu}{\sigma}\right]$$

$$= P\left[\frac{71-75}{10} < Z < \frac{79-75}{10}\right] = P[-0.4 < Z < 0.4]$$
$$= 2P(Z < 0.4) - 1 = 2(0.6554) - 1 = 0.3108$$

$\overline{X}$ دارای توزیع نرمال با میانگین ۷۵ و واریانس $\frac{100}{25}$ و استاندارد آن برابر است با:

$$Z = \frac{\overline{X} - 75}{2}$$

$$P[71 < \overline{X} < 79] = P\left[\frac{71-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{79-\mu}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$= P\left[\frac{71-75}{2} < Z < \frac{79-75}{2}\right] = P[-2 < Z < 2] = 0.9544$$

باتوجه به قضیه ۵-۴-۳، میانگین نمونه دارای توزیع نرمال است با میانگین μ و واریانس $\frac{\sigma^2}{n}$. اگر X_1، X_2،...، X_n نمونه تصادفی از جامعه ای یا توزیع نرمال با میانگین μ و واریانس σ^2 باشند سئوال این است که اگر جامعه مورد بررسی دارای توزیع نرمال نباشد مثلاً دارای توزیع یکنواخت، نمایی، گاما، پواسن و غیره باشد توزیع $\overline{X}$ چگونه بدست می‌آید. در چنین مواردی تحت شرایط خاص قضیه زیر دریافتن توزیع $\overline{X}$ برای تمام توزیع ها ما را یاری می‌کند.

۵-۵ قضیه حد مرکزی

اگر $\overline{X}$ میانگین نمونه تصادفی $X_1, X_2, \dots, X_n$ از توزیعی (جامعه‌ای) با میانگین μ و واریانس متناهی $\sigma^2 < \infty$ باشند آنگاه توزیع متغیر تصادفی $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ میل می‌کند به توزیع نرمال استاندارد اگر $n \to \infty$

این قضیه با استفاده از تابع مولد گشتاورها به راحتی اثبات می‌شود.

مثال ۵-۵-۱ فرض کنید $\overline{X}$ میانگین یک نمونه ۳۶ تایی از توزیع نمایی با پارامتر (میانگین) θ=۴ باشد. احتمال زیر را با استفاده از قضیه حد مرکزی حساب کنید.

$P\ [۳/۱ < \overline{X} < ۴/۶]$

در این مثال حجم نمونه n=۳۶ تا حدودی بزرگ و $\sigma^2 = V(X) =$ ۱۶ متناهی است. شرایط قضیه حد مرکزی برقرار است.

$$E(X) = \theta = \mu = 4 \qquad , \qquad V(X) = 16 = \sigma^2$$

$$P[3.1 < \overline{X} < 4.6] \approx P\left[\frac{3.1-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{4.6-\mu}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$= P\left[\frac{3.1-4}{\frac{4}{6}} < Z < \frac{4.6-4}{\frac{4}{6}}\right] = P[-1.35 < Z < 0.9]$$

$$= P[z < 0.9] - P[z < 1.35] = 0.7274$$

مثال ۵-۵-۲ فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع پواسن با $\lambda = ۰/۰۲$ و $n = ۱۰۰$ باشند. با استفاده از قضیه حد مرکزی احتمال $P\left[\sum_{i=1}^{n} X_i \geq 3\right]$ را محاسبه کنید.

$$P\left[\sum_{i=1}^{n} X_i > 3\right] = P[\overline{X} > 0.03\ ] \approx P\left[\frac{\overline{X}-\lambda}{\sqrt{\frac{\lambda}{n}}} > \frac{0.03-\lambda}{\sqrt{\frac{\lambda}{n}}}\right]$$

$$= P\left[Z > \frac{0.03-0.0\,2}{\sqrt{\frac{0.02}{100}}}\right] = P\left[Z > \frac{\sqrt{2}}{2}\right] = P[Z > 0.71\,] = 0.2389$$

مثال ۵-۵-۳ اگر $\overline{X}$ میانگین یک نمونه ۴۸ تایی از توزیع یکنواخت روی (۲، ۰) باشد احتمال تقریبی زیر را حساب کنید.

$P\,[۰/۹ < \overline{X} < ۱/۱]$

در این مثال، $f(x) = \frac{1}{2} \qquad 0 < x < ۲$

ابتدا میانگین و واریانس توزیع یکنواخت را حساب می‌کنیم.

$$E(X) = \int_0^2 xf(x)\,dx = \int_0^2 x.\frac{1}{2}\,dx = 1 = \mu$$

$$E(X^2) = \int_0^2 x^2 f(x)\,dx = \int_0^2 x^2.\frac{1}{2}\,dx = \frac{8}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{8}{6} - 1 = \frac{8-6}{6} = \frac{1}{3} = \sigma^2$$

$$P[0.9 < \overline{X} < 1.1] \approx P\left[\frac{0.9-1}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{1.1-1}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$= P\left[\frac{-0.1}{0.08} < Z < \frac{0.1}{0.08}\right] = P[-1.25 < Z < 1.25]$$
$$= 2P(Z < 1.25) - 1 = 2(0.8944) - 1 = 0.7888$$

۶-۵ تقریب نرمال برای توزیع دوجمله‌ای

در توزیع دوجمله‌ای با پارامترهای n و p برای nهای بزرگ محاسبه احتمال گاهی اوقات با استفاده از جدول ضمیمه (۱) خسته کننده و گاهی ممکن است جدولی با چنین nای در دسترس نباشد. به عنوان مثال اگر X دارای توزیع دوجمله‌ای با n=۳۰، $p = \frac{1}{2}$ باشد و بخواهیم احتمال پیشامد $X \geq$ ۱۵ را بدست آوریم، لازم است ۱۶ جمله زیر را محاسبه و جمع کنیم.

$$p(X \geq 15) = \sum_{x=15}^{30}\binom{30}{x}\left(\frac{1}{2}\right)^x = \binom{30}{15}\left(\frac{1}{2}\right)^{15} + \binom{30}{16}\left(\frac{1}{2}\right)^{16} + \dots + \binom{30}{30}\left(\frac{1}{2}\right)^{30}$$

در این مواقع قضیه حد مرکزی تقریب نرمال با میانگین $\mu = np$ و واریانس $\sigma^۲ = np(۱-p)$ تقریب خوبی برای توزیع دوجمله‌ای با پارامترهای n و p است وقتی که p در همسایگی $\frac{1}{2}$ باشد.

می‌دانیم اگر Y دارای توزیع دوجمله‌ای باشد، می‌توان Y را به صورت جمعی از متغیرهای برنولی یعنی $Y = \sum_{i=1}^{n} X_i$ نوشت که X_iها متغیرهای برنولی با میانگین p و واریانس$p(۱-p)$ می‌باشند و مقادیری که Y اختیار می‌کند اعداد صحیح ۰، ۱، ۲، ...، n است.

متغیر Y را که از نوع گسسته است می‌توان با توجه به نتیجه قضیه حد مرکزی به وسیله متغیر نرمال استاندارد تقریب زد. احتمال پیشامد $Y = k$ را می‌توان به صورت زیر تقریب زد.

$$p[Y = k] = p\left[k - \frac{1}{2} < Y < k + \frac{1}{2}\right] \approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} f(y)\,dy$$

یا

$$P[Y=k] \approx P\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}} < \frac{Y-np}{\sqrt{np(1-p)}} < \frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

$$P[Y=k] \approx P\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}} < Z < \frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

$$p[Y=k] = \phi\left[\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right] - \phi\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

مهم

که تابع $\phi(t)$ برابر است با:

$$\phi(t) = p[Z<t] = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

مثال ۵-۶-۱ اگر Y دارای توزیع دوجمله‌ای بـا n=۱۶ و $p=\frac{1}{2}$ باشـد مطلوبـست احتمالات $p\,[Y=۷]$ و $p\,[Y=۶,۷,۸,۹,۱۰]$:

$$P[Y=7] = P[6.5<Y<7.5] \approx \phi\left[\frac{7.5-np}{\sqrt{np(1-p)}}\right] - \phi\left[\frac{6.5-np}{\sqrt{np(1-p)}}\right]$$

$$= \phi\left[\frac{7.5-8}{2}\right] - \phi\left[\frac{6.5-8}{2}\right] = \phi(-0.25) - \phi(-0.75)$$

$$= 0.7734 - 0.5987 = 0.1747$$

مقدار دقیق احتمال با استفاده از جدول توزیع دوجمله‌ای برابر است با:

$$P(Y=7)=f(7)=P(Y\le 7)-P(Y\le 6)=0.4018-0.2272=0.1746$$

$$P[Y=6,7,8,9,10]=P[5.5<Y<10.5]\approx\phi\left(\frac{10/5-8}{2}\right)-\phi\left(\frac{5/5-8}{2}\right)=0.7888$$

مقدار دقیق احتمال با استفاده از جدول توزیع دوجمله‌ای برابر است با:

$$P[Y=6,7,8,9,10]=P[Y\le 10]-P[Y\le 5]=0.8949-0.1051=0.7898$$

۵-۷ توزیع واریانس نمونه

واریانس نمونه nتایی در آمار توصیفی به صورت $S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ تعریف شده بود. اکنون برای ناااریب بودن، آن را به صورت $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ تعریف می‌کنیم. هدف از این بخش بدست آوردن توزیع S^2 است. توزیع S^2 به جامعه ای که نمونه از آن استخراج می شود بستگی دارد. برای بدست آوردن توزیع میانگین و واریانس S^2 جامعه مورد بررسی را توزیع نرمال در نظر می‌گیریم و قضایای زیر را برای بدست آوردن توزیع S^2 بیان می‌کنیم.

قضیه ۵-۷-۱ اگر متغیر Z دارای توزیع نرمال استاندارد باشد آنگاه Z^2 دارای توزیع کی‌دو با یک درجه آزادی است.

برهان: با استفاده از تابع مولد گشتاورها

برای متغیر Z ، $M_z(t)=E[e^{tz}]=e^{\frac{1}{2}t^2}$

برای متغیر Z^2 داریم ،

$$M_{z^2}(t)=E[e^{tz^2}]=\int_{-\infty}^{\infty}e^{tz^2}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}dz=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}[1-2t]z^2}dz$$

با فرض $u=\sqrt{1-2t}\,z$ ،

$$M_{z^2}(t)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{u^2}{2}}\,\frac{du}{\sqrt{1-2t}}=(1-2t)^{-\frac{1}{2}}$$

$M_{z^2}(t)=(1-2t)^{-\frac{1}{2}}$ همان تابع مولد گشتاورها توزیع کی‌دو با یک درجه آزادی است.

قضیه ۵-۷-۲ اگر متغیرهای مستقل Z_1، Z_2، ...، Z_n دارای توزیع نرمال استاندارد باشند آنگاه $\sum_{i=1}^{n} Z_i^2$ دارای توزیع کی‌دو با n درجه آزادی است.

اثبات این قضیه با استفاده از تابع مولد گشتاورها آسان است که در اینجا بدون اثبات می‌پذیریم. از این قضیه استنتاج می‌شود که اگر دو متغیر مستقل دارای توزیع کی‌دو باشند جمع آنها نیز توزیع کی‌دو است. در مورد تفاضل هم در شرایط خاص درست است. یعنی اگر دو متغیر مستقل دارای توزیع کی‌دو باشند تفاضل آنها نیز دارای توزیع کی‌دو است با تفاضل درجه آزادی دو متغیر.

نتیجه ۵-۷-۳ اگر X_1، X_2، ...، X_n یک نمونه nتایی از توزیع نرمال با میانگین μ و واریانس σ^2 باشد آنگاه متغیر $\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$ دارای توزیع کی‌دو است با یک درجه آزادی.

این نتیجه از لم ۵-۴-۳ و قضیه ۵-۷-۱ به راحتی قابل نتیجه گیری است. باتوجه به قضایای گفته شده و نتایج، قضیه مهم زیر را بدون اثبات بیان می‌کنیم.

قضیه ۵-۷-۴ اگر $\overline{X}$ و S^2 به ترتیب میانگین و واریانس نمونه X_1، X_2، ...، X_n از جامعه نرمال با میانگین μ و واریانس σ^2 باشد آنگاه

الف- $\overline{X}$ و S^2 از هم مستقل‌اند.

ب- متغیر $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است.

مثال ۵-۷-۵ یک شرکت عینک سازی نوعی شیشه را برای ساخت عدسی خریداری می‌کند و از گذشته تجربه نشان داده است که واریانس ضریب شکست این شیشه $۱/۲۶ \times ۱۰^{-۴}$ است. اگر واریانس نمونه انتخاب شده که شامل ۲۰ قطعه می‌باشد، از $۲/۰۰\times۱۰^{-۴}$ تجاوز کند، شرکت محموله را رد می‌کند. احتمال اینکه شرکت محموله را رد کند چقدر است؟

در این مثال، واریانس جامعه $\sigma^۲=۱/۲۶\times۱۰^{-۴}$ و $n=۲۰$ است. احتمال رد برابر با:

$P[S^۲> ۲/۰۰ \times ۱۰^{-۴}]$

برای محاسبه این احتمال از قضیه ۵-۷-۴ استفاده می‌کنیم.

$$P[S^2 > 2.00\times10^{-4}] = P\left[\frac{(n-1)S^2}{\sigma^2} > \frac{2.00\times10^{-4}(n-1)}{\sigma^2}\right]$$

چون متغیر $\frac{(n-1)S^2}{\sigma^2}$ دارای کی‌دو است با ۱۹ درجه آزادی، لذا برای محاسبه احتمال از جدول ضمیمه (۵) استفاده می‌کنیم.

$$P\left[\frac{(n-1)S^2}{\sigma^2} > \frac{2.00\times10^{-4}\times19}{1.26\times10^{-4}}\right] = P[X > 30.2] \approx 0.05$$

مثال ۵-۷-۶ احتمال اینکه یک نمونه ۹ تایی از توزیع نرمال با میانگین $\mu=۰/۲۵$ و واریانس $\sigma^۲=۰/۴$ دارای میانگین بزرگتر از ۰/۲۸ و واریانس کمتر از ۰/۱۳۷ باشد چقدر است؟

$$P[\overline{X} > 0.28\ ,\quad S^2 < 0.137] = P[\overline{X} > 0.28]P[S^2 < 0.137]$$

طبق قضیه ۵-۷-۴، $\overline{X}$ و $S^۲$ از هم مستقل‌اند. پس:

$$P(\overline{X} > 0.28)P(S^2 < 0.137)$$

$$= P\left[\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{0.28-0.25}{0.21}\right] . P\left[\frac{(n-1)S^2}{\sigma^2} < \frac{0.137\,(8)}{0.4}\right]$$

$$= P(z > 1.43).P(z < 2.74) = (0.0764)(0.05) = 0.00382$$

مثال ۵-۷-۷ اگر $X_1, X_2, \ldots, X_n$ یک نمونه nتایی از جامعه نرمال با میانگین μ و واریانس σ^2 باشد امید ریاضی و واریانس S^2 را بدست آورید.

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \Rightarrow E(S^2) = \sigma^2$$

$$V\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \Rightarrow V(S^2) = \frac{2\sigma^4}{n-1}$$

۵-۸ توزیع t

فرض کنید که $X_1, X_2, \ldots, X_n$ یک نمونه nتایی از توزیع نرمال با میانگین μ و واریانس σ^2 باشد. می‌دانیم متغیر $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد و متغیر $\dfrac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است. متغیر T را که تابعی از دو متغیر است به صورت زیر تعریف می‌کنیم.

$$T = \frac{\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\dfrac{S}{\sigma}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

متغیر T نسبت دو متغیر است و به راحتی می‌توان ثابت کرد که T دارای توزیع استودنت است با $n-1$ درجه آزادی. متغیر T در مقایسه با متغیر $Z = \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای این حسن است که به واریانس جامعه یعنی σ^2 بستگی ندارد. لذا در مواقعی که واریانس جامعه مجهول و حجم نمونه کوچک است استفاده از متغیر T توصیه می‌شود به شرط آنکه نمونه گرفته شده از جامعه نرمال باشد. در حالت کلی با استفاده از قضیه حد مرکزی حتی وقتی که جامعه مورد بررسی نرمال نباشد برای حجم نمونه بزرگ نیز قابل استفاده است.

تبصره ۵-۸-۱ در توزیع استودنت اگر حجم نمونه یا درجه آزادی از حد تصور بزرگتر باشد $(n \to \infty)$ توزیع استودنت به توزیع نرمال استاندارد میل می‌کند.

مثال ۵-۸-۲ مدیر یک کارخانه تولید کننده لامپ ادعا می‌دارد که لامپ‌های او به طور متوسط بعد از ۵۰۰= μ ساعت کار خواهد سوخت. براساس یک نمونه ۲۵ تایی با انحراف معیار ۴۰ ساعت احتمال اینکه میانگین نمونه بیشتر از ۵۱۶/۵۱۲ باشد چقدر است؟

$$P[\overline{X} > 516.512] = P\left[\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{516.512-\mu}{\frac{S}{\sqrt{n}}}\right]$$

$$= P\left[T > \frac{516.512-500}{\frac{40}{\sqrt{25}}}\right] = P[T > 2.064]$$

باتوجه به جدول ضمیمه (۴) متغیر t دارای درجه آزادی ۲۴=۱-n،
$P[T>$۲/۰۶۴$]=$۰/۰۲۵

مثال ۵-۸-۳ زمان های لازم برای انجام یک کار توسط ۹ کارگر داده شده است مطلوبست: $P[\overline{X}-\mu>$۲/۴۰$]$

۲۹	۲۶	۳۸	۳۴	۳۲	۳۵	۴۲	۳۳	۲۷

$$P[\overline{X}-\mu > 2.40] = P\left[\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{2.40}{\frac{S}{\sqrt{n}}}\right]$$

که برای نمونه بالا، ۹=n ، ۵/۱۵۸=S و درجه آزادی برابر با ۸ است.

$$P\left[T > \frac{2.40}{\frac{5.158}{3}}\right] = P[T > 1.396] = 0.10$$

مثال ۵-۸-۴ برای متغیر $T = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$ امید ریاضی را حساب کنید.

T را می توان به صورت نسبت دو متغیر مستقل $\dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ و $\sqrt{\dfrac{(n-1)S^2}{(n-1)\sigma^2}}$ نوشت.

$$T = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} . \frac{1}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}}$$

$$E(T) = E\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) .\ E\left(\frac{1}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}}\right) = 0 \times E\left(\frac{1}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}}\right) = 0$$

۵-۹ توزیع نسبت واریانس دو نمونه

فرض کنید از جامعه اول که دارای توزیع نرمال با واریانس σ_1^2 است نمونه $X_1, X_2, \ldots, X_{n_1}$ به حجم n_1 و از جامعه دوم که دارای توزیع نرمال با واریانس σ_2^2 است نمونه $Y_1, Y_2, \ldots, Y_{n_2}$ به حجم n_2 موجود باشند. اگر $S_1^2 = \dfrac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2$ و $S_2^2 = \dfrac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$ به ترتیب واریانس های نمونه اول و دوم باشند متغیرهای $\dfrac{(n_1 - 1)S_1^2}{\sigma_1^2}$ و $\dfrac{(n_2 - 1)S_2^2}{\sigma_2^2}$ طبق قضیه ۵-۷-۴ دارای توزیع کی‌دو با درجه آزادی $n_1 - 1$ و $n_2 - 2$ هستند. برای بدست آوردن توزیع نسبت واریانس نمونه‌ای دو جامعه، قضیه زیر را بدون اثبات می‌پذیریم.

قضیه ۵-۹-۱ اگر متغیرهای تصادفی مستقل X و Y به ترتیب دارای توزیع کی‌دو با m و n درجه آزادی باشند آنگاه متغیر $F=\dfrac{X/m}{Y/n}$ دارای توزیع فیشر با m و n درجه آزادی است. معمولاً m را درجه آزادی صورت و n را درجه آزادی مخرج می‌گویند و آن را با علامت $F(m,n)$ نمایش می‌دهند.

برای بدست آوردن توزیع نسبت واریانس دو نمونه، متغیر F را به صورت زیر تعریف می‌کنیم.

$$F=\frac{\dfrac{S_1^2}{\sigma_1^2}}{\dfrac{S_2^2}{\sigma_2^2}}=\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2}$$

متغیر F شرایط قضیه ۵-۹-۱ را دارا می‌باشد و دارای توزیع فیشر با n_1-1 و n_2-1 درجه آزادی می‌باشد.

تبصره ۵-۹-۲ اگر متغیر تصادفی F دارای توزیع فیشر با m و n درجه آزادی باشد آنگاه متغیر $\frac{1}{F}$ دارای توزیع فیشر با n و m درجه آزادی است. یعنی:

$$F_{(\alpha,m,n)}=\frac{1}{F_{(1-\alpha,n,m)}}$$

این نتیجه به ما اجازه می‌دهد توزیع F را تنها برای دنباله سمت راست در جدول ضمیمه (۶) درج کنیم.

مثال ۵-۹-۳ با توجه به جدول ضمیمه (۶) مقدار C را برای احتمالات زیر محاسبه کنید.

$$P[F_{(9,10)}>C]=0.05\Rightarrow C=3.02$$

$$P[F_{(15,15)}>C]=0.01\Rightarrow C=3.52$$

$$P[F_{(8,10)}<C]=0.95\Rightarrow P[F_{(8,10)}>C]=0.05\Rightarrow C=3.07$$

مثال ۵-۹-۴ یک نمونه ۱۱ تایی از جامعه اول با واریانس $\sigma_1^2=10$ و یک نمونه ۱۶تایی از جامعه دوم با واریانس $\sigma_2^2=2$ انتخاب می‌شود احتمال $P\left[\frac{S_1^2}{S_2^2}<12.7\right]$ را حساب کنید (اگر $\alpha=0/05$ باشد).

$$P\left[\frac{S_1^2}{S_2^2}<12.7\right]=P\left[\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2}<12.7\left(\frac{\sigma_2}{\sigma_1}\right)^2\right]$$

$$=P\left[F_{(10,15)}<12.7\left(\frac{2}{10}\right)\right]=P\left[F_{(10,15)}<2.54\right]=0.05$$

خودآزمایی

۱- اگر مشاهدات نمونه تصادفی X_1, X_2, X_3, X_4, X_5 برابر با ۰/۲، ۰/۷، ۱/۰-، ۰، ۵ باشند مقدار آماره‌های زیر را بدست آورید.

$$\overline{X}=\frac{1}{5}\sum_{i=1}^{5}X_i \ , \quad S^2=\frac{1}{4}\sum_{i=1}^{5}(X_i-\overline{X})^2 \ , \quad M=\sum_{i=1}^{5}\frac{\left|X_i-\overline{X}\right|}{4}$$

۲- اگر X_1, X_2, X_3 یک نمونه سه تایی از جامعه‌ای با میانگین μ و واریانس σ^2 باشد کدام یک از آماره‌های زیر دارای کمترین واریانس است؟

$$T_1=\frac{2X_1+3X_2+2X_3}{6} \quad , \quad T_2=\frac{X_1+X_2+X_3}{3}$$

$$T_3=\frac{X_3+X_1}{2} \quad , \quad T_4=\frac{2X_1+X_2}{3}+\frac{1}{5}$$

۳- اگر $\overline{X}$ میانگین یک نمونه m تایی از جامعه‌ای با میانگین μ_1 و واریانس σ_1^2 و $\overline{Y}$ میانگین یک نمونه n تایی از جامعه‌ای دیگر با میانگین μ_2 و واریانس σ_2^2 باشند مطلوبست امید ریاضی و واریانس آماره‌های زیر.

$$T_1=\frac{m\overline{X}+n\overline{Y}}{m+n} \quad , \quad T_2=\frac{m\overline{X}}{m^2+1}+\frac{n\overline{Y}}{n^2+1}$$

۴- اگر $\overline{X}$ و S^2 به ترتیب میانگین و واریانس یک نمونه n تایی از توزیع نرمال با میانگین μ و واریانس σ^2 باشند امید ریاضی و واریانس آماره‌های زیر را بدست آورید (k و c ثابت هستند).

$$T_1=k\overline{X}+c \quad , \quad T_2=kS^2+c$$

$$T_3=\frac{\overline{X}}{k+c}-10 \quad , \quad T_4=\frac{S^2}{k+c}+10$$

۵- اگر X دارای توزیع پواسن با پارامتر λ و Y دارای توزیع دو جمله‌ای با پارامترهای n و p باشند توزیع مشترک X و Y را بدست آورید.

۶- اگر $X_1, X_2, \dots, X_n$ یک نمونه n تایی ا ز تابع چگالی زیر باشد توزیع مشترک $X_1, X_2, \dots, X_n$ را بدست آورید.

$$f(x) = \frac{x+1}{\theta(\theta+1)}\, e^{-x/\theta} \qquad x > 0$$

۷- جامعه شناسی ادعا می‌کند که قد افراد شهری دارای توزیع نرمال با میانگین ۱۷۰cm و انحراف معیار ۴ است. در یک نمونه ۲۵ تایی احتمال اینکه متوسط قد افراد نمونه بین ۱۷۵ تا ۱۶۵ باشد چقدر است؟

۸- مدیر کارخانه‌ای سازنده رایانه ادعا می‌کند که ۰/۲ درصد تولیدات رایانه کارخانه معیوب هستند. در یک نمونه ۱۰۰ تایی مطلوبست $P(\overline{X} > ۰/۳)$.

۹- اگر متغیر تصادفی X دارای تابع احتمال زیر باشد برای یک نمونه ۲۵ تایی احتمالات زیر را حساب کنید.

x	۱	۲	۳	۴
$f(x)$	۰/۲۵	۰/۲۵	۰/۲۵	۰/۲۵

$P[\overline{X} > ۳] \quad , \quad P[۱ < \overline{X} < ۲/۵] \quad , \quad P[\overline{X} > ۰/۷]$

۱۰- اگر یک نمونه ۷ تایی از جامعه‌ای نرمال با میانگین ۵ دارای انحراف معیار $S = ۱/۰۸$ باشد مطلوبست $P(\overline{X} > ۶)$. .

۱۱- تعداد دقایقی که پزشکی صرف معاینه هر بیمار می‌کند دارای توزیع نمایی با $\theta = ۹$ است. در یک نمونه ۲۵ تایی احتمال اینکه به طور متوسط بیش از ۱۵ دقیقه صرف معاینه بیماری کند چقدر است؟

۱۲- تعداد ماهی‌هایی که یک ماهیگیر در هر ساعت از دریاچه‌ای صید می‌کند دارای توزیع پواسن با $\lambda = ۱/۵$ است. احتمال اینکه ماهیگیر در ۴۸ ساعت کاری به طور متوسط بین ۱ تا ۳ ماهی صید کند چقدر است؟

۱۳- سوابق طبی نشان می‌دهند که در شهری یک نفر از هر ده نفر دچار بیماری تیروئید است. اگر بیست نفر را در این شهر به تصادف انتخاب کنیم و مورد آزمایش قرار دهیم احتمال اینکه حداقل یکی از آنها دچار تیروئید باشد چقدر است؟

الف- با استفاده از جدول توزیع دو جمله‌ای

ب- با استفاده از جدول توزیع پواسن

ج- با استفاده از تقریب نرمال

۱۴- اگر $X_۱, X_۲, \ldots, X_{۲۵}$ دارای توزیع نرمال با میانگین ۱۵۰۰ و واریانس ۱۰۰ باشند مطلوبست:

$P[\overline{X} < ۱۵۰, S^۲ > ۱۶۴]$

۱۵- یک فرایند تولید یاتاقانهای معین تحت کنترل است. اگر قطر یاتاقانها با میانگین ۰/۵ سانتیمتر باشد و یک نمونه ۱۰ تایی از این یاتاقانها دارای میانگین ۰/۵۰۶ سانتیمتر و انحراف معیار ۰/۰۰۴ باشند مقدار $T = \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}}$ را حساب کنید.

۱۶- مقادیر زیر را پیدا کنید.

الف- $F_{۰/۰۵}$ با درجه آزادی ۷ و ۳

ب- $F_{۰/۰۵}$ با درجه آزادی ۵ و ∞

ج- $F_{۰/۰۱}$ با درجه آزادی ۳۰ و ۱۲۰

۱۷- یک تست کنکور استاندارد در زمینه ریاضی به ۲۵ پسر و ۱۶ دختر داده شده است. میانگین نمره پسرها ۸۲ و انحراف ۸ و حال آنکه میانگین نمره دخترها ۷۸ و انحراف معیار ۷ شده است. اگر نسبت واریانس نمرات پسرها در جامعه نسبت به دخترها $\frac{1}{2}$ باشد مقدار آماره F را بدست آورید.

فصل ۶

برآورد نقطه‌ای و فاصله‌ای پارامتر

مقدمه

در فصل ۴ و ۵ با قالبهای احتمال که بعضاً شامل پارامتر بودند آشنا شدیم. در این فصل انواع برآورد پارامتر و روشهای برآورد را که شاخه‌ای از استنباط آماری است مورد بحث قرار می‌دهیم. پارامتر جامعه که یک مشخصه عددی است به طریق نقطه‌ای یا فاصله‌ای برآورد می‌شود. مجموعه تمام مقادیر ممکن پارامتر θ را فضای پارامتر می‌گویند و آن را به Θ نمایش می‌دهند.

$$\Theta = \{\theta \mid \theta \in B\}$$

فرض کنید $x_۱, x_۲, ..., x_n$ مشاهدات متناظر با نمونه تصادفی $X_۱, X_۲, ..., X_n$ از تابع چگالی احتمال $f_\theta(x)$ باشند. پارامتر θ یا تابعی از پارامتر θ یعنی $\tau(\theta)$ بر اساس مقادیر مشاهده نمونه به روش برآورد نقطه‌ای و فاصله‌ای برآورد می‌شود.

ارائه یک عدد معینی براساس مشاهدات به منظور تخمین پارامتر را برآورد نقطه‌ای و ارائه یک فاصله بین دو مقدار عددی را برآورد فاصله‌ای گویند. از بین روشهای مختلف برآورد پارامتر، دو روش، روش گشتاورها و روش تابع درستنمایی را با چند مثال ارائه می‌دهیم. در پایان، برآورد فاصله‌ای برای میانگین، نسبت، واریانس و نسبت واریانس جامعه ارائه می‌شود.

۶-۱ برآورد نقطه‌ای

فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال $f_\theta(x)$ باشد بـه طـوری کـه $\theta \in \Theta$. اگر $x_1, x_2, \ldots, x_n$ مشاهدات نمونه باشند برآورد نقطـه‌ای بـرای θ ، مقـدار عددی برآوردگر بر اساس نمونه مشاهده شده خواهد بود. اگـر فـرض کنـیم $\bar{x}$ مقـدار عددی برآوردگر $\bar{X}$ برای θ باشد احتمال اینکه $\bar{x}$ دقیقاً برابر با θ باشد صفر اسـت. به همین خاطر از $\bar{x}$ به عنوان یک برآورد برای θ یاد می‌کنند.

برای اینکه فرقی بین θ و برآورد آن قایل شویم، برآورد θ را به $\hat{\theta}$ نمایش می‌دهـیم. اختلاف $|\hat{\theta}-\theta|$ را خطای برآورد نقطه‌ای گویند. در ادامـه ایـن بخـش روش بـرآورد گشتاورها و روش درستنمایی ماکزیمم ارائه می‌شود.

۶-۲ روش گشتاورها

برآورد پارامتر به روش گشتاورها یکی از روشـهای قـدیمی اسـت کـه در سـال ۱۸۹۴ توسط کارل پیرسون پیشنهاد شد و اکنون هم در بـرآورد پـارامتر بیـشتر توزیعهـا قابـل استفاده است.

فرض کنید متغیر تـصادفی X دارای تـابع چگـالی احتمـال $f_\theta(x)$ و $x_1, x_2, \ldots, x_n$ مشاهدات نمونه تصادفی $X_1, X_2, \ldots, X_n$ باشند. برآورد پارامتر θ به روش گشتاورها از برابری گشتاورهای نمونه و جامعه به صورت زیر بدست می‌آید.

گشتاور مرتبه rام جامعه: $\mu_r = E(X^r)$

گشتاور غیرمرکزی مرتبه rام نمونه حول نقطه صفر: $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$

می‌دانیم برآورد θ با مقدار واقعی مساوی نیست لـذا گـشتاور مرتبـه rام نمونـه یـک برآوردی برای گشتاور مرتبه rام جامعه است. رابطه زیر همواره بین آنها برقرار است.

$\hat{\mu}_r = m'_r$

در هنگام برآورد با توجه به بعد پارامتر، می توان r را ۱، ۲، ... اختیار کرد. ذکر این نکته ضروری است که $\hat{\mu}_r$ باید عضوی از Θ باشد در غیر این صورت برآورد پذیرفته نمی‌شود.

مثال ۶-۲-۱ فرض کنید تعداد غلط‌هایی که در هر صفحه تایپ یافت می‌شود دارای توزیع پواسن با پارامتر مجهول λ باشد. اگر تعداد غلط‌ها در ۸ صفحه مشاهده شده به صورت زیر باشد λ را به روش گشتاورها برآورد کنید.

۱ ۳ ۰ ۱ ۲ ۰ ۱ ۲

$$f_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0,1,\ldots, \quad \Theta = \{\lambda \mid \lambda > 0\} = (0,\infty)$$

$$\mu_1 = E(X) = \lambda$$

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i$$

از برابری دو رابطه اخیر

$$\hat{\mu}_1 = \hat{\lambda} = m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\lambda} = \bar{x} = \frac{1}{8}\sum_{i=1}^{n} x_i = \frac{1}{8}(10) = 1.25$$

عدد ۱/۲۵ ، برآورد نقطه‌ای برای λ است. اگر نمونه عوض شود مقدار برآورد نیز عوض می‌شود ولی اگر حجم نمونه زیاد باشد و نحوه نمونه‌گیری تصادفی باشد برآورد در تکرار زیاد نمونه به مقدار واقعی پارامتر نزدیک می‌شود و خطای بین λ و $\hat{\lambda}$ به مینیمم می‌رسد.

مثال ۶-۲-۲ براساس یک نمونه nتایی میانگین جامعه‌ای با توزیع نرمال با میانگین μ و واریانس ۱ را به روش گشتاورها برآورد کنید.

$$f_\mu(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2} \qquad -\infty<x<\infty$$

$$-\infty<\mu<\infty$$

$$\left.\begin{array}{l}\mu_1=E(X)=\mu\\ m'_1=\bar{x}\end{array}\right\}\Rightarrow \hat{\mu}=\bar{x}$$

در این مثال اگر واریانس جامعه هم مجهول باشد برای برآورد μ و σ^2 لازم است از گشتاورهای مرتبه اول و دوم استفاده کنیم:

$$\begin{array}{ll}\mu_1=E(X)=\mu & m'_1=\bar{x}\\ \mu_2=E(X^2) & m'_2=\frac{1}{n}\sum_{i=1}^{n}x_i^2\end{array}$$

از برابری m'_1 با μ_1، میانگین جامعه μ برآورد می‌شود.
برای برآورد واریانس از رابطه زیر استفاده می‌کنیم.

$$\sigma^2=V(X)=E(X^2)-(E(X))^2=\mu_2-\mu_1^2$$

$$\hat{\sigma}^2=\hat{\mu}_2-\hat{\mu}_1^2=m'_2-m_1'^2$$

$$\hat{\sigma}^2=\frac{1}{n}\sum x_i^2-\bar{x}^2=\frac{n-1}{n}S^2$$

مثال ۶-۲-۳ اگر مقدار بارندگی در ایستگاه معینی در روز بخصوصی در فصل بارندگی دارای توزیع یکنواخت روی بازه $(0,\theta)$ توزیع شده باشد و مقدار بارندگی برای ۱۰ سال گذشته برحسب میلی متر به صورت زیر ثبت شده باشد θ را برآورد کنید.

۰ ۱۷/۵ ۲۵ ۰ ۲/۵ ۰ ۵ ۱۲/۵ ۰ ۱۵

$$f_\theta(x)=\frac{1}{\theta} \qquad 0<x<\theta \quad , \quad \theta>0$$

$$\mu_1=\mu=E(x)=\int_0^\theta \frac{x}{\theta}\,dx=\frac{\theta}{2}$$

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\frac{\hat{\theta}}{2} = \bar{x} \qquad , \qquad \hat{\theta} = 2\bar{x} = 2\left(\frac{1}{n}\sum x_i\right) = 2(7.75) = 15.5$$

۶-۳ روش درستنمایی ماکزیمم

روش درستنمایی ماکزیمم در سال ۱۹۱۲ توسط فیشر ارائه شد و در مواردی که برآورد به روش گشتاورها و روش درستنمایی ماکزیمم یکسان نیستند، برآورد به روش درستنمایی ماکزیمم را به روش گشتاورها ترجیح می‌دهند.

اگر $x_1، x_2،...، x_n$ مشاهدات متناظر نمونه تصادفی $X_1، X_2،...، X_n$ از تابع چگالی احتمال $f_\theta(x)$، $\theta \in \Theta$ باشند تابع چگالی توام x_iها برابر است با:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

در رابطه اخیر تنها متغیر θ است و می توان رابطه را فقط تابعی از θ نوشت.

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) \quad , \qquad \theta \in \Theta$$

در ادبیات آماری $L(\theta)$ را تابع درستنمایی گویند. هدف روش درستنمایی ماکزیمم این است که θ را طوری پیدا کند که $L(\theta)$ ماکزیمم شود. از آنجا که $L(\theta)$ و $\ln L(\theta)$ در مقدار یکسانی از θ ماکزیمم می‌شود برای راحتی θ را طوری پیدا می‌کنیم که $\ln L(\theta)$ را ماکزیمم کند.

مثال ۶-۳-۱ فرض کنید X دارای توزیع برنولی با پارامتر θ و $x_1، x_2،...، x_n$ مقادیر مشاهدات باشند. θ را به روش درستنمایی ماکزیمم برآورد کنید.

$$f_\theta(x) = \theta^x (1-\theta)^{1-x} \qquad x = 0,1 \qquad , \quad \Theta = (0,1)$$

تابع درستنمایی نمونه

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$L(\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$\ln L(\theta) = \sum x_i \ln\theta + (n - \sum x_i)\ln(1-\theta)$$

برای ماکزیمم کردن رابطه اخیر، از آن نسبت به θ مشتق می‌گیریم و برابر با صفر قرار می‌دهیم.

$$\frac{d\ln L(\theta)}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n-\sum x_i}{1-\theta} = 0$$

$$(1-\theta)\sum x_i = \theta(n - \sum x_i) \quad , \quad n\theta = \sum x_i \Rightarrow \hat{\theta} = \bar{x}$$

$\hat{\theta} = \bar{x}$ مقداری است که $\ln L(\hat{\theta})$ یا $L(\hat{\theta})$ را ماکزیمم می‌کند و حتماً $\frac{d^2 \ln L(\hat{\theta})}{d^2\hat{\theta}^2} < 0$.

مثال ۶-۳-۲ تعداد ساعاتی که یک لامپ الکترونی کار کند یک متغیر تصادفی نمایی با پارامتر θ است. اگر یک نمونه nتایی از طول عمر این لامپها در دست باشد θ را به روش درستنمایی ماکزیمم برآورد کنید.

$$f_\theta(x) = \frac{1}{\theta}e^{-x/\theta} \qquad x > 0 \quad , \quad \Theta = (0,\infty)$$

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n}\frac{1}{\theta}e^{-x_i/\theta} = \frac{1}{\theta^n}e^{-\sum x_i/\theta}$$

$$\ln L(\theta) = -n\ln\theta - \frac{\sum x_i}{\theta}$$

$$\frac{d\ln L(\theta)}{d\theta} = \frac{-n}{\theta} + \frac{\sum x_i}{\theta^2} = 0 \Rightarrow \hat{\theta} = \bar{x}$$

میانگین نمونه یک برآورد نقطه‌ای برای پارامتر مجهول θ است.

مثال ۶-۳-۳ اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد μ و σ^2 را براساس مشاهدات یک نمونه nتایی برآورد کنید.

$$f(x,\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad , \qquad -\infty<x<\infty$$

$$\Theta=\{(\mu,\sigma^2)\mid -\infty<\mu<\infty \;\; , \;\; \sigma^2>0\}=(-\infty,\infty)(0,\infty)$$

چون فضای پارامتر دو بعدی است تابع درستنمایی دو متغیره خواهد بود.

$$L(\mu,\sigma^2)=\prod_{i=1}^{n} f\,(x_i,\mu,\sigma^2)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^2 e^{-\frac{1}{2\sigma^2}\;\sum(x_i-\mu)^2}$$

$$\ln\,L(\mu,\sigma^2)=-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

برای برآورد μ و $\sigma^۲$ از تابع اخیر یک بار نسبت به μ و یک بار نسبت به $\sigma^۲$ مشتق می‌گیریم و برابر صفر قرار می‌دهیم.

$$\frac{d\ln\,L(\mu,\sigma^2)}{d\mu}=\frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0$$

$$\frac{d\ln\,L(\mu,\sigma^2)}{d\sigma^2}=-\frac{n}{2\sigma^2}+\frac{\sum(x_i-\mu)^2}{2\sigma^4}=0$$

از رابطه اول μ برآورد می شود و برای برآورد $\sigma^۲$ در رابطه دوم مورد استفاده قرار می‌گیرد.

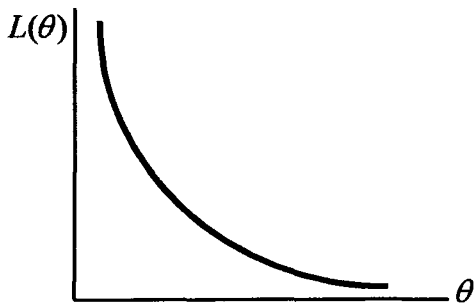
$$\hat{\mu}=\bar{x} \qquad , \qquad \hat{\sigma}^2=\frac{1}{n}\sum(x_i-\bar{x})^2$$

در هریک از مثالهای اخیر از مساوی قرار دادن مشتق $\ln L(\theta)$ با صفر، برآورد یا برآوردها بدست می‌آیند. در مثالهای اخیر تابع درستنمایی دارای ماکزیمم منحصر به فرد بود و مشتق در آن نقطه برابر صفر بود. ولی ممکن است تابع درستنمایی دارای مشتق در حوزه تعریف خود نباشد. در این حالت برای پیدا کردن برآورد، روش گشتاورها پیشنهاد می‌شود.

مثال ۶-۳-۴ فرض کنید X دارای تابع چگالی احتمال یکنواخت روی بازه $(0,\theta)$ باشد θ را براساس یک نمونه nتایی برآورد کنید.

$$f_\theta(x)=\frac{1}{\theta} \qquad , \quad 0<x<\theta \quad , \quad \Theta=(0,\infty)$$

$$L(\theta)=\prod_{i=1}^{n} f_\theta(x_i)=\frac{1}{\theta^n}$$



باتوجه به شیب $L(\theta)$ ، $L(\theta)$ در هیچ جا صفر نمی شود لذا احتیاجی به مشتق گرفتن از $L(\theta)$ و مساوی قراردادن آن با صفر نیست. θ به روش گشتاورها در مثال ۶-۲-۳ برآورد شده است.

۶-۴ برآورد فاصله‌ای

در بخش برآورد نقطه‌ای اشاره به این حقیقت شد که برآورد نقطه‌ای نمی‌تواند برابر با مقدار واقعی پارامتر باشد. به عنوان مثال فرض کنید مدیر کارخانه‌ای ادعا می‌کند که لامپهای تولیدی این کارخانه دارای عمر متوسط بین ۱۰ ± ۱۵۰۰ ساعت است. اگر براساس یک نمونه nتایی برآورد نقطه‌ای برای ادعای مدیر داشته باشیم مسلماً برآورد ما یک نقطه از بازه (۱۰+۱۵۰۰ و ۱۰−۱۵۰۰) خواهد بود که به خودی خود متضمن اطلاعاتی درباره میزان احتمال اینکه برآوردگر مقداری نزدیک به مقدار واقعی مجهول قبول کند، نیست. در صورتی که برآورد فاصله‌ای، هم تصویری از مقدار عددی واقعی پارامتر را می‌دهد وهم براساس نمونه، اشاره به این نکته می‌کند که تا چه حدی می‌توان به درستی حدسی که در باره مقدار عددی پارامتر زده شده است مطمئن بود.

تعریف ۶-۴-۱ اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $f_\theta(x)$ باشد و $T_1(X_1,\ldots,X_n)$ و $T_2(X_1,\ldots,X_n)$ دو آماره باشند به طوری که $T_1<T_2$ بازه (T_1,T_2) را یک بازه اطمینان با ضریب اطمینان $100(1-\alpha)$٪ برای θ گوییم. اگر احتمال اینکه دو آماره T_1، T_2، θ در برداشته باشند مستقل از θ باشد.

$$P[T_1(X_1,\ldots,X_n)<\theta<T_2(X_1,\ldots,X_n)]=1-\alpha$$

T_1 و T_2 را به ترتیب حدود اطمینان پایینی و بالایی θ و T_2-T_1 را طول بازه اطمینان گویند.

برای بدست آوردن بازه اطمینان برای θ یا تابعی از θ، $\tau(\theta)$ لازم است کمیت محوری تعریف کنیم.

تعریف ۶-۴-۲ هر تابعی از نمونه و پارامتر را که توزیع آن مستقل از پارامتر باشد کمیت محوری گویند. به عنوان مثال $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ برای نمونه nتایی از جامعه نرمال با میانگین μ و واریانس معلوم σ^2 یک کمیت محوری است. چون $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است. از آنجا که برای یافتن فاصله اطمینان نیاز به داشتن آماره است، دانستن توزیع جامعه‌ای که نمونه ازآن استخراج شده است ضروری است. فرض کنید متغیر تصادفی Z دارای توزیع نرمال استاندارد باشد. احتمال اینکه دو عدد ۱/۹۶ و ۱/۹۶- متغیر تصادفی Z را در برداشته باشند برابر با ۰/۹۵ است.

$$P[-1/96<Z<1/96]=0/95$$

عکس آن نیز درست است. اگر دو عدد a و b متغیر تصادفی Z را با احتمال ۰/۹۵ در برداشته باشند a و b به ترتیب برابرند با ۱/۹۶- و ۱/۹۶ . در حالت کلی برای متغیر تصادفی Z رابطه زیر همواره برقرار است.

$$P\left[-z_{\frac{\alpha}{2}}<Z<z_{\frac{\alpha}{2}}\right]=(1-\alpha)$$

که برای $\alpha = ۰/۰۵$ رابطه اخیر برابر است با:

$$P\,[-۱/۹۶ < Z < ۱/۹۶] = ۰/۹۵$$

۶–۵ فاصله اطمینان برای میانگین توزیع نرمال

حالت اول: اگرمتغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس معلوم $\sigma^۲$ باشد براساس یک نمونه nتایی یک فاصله اطمینان، با ضریب اطمینان $۱۰۰(۱-\alpha)٪$ برای μ به صورت زیر بدست می‌آید.

در این حالت می‌دانیم که متغیر $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ یک کمیت محوری و دارای توزیع نرمال استاندارد است. لذا:

$$P=\left[-z_{\frac{\alpha}{2}}<\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<z_{\frac{\alpha}{2}}\right]=1-\alpha$$

مجهول μ

یعنی احتمال رخداد پیشامد $-z_{\frac{\alpha}{2}}<\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<z_{\frac{\alpha}{2}}$ برابر با $(۱-\alpha)$ است. اگر پیشامد فوق را نسبت به μ حل کنیم داریم:

$$\overline{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

برای یک نمونه nتایی بازه اطمینان برای μ برابر است با:

$$\overline{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

برای $\alpha = ۰/۰۵$

$$\overline{X}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+1.96\frac{\sigma}{\sqrt{n}}$$

مثال ۶-۵-۱ اگر عمر رایانه‌ها دارای توزیع نرمال با میانگین μ و واریانس ۰/۲۵ سال باشد براساس مشاهدات زیر یک فاصله اطمینان ۰/۹۵ برای μ بدست آورید.

۱۰   ۷   ۵   ۶   ۸   ۵   ۹   ۶   ۷

$n=۹$, $\alpha=۰/۰۵$, $z_{\frac{\alpha}{2}}=z_{۰/۰۲۵}=۱/۹۶$, $\sigma=۰/۵$

$$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{9}\sum_{i=1}^{9}x_i=\frac{1}{9}(63)=7$$

$$7-1.96\frac{0.5}{\sqrt{9}}<\mu<7+1.96\frac{0.5}{\sqrt{9}}$$
$$7-0.33<\mu<7+0.33$$
$$6.67<\mu<7.33$$

حالت دوم:اگرمتغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس مجهول σ^2 باشد براساس یک نمونه n تایی یک فاصله اطمینان، با ضریب اطمینان $(۱-\alpha)۱۰۰$٪ برای μ به صورت زیر بدست می‌آید:

در این حالت می‌دانیم که متغیر $\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}$ دارای توزیع استونت با n درجه آزادی و یک کمیت محوری است لذا:

مجهول σ

$$P\left[-t_{(\frac{\alpha}{2},n-1)}<\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}<t_{(\frac{\alpha}{2},n-1)}\right]=1-\alpha$$

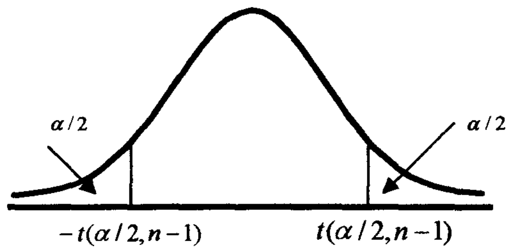
از رابطه اخیر داریم:

$$\overline{X}-t_{(\frac{\alpha}{2},n-1)}\frac{s}{\sqrt{n}}<\mu<\overline{X}+t_{(\frac{\alpha}{2},n-1)}\frac{s}{\sqrt{n}}$$

برای یک نمونه nتایی بازه اطمینان برای μ برابر است با:

$$\bar{x}-t_{(\frac{\alpha}{2},n-1)}\cdot\frac{s}{\sqrt{n}}<\mu<\bar{x}+t_{(\frac{\alpha}{2},n-1)}\cdot\frac{s}{\sqrt{n}}$$

مقدار $t_{(\frac{\alpha}{2},n-1)}$ از جدول استودنت برای α و n درجه آزادی بدست می‌آید.



مثال ۶-۵-۲ میانگین کاهش وزن توپها به دلیل ساییدگی در یک زمان معین در محلولی برای n=۱۶ توپ برابر با ۳/۴۲ گرم با انحراف معیار ۰/۶۸ گرم می‌باشد. برای میانگین واقعی کاهش وزن چنین توپهایی تحت شرایط تعیین شده، یک فاصله اطمینان ۰/۹۹ برای μ بدست آورید.

n=۱۶ , $\bar{x}$=۳/۴۲ , s = ۰/۶۸ , ۱−α = ۰/۹۹ , α = ۰/۰۱

$\frac{\alpha}{2}$ = ۰/۰۰۵ , t(۰/۰۰۵,۱۵)=۲/۹۴۷

$$\bar{x}-t_{(\frac{\alpha}{2},n-1)}\frac{s}{\sqrt{n}}<\mu<\bar{x}+t_{(\frac{\alpha}{2},n-1)}\frac{s}{\sqrt{n}}$$

$$3.42-2.947\frac{0.68}{\sqrt{16}}<\mu<3.42+2.947\frac{0.68}{\sqrt{16}}$$

$$2.92<\mu<3.92$$

۶-۶ فاصله اطمینان برای P در توزیع دو جمله‌ای

اگر متغیر تصادفی X دارای توزیع دو جمله‌ای با پارامتر مجهول P و n معلوم باشد $E(X)=nP$ و $V(X)=nP(1-P)$ است.

از رابطه $E(X)=nP$ داریم:

$$E\left(\frac{X}{n}\right)=P \quad , \qquad V\left(\frac{X}{n}\right)=\frac{P(1-P)}{n}$$

متغیر تصادفی X مجموع چند متغیر برنولی است، $\frac{X}{n}$ یک برآوردگر برای P جامعه است لذا:

$$\hat{P}=\frac{x}{n} \quad , \qquad E(\hat{P})=P \qquad , \qquad V(\hat{P})=\frac{P(1-P)}{n}$$

باتوجه به روابط اخیر اگر حجم نمونه از حدی بزرگتر باشد کمیت محوری زیـر دارای توزیع تقریبی نرمال استاندارد خواهد بود.

$$z=\frac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}}$$

$$P\left[-z_{\frac{\alpha}{2}}<\frac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}}<z_{\frac{\alpha}{2}}\right]\approx 1-\alpha$$

از رابطه اخیر:

$$\frac{X}{n}-z_{\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}<P<\frac{X}{n}+z_{\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}$$

اگر x مقدار مشاهده X باشد فاصله اطمینان اخیر قابل ارزیابی نیست چون واریانس $\frac{X}{n}$ شامل پارامتر P است. چون $\hat{P}=\frac{x}{n}$ یک برآورد نقطه‌ای برای P است. بـرآورد نقطه‌ای واریانس $\frac{X}{P}$ برابر است با:

$$\frac{\hat{P}(1-\hat{P})}{n}=\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}$$

فاصله اطمینان تقریبی برای P برابر است با:

$$\left(\frac{x}{n}-z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}<P<\frac{x}{n}+z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}\right)$$

مثال ۶-۶-۱ دریک نظرخواهی از ۲۰۰ نفر در مورد احداث پارک علوم، ۱۰۴ نفر از این طرح استقبال کرده‌اند. یک فاصله اطمینان ۹۹ درصد برای نسبت کسانی که از این طرح استقبال کرده‌اند بدست آورید.

$$n=۲۰۰ \quad , \quad x=۱۰۴ \quad , \quad \hat{P}=\frac{x}{n}=\frac{104}{200} \quad , \quad \alpha=۰/۰۱$$

$$\frac{\alpha}{2}=۰/۰۰۵ \quad , \quad z_{\frac{\alpha}{2}}=۱/۶۴۵$$

$$\frac{x}{n}-z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}<P<\frac{x}{n}+z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

$$\frac{104}{200}-1.645\sqrt{\frac{(0.52)(0.48)}{200}}<P<\frac{104}{200}+1.645\sqrt{\frac{(0.52)(0.48)}{200}}$$

$$0.462<P<0.578$$

۶-۷ فاصله اطمینان برای تفاضل میانگین دو جامعه

فرض کنید $X_1, X_2, \dots, X_n$ یک نمونه m تایی از جامعه ای نرمال با میانگین μ_1 و واریانس معلوم σ_1^2 و $Y_1, Y_2, \dots, Y_n$ یک نمونه n تایی از جامعه دیگر نرمال با میانگین μ_2 و واریانس معلوم σ_2^2 باشند. می دانیم که $\overline{X}$ و $\overline{Y}$ به ترتیب برآوردگرهای μ_1 و μ_2 می باشند. لذا هرگونه استنباط روی μ_1 و μ_2 براساس $\overline{X}$ و $\overline{Y}$ خواهد بود. در این بخش هدف، برآورد فاصله اطمینان برای $\mu_1-\mu_2$ است.

باتوجه به خواص برآوردگرها، $\overline{Y}-\overline{X}$ یک برآوردگر برای $\mu_۲-\mu_۱$ است. چون $\overline{X}$ و $\overline{Y}$ تک تک دارای توزیع نرمال هستند تفاصل آنها نیز نرمال است که میانگین و واریانس آن به صورت زیر محاسبه می‌شود.

$$E(\overline{Y}-\overline{X})=E(\overline{Y})-E(\overline{X})=\mu_2-\mu_1$$

$$v(\overline{Y}-\overline{X})=v(\overline{Y})+v(\overline{X})=\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}$$

پس متغیر $\overline{Y}-\overline{X}$ دارای نرمال با میانگین $\mu_۲-\mu_۱$ و واریانس $\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}$ است. اگر متغیر Z متغیر استاندارد شده $\overline{Y}-\overline{X}$ باشد، Z برابر خواهد بود با:

$$Z=\frac{\overline{Y}-\overline{X}-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}}}$$

چون Z نرمال استاندارد است لذا داریم:

$$P\left[-z_{\frac{\alpha}{2}}<\frac{\overline{Y}-\overline{X}-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}}}<z_{\frac{\alpha}{2}}\right]=1-\alpha$$

از رابطه اخیر داریم:

$$(\overline{Y}-\overline{X})-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}<\mu_2-\mu_1<(\overline{Y}-\overline{X})+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$$

این رابطه، یک فاصله اطمینان $(1-\alpha)100\%$ برای $\mu_۲-\mu_۱$ است. اگر دو جامعه دارای واریانس مشترک $\sigma^۲$ و حجم نمونه ها مساوی n باشند فاصله اطمینان برای $\mu_۲-\mu_۱$ برابر است با:

$$(\overline{Y}-\overline{X})-z_{\frac{\alpha}{2}}\sqrt{\frac{2\sigma^2}{n}}<\mu_2-\mu_1<(\overline{Y}-\overline{X})+z_{\frac{\alpha}{2}}\sqrt{\frac{2\sigma^2}{n}}$$

مثال ۶-۷-۱ برای برآورد تفاصل زمان متوسط دو گروه که صرف مونتاژ اجزای رایانه‌ای معین می‌کنند مهندسی صنعتی دریک شرکت الکترونیکی متوسط زمان انجام کار برای گروه اول که ۱۶ نفر هستند را ۱۲/۷۳ و برای گروه دوم که ۲۵ نفر هستند را ۱۴/۲۳ دقیقه بدست می‌آورد. اگر واریانس دو گروه یکسان و برابر با ۴ دقیقه باشد یک فاصله اطمینان ٪۹۵ برای $\mu_۲-\mu_۱$ بدست آورید.

$m=۱۶ \quad , \quad n=۲۵ \quad , \quad \bar{x}=۱۲/۷۳ \quad , \quad \bar{y}=۱۴/۲۳ \quad , \quad \alpha=۰/۰۵$

$\sigma_۱^۲=\sigma_۲^۲=۴ \quad , \quad z_{\frac{\alpha}{2}}=۱/۹۶$

$$(\bar{y}-\bar{x})-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{m}+\frac{\sigma^2}{n}}<\mu_2-\mu_1<(\bar{y}-\bar{x})+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{m}+\frac{\sigma^2}{n}}$$

$$(14.23-12.73)-1.96\sqrt{\frac{4}{16}+\frac{4}{25}}<\mu_2-\mu_1<(14.23-12.73)+1.96\sqrt{\frac{4}{16}+\frac{4}{25}}$$

$$1.5-1.25<\mu_2-\mu_1<1.5+1.255$$
$$0.25<\mu_2-\mu_1<2.755$$

در فاصله اطمینان برای $\mu_۲-\mu_۱$ فرض شده است که $\sigma_۱^۲$ و $\sigma_۲^۲$ معلوم هستند. اگر $\sigma_۱^۲$ و $\sigma_۲^۲$ مجهول باشند برای حجم نمونه‌های بزرگ m و n می توان از واریانس نمونه $S_x^2=\frac{1}{m-1}\sum(x_i-\bar{x})^2$ به جای $\sigma_۱^۲$ و از $S_y^2=\frac{1}{n-1}\sum(y_i-\bar{y})^2$ به جای $\sigma_۲^۲$ استفاده کرد. در این حالت فاصله اطمینان برای $\mu_۲-\mu_۱$ برابر است با:

$$(\bar{y}-\bar{x})-z_{\frac{\alpha}{2}}\sqrt{\frac{S_x^2}{m}+\frac{S_y^2}{n}}<\mu_2-\mu_1<(\bar{y}-\bar{x})+z_{\frac{\alpha}{2}}\sqrt{\frac{S_x^2}{m}+\frac{S_y^2}{n}}$$

مثال ۶-۷-۲ برای مقایسه عمر متوسط دو نوع لامپ، اطلاعات زیر با نمونه‌گیری از این دو نوع لامپ بدست آمده است. یک فاصله اطمینان ۰/۹۹ برای تفاضل متوسط دو نوع لامپ بدست آورید.

نوع لامپ	حجم نمونه	میانگین	واریانس
X	۴۵	۹۸۴	۸۷۴۲
Y	۵۲	۱۱۲۱	۹۴۱۱

$$(1121-984)-1.645\sqrt{\frac{8742}{45}+\frac{9411}{52}}<(\mu_2-\mu_1)<(1121-984)+1.645\sqrt{\frac{8742}{45}+\frac{9411}{52}}$$

$$105.13<\mu_2-\mu_1<168.87$$

اگر در مفروضات پیدا کردن فاصله اطمینـان بـرای μ_1 ـ μ_2 ، واریـانس دو جامعـه را مجهول و برابر هم یا نزدیک به هم فرض کنیم و حجـم نمونـه هـا m و n کوچـک باشند توزیع $\overline{Y}-\overline{X}$ دارای توزیع استودنت است با $m+n-۲$ درجه آزادی (دلیل بـه عهده خواننده) کمیت محوری در این حالت برابر است با:

$$T=\frac{(\overline{Y}-\overline{X})-(\mu_2-\mu_1)}{\sqrt{S_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}}$$

که درآن:

$$S_p^2=\frac{(m-1)S_x^2+(n-1)S_y^2}{m+n-2}$$

و فاصله اطمینان از رابطه زیر بدست می‌آید.

$$(\overline{y}-\overline{x})-t_{(\frac{\alpha}{2},\ m+n-2)}S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\le\mu_2-\mu_1\le(\overline{y}-\overline{x})+t_{(\frac{\alpha}{2},\ m+n-2)}S_p\sqrt{\frac{1}{m}+\frac{1}{n}}$$

مثال ۶-۷-۳ برای مقایسه دو نوع مدل رایانه A و B از مدل A یک نمونه ۷ تایی و از مدل B یک نمونه ۵ تایی انتخاب شده و مدت عمر مفید آنها به ماه به صورت زیـر ثبت شده است. اگر توزیع جوامع مورد بررسی نرمال باشند یـک فاصـله اطمینـان ۹۵٪ برای عمر مفید تفاضل میانگین ها بدست آورید.

A :	۳۲	۳۰	۳۳	۳۲	۲۹	۳۴	۳۲
B :	۳۳	۳۵	۳۶	۳۷	۳۵		

واریانس مجهول و حجم نمونه‌ها کوچک، پس توزیع کمیـت محـوری اسـتودنت بـا $m+n-۲$ درجه آزادی:

$m=۷$, $n=۵$, $\alpha=۰/۰۵$, $t_{\left(\frac{\alpha}{2},10\right)}=۲/۲۲۸$

$$\bar{x} = ۳۱/۷۱ \quad , \quad S_x^2 = ۲/۹۰ \quad , \quad \bar{y} = ۳۵/۲ \quad , \quad S_y^2 = ۲/۲۰$$

$$S_p^2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2} = \frac{6 \times 2.90 + 4 \times 2.20}{10} = 2.62$$

$$(\bar{y}-\bar{x}) - 2.228\, S_p \sqrt{\frac{1}{m}+\frac{1}{n}} < \mu_2 - \mu_1 \le (\bar{y}-\bar{x}) + 2.228\, S_p \sqrt{\frac{1}{m}+\frac{1}{n}}$$

$$3.49 - 2.11 \le \mu_2 - \mu_1 \le 3.49 + 2.11$$
$$1.38 \le \mu_2 - \mu_1 \le 5.6$$

۶-۸ فاصله اطمینان برای واریانس جامعه

فرض کنید که متغیر تصادفی X دارای توزیع نرمال با میانگین مجهول μ و واریانس مجهول $\sigma^۲$ باشد. اگر $X_۱$، $X_۲$، ...، X_n یک نمونه از این جامعه باشد طبق قضیه ۵-۷-۴، $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی دو با $n-۱$ درجه آزادی است. برای یـافتن فاصـله اطمینان می‌توان از $\frac{(n-1)S^2}{\sigma^2}$ به عنوان کمیت محوری استفاده کرد.

$$P\left[\chi^2_{[1-\frac{\alpha}{2},\, n-1]} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{[\frac{\alpha}{2},\, n-1]}\right] = 1-\alpha$$

که اعداد $\chi^2_{[\frac{\alpha}{2},\, n-1]}$ و $\chi^2_{[1-\frac{\alpha}{2},\, n-1]}$ از جـدول کـی دو قابـل محاسـبه اسـت. از پیـشامد احتمال اخیر می توان $\sigma^۲$ را برحسب $S^۲$ به صورت زیر بدست آورد.

$$\boxed{\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},\, n-1]}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},\, n-1]}}}$$

رابطه اخیریک فاصله اطمینان$(۱-\alpha)$۱۰۰٪ برای $\sigma^۲$است.فاصله اطمینـان$(۱-\alpha)$۱۰۰٪ برای σ یا انحراف معیار برابر خواهد بود با:

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},\, n-1]}}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},\, n-1]}}}$$

مثال ۶-۸-۱ فرض کنید که از ماشینی برای بریدن یک قطعه بزرگ فلز به قطعات کوچکتر استفاده گردد و تغییرپذیری طبیعی ماشین سبب شود که طول یک قطعه کوچک بریده شده یک متغیر تصادفی نرمال با میانگین مجهول μ و واریانس σ^2 باشد. اگر σ^2 بیش از حد بزرگ شود باید ماشین را تعمیر کرد. فرض کنید که یک نمونه تصادفی به اندازه ۲۵، از قطعاتی که به وسیله این ماشین بریده شده‌اند، انتخاب $\sum_{i=1}^{25} x_i = ۲۵۲/۷۵$ و $\sum x_i^2 = ۲۵۷۹/۶۸۷۵$ را بدست آورده باشیم، یک فاصله اطمینان ۹۰٪ برای σ^2 بدست آورید.

$$(n-1)S^2 = \sum x_i^2 - n\bar{x}^2 = 2579.6875 - 25\left(\frac{252.75}{25}\right)^2$$

$$(n-1)S^2 = 24.385$$

$$X^2_{(\frac{\alpha}{2},\, n-1)} = X^2_{(0/05,\, 24)} = 36.415$$

$$\chi^2_{(1-\frac{\alpha}{2},\, n-1)} = \chi^2_{(0/95,\, 24)} = 13.848$$

$$\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},\, n-1]}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},\, n-1]}}$$

$$\frac{24.385}{36.415} \le \sigma^2 \le \frac{24.385}{13.848}$$

$$0.669 \le \sigma^2 \le 1.761$$

۶-۹ فاصله اطمینان برای نسبت دو واریانس

اگر $X_1, X_2, \dots, X_m$ یک نمونه تصادفی از جامعه نرمال با میانگین μ_1 و واریانس σ_1^2 و $Y_1, Y_2, \dots, Y_n$ یک نمونه تصادفی از جامعه نرمال دیگر با میانگین μ_2 و واریانس σ_2^2 باشند در بخش ۵-۹ فصل پنجم نشان دادیم که متغیر $F = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$

دارای توزیع فیشر با ۱- m و ۱- n درجه آزادی است. متغیر F یک کمیت محوری است چون توزیع آن مستقل از σ_1^2 و σ_2^2 است لذا،

$$P\left[f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} < F < f_{(\frac{\alpha}{2},\, m-1,\, n-1)}\right] = 1-\alpha$$

$$P\left[f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} < \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2} < f_{(\frac{\alpha}{2},\, m-1,\, n-1)}\right] = 1-\alpha$$

که $f_{(\frac{\alpha}{2},\, m-1,\, n-1)}$ و $f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)}$ از جدول توزیع فیشر بدست می‌آیند. اگر $S_1 = \frac{1}{m-1}\sum(x_i-\bar{x})^2$ و $S_2 = \frac{1}{n-1}\sum(y_i-\bar{y})^2$ به ترتیب واریانس نمونه اول و دوم باشند فاصله اطمینان برای $\frac{\sigma_1^2}{\sigma_2^2}$ از رابطه زیر بدست می‌آید.

$$\left[\frac{1}{f_{(\frac{\alpha}{2},m-1,n-1)}}\,\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{f_{(1-\frac{\alpha}{2},m-1,n-1)}}\,\frac{S_1^2}{S_2^2}\right]$$

مثال ۶-۹-۱ نمرات زیر نمونه‌ای از نمرات برنامه نویس در دو گروه ۱ و ۲ می‌باشد. اگر فرض نرمال بودن نمرات در دو گروه پذیرفته شود یک فاصله اطمینان ۹۰٪ برای نسبت واریانس دو جامعه بدست آورید.

گروه اول	۱۲	۱۰	۱۴	۱۳	۱۱		
گروه دوم	۱۷	۱۵	۱۴	۱۶	۱۷	۱۷	۱۶

$$m = 5 \quad , \quad n = 7$$

$$S_1^2 = \frac{1}{m-1}\sum(x_i-\bar{x})^2 = 2.5$$

$$S_2^2 = \frac{1}{n-1}\sum(y_i-\bar{y})^2 = 1.33$$

$$\alpha = 0.1 \quad , \quad \frac{\alpha}{2} = 0.05 \quad , \quad 1-\frac{\alpha}{2} = 0.95$$

$$f_{(\frac{\alpha}{2},\, m-1,\, n-1)} = f_{(0.05,4,6)} = 4.53$$

$$f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} = f_{(0.95,4,6)} = \frac{1}{f_{(0.05,6,4)}} = \frac{1}{6.16} = 0.162$$

$$\frac{1}{f_{(\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2}\leq\frac{\sigma_1^2}{\sigma_2^2}\leq\frac{1}{f_{(1-\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2}$$

$$\frac{1}{4.53}\frac{2.5}{1.33}\leq\frac{\sigma_1^2}{\sigma_2^2}\leq\frac{1}{0.162}\frac{2.5}{1.33}$$

$$0.415\leq\frac{\sigma_1^2}{\sigma_2^2}\leq 11.603$$

خودآزمایی

۱- برای توابع چگالی احتمال زیر فضای نمونه را بنویسید.

$$f(x)=\theta(1-\theta)^x \qquad x=0,1,2,\ldots \qquad , \qquad 0<\theta<1$$

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{1-\beta} \qquad 0<x<1 \qquad 0<\alpha<1 \qquad 0<\beta<1$$

$$f(x)=\frac{1}{\theta-7} \qquad 7<x<\theta$$

۲- دریک ایستگاه هواشناسی ده روز در فصل بارندگی، روزها از نظر بارانی بودن یا نبودن به صورت زیر ثبت شده‌اند. اگر روزهای بارانی را ازهم مستقل فرض کنیم، چه قالب احتمالی می‌توان ارائه داد؟ پارامتر مدل ارائه شده را برآورد کنید.

۰ ۱ ۰ ۱ ۱ ۰ ۰ ۱ ۰ ۱

۳- برآوردگرهای گشتاورها را براساس یک نمونه تصادفی $X_1, X_2, \ldots, X_n$ برای هریک از توابع چگالی احتمال زیر پیدا کنید.

$$f(x)=\theta x^{\theta-1} \qquad , \qquad 0<x<1$$

$$f(x)=(\theta+1)\ x^{-(\theta+2)} \qquad x>1$$

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ x^{\alpha-1}(1-x)^{1-\beta} \qquad 0<x<1$$

$$f(x)=\theta(1-\theta)^{x-1} \qquad x=1,2,3,\ldots$$

$$f(x)=\frac{1}{b-a} \qquad a<x<b$$

۴- فرض کنید برای برآورد پارامتر θ در تابع احتمال سری لگاریتمی، مشاهدات به صورت جدول فراوانی زیر داده شده باشد. اگر $\overline{X}$ یک برآورد نقطه‌ای برای θ باشد

θ را برآورد کنید.

x	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
f	۱۰۶۲	۲۶۳	۱۲۰	۵۰	۲۲	۷	۶	۲	۰	۱	۱

۵- تعداد ساعاتی که یک لامپ الکترونی کار می کند یک متغیر تصادفی نمایی با پارامتر θ است. θ را براساس یک نمونه تصادفی $X_1، X_2، ...، X_n$ به روش گشتاورها برآورد کنید.

۶- اگر متغیر تصادفی X روی بازه $(\theta-1,\theta+2)$ تعریف شده باشد. θ را به روش گشتاورها برآورد کنید.

۷- برآوردگرهای درستنمایی ماکزیمم را براساس یک نمونه تصادفی $X_1، X_2، ...، X_n$ برای هریک از توابع چگالی احتمال زیر پیدا کنید.

$$f(x)=\theta e^{-\theta x} \qquad , \qquad x>0$$

$$f(x)=\binom{x+r-1}{r-1}\theta^r(1-\theta)^x \qquad x=0,1,2,....$$

$$f(x)=\frac{1}{\Gamma(n)}x^{n-1}e^{-x} \qquad x>0$$

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{1-\beta} \qquad 0<x<1$$

$$f(x)=\beta x^{\beta-1}e^{-x^{\beta}} \qquad x>0$$

۸- برای پارامتر θ فواصل اطمینان زیر پیشنهاد شده است کدام یک بهترین فاصله اطمینان برای θ است؟

(-۱ ، ۲) , (۰/۵ ، ۱) , (۰ ، ۲) , (۰/۷۵ ، ۱/۲۵)

۹- یک مطالعه آلودگی هوا در یک ایستگاه آزمایشی انجام شده است. مقادیر مواد آلی محلول در بنزین که در هوا معلق می‌باشد در ۸ نمونه مختلف هوا به صورت ۲/۲، ۱/۸، ۳/۱، ۲/۰، ۲/۴، ۲/۵، ۲/۱ و ۱/۲ بوده است. فرض کنید جامعه‌ای که از آن نمونه گرفتـه شده نرمال باشد. یک فاصله اطمینان ٪۹۵ برای میانگین واقعی بدست آورید.

۱۰- فرض کنید که ۱۲ پرتقال شهسواری بخرید، آنها را وزن کنید و بدست آورید.

$$\sum_{i=1}^{12} x_i = ۱/۹۸ \qquad \text{و} \qquad \sum_{i=1}^{12} x_i^2 = ۰/۳۴۵$$

اگر وزن پرتقال‌ها (برحسب کیلوگرم) دارای توزیع نرمـال بـا میـانگین μ و واریـانس $\sigma^۲$ باشند یک فاصله اطمینان ۹۵ درصد برای μ و $\sigma^۲$ بدست آورید.

۱۱- استادی یک امتحان نهایی برای درس مرکز کنترل اجرا و طراحی رایانه برای گروه بزرگی از دانشجویان درنظر گرفته است. تصور او راجع به نمره آنها به طور ذهنی یـک توزیع نرمال با $\mu=$۶۷/۲ و $\sigma=$۱/۵ می‌باشد.

الف- چه احتمال پیشین اختصاص دهد تا نمره متوسط ۲۵ نفر در فاصله ۶۵ تا ۷۰ قرار گیرد.

ب- اگر امتحان در یک گروه ۴۰ نفری از دانشجویان به عنوان یک نمونه تصادفی انجام شود و میانگین و انحراف معیار نمره‌ها به ترتیب ۷۴/۹ و ۷/۴ باشد مقادیر μ و $\sigma^۲$ را برآورد کنید و خطای حاصل برآوردها را بدست آورید.

۱۲- فرض کنید متغیر Y دارای توزیع دوجمله ای با پارامتر P و $n=$۳۰۰ باشد. اگر $y=$۷۵ مقدار مشاهده برای متغیر Y باشد یک فاصـله اطمینـان ۹۰ درصـد بـرای P بدست آورید.

۱۳- اگـر $\overline{X}$ و $\overline{Y}$ میـانگین هـای دو نمونـه تـصادفی nتـایی بـه ترتیـب از توزیـع $N(\mu_۱,\sigma^۲)$ و $N(\mu_۲,\sigma^۲)$ باشند n را طوری پیدا کنید که:

$$P\left[\overline{X}-\overline{Y}-\frac{\sigma}{2}<\mu_1-\mu_2<\overline{X}-\overline{Y}-\frac{\sigma}{2}\right]=۰/۹$$

۱۴- در یک شرکت پردازش، نمونه‌ای ۱۲ تایی از رایانه‌ها دارای عمر متوسط ۴/۵ سال و در شرکت دیگر عمر متوسط برای نمونه ۱۵ تایی ۳/۴ سال بدست آمده است. اگر توزیع مشاهدات نرمال با واریانس مشترک ۱ باشد یک فاصله اطمینان ۹۵ درصد برای $\mu_۱-\mu_۲$ بدست آورید.

۱۵- اگر $\overline{X}$ میانگین یک نمونه تصادفی از توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ باشد احتمال اینکه فاصله تصادفی $\left[\frac{n(\overline{X}-\mu)^2}{5.024}, \frac{n(\overline{X}-\mu)^2}{0.004}\right]$ ، $\sigma^۲$ را در برداشته باشد چقدر است؟

۱۶- زمان لازم برای انجام یک کار به دو روش I و II توسط کاربران رایانه برحسب ثانیه به صورت زیر بدست آمده است:

I :	۲۰	۱۶	۲۶	۲۷	۲۳	۲۲	
II :	۲۷	۳۲	۴۲	۳۵	۳۲	۳۴	۳۸

الف- یک فاصله اطمینان ۹۰ درصد برای واریانس روش اول بدست آورید.
ب- یک فاصله اطمینان ۹۰ درصد برای نسبت واریانس‌ها بدست آورید.

۱۷- اعداد زیر مقادیر سختی بدست آمده برای نمونه‌های دو آلیاژ منیزیم می‌باشند.

آلیاژ I : ۶۶/۳	۶۳/۵	۶۴/۹	۶۱/۸	۶۴/۳	۶۴/۷	۶۵/۱	۶۴/۵	۶۸/۴	۶۳/۲
آلیاژ II : ۷۱/۳	۶۰/۴	۶۲/۶	۶۳/۹	۶۸/۸	۷۰/۱	۶۴/۸	۶۸/۹	۶۵/۸	۶۶/۲

با فرض نرمال بودن داده‌های دو نمونه یک فاصله اطمینان ۹۵ درصد برای نسبت واریانس‌ها بدست آورید.

فصل ۷

آزمون فرض‌های آماری

مقدمه

آزمون فرض‌های آماری یکی از شاخه‌های آمار استنباطی است که از آن به عنوان مهمترین تئوری تصمیم‌گیری یاد می‌کنند. در ادبیات آماری هرگونه حدس یا تخمین برای پارامترجامعه را فرض آماری تلقی می‌کنند و برای آزمون آن روشهای متفاوتی وجود دارد که بعضی از آنها در این فصل مورد بحث قرار می‌گیرد.

۷-۱ مفاهیم اولیه

در فرض‌هایی که روی پارامتر جامعه اتخاذ می‌شود ممکن است فرض ساده یا مرکب باشد و برای آزمون آنها ممکن است مرتکب خطای نوع اول یا نوع دوم یا تواماً نوع اول و دوم شویم که در این فصل فقط خطای نوع اول و دوم را مورد بحث قرار می‌دهیم. برای روشن شدن مطلب، فرض ساده، فرض مرکب، خطای نوع اول، دوم و توان آزمون را تعریف می‌کنیم.

تعریف ۷-۱-۱ هر فرض آماری که توزیع جامعه را کاملاً مشخص کند فرض ساده و فرضی که توزیع جامعه را مشخص نکند فرض مرکب گویند.

تعریف ۷-۱-۲ در انجام آزمون فرض، اگر فرض آماری درست باشد و ما آن را رد کنیم مرتکب خطای نوع اول و اگر فرض آماری را که پذیرفتیم نادرست باشد مرتکب خطای نوع دوم می‌شویم.
در آزمون فرضها یکی از کارهای مهمی که باید انجام داد نوشتن فرض‌ها به زبان آمار است. در این بخش با ارائه یک مثال، نحوه نوشتن فرض به زبان آمار را تجربه می‌کنیم.

مثال ۷-۱-۳ فرض کنید مدیر کارخانه‌ای سازنده رایانه، ادعا می‌کند که عمر متوسط رایانه‌های تولیدی کارخانه بیش از ۵ سال است.
در این مثال بحث اصلی، میانگین جامعه (μ) است و ادعای مدیر یک ادعا یا فرض مرکب است چون بیش از ۵ سال، ممکن است ۵/۵ ، ۶ و ... سال باشد. ادعای او را که فرض در مورد μ است به صورت زیر می‌نویسیم.

$$H_1 : \mu > ۵$$

در مقابل ادعا مدیر ممکن است شما ادعا کنید که میانگین متوسط عمر رایانه‌های تولیدی مساوی ۵ سال باشد.

$$H_\circ : \mu = ۵$$

ادعای شما در مقابل ادعای مدیر یک فرض ساده است و براساس ادعای شما توزیع جامعه کاملاً مشخص است. اگر فرض کنیم عمر رایانه دارای توزیع نمایی با پارامتر $\theta = \mu$ باشد براساس ادعای شما تابع چگالی نمایی برابر است با:

$$f(x) = \frac{1}{5} e^{-x/5} \qquad x > 0$$

که این توزیع یا تابع چگالی فاقد پارامتر است. در آزمون فرض $H_\circ : \mu = ۵$ در مقابل $H_1 : \mu > ۵$ است اگر یکی از آنها قبول شود دیگری رد می‌شود.
فرضهای فوق را تواماً به صورت زیر نیز می‌نویسند:

$$\begin{vmatrix} H_\circ : \mu = 5 \\ H_1 : \mu > 5 \end{vmatrix}$$

در حالت کلی فرض در مورد میانگین را می‌توان به صورتهای زیر نوشت:

مرکب یکطرفه مرکب دوطرفه

$$1)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad 2)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad 3)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

که μ_0 یک عدد یا همان تخمین برای μ است. فرضهای ۱ و ۲ فرض ساده در مقابل فرض مرکب یکطرفه است و فرض ۳ فرض ساده در مقابل فرض مرکب دوطرفه است.

از آنجایی که به دلایلی دسترسی به جامعه نیست برای انجام آزمون فوق نیاز است که یک نمونه تصادفی از جامعه مورد بررسی اتخاذ کنیم و نسبت به قبول یا رد یکی از فرضهای H_0 در مقابل H_1 تصمیم گیری کنیم. فرض کنید می‌خواهیم فرض زیر را آزمون کنیم.

$$\begin{cases} H_0 : \mu = 5 \\ H_1 : \mu > 5 \end{cases}$$

اگر $X_1, X_2, \dots, X_n$ یک نمونه تصادفی از جامعه مورد بررسی باشد میانگین نمونه برابر است با:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

از آنجایی که در اکثر توزیع‌ها میانگین نمونه یک برآورد نقطه‌ای برای میانگین جامعه است. می‌توان در این فرایند از مقایسه میانگین نمونه با میانگین جامعه μ نسبت به قبول یا رد H_0 اقدام کرد. در نمونه‌هایی که مقدار $\overline{X}$ بیشتر از ۵ است یک پشتیبان برای فرض H_1 است. بنابراین می‌توان گفت که مقادیر بزرگ $\overline{X}$ باعث قبول H_1 و رد H_0 می‌شود. از این نکته استفاده می‌کنیم و خطای نوع اول و دوم را به زبان آماری به صورت زیر می‌نویسیم.

خطای نوع اول: رد H_0 وقتی که H_0 درست است.

$[RH_0 | H_0$ درست $]$

خطای نوع دوم: قبول H_0 وقتی که H_0 غلط است.

H_1 درست بعد H_0 را رد می‌کنه

$[AH_\circ | H_\circ$ غلط $]$

احتمال ارتکاب خطای نوع اول را سطح معنی‌دار آزمون گویند و آن را بـه α نمـایش می‌دهند.

$\alpha = P[RH_\circ | H_\circ$ درست $]$

Rejectiv

احتمال ارتکاب خطای نوع دوم را به β نمایش می‌دهند.

$\beta = P[AH_\circ | H_\circ$ غلط $]$

Acceptens

برای فرایند اخیر می‌توان α و β را به صورت زیر نوشت:

$$\alpha = P[\overline{X} > \overline{X}_\circ \mid \mu = 5]$$

$$\beta = P[\overline{X} < \overline{X}_\circ \mid \mu \neq 5]$$

که $\overline{X}_\circ$ یک عدد است و می‌توان تحت شرایطی آن را بدست آورد. β-۱ را توان آزمون گویند.

$$\text{توان آزمون} = 1 - \beta = 1 - p[\overline{X} < \overline{X}_\circ \mid \mu \neq 5]$$
$$= p[\overline{X} \geq \overline{X}_\circ \mid \mu \neq 5]$$

از توان آزمون برای مقایسه دو آزمون استفاده می‌شود. در شرایطی که دو آزمون دارای αی یکسان باشند آزمونی بهتر است که توان آن بزرگتر باشد. یک آزمون خوب آن است که در آن α و β هر دو کوچک باشند. تنها راهی که می‌توان α و β را کم کرد افزایش حجم نمونه است.

برای آزمون فرض $H_\circ$ در مقابل H_1 ، با اتکا به نمونه تصادفی گرفته شده، آماره‌ای مناسب تحت فرض $H_\circ$ تعریف می‌شود. آماره تعریف شده روی فضای نمونه، فضای نمونه را به دو ناحیه، ناحیه بحرانی و ناحیه قبول افراز می‌کند.

مثال ۷-۱-۴ فرض کنید X_1 و X_2 یک نمونه دو تایی از توزیع برنولی با پارامتر θ باشند. برای آزمون فرض $H_0: \theta = \frac{1}{2}$ در مقابل $H_1: \theta = \frac{1}{3}$ آماره آزمون را به صورت زیر تعریف می‌کنیم.

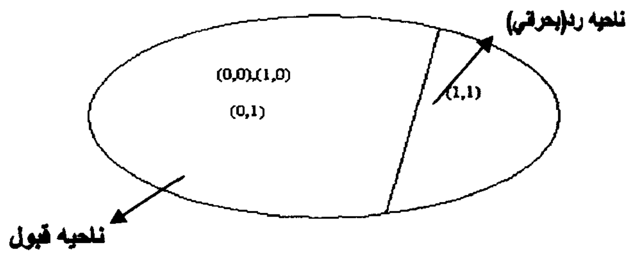
$$T = \frac{X_1 + X_2}{2}$$

اگر آماره T مقدار ۱ را اختیار کند، فرض H_0 را رد می‌کنیم. با این فرض داریم:

$$f(x) = \theta^x (1-\theta)^{1-x} \qquad x = 0,1$$

$$S = \{(X_1, X_2) | X_1 = 0,1\ ,\ X_2 = 0,1\} = \{\underbrace{(0,0),(0,1),(1,0)}_{\text{قبول } H_0},\underbrace{(1,1)}_{\text{رد } H_0}\}$$

آماره T، S را به صورت زیر افراز می‌کند.



برای مثال اخیر α ، β و توان آزمون به صورت زیر محاسبه می‌شود:

$$\alpha = P\{RH_0 | H_0\} = P\left[T = 1 \middle| \theta = \frac{1}{2}\right]$$

$$= P\left[X_1 = 1,\ X_2 = 1 \middle| \theta = \frac{1}{2}\right] = P\left(X_1 = 1 \middle| \theta = \frac{1}{2}\right) P\left(X_2 = 1 \middle| \theta = \frac{1}{2}\right)$$

$$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\beta = [AH_0 | H_1] = P\left[T \neq 1 \middle| \theta = \frac{1}{3}\right]$$

$$=P\left[T=0 \text{ or } T=\frac{1}{2}\middle|\theta=\frac{1}{3}\right]=P\left[T=0\middle|\theta=\frac{1}{3}\right]+P\left[T=\frac{1}{2}\middle|\theta=\frac{1}{3}\right]$$

$$=P\left[X_1=0, X_2=0\middle|\theta=\frac{1}{3}\right]+P\left[X_1=1, X_2=0\middle|\theta=\frac{1}{3}\right]$$
$$+P\left[X_1=1, X_2=0\middle|\theta=\frac{1}{3}\right]=\frac{2}{3}\times\frac{2}{3}+\frac{2}{3}\times\frac{1}{3}+\frac{1}{3}\times\frac{2}{3}=\frac{8}{9}$$

توان آزمون $=1-\beta=1-\frac{8}{9}=\frac{1}{9}$

در دو حالت کلی اگر C و $\overline{C}$ به ترتیب ناحیه رد و قبول باشند. فضای نمونه برابر با اجتماع C و $\overline{C}$ است. یعنی

$S=\overline{C}UC$

برای مثال اخیر،

$S=\{(0,0),(0,1),(1,0),(1,1)\}$
$C=\{(1,1)\}\ ,\ \overline{C}=\{(0,1),(1,0),(0,0)\}$

۷-۲ آزمون فرض برای میانگین توزیع نرمال

فرض کنید $X_۱$، $X_۲$،...، X_n یک نمونه nتایی از توزیع نرمال با میانگین مجهول μ و واریانس معلوم $\sigma^۲$ باشند. هدف آزمون فرض زیر در مورد میانگین جامعه است.

$H_\circ:\mu=\mu_\circ$ در مقابل $H_1:\mu>\mu_\circ$

فرض $H_\circ$ یک فرض ساده در مقابل فرض $H_۱$ که یک فرض مرکب یک طرفه است قرار دارد. در بحث برآوردها، مشخص شده که میانگین نمونه $\overline{X}$ یک برآورد نقطه‌ای برای μ است لذا هرگونه تصمیم‌گیری در مورد فرض فوق براساس $\overline{X}$ خواهد بود. مقادیر بزرگ $\overline{X}$ پشتیبانی برای فرض $H_۱$ یا تأییدی $H_۱$ است و باعث رد فرض $H_\circ$ می‌شود. رد فرض $H_\circ$ را می‌توان به صورتهای زیر نیز نوشت:

رد $H_\circ$ هم ارز است با : $\bar{X} >$ عدد

رد $H_\circ$ هم ارز است با : $\bar{X} - \mu_\circ >$ عدد

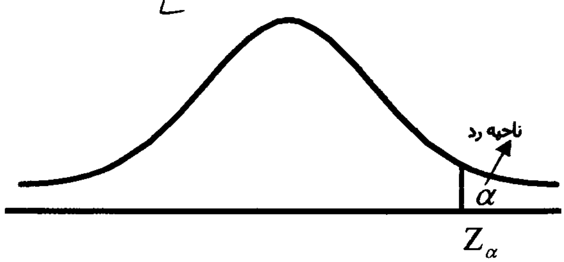
رد $H_\circ$ هم ارز است با: $\dfrac{\bar{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}} >$ عدد

اما $\dfrac{\bar{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}}$ یک کمیت محوری است و تحت فرض $H_\circ$ یک آماره است و از آن به عنوان آماره آزمون یاد می‌کنند.

اگر α سطح معنی دار آزمون باشد با توجه به توزیع آماره آزمون داریم:

$$\alpha = P\left[\frac{\bar{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha \,\middle|\, H_\circ \text{ درست}\right]$$

چون $\dfrac{\bar{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha$ تحت فرض $H_\circ$ دارای توزیع نرمال استاندارد است ناحیه بحرانی و ناحیه قبول برای آزمون فرض فوق به صورت زیر خواهد بود.



مقدار Z_α با توجه به مقدار α از جدول نرمال استاندارد محاسبه می‌شود. برای تصمیم‌گیری در مورد فرض $H_\circ : \mu = \mu_\circ$ در مقابل $H_1 : \mu > \mu_\circ$ مراحل زیر را انجام می‌دهیم.

۱- مقدار آماره آزمون را تحت فرض $H_\circ$ و مشاهدات محاسبه می‌کنیم.

$$Z_{\circ}=\frac{\bar{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}$$

۲- مقدار Z_{α} را از جدول نرمال استاندارد بدست می‌آوریم.

۳- اگر مقدار $Z_{\circ}$ از مقدار Z_{α} بزرگتر باشد فرض $H_{\circ}$ رد می‌شود در غیراینصورت پذیرفته می‌شود.

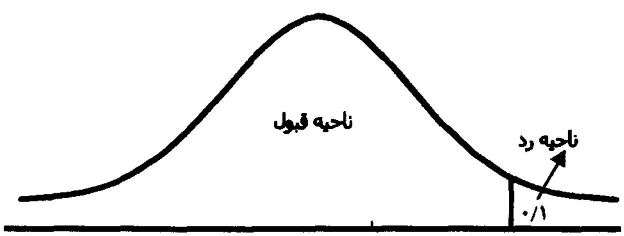
مثال ۷-۲-۱ اگر مشاهدات $x_{۱}, x_{۲}, \dots, x_{۲۵}$ از توزیع نرمال با میانگین μ و واریانس ۱۰۰ دارای میانگین $\bar{x}=۷۶$ باشند فرض زیر را در سطح $\alpha=۰/۱$ آزمون کنید.

$$H_{\circ}:\mu=۷۵ \qquad \text{در مقابل} \qquad H_{1}:\mu>۷۵$$

$$n=۲۵ \quad , \quad \alpha=۰/۱ \quad , \quad \sigma=۱۰$$

$$Z_{\circ}=\frac{\bar{x}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}=\frac{76-75}{\frac{10}{\sqrt{25}}}=0.5$$ آماره آزمون

مقدار جدول $Z_{\alpha}=۱/۲۹ \Leftarrow P[Z>Z_{۰/۱}]=۰/۱$ ، $P[Z>Z_{\alpha}]=\alpha$ چون $Z_{\circ}<Z_{\alpha}$ است فرض $H_{\circ}$ پذیرفته می‌شود.



مراحل انجام آزمون فرض $H_{\circ}:\mu=\mu_{\circ}$ در مقابل $H_{1}:\mu<\mu_{\circ}$ همانند مثال اخیر است و فقط ناحیه رد یا بحرانی به جای اینکه در سمت راست باشد در سمت چپ آن قرار دارد. مراحل انجام آزمون فرض ساده $H_{\circ}:\mu=\mu_{\circ}$ در مقابل فرض مرکب دوطرفه $H_{1}:\mu\neq\mu_{\circ}$ به صورت زیر است.

دراین فرض یا آزمون مقادیر بزرگ یا کوچک $\bar{X}$ پشتیبانی برای H_1 است یا باعث رد H_0 می‌شود. بنابراین

رد H_0 هم ارز است با : $\bar{X} > K_0$ یا $\bar{X} < K'_0$

رد H_0 هم ارز است با : $\bar{X}-\mu_0 > K_1$ یا $\bar{X} - \mu_0 < K'_1$

رد H_0 هم ارز است با: $\dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < K_2$ یا $\dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < K'_2$

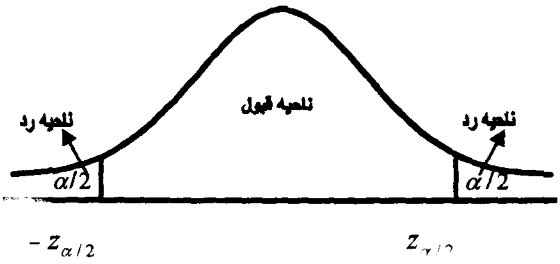
که K_0، K'_0، K_1، K'_1، K_2 و K'_2 اعداد ثابت هستند.

$\dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ تحت فرض H_0 آماره آزمون است و برای αی معین داریم.

$$\left(P\left[\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{\alpha/2} \middle| H_0 \quad \text{درست} \right] = \alpha/2 \right)$$

$$\left(P\left[\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{\alpha/2} \middle| H_0 \quad \text{درست} \right] = \alpha/2 \right)$$

ناحیه بحرانی یا رد برای آزمون H_0 در مقابل H_1 به صورت زیر خواهد بود.



مقادیر $z_{\alpha/2}$ و $-z_{\alpha/2}$ که مقادیر متفاوت هستند از جدول نرمال استاندارد بدست می‌آیند.
برای تصمیم‌گیری در مورد آزمون فرض H_0 در مقابل H_1 مراحل زیر را انجام می‌دهیم.

۱- مقدار آماره را تحت فرض H_0 و مشاهدات محاسبه می‌کنیم.

$$Z_0 = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

۲- مقادیر $z_{\alpha/2}$ و $-z_{\alpha/2}$ را از جدول نرمال استاندارد بدست می‌آوریم.

۳- اگر بازه ($z_{\alpha/2}$ و $-z_{\alpha/2}$) عدد Z_0 را دربرداشته باشد فرض H_0 را می‌پذیریم در غیراینصورت آن را رد می‌کنیم.

مثال ۷-۲-۲ یک نمونه از ۶ میله فولادی دارای قدرت تراکم متوسط ۵۸۳۲۹ پوند بر اینچ مربع است. اگر توزیع جامع نرمال با انحراف معیار ۶۴۸ باشد آیا در سطح ۵ درصد می‌توان گفت قدرت تراکم متوسط واقعی فولاد که نمونه از آن گرفته شده است برابر با ۵۸۰۰۰ است.

$$\begin{cases} H_0 : \mu = 58000 \\ H_1 : \mu \neq 58000 \end{cases}$$

$$n = ۶ \quad , \quad \overline{x} = ۵۸۳۹۲ \quad , \quad \sigma = ۶۴۸ \quad , \quad \alpha = ۰/۰۵ \quad , \quad \mu = ۵۸۰۰۰$$

آماره آزمون

$$Z_0 = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{58392 - 58000}{\frac{648}{\sqrt{6}}} = 1.48$$

مقدار جدول

$$P\left(z > z_{\alpha/2}\right) = ۰/۰۲۵ \Rightarrow z_{\alpha/2} = ۱/۹۶$$
$$P\left(z < -z_{\alpha/2}\right) = ۰/۰۲۵ \Rightarrow -z_{\alpha/2} = -۱/۹۶$$

چون بازه (۱/۹۶ و ۱/۹۶-) عدد ۱/۴۸ را در بردارد فرض H_0 پذیرفته می‌شود. یعنی جامعه دارای میانگین μ=۵۸۰۰۰ است.

در مفروضات و مثالهای اخیر واریانس جامعه، معلوم فرض شده بود. در حالتی که واریانس جامعه نرمال مجهول باشد. فرضهای زیر چگونه آزمون می‌شود.

$$1)\begin{cases} H_\circ : \mu = \mu_\circ \\ H_1 : \mu > \mu_\circ \end{cases} \qquad 2)\begin{cases} H_\circ : \mu = \mu_\circ \\ H_1 : \mu < \mu_\circ \end{cases} \qquad 3)\begin{cases} H_\circ : \mu = \mu_\circ \\ H_1 : \mu \neq \mu_\circ \end{cases}$$

برای آزمون فرضهای فوق یک نمونه تصادفی از توزیع نرمال با میانگین μ و واریانس مجهول $\sigma^۲$ می‌گیریم. توزیع $\overline{X}$ با توجه به بحث بخش ۶-۵ فصل ۶ دارای استودنت با $n-۱$ درجه است.

برای آزمون فرض $H_\circ : \mu = \mu_\circ$ در مقابل $H_1 : \mu < \mu_\circ$ مراحل زیر را انجام می‌دهیم:

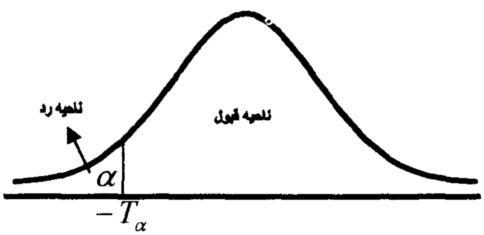
۱- مقدار آماره را تحت فرض $H_\circ$ و مشاهدات محاسبه می‌کنیم.

$$T_\circ = \frac{\bar{x} - \mu_\circ}{\frac{S}{\sqrt{n}}}$$

۲- مقدار T_α را از جدول استودنت با $n-۱$ درجه آزادی بدست می‌آوریم.

$$P[T < -t_\alpha] = \alpha$$

۳- اگر مقدار $T_\circ$ از t_α کوچکتر باشد فرض $H_\circ$ را رد می‌کنیم در غیراینصورت آن را می‌پذیریم.



مثال ۷-۲-۳ آزمون کارکردها با ۶ مدل از یک موتور آزمایشی نشان داده که عملکرد آنها به صورت ۲۴، ۲۸، ۲۱، ۲۳، ۳۲، ۲۲ دقیقه با یک گالن از نوع معینی سوخت بوده است. آیا در سطح ۵ درصد می‌توان گفت که متوسط کارکرد این نوع موتور حداکثر برای هر گالن با این سوخت ۲۹ دقیقه می‌باشد؟

$$\boxed{\mu_0 = ۲۹}$$

$$H_0 : \mu = ۲۹ \qquad\qquad H_1 : \mu < ۲۹$$

۵=درجه آزادی ، $S^۲ = ۱۷/۶$ ، $\bar{x} = ۲۵$ ، $\alpha = ۰/۰۵$ ، $n = ۶$.

$S = ۴٫۱۹$

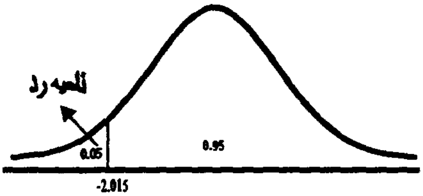
آماره آزمون

$$T_0 = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{25 - 29}{\frac{4.19}{\sqrt{6}}} = -2.34$$

مقدار جدول

$$P[T < t_\alpha] = \alpha$$
$$P[T < t_\alpha] = ۰/۰۵ \Rightarrow T_\alpha = -۲/۰۱۵$$

چون $T_0 < t_\alpha$ است فرض H_0 رد می‌شود.



مثال ۷-۲-۴ پنج اندازه‌گیری از ظرفیت تحریک نوع معین از سیگار منجر به ۱۴/۵، ۱۴/۲، ۱۴/۴، ۱۴/۳ و ۱۴/۶ میلیگرم در هر سیگار شده است. فرض $H_0 : \mu = ۱۴$ را در مقابل $H_1 : \mu \neq ۱۴$ آزمون کنید اگر $\alpha = ۰/۰۵$ باشد.

$$n = ۵ \quad , \quad \bar{x} = ۱۴/۴ \quad , \quad S^۲ = ۰/۰۲۵ \quad , \quad \alpha = ۰/۰۵ \quad , \quad \frac{\alpha}{2} = ۰/۰۲۵$$

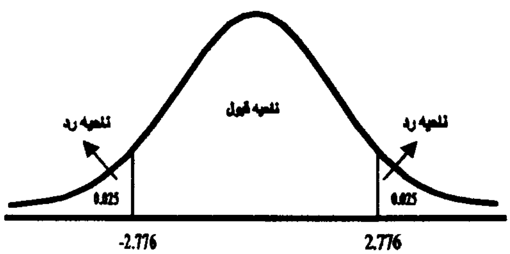
آماره آزمون

$$T_0 = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{14.4 - 14}{\frac{0.158}{\sqrt{5}}} = 5.66$$

مقدار جدول

$$P\left[T < -t_{\frac{\alpha}{2}}\right] = ۰/۰۲۵ \Rightarrow -t_{\frac{\alpha}{2}} = -۲/۹۹۴$$
$$P\left[T > t_{\frac{\alpha}{2}}\right] = ۰/۰۲۵ \Rightarrow t_{\frac{\alpha}{2}} = ۲/۹۹۴$$

چون $\left(-t_{\alpha/2}, t_{\alpha/2}\right)$ عدد $T_{\circ}$=۵/۶۶ را دربرندارد فرض $H_{\circ}$ رد می‌شود.



۷-۳ آزمون فرضهای دو طرفه با استفاده از فاصله اطمینان

در مواقع آزمون فرض ساده در مقابل فرض مرکب دو طرفه می‌توان با استفاده از فاصله اطمینان بدست آمده برای پارامتر مفروض جامعه نسبت به قبول یا رد فرض $H_{\circ}$ اقدام کرد. روش آزمون در حالت خاص برای مثال اخیر ارائه می‌شود.

در مثال اخیر، نواحی رد یا بحرانی از دو رابطه زیر بدست می‌آید.

$$P\left[T < -t_{\alpha/2}\right] = \alpha/2$$
$$P\left[T > t_{\alpha/2}\right] = \alpha/2$$

با استفاده از دو رابطه اخیر احتمال اینکه متغیر T تحت $H_{\circ}$ در ناحیه قبول قرار گیرد برابر است با

$$P\left[-t_{\alpha/2} < T < t_{\alpha/2}\right] = 1 - \alpha$$

اما

$$T = \frac{\overline{X} - \mu_{\circ}}{\frac{S}{\sqrt{n}}}$$

پس

$$P\left[-t_{\alpha/2} < \frac{\overline{X}-\mu_\circ}{\frac{s}{\sqrt{n}}} < t_{\alpha/2}\right] = 1-\alpha$$

اگر پیشامد احتمال اخیر را نسبت به $\mu_\circ$ حل کنیم داریم.

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu_\circ < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

رابطه اخیر ناحیه اطمینان دو طرفه برای $\mu_\circ$ تحت فرض $H_\circ$ است. برای آزمون فرض $H_\circ : \mu = \mu_\circ$ در مقابل $H_1 : \mu \neq \mu_\circ$ با استفاده از فاصله اطمینان، ابتدا فاصله اطمینان را برای $\mu_\circ$ بدست می‌آوریم. اگر $\mu = \mu_\circ$ تحت فرض $H_\circ$ در این فاصله قرار گرفت فرض $H_\circ$ پذیرفته می‌شود در غیراینصورت رد می‌شود. برای مثال اخیر، فاصله اطمینان برای $\mu_\circ$ برابر است با

$$\bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$14.4 - 2.776\frac{0.158}{\sqrt{5}} < \mu < 14.4 + 2.776\frac{0.158}{\sqrt{5}}$$

$$14.204 < \mu < 14.596$$

چون تحت فرض $H_\circ$ ، $\mu = \mu_\circ = ۱۴$ در این فاصله قرار دارد. فرض $H_\circ$ پذیرفته می‌شود.

۴-۷ آزمون فرض آماری با استفاده از P-مقدار

یکی از روشهای آزمون فرض آماری استفاده از P-مقدار یا مقدار احتمال است. اکنون که نرم افزارهای آماری ملاک آزمون فرض آماری را برحسب P-مقدار ارائه می‌کنند، لازم دیدیم که این بحث را در بخش جداگانه مورد بحث قرار دهیم.

فرض کنید $\overline{X}$ میانگین یک نمونه تصادفی از توزیع نرمال با میانگین μ، واریانس معلوم $\sigma^۲$ باشد و بخواهیم فرض $H_\circ : \mu = \mu_\circ$ را در مقابل فرض $H_1 : \mu < \mu_\circ$ آزمون کنیم. در این آزمون مقادیر کوچک $\overline{X}$ باعث رد فرض $H_\circ$ می‌شود. اگر α

سطح معنی‌دار آزمون باشد فرض $H_\circ$ رد می‌شود اگر احتمال پیشامد $\overline{X} \le \bar{x}$ تحت فرض $H_\circ$ کمتر یا مساوی α باشد یعنی P-مقدار برابر است با

$$\text{مقدار} - P = p\left[\overline{X} \le \bar{x} \middle| \quad H_\circ\right] \le \alpha$$

یا

$$\text{مقدار}-P = P\left[\frac{\overline{X}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right] = \left[Z < \frac{\bar{x}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right] \le \alpha$$

در استفاده از روش P-مقدار فرض $H_\circ$ رد می‌شود اگر P-مقدار کمتر یا مساوی α تعیین شده باشند.

مثال ۷-۴-۱ یک نمونه ۳۶ تایی از یک ماشین سازنده نوشابه‌های غیر الکلی به طور متوسط دارای ۷/۴ انس مواد اصلی است. اگر توزیع جامعه مورد بررسی دارای نرمال با انحراف معیار $\sigma =$ ۰/۴۸ باشد.

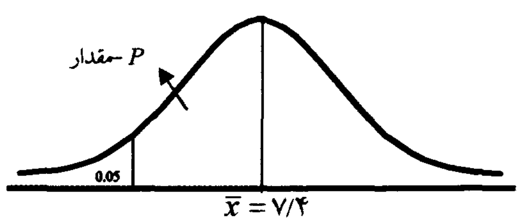
فرض $H_0 : \mu =$ ۷/۵ را در مقابل $\mu <$ ۷/۵ در سطح ۵ درصد آزمون کنید.

$$n = ۳۶ \quad , \quad \bar{x} = ۷/۴ \quad , \quad \sigma = ۰/۴۸ \quad , \quad \alpha = ۰/۰۵$$

$$\text{مقدار}-P = P\ \left[\overline{X} \le \bar{x} | H_\circ\right] = P\left[\frac{\overline{X}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right] = P\left[Z < \frac{7.407-7.5}{\frac{0.48}{\sqrt{36}}}\right]$$

$$= p[Z < -1.25] = 0.1056$$

چون P-مقدار کمتر از ۰/۰۵ نیست فرض $H_\circ$ پذیرفته می‌شود.



مثال ۷-۴-۲ یک کارخانه تولید کننده لامپهای روشنایی ادعا می‌کند که طول عمر لامپهای تولیدی دارای توزیع نرمال با میانگین $\mu=۸۰۰$ و انحراف معیار ۴۰ ساعت است. اگر یک نمونه ۳۰ تایی دارای متوسط ۷۸۸ ساعت باشد فرض $H_{\circ}: \mu=۸۰۰$ را در مقابل $H_1: \mu \neq ۸۰۰$ در سطح ۵ درصد آزمون کنید.

$$n=۳۰ \quad , \quad \bar{x}=۷۸۸ \quad , \quad \alpha=۰/۰۵ \quad , \quad \frac{\alpha}{2}=۰/۰۲۵$$

در این مثال، مقدار بزرگ یا کوچک $\overline{X}$ باعث رد $H_{\circ}$ می‌شود.

P–مقدار در این مثال از دو رابطه زیر بدست می‌آید.

$$\text{مقدار–}P_1 = P\left[\overline{X} > \bar{x} \quad \text{درست} \quad H_{\circ}\right] = P\left[Z > \frac{\bar{x}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$\text{مقدار–}P_1 = P\left[\overline{X} < \bar{x} \quad \text{درست} \quad H_{\circ}\right] = P\left[Z < \frac{\bar{x}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right]$$

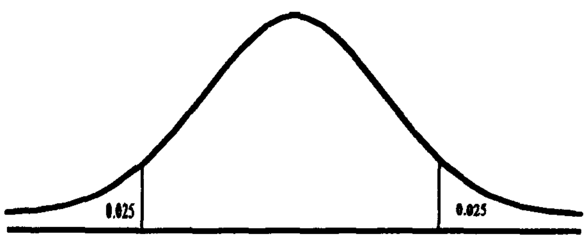
$$\text{مقدار–}P_1 = P\left[Z > \frac{788-800}{\frac{40}{\sqrt{30}}}\right] = P[Z > -1.643] = 0.9495$$

$$\text{مقدار–}P_2 = P\left[Z < \frac{788-800}{\frac{40}{\sqrt{30}}}\right] = P[Z < -1.643] = 0.0505$$

$P_۱$–مقدار و $P_۲$–مقدار را با $\frac{\alpha}{2}=۰/۰۲۵$ مقایسه می‌کنیم.

اگر مقدار $P_۱$–مقدار یا $P_۲$–مقدار کمتر از ۰/۰۲۵ باشد فرض $H_{\circ}$ رد می‌شود.

چون $P_۱$–مقدار $<۰/۰۲۵$ یا $P_۲$–مقدار $<۰/۰۲۵$ است پس فرض $H_{\circ}$ پذیرفته می‌شود.



مثال ۷-۴-۳ یک نمونه ۹ تایی از یک جامعه نرمال دارای میانگین ۳۲/۸= $\bar{x}$ و انحراف معیار ۴/۵۱= s است. اگر ۰/۱= α باشد فرض ۳۰= $H_{\circ} = \mu$ را در مقابل ۳۰< $H_{\circ} = \mu$ آزمون کنید.

n=۹ ، α =۰/۱ ، $\bar{x}$ =۳۲/۸ ، s =۴/۵۱ ، ۸=درجه آزادی

$$P\text{-مقدار} = P\left[\bar{X} > \bar{x} \mid H_{\circ}\ \text{درست}\right] = P\left[\frac{\bar{X}-\mu_{\circ}}{\frac{s}{\sqrt{n}}} > \frac{\bar{x}-\mu_{\circ}}{\frac{s}{\sqrt{n}}}\right]$$

$$= P\left[T > \frac{32.8-30}{\frac{4.51}{\sqrt{9}}}\right] = P[T > 1.86]$$

از اینکه آماره T دارای توزیع استودنت با n-۱ درجه آزادی است از جدول استودنت P-مقدار بدست می‌آید.

P-مقدار $= P\,[T >$ ۱/۸۶]=۰/۰۵

چون P-مقدار کمتر از ۰/۰۱ است فرض $H_{\circ}$ رد می‌شود.

۷-۵ آزمون فرض برای پارامتر توزیع دوجمله‌ای

فرض کنید متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامتر مجهول P و n معلوم باشد. انواع فرضها در مورد پارامتر P عبارتند از

$$1)\begin{vmatrix} H_{\circ}: P = P_{\circ} \\ H_1: P > P_{\circ} \end{vmatrix} \qquad 2)\begin{vmatrix} H_{\circ}: P = P_{\circ} \\ H_1: P < P_{\circ} \end{vmatrix} \qquad 3)\begin{vmatrix} H_{\circ}: P = P_{\circ} \\ H_1: P \neq P_{\circ} \end{vmatrix}$$

برای انجام آزمون فرضهای فوق از کمیت محوری $\left(Z = \dfrac{\frac{X}{n} - P}{\sqrt{\frac{P(1-P)}{n}}}\right)$ که برای n بزرگ تحت فرض $H_{\circ}$ دارای توزیع نرمال استاندارد است استفاده می‌کنیم.

چون Z دارای توزیع نرمال استاندارد است نواحی قبول یا رد همانند آزمون برای میانگین توزیع نرمال است.

مثال ۷-۵-۱ محموله‌ای شامل ۵۰ رایانه است، اگر ۸ رایانه در این محموله معیوب باشد آیا در سطح ۵ درصد می‌توان گفت نسبت معیوب در جامعه کمتر از ۲۰ درصد است؟

$$n=۵۰ \quad , \quad x=۸ \quad , \quad \alpha=۰/۰۵ \quad ، \quad P=۰/۲$$

$$\begin{cases} H_\circ : P = 0.2 \\ H_1 : P < 0.2 \end{cases}$$

آماره آزمون

$$Z_\circ = \frac{\frac{x}{n} - P_\circ}{\sqrt{\frac{x}{n}(1-\frac{x}{n})}} = \frac{\frac{8}{50} - 0.2}{\sqrt{\frac{8}{50}(1-\frac{8}{50})}} = -0.109$$

مقدار جدول $P[Z < -Z_\alpha] = ۰/۰۵ \Rightarrow -Z_\alpha = -۱/۶۴$ چون $Z_\circ > -Z_\alpha$ پس فرض $H_\circ$ پذیرفته می‌شود.

۷-۶ آزمون فرض برای تفاضل میانگین دو جامعه

در تحقیقات کاربردی، مسائل متعددی موجودند که در آنها فرضهایی درباره تفاضل بین دو میانگین دو جامعه مورد توجه است. برای مثال ممکن است بخواهیم متوسط سرعت یا عمر متوسط رایانه‌هایی را که توسط دو سازنده تولید می‌شود باهم مقایسه کنیم. اگر توزیع دو جامعه مورد بررسی معلوم باشد، آزمون فرض برابری میانگین دو جامعه امکان پذیر است.

فرض کنید دو جامعه مورد بررسی مستقل و دارای توزیع نرمال وبه ترتیب دارای میانگین‌های $\mu_۱$ و $\mu_۲$ و واریانس $\sigma_۱^۲$ و $\sigma_۲^۲$ باشند. انواع فرضها در مورد مقایسه میانگین‌های دو جامعه عبارتند از:

$$1)\begin{cases} H_\circ : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \qquad 2)\begin{cases} H_\circ : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \qquad 3)\begin{cases} H_\circ : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

یا

$$1)\begin{vmatrix} H_\circ : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{vmatrix} \qquad 2)\begin{vmatrix} H_\circ : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 < 0 \end{vmatrix} \qquad 3)\begin{vmatrix} H_\circ : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{vmatrix}$$

آماره آزمون برای مقایسه تفاضل $\mu_۱ - \mu_۲$ براساس نمونه تصادفی از جامعه اول و دوم خواهد بود. اگر $\bar{X}_۱, \bar{X}_۲$ به ترتیب میانگین نمونه اول و دوم باشند کمیت محوری، همان کمیت محوری بحث شده در فصل ۶ خواهد بود.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

که Z تحت فرض $H_\circ$ دارای توزیع نرمال استاندارد خواهد بود $Z_\circ$ مشاهده شده براساس مشاهده نمونه از دو جامعه برابر است با

$$Z_\circ = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

روش آزمون فرض $H_\circ$ در مقابل $H_۱$ همانند آزمون برای میانگین توزیع نرمال با واریانس معلوم است. برای آزمون فرض $H_\circ : \mu_۱ = \mu_۲$ در مقابل $H_\circ : \mu_۱ > \mu_۲$ مراحل زیر را انجام می‌دهیم.

۱- آماره آزمون را تحت فرض $H_\circ$ و مشاهدات محاسبه می‌کنیم ($Z_\circ$).

۲- مقدار Z_α را از رابطه $P(Z < Z_\alpha) = \alpha$ بدست می‌آوریم.

۳- اگر $Z_\circ > Z_\alpha$ باشد فرض $H_\circ$ را رد می‌کنیم.

مثال ۷-۶-۱ برای آزمون کردن ادعایی که مقاومت سیم برق بیش از۰/۰۵ اهم به وسیله آلیاژ کردن حاصل می‌شود، ۳۲ مقدار برای سیم استاندارد دارای میانگین $\bar{x}_۱$=۰/۱۳۶ و ۳۲ مقدار برای سیم آلیاژ شده دارای میانگین $\bar{x}_۲$=۰/۰۸۳ بدست آمده است. اگر $\sigma_۱$=۰/۰۰۴ و $\sigma_۲$=۰/۰۰۵ باشند در سطح ۵ درصد فرض $H_\circ : \mu_۱ - \mu_۲$=۰/۰۵ را در مقابل $H_۱ : \mu_۱ - \mu_۲$>۰/۰۵ آزمون کنید.

$m = n =$۳۲ , $\bar{x}_۱$=۰/۱۳۶ , $\bar{x}_۲$=۰/۰۸۳ , $\sigma_۱$=۰/۰۰۴ , $\sigma_۲$=۰/۰۰۵

α=۰/۰۵ , $\mu_۱ - \mu_۲$=۰/۰۵

۱- آماره آزمون

$$Z_0=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}=\frac{(0.136-0.083)-0.05}{\sqrt{\frac{0.000016}{32}+\frac{0.000025}{32}}}$$

$$Z_\circ=۲/۶۵$$

۲- مقدار جدول $P[Z<Z_\alpha]=۰/۰۵\Rightarrow Z_\alpha=۱/۶۴۵$

چون $Z_\circ>Z_\alpha$ پس فرض $H_\circ$ رد می‌شود.

اگر در آزمون فرض برابری میانگین دو جامعه نرمال، واریانس‌ها مجهول و برابر هم یا نزدیک به هم باشند و حجم نمونه‌ها کوچک باشند آماره آزمون تحت فرض $H_\circ$ دارای توزیع استودنت با $m+n-۲$ درجه آزادی خواهد بود.

$$T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{S_P^2(\frac{1}{m}+\frac{1}{n})}}$$

که

$$S_P^2=\frac{(m-1)s_1^2+(n-1)s_2^2}{m+n-2}$$

مثال ۷-۶-۲ برای مثال ۶-۷-۳ فرض $H_\circ:\mu_A=\mu_B$ را در مقابل $H_1:\mu_A\neq\mu_B$ را در سطح ۵ درصد آزمون کنید.

$m=۷$ ، $n=۵$ ، $\alpha=۰/۰۵$ ، $\frac{\alpha}{2}=۰/۰۲۵$ ، $\bar{x}_۱=۳۱/۷۱$

$\bar{x}_۲=۳۵/۲$ ، $s_۱^۲=۲/۹۰$ ، $s_۲^۲=۲/۲۰$ ، $S_P^۲=۲/۶۲$ ، ۱۰=درجه آزادی

۱- آماره آزمون

$$T_\circ=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{S_p^2(\frac{1}{m}+\frac{1}{n})}}=\frac{(31.71-35.2)-0}{\sqrt{2.62(\frac{1}{7}+\frac{1}{5})}}=-3.68$$

۲- مقدار جدول

$$P\left[T<-t_{\frac{\alpha}{2}}\right]=۰/۰۲۵\Rightarrow -t_{\frac{\alpha}{2}}=-۲/۲۲۸$$

$$P\left[T>t_{\frac{\alpha}{2}}\right]=۰/۰۲۵\Rightarrow t_{\frac{\alpha}{2}}=۲/۲۲۸$$

چون $T < -t_{\alpha/2}$ است فرض H_0 رد می‌شود. یعنی میانگین دو جامعه مساوی نیستند.

۷-۷ آزمون فرض برای واریانس جامعه

یکی از راههای بررسی تغییرپذیری جامعه، بررسی یا آزمون فرض درباره واریانس جامعه است. به عنوان مثال مهندس کنترل کیفیت باید مراقبت نماید که تغییرپذیری اندازه‌ها از حد معینی بیشتر نشود یا یک داروساز ممکن بخواهد بداند که آیا میزان تغییر در اثر بخشی یک دارو در حدود قابل قبول است یا نه.
انواع فرضهایی که می‌توان درباره واریانس داشت عبارتند از:

$$1)\left|\begin{array}{l}H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2\end{array}\right. \qquad 2)\left|\begin{array}{l}H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2\end{array}\right. \qquad 3)\left|\begin{array}{l}H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2\end{array}\right.$$

اگر جامعه مورد بررسی دارای توزیع نرمال با میانگین μ مجهول و واریانس σ^2 باشد درفصل ۶ نشان داده شد که کمیت $\frac{(n-1)s^2}{\sigma^2}$ یک کمیت محوری است و تحت فرض H_0 دارای توزیع کی دو با $(n-1)$ درجه آزادی است. لذا برای انجام آزمون فرضها از آماره فوق استفاده می‌کنیم. برای آزمون فرض $H_0 = \sigma^2 = \sigma_0^2$ در مقابل $H_1 = \sigma^2 > \sigma_0^2$ مراحل زیر را انجام می‌دهیم.
چون s^2 یک برآورد نقطه‌ای برای σ^2 است لذا مقادیر بزرگ s^2 باعث رد H_0 می‌شود.

رد فرض H_0 هم ارز است با $s^2 >$ عدد
رد فرض H_0 هم ارز است با $(n-1)s^2 >$ عدد
رد فرض H_0 هم ارز است با $\frac{(n-1)s^2}{\sigma^2} >$ عدد

آماره $\frac{(n-1)s^2}{\sigma_0^2}$ تحت فرض H_0 دارای توزیع کی دو با $n-1$ درجه آزادی است. اگر α سطح معنی دار آزمون باشد.

$$\alpha = P\left[\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{(n-1,\alpha)}\right]$$

مقدار $\chi^2_{(n-1,\alpha)}$ از جدول توزیع کی دو بدست می‌آید.

برای آزمون فرض $H_0 = \sigma^2 = \sigma_0^2$ در مقابل $H_1 = \sigma^2 > \sigma_0^2$ کافی است آماره $\chi^2_\circ = \frac{(n-1)s^2}{\sigma^2_\circ}$ تحت فرض $H_\circ$ و مشاهدات محاسبه و با مقدار جدول $\chi^2_{(n-1,\alpha)}$ مقایسه نمود. اگر $\chi^2_\circ > \chi^2_{(n-1,\alpha)}$ باشد فرض $H_\circ$ رد می‌شود.

مثال ۷-۷-۱ بسته های ۵۰۰ گرمی چای به وسیله یک دستگاه اتوماتیک پر می‌شوند. اگر وزن بسته ها از توزیع نرمال با میانگین μ و واریانس σ^2 پیروی کنند برای یک نمونه ۵ تایی مشاهده شده، ۴۹۹، ۵۰۲، ۵۰۱، ۴۹۸ و ۵۰۰ فرض $H_\circ : \sigma^2 = ۲/۴۵$ را در مقابل $H_1 = \sigma^2 > ۲/۴۵$ در سطح ۵ درصد آزمون کنید.

$$n = ۵ \quad , \quad \alpha = ۰/۰۵ \quad , \quad s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = 2.5$$

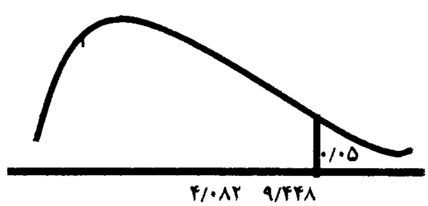
آماره آزمون

$$\chi^2_\circ = \frac{(n-1)s^2}{\sigma^2_\circ} = \frac{4 \times 2.5}{2.45} = 4.082$$

مقدار جدول

$$P\left[\chi^2 > \chi^2_{(n-1,\alpha)}\right] = ۰/۰۵ \Rightarrow \chi^2_{(4,\alpha)} = ۹/۴۴۸$$

چون $\chi^2_\circ < \chi^2_{(4,\alpha)}$ است فرض $H_\circ$ پذیرفته می‌شود.



۷-۸ آزمون فرض برای نسبت دو واریانس

همچنان که دیدیم، وقتی که تفاوت بین میانگین‌های دو جامعه تحت مطالعه باشد و استفاده از توزیع نرمال یا استودنت در تعیین فاصله اطمینان و آزمون فرض مورد نظر

باشد، فرض تساوی واریانس دو جامعه مزبور را باید در نظر داشت. سئوال این است که آیا تفاوت بین واریانس‌های دو نمونه بر تفاوت واقعی میان واریانس‌های دو جامعه دلالت می‌کند یا اگر تفاوت بین واریانس‌های نمونه وجود داشته باشد، آیا تفاوت بر اثر شانس و برحسب تصادف است یا واقعاً واریانسهای دو جامعه با هم اختلاف دارند.

در این بخش به دنبال شیوه‌ای می‌گردیم که پاسخ مناسب آماری را در مورد صحت فرض اختلاف و یا عدم اختلاف واقعی بین واریانسهای دو جامعه ارائه دهد. تصمیم‌گیریهایی که به سازگاری یا عدم سازگاری پراش(واریانس) دو جامعه مربوط می‌شود معمولاً براساس آزمون نسبت واریانسها قرار دارد.

در آزمون فرض، این فرض را آزمون می‌کنیم که نسبت واریانس‌های دو جامعه برابر یک است.

اگر متغیرتصادفی X دارای توزیع نرمال با میانگین μ_1 و واریانس σ_1^2 و متغیر تصادفی Y دارای توزیع نرمال با میانگین μ_2 و واریانس σ_2^2 و متغیرهای X و Y از هم مستقل باشند آنگاه فرضهایی می‌توان برای مقایسه واریانس دو جامعه به صورت زیر نوشت:

$$1)\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{cases} \qquad 2)\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{cases} \qquad 3)\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

اگر S_1^2 واریانس یک نمونه mتایی از جامعه اول و S_2^2 واریانس یک نمونه nتایی از جامعه دوم باشد در فصل ۶ دیدیم که آماره $\left(F = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}\right)$ دارای توزیع فیشر با $m-1$ و $n-1$ درجه آزادی است. برای آزمون فرضهای اخیر از آماره F استفاده می‌کنیم.

در مقایسه واریانس‌های دو جامعه برای سهولت این قرارداد را همیشه درنظر داریم که واریانس نمونه بزرگتر را در صورت قرار می‌دهیم به قسمی که نسبت واریانسهای نمونه همیشه بزرگتر یا مساوی ۱ باشد.

برای نمونه برای آزمون فرض $H_0 = \sigma_1^2 = \sigma_2^2$ در مقابل $H_1 = \sigma_1^2 > \sigma_2^2$ مراحل زیر را انجام می‌دهیم.

۱- مقدار آماره را تحت $H_{\circ}$ و نمونه از رابطه زیر بدست می‌آوریم.

$$F_{\circ}=\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2}$$

۲- مقدار $F_{(m-1,n-1,\alpha)}$ را از جدول فیشر بدست می‌آوریم.

۳- اگر $F_0>F_{(m-1,n-1,\alpha)}$ باشد فرض $H_{\circ}$ را رد می‌کنیم در غیر اینصورت آن را می‌پذیریم.

مثال ۷-۸-۱ آزمونی به منظور اندازه‌گیری سطح اضطراب برای نمونه‌ای از مردان و زنان قبل از انجام یک نوع عمل جراحی ترتیب داده شده است. اگر اطلاعات بدست آمده به صورت جدول زیر باشد فرض $H_0=\sigma_1^2=\sigma_2^2$ را در مقابل $H_1=\sigma_1^2>\sigma_2^2$ در سطح ۵ درصد آزمون کنید.

زنان	$n_1=21$	$s_1^2=275$
مردان	$n_2=16$	$s_2^2=150$

۱- محاسبه مقدار آماره

$$F_{\circ}=\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{s_1^2}{s_2^2}=1\times\frac{275}{150}=1.833$$

۲- مقدار جدول

$$F_{(m-1,n-1,\alpha)}=F_{(20,15,0/05)}=2/33=F_{\alpha}$$

۳- چون $F_0<2/33$ فرض فرض $H_{\circ}$ پذیرفته می‌شود.

خودآزمایی

۱- الف: فرق بین فرض ساده و مرکب چیست؟

ب: خطای نوع اول و دوم را تعریف کنید.

۲- اگر متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای P و n=۵ باشد کدام یک از فرضهای زیر در مورد P ساده و کدام مرکب است؟

$$H_\circ : p = \frac{1}{2} \quad , \quad H_\circ : p = 0.9 \quad , \quad H_\circ : p \le \frac{1}{4}$$

$$H_\circ : p \ne \frac{1}{2} \quad , \quad H_\circ : p > \frac{1}{5} \quad , \quad H_\circ : 0 \le p \le 0.2$$

۳- فرض کنید متغیر تصادفی X دارای توزیع پواسن با پارامتر λ باشد. برای آزمون فرض $H_0 = \lambda = ۲$ در مقابل λ=۱/۵ اگر ناحیه بحرانی به صورت $C = \{x | x \le 1\}$ باشد α ، β و توان آزمون را حساب کنید.

۴- فرض کنید X در فاصله θ تا θ- به طور یکنواخت توزیع شود. یک مقدار x را مشاهده می‌کنیم و می‌خواهیم فرض $H_0 : \theta = ۱$ را در مقابل $H_1 : \theta = ۱/۵$ آزمون کنیم. تصمیم می‌گیریم که اگر مقدار نمونه از ۰/۹۹ تجاوز کند $H_\circ$ را رد کنیم، α ، β و β-۱ را حساب کنید.

۵- فرض کنید $\overline{X}$ میانگین یک نمونه تصادفی با حجم n=۳۶ از جامعه نرمال با میانگین μ و واریانس $\sigma^۲$=۹ باشد. برای آزمون فرض $H_0 : \mu = ۵۰$ در مقابل $H_1 : \mu > ۵۰$ اگر $\overline{X} \ge ۵۰/۸$ باشد فرض $H_\circ$ رد می‌شود. احتمال ارتکاب خطای نوع را حساب کنید.

۶- اگر مقدار تشعشع کیهانی که یک خلبان در مدت پرواز با یک جت در معرض آن قرار می‌گیرد دارای توزیع نرمال با میانگین μ و واریانس $\sigma^۲$ باشد با استفاده از نمونه زیر فرض $H_0 : \mu = ۳/۳$ را درمقابل $H_1 : \mu < ۳/۳$ درسطح ۵ درصدآزمون کنید.

۲/۴۵ ۴/۲۲ ۴/۷۵ ۴/۴۷ ۲/۵۸ ۳/۴۰ ۱/۱۲ ۱/۴۳ ۳/۶۵

۷- اگر $\overline{X}$ میانگین یک نمونه ۱۶ تایی از توزیع نرمال با میانگین μ و انحراف معیار $\sigma=۸$ باشد برای آزمون $H_0: \mu=۳۵$ در مقابل $H_1: \mu>۳۵$ فرض H_0 را رد می‌کنیم اگر $\overline{X}>۳۶/۵$ باشد. مقدار P-مقدار را حساب کنید.

۸- در یک کارخانه، تجربه نشان داده است که ۱۵ درصد از کارگران به علت وجود ذرات سمی معلق در کارخانه پس از ۲۰ سال مبتلا به بیماری خاص می‌شوند. مدیریت جدید با ایجاد شرایط ادعا می‌کند که از ۱۴۰ کارگر حداکثر ۱۹ کارگر مبتلا به بیماری می‌شوند. آیا در سطح ۵ درصد ادعای مدیریت جدید پذیرفته می‌شود؟

۹- اگر $\overline{X}$ و $\overline{Y}$ به ترتیب میانگین‌های نمونه‌های تصادفی به حجم $n_۱=۱۴$ و $n_۲=۱۸$ از جامعه های نرمال بامیانگین $\mu_۱$،واریانس $\sigma_۱^۲=۲۶$ و میانگین $\mu_۲$ و واریانس $\sigma_۲^۲=۲۱$ باشند چگونگی آزمون $H_0: \mu_۱-\mu_۲=c$ را در مقابل $H_1: \mu_۱-\mu_۲<c$ در سطح ۰/۰۲۵ برای $c=0$ و $c=۱$ شرح دهید.

۱۰- دو تولیدکننده لامپ ادعا می‌کنند که بر اساس فرایند جدید در تولید، عمر لامپهای تولیدی نسبت به گذشته افزایش یافته است. در یک بررسی نمونه از این دو تولید کننده اطلاعات زیر بدست آمده است. آیا در سطح ۵ درصد می توان فرض $H_0: \mu_۱=\mu_۲$ را در مقابل $H_1: \mu_۱\neq\mu_۲$ پذیرفت؟

	تولید کننده اول	تولید کننده دوم
حجم نمونه	۱۰۰	۱۰۰
میانگین	۷۹۸	۸۲۶
واریانس	۷۹۸۲	۹۰۰۱

۱۱- یک مهندس علاقه‌مند است قدرت استحکام دو نوع میلگرد از نوع A و B را باهم مقایسه کند. برای این منظور یک نمونه $n_۱=۱۰$ از نوع A و یک نمونه $n_۱=۱۲$

از نوع B انتخاب می‌کند و قدرت استحکام آنها را اندازه‌گیری می‌کند و اطلاعات زیر را بدست می‌آورد.

$n_۱=۱۰ \quad , \quad \bar{x}_۱=۸۲/۶ \quad , \quad s_۱^۲=۶/۵۲$

$n_۲=۱۲ \quad , \quad \bar{x}_۲=۷۸/۱ \quad , \quad s_۲^۲=۷/۰۲$

در سطح ۵ درصد فرض $H_0 : \mu_۱=\mu_۲$ را در مقابل $H_1 : \mu_۲<\mu_۱$ آزمون کنید.

۱۲- در یک نمونه تصادفی از آزمون دانشجویان، نمرات ۱۰ نفر عبارتند از:

۵۰ ۶۵ ۵۸ ۸۲ ۸۰ ۶۰ ۵۲ ۴۰ ۸۲ ۹۵

در سطح ۵ درصد فرض $H_0 : \sigma = ۱۷$ را در مقابل $H_۱ : \sigma > ۱۷$ آزمون کنید.

۱۳- برای بررسی میزان آلودگی هوای خانه در فاصله زمانی ۲۴ ساعته در یک نمونه ۱۶ تایی متوسط و واریانس میزان آلودگی در خانه‌هایی که فاقد افراد سیگاری است به ترتیب $\bar{x}=۶۷/۱$ ، $s_x^۲=۷/۸۲$ و در یک نمونه ۱۳ تایی متوسط و واریانس میزان آلودگی در خانه‌هایی که دارای حداقل یک فرد سیگاری است به ترتیب $\bar{y}=۱۳۲/۳$ و $s_y^۲=۲۴/۱۲$ بدست آمده است.

الف- فرض $H_0 = \sigma_۱^۲=\sigma_۲^۲$ را در مقابل $H_1 = \sigma_۱^۲<\sigma_۲^۲$ در سطح ۵ درصد آزمون کنید.

ب- اگر فرض $H_۰$ رد شد. فرض $H_0 : \mu_۱ = \mu_۲$ را در مقابل $H_1 : \mu_۱ < \mu_۲$ در سطح ۵ درصد آزمون کنید.

[راهنمایی از آماره $Z=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_x^2}{m}+\dfrac{S_y^2}{n}}}$ استفاده کنید، علیرغم اینکه حجم نمونه‌ها زیاد بزرگ نیستند]

فصل ۸

همبستگی و رگرسیون

مقدمه

در فصل اول با جامعه آماری که از مقادیر یک یا چند صفت متغیر تشکیل شده بود آشنا شدیم. گاهی اوقات تحلیل‌گر آماری علاقمند به مطالعه همزمان دو یا چند صفت آماری در جامعه است. اگر دو متغیر X و Y، صفات متغیر مورد بررسی باشند همبستگی بین آنها، برازش خط رگرسیون، پیش‌بینی، و آزمون فرضها روی پارامترهای خط رگرسیون در این فصل مورد بحث قرار می‌گیرند.

۸-۱ ضریب همبستگی

در فصل سوم ضریب همبستگی دو متغیر تصادفی X و Y را به صورت زیر تعریف کردیم:

$$\rho = \frac{E(XY) - E(X)E(Y)^{=cov}}{\sqrt{E(X-\mu_x)^2 . E(Y-\mu_y)^2}}$$

بررسی نمودار پراکنش و محاسبه ضریب همبستگی بین دو متغیر تصادفی X و Y، و تجزیه و تحلیل مقدار ضریب همبستگی در نقاط $\rho=1$ یا $\rho=-1$ این نکته را در ذهن متصور می‌کند که بین دو متغیر X و Y یک رابطه خطی وجود دارد. اگر ρ یکی از دو مقدار ۱ یا ۱- را اختیار نکند آیا در صفحه $X\,Y$ خطی وجود دارد که احتمالات

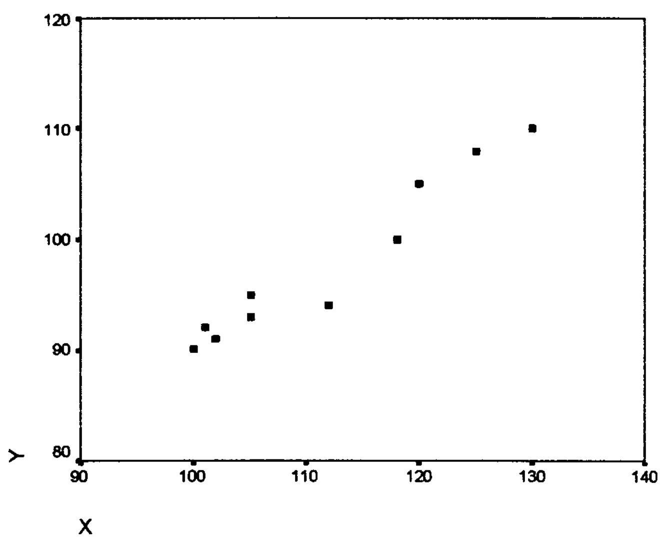
برای X و Y در نواری در اطراف این خط متمرکز باشد؟ تحت برخی شرایط جواب این سئوال مثبت است. تحت این شرایط می‌توان ρ را به عنوان اندازه‌ای برای شدت تمرکز احتمالات X و Y در اطراف این خط دانست.

بررسی همبستگی دو متغیر تصادفی X و Y در جامعه مستلزم دانستن توزیع توأم آنها یا بررسی تمام اعضاء جامع دو بعدی است که همراه با هزینه و صرف وقت زیاد است. لذا به جای بررسی رابطه بین X و Y در جامعه به بررسی در نمونه اکتفا می‌کنیم. یکی از روشهایی که بوسیله آن می‌توان همبستگی بین دو متغیر را نشان داد نمودار پراکنش یا پراکندگی است که بوسیله آن نوع ارتباط دو متغیر مشخص می‌شود.

مثال ۸-۱-۱ متوسط درآمد (X) و هزینه (Y) ده خانوار عبارتند از:

X	۱۰۰	۱۰۲	۱۰۵	۱۰۵	۱۰۱	۱۱۲	۱۱۸	۱۲۰	۱۲۵	۱۳۰
Y	۹۰	۹۱	۹۳	۹۵	۹۲	۹۴	۱۰۰	۱۰۵	۱۰۸	۱۱۰

نمودار پراکنش را رسم کنید



نمودار فوق همبستگی دو متغیر X و Y را نشان می‌دهد که با افزایش درآمد، هزینه نیز افزایش می‌یابد. برای بیان میزان همبستگی، معیار ضریب همبستگی پیرسن را در نمونه که برآوردی برای ضریب همبستگی جامعه است به صورت زیر تعریف می‌کنیم.
فرض کنید (x_1, y_1)، (x_2, y_2)، ...، (x_n, y_n) زوج‌های نمونه به حجم n باشند. ضریب همبستگی نمونه‌ای پیرسن عبارتست از

$$\left(r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}} \right)$$

۸-۱-۲ خصوصیات ضریب همبستگی نمونه‌ای پیرسن

۱- همواره $-1 \le r \le 1$ و مستقل از واحد اندازه‌گیری است.

۲- هنگامی که $r=1$ است، همبستگی دو متغیر X و Y شدید وهمسو است افزایش یکی باعث افزایش دیگری می‌شود.

۳- هنگامی که $r=-1$ است، همبستگی دو متغیر X و Y شدید وخلاف هم و افزایش یکی باعث کاهش دیگری می‌شود.

۴- هنگامی که r در همسایگی صفر است، همبستگی دو متغیر ضعیف است.

مثال ۸-۱-۳ ضریب همبستگی نمونه‌ای جدول زیر را بدست آورید.

i	۱	۲	۳	۴	۵	۶
x_i	۸	۶	۱۲	۱۴	۱۶	۱۰
y_i	۱۵	۱۰	۲۰	۲۵	۳۰	۲۰

برای محاسبه r از جدول زیر استفاده می‌کنیم.

i	۱	۲	۳	۴	۵	۶	جمع
x_i	۸	۶	۱۲	۱۴	۱۶	۱۰	۶۶
y_i	۱۵	۱۰	۲۰	۲۵	۳۰	۲۰	۱۲۰
$x_i\ y_i$	۱۲۰	۶۰	۲۴۰	۳۵۰	۴۸۰	۲۰۰	۱۴۵۰
$(x_i-\bar{x})^2$	۹	۲۵	۱	۹	۲۵	۱	۷۰
$(y_i-\bar{y})^2$	۲۵	۱۱۰	۰	۲۵	۱۰۰	۰	۲۵۰

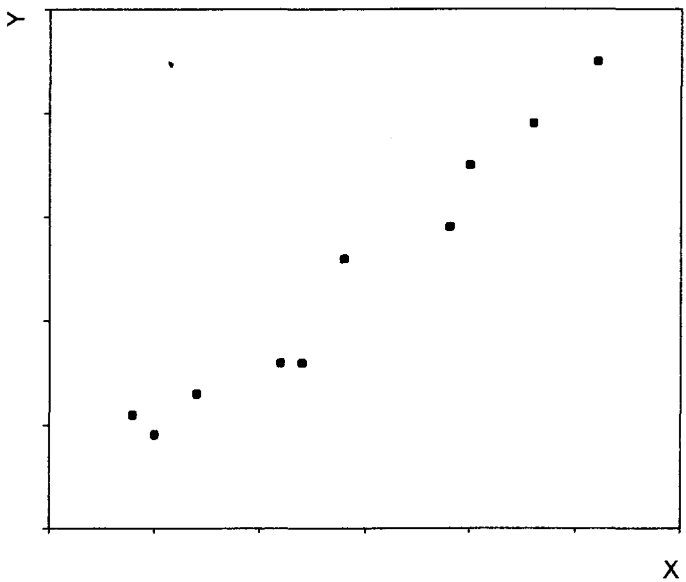
$$r=\frac{\sum x_i y_i - n\bar{x}\ \bar{y}}{\sqrt{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right]}}=\frac{1450-6(11)(20)}{\sqrt{(70)(250)}}=0.9827$$

مثال کاربردی ۸-۱-۴ وزن و مقدار مصرف بنزین ۱۰ مدل اتومبیل در جدول زیر داده شده است.

الف- نمودار پراکنش را رسم کنید

ب- ضریب همبستگی نمونه‌ای را بدست آورید ونتایج بدست آمده راتحلیل کنید.

مدل	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
وزن (X)	۳/۵	۳/۸	۴/۱	۲/۲	۲/۶	۲/۹	۲/۰	۲/۷	۱/۹	۳/۴
مقدار مصرف بنزین (Y)	۵/۵	۵/۹	۶/۵	۳/۳	۳/۶	۴/۶	۲/۹	۳/۶	۳/۱	۴/۹



$$n = ۱۰ \quad ، \quad \bar{x} = ۲/۹۱ \quad ، \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = ۵/۲۸۹$$

$$\bar{y} = ۴/۳۹ \quad ، \quad \sum_{i=1}^{n}(y_i - \bar{y})^2 = ۱۴/۵۸۹ \quad ، \quad \sum_{i=1}^{n}x_i y_i = ۱۳۶/۳۵$$

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\ \bar{y}}{\sqrt{\left(\sum (x_i - \bar{x})^2\right)\left(\sum (y_i - \bar{y})^2\right)}} = \frac{136.35 - 10(2.91)(4.39)}{\sqrt{(5.289\ )(14.589\ )}} = 0.98$$

با توجه به نمودار پراکنش و مقدار همبستگی، نتیجه می‌شود که همبستگی مثبت است. با افزایش وزن ماشین، مقدار مصرف بنزین نیز افزایش می‌یابد.

۸-۲ خط رگرسیون

فرض کنید $f(x,y)$ تابع چگالی احتمال توام دو متغیر X و Y و $f(x)$ تابع چگالی احتمال حاشیه‌ای X باشد. تابع چگالی احتمالی شرطی Y به شرط $X = x$ برابر است با:

$$f(y \mid X = x) = \frac{f(x,y)}{f(x)}$$

و میانگین شرطی Y به شرط $X=x$ برابر است با:

$$E(Y \mid X=x)=\int_{-\infty}^{\infty} y f(y|x) dy=\frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f(x)}$$

میانگین Y به شرط $X=x$ فقط تابعی از x خواهد بود که این تابع تحت شرایطی خطی است و به صورت زیر تعریف می‌شود:

$$E(Y|X)=\alpha+\beta X$$

برای راحتی $E(Y|X)$ را به Y نمایش می‌دهند و معادله خط رگرسیون را در جامعه به صورت زیر می‌نویسند:

$$Y=\alpha+\beta X$$ معادله رگرسیون مهم

که α و β به ترتیب عرض از مبداء و شیب خط رگرسیون و X را متغیر مستقل و Y را متغیر وابسته می‌نامند.

معادله خط رگرسیون در جامعه را می‌توان برای تک تک زوج‌ها به صورت زیر نوشت:

$$Y_i=\alpha+\beta X_i \qquad i=1,2,\ldots,N$$

اگر همه X_iها و Y_iها را دقیقاً مشاهده یا اندازه بگیریم رابطه اخیر همواره برقرار است و می‌توان با مشاهده همه زوج‌های جامعه پارامترهای α و β را محاسبه کرد. اما اغلب در اندازه‌گیریها یا محاسبه X_iها و Y_iها دچار اشتباه و خطا هستیم و ممکن است مقدار X_iها یا Y_iها را کمتر یا بیشتر اندازه بگیریم. در این حالت دچار خطایی به اندازه ε_i خواهیم شد. لذا معادله خط رگرسیون را به صورت زیر در نظر می‌گیریم.

$$Y_i=\alpha+\beta x_i+\varepsilon_i$$ خطای اندازه گیری

اگر در اندازه‌گیریها همه ε_iها صفر باشند معادله اخیر همان معادله $Y_i=\alpha+\beta X_i$ خواهد بود. فرض می‌شود ε_i یک متغیر تصادفی نرمال با میانگین صفر و واریانس σ^2 است. چون Y_i تابع خطی از ε_i است. Y_iها نیز متغیر تصادفی نرمال با میانگین

$\alpha+\beta X_i$ و واریانس $\sigma^{۲}$ می‌باشند. یادآوری می‌شود که X_iها متغیرهای تصادفی نیستند.

اما به دلیلی که قبلاً گفتیم، بررسی کل جامعه مقرون به صرفه نیست. لذا به جای بررسی جامعه به بررسی یک نمونه nتایی می‌پردازیم. دراین حالت خط رگرسیون در نمونه را به صورت زیر تعریف می‌کنیم.

$$y_i=a+bx_i+e_i \quad , \quad i=۱,۲,\ldots,n$$

که e_i اندازه خطا در مشاهده iام است. مقادیر بدست آمده برای a و b براساس نمونه nتایی، به ترتیب تخمینی برای α و β خواهد بود. در ادامه بحث فرض می‌شود. e_iها از هم مستقل و دارای توزیع نرمال با میانگین صفر و واریانس $\sigma_e^{۲}$ می‌باشند. از رابطه زیر می‌توان e_iها را به صورت زیر بدست آورد:

$$e_i=y_i-a-bx_i$$

می‌توان با می‌نیمم کردن $\sum_{i=1}^{n}e_i$ یا $\sum_{i=1}^{n}e_i^2$ ، مقادیر a و b را بدست آورد. ولی بهتر است با می‌نیمم کردن $\sum_{i=1}^{n}e_i^2$، a و b را بدست آورد (چرا؟)

لذا:

$$e_i=y_i-a-bx_i$$
$$e_i^2=(y_i-a-bx_i)^2$$
$$\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

با فرض برابری این رابطه

$$\varphi(a,b)=\sum_{i=1}^{n}e_i^2$$

تابع $\varphi(a,b)$ تابعی از a و b خواهد بود. برای بدست آوردن a و b از تابع $\varphi(a,b)$ یکبار نسبت به a و یکبار نسبت به b مشتق می‌گیریم و برابر صفر قرار می‌دهیم.

$$\frac{\partial\varphi(a,b)}{\partial a}=\sum_{i=1}^{n}(y_i-a-bx_i)=0$$
$$\frac{\partial\varphi(a,b)}{\partial b}=\sum_{i=1}^{n}x_i(y_i-a-bx_i)=0$$

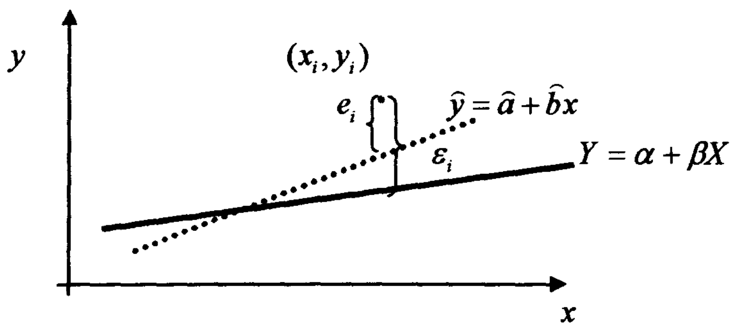
از حل دو معادله اخیر a و b عبارتند از

$$\boxed{\hat{b}=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\ \bar{y}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\quad ,\quad \hat{a}=\bar{y}-\hat{b}\bar{x}}$$

که $\hat{a}$ و $\hat{b}$ به ترتیب عرض از مبداء و شیب خط رگرسیون در نمونه می‌باشند. و معادله خط برازش شده عبارتست از

$$\hat{y}_i=\hat{a}+\hat{b}x_i\quad ,\quad i=۱,۲,\dots,n$$

در نمودار زیر مقادیر ε_i و e_i برای نقطه‌ای مشخصی با هم مقایسه می‌شوند.



مثال ۸-۲-۱ درصد ظرفیت تولید و استخدام کارگر در یکی از صنایع در ۶ سال گذشته در جدول زیر آمده است. ضریب همبستگی را محاسبه کنید و اگر خط رگرسیون برازش شده در نمونه به صورت $\hat{y}=\hat{a}+\hat{b}x$ باشد مقادیر $\hat{a}$ و $\hat{b}$ را بدست آورید.

سال	۱۳۷۵	۱۳۷۶	۱۳۷۷	۱۳۷۸	۱۳۷۹	۱۳۸۰
درصد ظرفیت تولید (X)	۹۳/۰	۸۹/۸	۸۴/۵	۶۰/۶	۶۳/۳	۶۶/۸
تعداد کارگران (Y)	۵۱۹	۵۰۹	۵۰۸	۴۱۲	۴۰۰	۴۵۰

n=۶ ، $\bar{x}$=۷۶/۳۳ ، $\bar{y}$=۴۶۶/۳۳ ، $\sum_{i=1}^{6} x_i y_i$ = ۲۱۷۲۴۸/۴

$\sum_{i=1}^{n}(x_i-\bar{x})^2$ = ۱۰۳۴/۱۱ ، $\sum_{i=1}^{n}(y_i-\bar{y})^2$ = ۱۳۹۴۹/۳۴

$$r=\frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\left(\sum(x_i-\bar{x})^2\right)\left(\sum(y_i-\bar{y})^2\right)}}=\frac{217248.4-6(76.33)(466.33)}{\sqrt{(1034.11)(13949.34)}}=0.97$$

$$\hat{b}=\frac{\sum_{i=1}^{n} x_i y_i - h\overline{xy}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{217248.4-6(76.33)(466.33)}{1034.11}=3.557$$

$$\hat{a}=\bar{y}-\hat{b}\bar{x}=466.33-(3.557)(76.33)=194.824$$

خط برازش شده:

$\hat{y}$ = ۱۹۴/۸۲۴+۳/۵۵۷ x

۸-۳ پیش بینی

با محاسبه مقادیر $\hat{a}$ ، $\hat{b}$ خط رگرسیون نمونه $\hat{y}=\hat{a}+\hat{b}x$ کاملاً مشخص می‌شود. با معلوم بودن خط رگرسیون برازش شده در نمونه، مقدار $\hat{y}_\circ$ برای مقدار مشاهده شده $x_\circ$ را، مقدار پیش بینی گویند.

مقدار پیش بینی $\hat{y}_\circ$ وقتی از دقت بالا برخورد است که مقدار مشاهده شده $x_\circ$ در همسایگی میانگین نمونه باشد.

مثال ۸-۳-۱ در مثال اخیر مقادیر پیش بینی را برای مقادیر $x_\circ$=۷۰/۳۳ ، $x_\circ$=۷۶/۳۳ و $x_\circ$=۸۰/۱۲ بدست آورید.

خط برازش داده شده: $\hat{y}$ = ۱۹۴/۸۲۴+۳/۵۵۷ x

برای $x_\circ$=۷۰/۳۳: $\hat{y}_\circ$ = ۱۹۴/۸۲۴+۳/۵۵۷(۷۰/۳۳)=۴۴۴/۹۸۷۸

برای $x_\circ$=۷۶/۳۳: $\hat{y}_\circ$ = ۱۹۴/۸۲۴+۳/۵۵۷(۷۶/۳۳)=۴۶۶/۳۲۹۸

برای $x_\circ$=۸۰/۱۲: $\hat{y}_\circ$ = ۱۹۴/۸۲۴+۳/۵۵۷(۸۰/۱۲)=۴۷۹/۸۱۰۸

۸-۴ آزمون فرض برای β

در بخش ۸-۲ متذکر شدیم که مقدار بدست آمده برای b براساس یک نمونه تصادفی nتایی یک برآورد گر برای β است به طوری که

$$\hat{\beta}=\hat{b}=\frac{\sum X_iY_i-n\overline{X}\,\overline{Y}}{\sum(X_i-\overline{X})^2}=\frac{\sum(X_i-\overline{X})Y_i}{\sum(X_i-\overline{X})^2}=\sum_{i=1}^{n}c_iY_i$$

$$c_i=\frac{X_i-\overline{X}}{\sum(X_i-\overline{X})^2}$$

از معادله اخیر نتیجه می‌شود که $\hat{\beta}$ یک تابع خطی از Y_iهاست. و چون Y_iها دارای توزیع نرمال هستند $\hat{\beta}$ نیز دارای توزیع نرمال با میانگین و واریانس زیر است:

$$E(\hat{\beta})=E\left[\sum_{i=1}^{n}c_iY_i\right]=\sum_{i=1}^{n}c_iE(Y_i)=\sum_{i=1}^{n}c_i(\alpha+\beta X_i)$$

$$=\alpha\sum_{i=1}^{n}c_i+\beta\sum_{i=1}^{n}c_iX_i=\beta$$

$$V(\hat{\beta})=V\left(\sum_{i=1}^{n}c_iY_i\right)=\sum_{i=1}^{n}c_i^2V(Y_i)=\sum_{i=1}^{n}c_i^2\sigma^2=\frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

مقادیر مثبت یا منفی β در معادله خط رگرسیون $Y=\alpha+\beta X$ بیانگر این واقعیت است که بین X و Y رابطه خطی مستقیم وجود دارد. در مواقعی که β نامعلوم است. تنها می‌توان از داده‌ها نمونه استنباطی روی β انجام داد. با آزمون فرض روی β می‌توان پی به رابطه خطی X و Y برد. رد فرض $H_0:\beta=0$ در مقابل $H_1:\beta\neq 0$ دلیل کافی برای وجود رابطه خطی مستقیم بین X و Y می‌باشد. برای آزمون فرض $H_0:\beta=0$ در مقابل $H_1:\beta\neq 0$ از آماره آزمون زیر تحت فرض H_0 استفاده می‌کنیم

$$\boxed{Z=\frac{\hat{b}-\beta_0}{\sqrt{\frac{\sigma^2}{\sum(X_i-\overline{X})^2}}}}$$

قبول H_0 $\quad -Z_{\frac{\alpha}{2}}<Z<Z_{\frac{\alpha}{2}}$

رد H_0 $\quad Z<-Z_{\frac{\alpha}{2}}$ یا $Z>Z_{\frac{\alpha}{2}}$

با فرض معلوم بودن σ^2 ، متغیر Z دارای توزیع نرمال استاندارد است.
در مواقعی که σ^2 نامعلوم است از برآوردگر نمونه‌ای آن استفاده می‌کنیم.

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=2}^{n}\left[y_i - \hat{a} - \hat{b}x_i\right]^2$$

در این حالت آماره آزمون تحت فرض H_0 عبارتست از:

$$T = \frac{\hat{b} - \beta}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum (X_i - \overline{X})^2}}}$$

$e_i = y_i - \hat{y}$ (واقعی) (پیش‌بینی)

که دارای توزیع استودنت با $n-2$ درجه آزادی است. برای آزمون فرض مراحل زیر را انجام می‌دهیم.

۱- مقدار $T_0 = \dfrac{\hat{b} - \beta}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum (x_i - \overline{x})^2}}}$ را تحت فرض H_0 و نمونه محاسبه می‌کنیم.

۲- مقدار $t(n-2, \alpha/2)$ را از جدول استودنت بدست می‌آوریم.

۳- اگر $T_\circ > t(n-2, \alpha/2)$ یا $T_\circ < -t(n-2, \alpha/2)$ باشد فرض H_0 را رد می‌کنیم.

۸-۵ آزمون فرض برای α

در بخش ۸-۲ از $\hat{a}$ به عنوان یک برآوردگر نقطه‌ای برای α یاد کردیم به طوری که:

$$\hat{\alpha} = \hat{a} = \overline{y} - \hat{b}\,\overline{x} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \overline{X}\sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n}\left[\frac{1}{n} - \overline{X}c_i\right] Y_i$$

$$E(\hat{\alpha}) = \sum_{i=1}^{n}\left[\frac{1}{n} - \overline{X}c_i\right] E(Y_i) = \sum_{i=1}^{n}\left[\frac{1}{n} - \overline{X}c_i\right](\alpha + \beta X_i)$$
$$= \sum_{i=1}^{n}\left[\frac{\alpha}{n} + \frac{\beta X_i}{n} - \alpha\overline{X}c_i - \beta\overline{X}c_i X_i\right] = \alpha$$

$$V(\hat{\alpha}) = \sum_{i=1}^{n} \left[\frac{1}{n} - \overline{X}c_i\right]^2 V(Y_i) = \sigma^2 \sum_{i=1}^{n} \left[\frac{1}{n} - \overline{X}c_i\right]^2$$

$$= \sigma^2 \sum_{i=1}^{n} \left[\frac{1}{n} + \overline{X}^2 c_i^2 - \frac{2\overline{X}c_i}{n}\right] = \sigma^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}\right)$$

همانند $\hat{\beta}$ ، $\hat{\alpha}$ هم دارای توزیع نرمال با میانگین α و واریانس $\sigma^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}\right)$ است.

در خط رگرسیون، از عرض از مبدأ α می‌توان به عنوان یک مؤلفه ثابت در رابطه بین X و Y یاد کرد. مثلاً وقتی که Y هزینه یک نوبت تولید و X تعداد واحدها در هر نوبت باشد، α هزینه ثابت شروع به کار نوبت تولید و β هزینه متغیر به ازای هر واحد تولید می‌باشد. اگر $\alpha = 0$ باشد خط رگرسیون از مبدأ می‌گذرد و مدل رگرسیونی به مدل ساده $Y = \beta X$ تبدیل می‌شود. آزمون فرض روی α همانند آزمون فرض روی β است.

برای آزمون فرض $H_0 : \alpha = 0$ در مقابل $H_1 : \alpha \neq 0$ آماره آزمون تحت فرض H_0 برابر است با:

$$Z = \frac{\hat{a} - \alpha}{\sqrt{\sigma^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}\right]}}$$

$\hat{\sigma}^۲$ واریانس نمونه

$\sigma^۲$ واریانس جامعه

که دارای توزیع نرمال استاندارد است.

اگر $\sigma^۲$ مجهول باشد از برآوردگر آن $\sigma_e^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$ استفاده می‌کنیم و آماره آزمون تحت فرض H_0 در این حالت برابر است با:

$$T = \frac{\hat{a} - \alpha}{\sqrt{\hat{\sigma}_e^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}\right]}}$$

و دارای توزیع استودنت با n-۲ درجه آزادی است. برای آزمون فرض مراحل زیر را انجام می‌دهیم:

۱- مقدار $T_{\circ}=\dfrac{\hat{a}-\alpha}{\sqrt{\hat{\sigma}_e^2\left[\dfrac{1}{n}+\dfrac{\overline{X}^2}{\sum(X_i-\overline{X})^2}\right]}}$ راتحت فرض H_0 و نمونه محاسبه می‌کنیم.

۲- مقدار $t(n-2,\alpha/2)$ را از جدول استودنت بدست می‌آوریم.

۳- اگر $T_{\circ}>t(n-2,\alpha/2)$ یا $T_{\circ}<-t(n-2,\alpha/2)$ باشد فرض H_0 را رد می‌کنیم.

مثال ۸-۵-۱ برای برآورد رابطه بین قیمت رایانه و عمر رایانه نمونه‌ای ۵ تایی به طور تصادفی انتخاب شده و نتایج زیر بدست آمده است.

قیمت (X)	۲	۳	۴	۵	۶
عمر (Y)	۵	۶	۶	۷	۸

الف- ضریب همبستگی نمونه‌ای را محاسبه کنید

ب- اگر $y=\hat{a}+\hat{b}x$ یک خط برازش برای X و Y باشد $\hat{a}$ و $\hat{b}$ را بدست آورید.

ج- اگر $Y=\alpha+\beta X$ خط رگرسیون در جامعه باشد فرضهای زیر را در سطح ۵ درصد آزمون کنید.

$$\begin{vmatrix}H_{\circ}:\alpha=0\\ H_1:\alpha\neq 0\end{vmatrix}\quad,\quad\begin{vmatrix}H_{\circ}:\beta=0\\ H_{\circ}:\beta\neq 0\end{vmatrix}$$

د- اگر رایانه‌ای به قیمت ۵/۵ خریداری شود چقدر عمر می‌کند؟

i	x_i	y_i	x_i-y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	e_i	e_i^2
۱	۲	۵	۱۰	−۲	۴	−۱/۴	۱/۹۶	۰	۰
۲	۳	۶	۱۸	−۱	۱	−۰/۴	۰/۱۶	۰/۳	۰/۰۹
۳	۴	۶	۲۴	۰	۰	−۰/۴	۰/۱۶	−۰/۴	۰/۱۶
۴	۵	۷	۳۵	۱	۱	۰/۶	۰/۳۶	−۰/۱	۰/۰۱
۵	۶	۸	۴۸	۲	۴	۱/۶	۲/۵۶	۰/۲	۰/۰۴
جمع	۲۰	۳۲	۱۳۵	۰	۱۰	۰	۵/۲	۰	۰/۳

$$\bar{x} = ۴ \quad , \quad \bar{y} = ۶/۴ \quad , \quad r = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\sum (x_i - \bar{x})^2 . \sum (y_i - \bar{y})^2}} = \frac{135 - 5(4)(6.4)}{\sqrt{10 \times 5.2}} = 0.9707$$

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{135 - 5(4)(6.4)}{10} = 0.7$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = ۶/۴ - ۰/۷(۴) = ۳/۶$$

$$\hat{y} = ۳/۶ + ۰/۷\, x$$

$$e_i = y_i - \hat{y}_i = y_i - ۳/۶ + ۰/۷\, x$$

$$e_1 = y_1 - (۳/۶ + ۰/۷(۲)) = ۵ - ۳/۶ - ۱/۴ = ۰$$

$$e_2 = y_2 - (۳/۶ + ۰/۷(۳)) = ۶ - ۳/۶ - ۲/۱ = ۰/۳$$

$$e_3 = -۰/۴ \quad , \quad e_4 = -۰/۱ \quad , \quad e_5 = ۰/۲$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{5} e_i^2 = \frac{1}{3}(0.3) = 0.1$$

مقدار آماره آزمون برای آزمون فرض $H_{\circ} : \beta = 0$ در مقابل $H_1 : \beta \neq 0$

$$T_{\circ} = \frac{\hat{b} - \beta}{\sqrt{\hat{\sigma}^2 / \sum (x_i - \bar{x})^2}} = \frac{0.7 - 0}{\sqrt{\frac{0.1}{10}}} = 7$$

مقدار جدول: $3.182 = t(3, 0.025) = t(n-2, \alpha/2)$

چون $T_{\circ} > ۳/۱۸۲$ پس فرض $H_{\circ}$ رد می‌شود. یعنی $\beta \neq 0$ است.

مقدار آماره آزمون برای آزمون فرض $H_{\circ} : \alpha = 0$ در مقابل $H_1 : \alpha \neq 0$

$$T_{\circ} = \frac{\hat{a} - \alpha}{\sqrt{\hat{\sigma}_e^2 \left[\frac{1}{n} + \frac{x^{-2}}{\sum (x_i - \bar{x})^2}\right]}} = \frac{3.6 - 0}{\sqrt{0.1\left[\frac{1}{5} + \frac{16}{10}\right]}} = 8.485$$

مقدار جدول برابر است با $t_{(3,0/025)}=3/182$

چون $T_0>3/182$ پس فرض H_0 رد می‌شود یعنی $\alpha\neq 0$ است.

خط رگرسیون برازش شده : $\hat{y}=3/6+0/7x$

مقدار پیش بینی برای $x=5/5$ برابر است با $\hat{y}=3/6+0/7(5/5)=7/45$

بیشتر مواقع با رسم مشاهدات $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ و بررسی نمودار پراکنش، ملاحظه می‌شود که علیرغم اینکه فرض خطی بودن بین X و Y بوسیله آزمون‌های آماری محرز شده، ولی نمودار پراکنش غیرخطی بودن بین X و Y را نشان می‌دهد. در این مواقع امکان تبدیل متغیرهای X یا Y وجود دارد. در مواقعی که واریانس خط ثابت نباشد، تبدیل متغیر مستقل ممکن و مفید است. ممکن است حالتهایی وجود داشته باشد که هر دو متغیر نیاز به تبدیل داشته باشند. تبدیلهایی که معمولاً مورد استفاده قرار می‌گیرند، تبدیل لگاریتمی، جذر و تبدیلهای وارون است.

با تبدیل متغیر یا متغیرها، رابطه جدید بین X و Y به خطی بودن نزدیک می‌شود. در این صورت مدل رگرسیون خطی را می‌توان برحسب متغیرهای تبدیل یافته فرمول‌بندی کرد و تجزیه و تحلیل مناسب را بر مبنای داده‌های تبدیل یافته انجام داد.

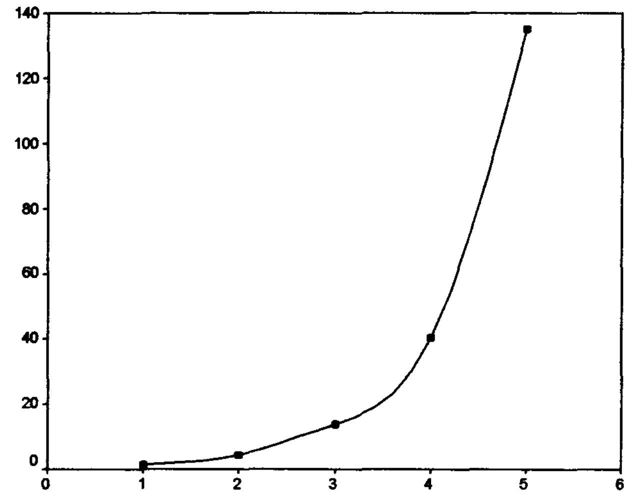
مثال ۸-۵-۲ جدول زیر را در نظر بگیرید

x_1	۱	۲	۳	۴	۵
y_1	۱/۶	۴/۵	۱۳/۸	۴۰/۲	۱۳۵

الف- نمودار پراکنش را رسم کنید.

ب- اگر منحنی حاصل از نمودار پراکنش دارای معادله هندسی $y_1=cx_1^d$ باشد مقادیر c و d را بدست آورید.

ج- برای $x_1=2/5$ مقدار y_1 را پیش بینی کنید.



معادله $y_1 = cx_1^d$ با گرفتن لگاریتم

$$\log(y_1) = \log(c) + d\log(x_1)$$

به معادله $y = a + bx$ تبدیل می‌شود کـه در آن $y = \log(y_1)$ ، $a = \log(c)$، $b = d$ و $x = \log(x_1)$ است. با توجه به تبدیلات

x	۰	۰/۳۰	۰/۴۸	۰/۶۰	۰/۶۹
y	۰/۲۰	۰/۶۵	۱/۱۴	۱/۶۰	۲/۱۳

ضرایب a و b با توجه به تبدیلات محاسبه می‌شود.

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\ \bar{y}}{\sum (x_i - \bar{x})^2} = ۲/۶۸۴$$

$$\hat{a} = \bar{y} - b\bar{x} = ۰/۰۳۳$$

معادله رگرسیون تبدیل یافته: $\hat{y} = ۰/۰۳۳ + ۲/۶۸۴\, x$

معادله رگرسیون قبل از تبدیل: $\hat{y}_۱ = ۱/۰۷۹ \times x_۱^{۲/۶۸۴}$

چون $C = (10)^{\hat{a}} = (10\ )^{0.033} = 1.07\ 9$

مقدار پیش بینی برای $x_1=۲/۵$ برابر است با

$$\hat{y}_1=۱/۰۷۹\times(۲/۵)^{۲/۶۸۴}=۱۲/۷۸$$

خودآزمایی

۱- مقدار کود مصرفی (X) برحسب تن در هکتار و محصول بدست آمده از هر هکتار (Y) در جدول زیر آمده است. نمودار پراکنش بین X و Y را رسم کنید و ضریب همبستگی نمونه‌ای را بدست آورید.

x	۱	۲	۵	۳	۲	۳	۳	۵	۴
y	۱۴	۱۴	۱۸	۱۵	۱۵	۱۷	۱۶	۱۸	۱۷

۲- نشان دهید که

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} \quad , \quad b=\frac{\sum x_iy_i-n\bar{x}\,\bar{y}}{\sum x_i^2-n\bar{x}^2}$$

۳- اگر

$$b_{xy}=\frac{\sum x_iy_i-\left(\sum x_i\right)\left(\sum y_i\right)}{n\sum y_i^2-\left(\sum y_i\right)^2}\,,\; b_{yx}=\frac{n\sum x_iy_i-\left(\sum x_i\right)\left(\sum y_i\right)}{n\sum x_i^2-\left(\sum x_i\right)^2}$$

تعریف شوند. نشان دهید که r برابر با میانگین هندسی b_{xy} و b_{yx} است.

۴- فرض کنید که $E[y|X=x]=\beta X$ و (X_1,Y_1)، (X_2,Y_2)، ...، (X_n,Y_n) یک نمونه تصادفی از Y با مقادیر X وابسته به آنها باشد. فرض کنید $V(y_i)=\sigma_i^2$ نشان دهید که برآوردگر:

$$\hat{b}=\frac{\sum\frac{X_iy_i}{\sigma_i^2}}{\sum\frac{x_i^2}{\sigma_i^2}}$$

یک برآوردگر نااریب برای β است.

۵- اطلاعات زیر مربوط به متوسط فروش روزانه رایانه و هزینه تبلیغات روزانه یکی از نمایندگی‌هاست.

هزینه تبلیغات (هزار ریال)	X	۲	۱	۴	۳	۵	۷	۹	۶	۸	۱۰
فروش روزانه (ده هزار ریال)	Y	۲۸	۲۰	۳۲	۲۹	۳۶	۳۳	۲۶	۳۴	۳۳	۴۰

الف- ضریب همبستگی بین هزینه تبلیغات و فروش روزانه را بدست آورید.

ب- اگر $\hat{y}=\hat{a}+\hat{b}x$ خط برازش شده باشد $\hat{a}$ و $\hat{b}$ را بدست آورید.

ج- اگر ۱۱۰۰۰ ریال هزینه تبلیغات شود تعداد فروش چقدر است؟

۶- اگر در یک نمونه ۱۰ تایی داشته باشیم

$$\sum x_i = ۵۰۰ \quad , \quad \sum y_i = ۱۱۰۰ \quad , \quad \sum x_i y_i = ۶۱۸۰۰ \quad , \quad \sum x_i^2 = ۲۸۴۰۰$$

خط رگرسیون را بدست آورید.

۷- با توجه به دو معادله رگرسیون نمونه‌ای، میانگین $r, \bar{y}, \bar{x}$ را بدست آورید.

$$۲۵x = ۶y + ۷ \quad \text{و} \quad ۴y = ۹x + ۱۵$$

۸- نمره‌های میان ترم (X) و پایان ترم (Y) یک کلاس ۹ نفری به صورت جدول زیر است:

X	۷۷	۵۰	۷۱	۷۲	۸۱	۹۴	۹۶	۹۹	۶۷
Y	۸۲	۶۶	۷۸	۳۴	۴۷	۸۵	۹۹	۹۹	۶۸

الف- معادله خط رگرسیون را پیدا کنید.

ب- نمره پایان ترم دانشجویی که در میان ترم ۸۵ گرفته است و در پایان ترم بـه علـت بیماری نتوانسته است در امتحان پایان ترم شرکت کند تخمین بزنید.

۹- اگر متغیرهای

$$T_\alpha = \frac{\hat{a}-\alpha}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\overline{X}^2}{\sum(X_i-\overline{X})^2}\right]}} \ , \ T_\beta = \frac{\hat{b}-\beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum(X_i-\overline{X})^2}}}$$

کمیت‌های محوری باشند فاصله‌های اطمینان $(1-\alpha)$ درصد برای β و α را بدست آورید.

۱۰- اگر $C_i = \frac{X_i-\overline{X}}{\sum(X_i-\overline{X})^2}$ باشد نشان دهید که

$$\sum_{i=1}^{n} C_i = 0 \quad ، \quad \sum_{i=1}^{n} C_i X_i = ۱ \quad ، \quad \sum_{i=1}^{n} C_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

۱۱- اگر متغیرهای X و Y دارای تابع چگالی توأم زیر باشند نشان دهید که

$$E(Y|X=x) = \frac{2}{3}-\frac{2}{3}x$$

$$f(x,y) = 24xy \qquad x>0 \qquad y>0 \qquad x+y<۱$$

۱۲- جدول زیر مقادیر ارزیابی شد و قیمتهای فروش ۸ خانه را نشان می‌دهد.

مقدار ارزیابی شده (X)	۷۰/۳	۱۰۲	۶۲/۵	۷۴/۸	۵۷/۹	۸۱/۶	۱۱۰/۴	۸۸
قیمت فروش (Y)	۱۱۴/۴	۱۶۹/۳	۱۰۶/۲	۱۲۵	۹۹/۸	۱۳۲/۱	۱۷۴/۲	۱۴۳/۵

الف- یک خط رگرسیون کمترین مربعات برای داده‌ها برازش دهید.

ب- فرض $H_0 = \beta = ۱/۳$ را درمقابل $H_1 = \beta \neq ۱/۳$ درسطح ۵ درصد آزمون کنید.

۱۳- درجدول زیر نمرات تست هوش و برنامه نویسی ۱۲ دانشجو ثبت شده است.

الف- ضرایب خط رگرسیون نمونه را بدست آورید.

ب- اگر خطای مشاهدات دارای واریانس $\sigma_e^۲$ باشد آن را برآورد کنید.

ج- فرض $H_{\circ}: \beta = 0$ را در مقابل $H_1: \beta \neq 0$ در سطح ۵ درصد آزمون کنید.
د- فرض $H_{\circ}: \alpha = 0$ را در مقابل $H_1: \alpha \neq 0$ با استفاده از فاصله اطمینان برای α آزمون کنید

دانشجو	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
نمره تست هوش	۶۵	۵۰	۵۵	۶۵	۵۵	۷۰	۶۵	۷۰	۵۵	۷۰	۵۰	۵۵
نمره برنامه نویسی	۸۵	۷۴	۷۶	۹۰	۸۵	۸۷	۹۴	۹۸	۸۱	۹۱	۷۶	۷۴

۱۴- جدول زیر تعداد ثبت نام شدگان در یک دانشگاه را در ۷ سال گذشته نشان می‌دهد. با استفاده از روش حداقل مربعات منحنی‌ای به شکل $Y = \alpha\beta^X$ را برآورد کرده و تعداد ثبت‌نام را برای ۵ سال آینده پیش‌بینی کنید.

سال (X)	۱	۲	۳	۴	۵	۶	۷
تعداد ثبت نامه (Y)	۳۰۴	۳۴۱	۳۹۳	۴۵۷	۵۴۸	۶۷۰	۸۸۲

۱۵- یک مهندس تولید رابطه بین سرعت قالب‌گیری لوله پلاستیکی (X) و تعداد عیبها در ۵۰۰ فوت لوله (Y) را که با ماشینی جدید تولید می‌شود بررسی کرده است. ۱۲ امتحان با ماشین انجام شده و برد سرعتهای قالب‌گیری از ۰/۲ فوت تا ۱/۲ فوت در ثانیه بوده است. در هر نوبت امتحان، ۵۰۰ فوت لوله تولید شده است نتایج به شرح زیر بوده‌اند:

i	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
X_i	۰/۲	۰/۲	۰/۴	۰/۴	۰/۶	۰/۶	۰/۸	۰/۸	۱/۰	۱/۰	۱/۲	۱/۲
Y_i	۳	۰	۶	۳	۱۲	۷	۱۹	۱۵	۲۴	۲۹	۳۹	۳۴

الف- نمودار پراکنش را رسم کنید و ضرایب خط رگرسیون $\hat{y} = \hat{a} + \hat{b}x$ را بدست آورید.
ب- پس از تحلیل بیشتر، مهندس تصمیم می‌گیرد که از تبدل $Y' = \sqrt{Y}$ استفاده کند. نمودار پراکنش Y' را در مقابل X رسم کنید. آیا در اینجا تبدیل فوق در خطی کردن رابطه بین متغیرها موفق به نظر می‌رسد. خط برازش شده را بنویسید.

جدول تابع توزیع دوجمله‌ای

$$P(X \le x; p) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$P(X \le x; p) = p(X \ge n-x; 1-p) = 1 - p(X \le n-x-1; 1-p)$$

جدول ۱

		p									
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9760	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6553	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8508	0.7447	0.6562
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ادامه جدول ۱

n	x	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0750	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.999[illegible]	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9423
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
	2	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
	3	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ادامه جدول ۱

		p									
n	*x*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
	5	1.0000	0.9991	0.9924	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	13	1.0000	1 0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	0.9958	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ادامه جدول ۱

n	x	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1927	0.0905	0.0424	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9278	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9571	0.7892	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0063
	4	0.9988	0.9776	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855

ادامه جدول

n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	10	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1159
	7	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927
	10	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9819
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
19	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	0.9868	0.8850	0.6841	0.4551	0.2630	0.1332	0.0591	0.0230	0.0077	0.0022
	4	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ادامه جدول ۱

		p									
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1185	0.0510	0.0189	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

جدول توزیع پواسن

جدول ۲

$$P(X \le x)=\sum_{k=0}^{x}\frac{\lambda^k e^{-\lambda}}{k!}$$

x	$\lambda = E(X)$									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.905	0.819	0.741	0.670	0.607	0.549	0.497	0.449	0.407	0.968
1	0.995	0.982	0.963	0.938	0.910	0.878	0.844	0.809	0.772	0.636
2	1.000	0.999	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.920
3	1.000	1.000	1.000	0.999	0.998	0.997	0.994	0.991	0.987	0.981
4	1.000	1.000	1.000	1.000	1.000	1.0001	0.999	0.999	0.998	0.996
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0.333	0.301	0.273	0.247	0.223	0.202	0.183	0.165	0.150	0.135
1	0.699	0.663	0.627	0.592	0.558	0.525	0.493	0.463	0.434	0.406
2	0.900	0.879	0.857	0.833	0.809	0.783	0.757	0.731	0.604	0.677
3.	0.947	0.966	0.957	0.946	0.934	0.921	0.907	0.891	0.875	0.857
4	0.995	0.992	0.989	0.986	0.981	0.976	0.970	0.964	0.956	0.947
5	0.999	0.998	0.998	0.997	0.996	0.994	0.992	0.990	0.987	0.983
6	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.997	0.995
7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

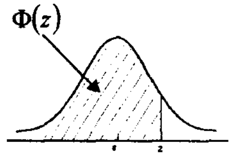
x	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
0	0.111	0.091	0.074	0.061	0.050	0.041	0.033	0.027	0.022	0.018
1	0.355	0.308	0.267	0.231	0.199	0.171	0.147	0.126	0.107	0.092
2	0.623	0.570	0.518	0.469	0.423	0.380	0.340	0.303	0.269	0.238
3	0.819	0.779	0.736	0.692	0.647	0.603	0.558	0.515	0.473	0.433
4	0.928	0.904	0.877	0.848	0.815	0.781	0.744	0.706	0.668	0.629
5	0.975	0.964	0.957	0.935	0.916	0.895	0.871	0.844	0.816	0.785
6	0.993	0.988	0.983	0.976	0.966	0.955	0.942	0.927	0.909	0.889
7	0.998	0.997	0.995	0.992	0.988	0.983	0.977	0.969	0.960	0.949
8	1.000	0.999	0.999	0.998	0.986	0.994	0.992	0.988	0.984	0.979
9	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.996	0.994	0.992
10	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ادامه جدول ۲

x	$\lambda = E(X)$ 4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0
0	0.015	0.012	0.010	0.008	0.007	0.006	0.005	0.005	0.003	0.002
1	0.078	0.066	0.056	0.048	0.040	0.034	0.029	0.024	0.021	0.017
2	0.210	0.185	0.163	0.143	0.125	0.109	0.095	0.082	0.072	0.062
3	0.395	0.359	0.326	0.294	0.265	0.238	0.213	0.191	0.170	0.151
4	0.590	0.551	0.513	0.476	0.440	0.406	0.373	0.342	0.313	0.285
5	0.753	0.720	0.686	0.651	0.616	0.581	0.546	0.512	0.478	0.446
6	0.867	0.844	0.818	0.791	0.762	0.732	0.702	0.670	0.638	0.606
7	0.936	0.921	0.905	0.887	0.867	0.845	0.822	0.797	0.771	0.744
8	0.972	0.964	0.955	0.944	0.932	0.918	0.903	0.886	0.867	0.847
9	0.989	0.985	0.980	0.975	0.968	0.960	0.951	0.941	0.929	0.916
10	0.996	0.994	0.992	0.990	0.986	0.982	0.977	0.972	0.965	0.957
11	0.999	0.998	0.997	0.996	0.995	0.993	0.990	0.988	0.984	0.980
12	1.000	0.999	0.999	0.999	0.998	0.997	0.996	0.995	0.993	0.991
13	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.996
14	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

x	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	11.0	12.0
0	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.011	0.007	0.005	0.003	0.002	0.001	0.001	0.000	0.000	0.000
2	0.043	0.030	0.020	0.014	0.009	0.006	0.004	0.003	0.001	0.001
3	0.112	0.082	0.059	0.042	0.030	0.021	0.015	0.010	0.005	0.002
4	0.224	0.173	0.132	0.100	0.074	0.055	0.040	0.029	0.015	0.008
5	0.369	0.301	0.241	0.191	0.150	0.116	0.089	0.067	0.038	0.020
6	0.527	0.450	0.378	0.313	0.256	0.207	0.165	0.130	0.079	0.046
7	0.673	0.599	0.525	0.453	0.386	0.324	0.269	0.220	0.143	0.090
8	0.792	0.729	0.662	0.593	0.523	0.456	0.392	0.333	0.232	0.155
9	0.877	0.830	0.776	0.717	0.653	0.587	0.522	0.458	0.341	0.242
10	0.933	0.901	0.862	0.816	0.763	0.706	0.645	0.583	0.460	0.347
11	0.966	0.947	0.921	0.888	0.849	0.803	0.752	0.697	0.579	0.462
12	0.984	0.973	0.957	0.936	0.909	0.876	0.836	0.792	0.689	0.576
13	0.993	0.987	0.978	0.966	0.946	0.926	0.898	0.864	0.781	0.682
14	0.997	0.994	0.990	0.983	0.973	0.959	0.940	0.917	0.854	0.772
15	0.999	0.998	0.995	0.992	0.986	0.978	0.967	0.951	0.907	0.844
16	1.000	0.999	0.998	0.996	0.993	0.989	0.982	0.973	0.944	0.899
17	1.000	1.000	0.999	0.998	0.997	0.995	0.991	0.986	0.968	0.937
18	1.000	1.000	1.000	0.999	0.999	0.998	0.996	0.993	0.982	0.963
19	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.991	0.979
20	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.988
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.994
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999

جدول توزیع نرمال استاندارد



$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-w^2}{2}}$$

$$\Phi(-z) = 1 - \Phi(z)$$

جدول ۳

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	0.7703	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

Selected Upper Precntage Points

Tail probability x	0.100	0.050	0.025	0.010	0.005
Upper percentage Point z (x)	1.282	1.645	1.960	2.326	2.576

Source : Reproduced in abridged form from Table 1 of E.S. pearson and H .O. Hartely , Biometrika Tables for Statisticians, Vol. 1 (Cambridge : Cambridge Univerity Press ,1954).

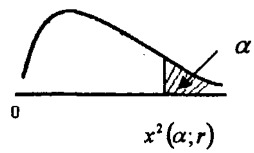
جدول توزیع استودنت

جدول ۴

r	α =0.10	α =0.05	α =0.025	α =0.01	α =0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.635	2.353	3.182	4.541	5.841
4	1.533	2.132	2.996	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1. 328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Source : Reproduced with permission from Table 12 of E. S. pearson and H .O. Hartely , Biometrika Tables for Statisticians, Vol. 1 (Cambridge : Cambridge Univerity Press ,1954).

جدول توزیع کی‌دو

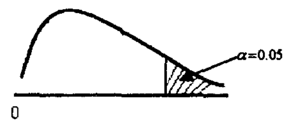


جدول ۵

r	α =0.995	α =0.99	α =0.975	α =0.95	α =0.05	α =0.025	α =0.01	α =0.005	r
1	0.0^4393	0.0^3157	0.0^3982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	23.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	380932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.688	13091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	30

Source : Reproduced with permission from Table 8 of E. S. pearson and H .O. Hartely , Biometrika Tables for Statisticians, Vol. 1 (Cambridge : Cambridge Univerity Press ,1954).

جدول توزیع فیشر

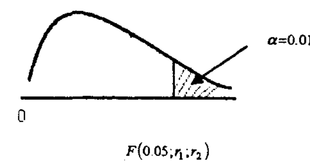


$F(0.05;r_1;r_2)$

جدول ۶

r_2 =Degrees of freedom for denominator	r_1 =Degrees of freedom for unmerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.09	251.14	252.20	253.25	254.32
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	10.128	9.5521	9.2766	9.1172	9.0135	8.946	8.8868	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5265
4	7.7086	6.9443	6.5914	6.3883	6.2560	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6878	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3984	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2066	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6688
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5108	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8378	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2840	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8626	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.601	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6021	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2230	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4035	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3808	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9796	1.9302	1.8780
20	4.3513	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3661	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8895	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8649	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7897	1.7331
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9751	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7307	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2782	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6377
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0848	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2540	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2900	2.1750	2.0867	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	18307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

Source : Reproduced with permission from Table 8 of E. S. pearson and H .O. Hartely , Biometrika Tables for Statisticians, Vol. 1 (Cambridge: Cambridge Univerity Press 1954).



ادامه جدول ۶

r_2 = Degrees of freedom for denominator	r_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052.2	4999.5	5403.3	5624.6	5763.7	5859.0	5928.3	5981.6	6022.5	6055.8	6106.3	6157.3	6208.7	6234.6	6260.7	6286.8	6313.0	6339.4	6366.0
2	98.503	99.000	99.166	99.249	99.299	99.332	99.356	99.374	99.388	99.399	99.416	99.432	99.449	99.458	99.466	99.474	99.483	99.491	99.501
3	34.116	30.817	29.457	28.710	28.237	27911	27.672	27.489	27.345	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.952	13.558	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10289	10.158	10.051	9.8883	9.7222	9.5527	9.4665	9.3793	9.2912	9.2020	9.1118	9.0204
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1016	7.9761	7.8741	7.7183	7.5590	7.3958	7.3127	7.2285	7.1432	7.0568	6.9690	6.8801
7	12.246	9.5466	8.4513	7.8467	7.4604	7.1914	6.9928	6.8401	6.7188	6.6201	6.4691	6.3143	6.1554	6.0743	5.9921	5.9084	5.8236	5.7372	5.6495
8	11.259	8.6491	7.5910	7.0060	6.6318	6.3707	6.1776	6.0289	5.9106	5.8143	5.6668	5.5151	5.3591	5.2793	5.1981	5.1156	5.0316	4.9460	4.8588
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511	5.2565	5.1114	4.9651	4.8080	4.7290	4.6486	4.5667	4.4831	4.3978	4.3104
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424	4.8492	4.7059	4.5582	4.4045	4.3269	4.2469	4.1653	4.0819	3.9965	3.9090
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315	4.5393	4.3974	4.2509	4.0990	4.0209	3.9411	3.8596	3.7761	3.9604	3.6025
12	9.3302	6.9266	5.9526	5.4119	5.0643	4.8206	4.6395	4.4994	4.3875	4.2961	4.1553	4.0096	3.8584	3.7805	3.7008	3.6192	3.5355	3.4494	3.3608
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911	4.1003	3.9603	3.8154	3.6646	3.5868	3.5070	3.4253	3.3413	3.2548	3.1654
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297	3.9394	3.8001	3.6557	3.5052	3.4274	3.3476	3.2656	3.1813	3.0942	3.0040
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948	3.8049	3.6662	3.5222	3.3719	3.2940	3.2141	3.1319	3.0471	2.9595	2.8684
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804	3.6909	3.5527	3.4089	3.2788	3.1808	3.1007	3.0182	2.9330	2.8447	2.7528
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822	3.5931	3.4552	3.3117	3.1615	3.0835	3.0032	2.9205	2.8348	2.7459	2.6530
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971	3.5082	3.3706	3.2273	3.0771	2.9990	2.9185	2.8354	2.7493	2.6597	2.5660
19	8.1850	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225	3.4338	3.2965	3.1533	3.0031	2.9949	2.8442	2.7608	2.6742	2.5839	2.4893
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567	3.3682	3.2311	3.0880	2.9377	2.8594	2.7785	2.6947	2.6077	2.5168	2.4212
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981	3.3098	3.1729	3.0299	2.8796	2.8011	2.7200	2.6359	2.5484	2.4568	2.3603
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458	3.2576	3.1209	2.9780	2.8274	2.7488	2.6675	2.5831	2.4951	2.4029	2.3055
23	7.8811	5.6637	4.7649	4.2635	3.9392	3.7102	3.5390	3.4057	3.2986	3.2106	3.0740	2.9311	2.7805	2.7017	2.6202	2.5355	2.4471	2.3542	2.2559
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560	3.1681	3.0316	2.8887	2.7380	2.6591	2.5773	2.4932	2.4035	2.3099	2.2107
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172	3.1294	2.9931	2.8502	2.6993	2.6203	2.5383	2.4530	2.3637	2.2695	2.1694
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818	3.0941	2.9579	2.8150	2.6640	2.5848	2.5026	2.4170	2.3273	2.2325	2.1315
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494	3.0618	2.9256	2.7827	2.6316	2.5522	2.4699	2.3840	2.2938	2.1984	2.0965
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195	3.0320	2.8959	2.7530	2.6017	2.5223	2.4397	2.3535	2.2629	2.1670	2.0642
29	7.5976	5.4205	4.5378	4.0449	3.7254	3.4995	3.3302	3.1982	3.0920	3.0045	2.8685	2.7256	2.5742	2.4946	2.4118	2.3253	2.2344	2.1378	2.0342
30	7.5625	5.3904	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665	2.9791	2.8431	2.7002	2.5487	2.4689	2.3860	2.2992	2.2079	2.1107	2.0062
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876	2.8005	2.6648	2.5216	2.3689	2.2880	2.2034	2.1142	2.0194	1.9172	1.8047
60	7.0771	4.9774	4.1259	3.6491	3.3389	3.1187	2.9530	2.8233	2.7185	2.6318	2.4961	2.3523	2.1978	2.1154	2.0285	1.9360	1.8363	1.7263	1.6006
120	6.8510	4.7865	3.9493	3.4796	3.1735	2.9559	2.7918	2.6629	2.5586	2.4721	2.3363	2.1915	2.0346	1.9500	1.8600	1.7628	1.6557	1.5330	1.3805
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073	2.3209	2.1848	2.0385	1.8783	1.7908	1.6964	1.5923	1.4730	1.3246	1.0000

Source : Reproduced with permission from Table 8 of E. S. pearson and H .O. Hartely , Biometrika Tables for Statisticians, Vol. 1 (Cambridge: Cambridge Univerity Press 1954).

فواصل اطمینان

پارامتر	مفروضات	فاصله اطمینان
μ	$N(\mu;\sigma^2)$ با حجم نمونه بزرگ با واریانس معلوم	$\bar{x} \pm z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$
μ	$N(\mu;\sigma^2)$ واریانس مجهول	$\bar{x} \pm t_{\left(\frac{\alpha}{2};n-1\right)} \dfrac{S}{\sqrt{n}}$
$\mu_1 - \mu_2$	توزیع جامعه‌ها مستقل با واریانس‌های مجهول برای حجم نمونه‌های بزرگ n_1, n_2	$\bar{x} - \bar{y} \pm t_{\left(\frac{\alpha}{2};n_1+n_2-2\right)} \sqrt{\dfrac{(n_1-1)S_x^2 + (n_2-1)S_y^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}$
$\mu_1 - \mu_2$	توزیع جامعه‌های مستقل با واریانس‌های معلوم برای حجم نمونه‌های بزرگ n_1, n_2	$\bar{x} - \bar{y} \pm z_{\frac{\alpha}{2}} \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
p	$Binomial \;\; b(n,p)$ برای nهای بزرگ	$\dfrac{x}{n} \pm z_{\frac{\alpha}{2}} \sqrt{\dfrac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$

آزمون فرض ها

فرض ها	ناحیه رد H_0
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$\dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \ge z_\alpha$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$\dfrac{\bar{x} - \mu_0}{S/\sqrt{n}} \ge t(\alpha, n-1)$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$	$\dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \ge t(\alpha, n_1 + n_2 - 2)$
$H_0 : p = p_0$ $H_1 : p > p_0$	$\dfrac{\dfrac{x}{n} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} > z_\alpha$
$H_0 : \beta = 0$ $H_1 : \beta > 0$	$\dfrac{\hat{b}}{\sqrt{\dfrac{\sum\left(y_i - \hat{y}_i\right)/(n-2)}{\sum(x_i - \bar{x})^2}}} \ge t(\alpha, n-2)$
$H_0 : \alpha = 0$ $H_1 : \alpha > 0$	$\dfrac{\hat{a}}{\sqrt{\dfrac{\sum\left(y_i - \hat{y}_i\right)}{n-2}\left[\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right]}} \ge t(\alpha, n-2)$

منابع

۱ - آمار و کاربرد آن در مدیریت، تألیف دکتر پرویز کیقبادی، دکتر حسن ستاری (۱۳۴۴)، انتشارات دانشگاه تهران.

۲ - اصول و روشهای آمارزیستی، تألیف واین . و . دانیل (۱۳۶۳)، ترجمه دکتر سید محمد- تقی آیت اللهی، انتشارات امیر کبیر.

۳ - مقدمه‌ای بر آمار ریاضی، تألیف هوگ و کرایج (۱۳۶۳) ، ترجمه دکتر نوروز ایزد دوستدار، انتشارات دانشگاه تهران .

۴ - مفاهیم و روشهای آماری جلد I و II، تألیف گوری . ک. باتاچاریا و ریچارد ا.جانسون (۱۳۶۹)، ترجمه مرتضی ابن اشوب و فتاح میکائیلی، مرکز نشر دانشگاهی.

۵ - شرح و حل مسایل احتمالات، تألیف دکتر بیژن شمس، دکتر محمد رضا سلطانپور و مهندس رسول صادقی حریری (۱۳۷۳)، نشر کتابخانه فروردین .

۶ - آمار کاربردی جلد دوم، تألیف جان نتر، ویلیام واسرمن و تیمور (۱۳۷۴)، ترجمه دکتر علی عمیدی، مرکز نشر دانشگاهی .

۷ - مقدمه‌ای بر احتمالات و آمار کاربردی، تألیف رونالد والپول (۱۳۷۵)، ترجمه دکتر میر- بهادر قلی آریا نژاد و مهندس محمد ذهبیون، انتشارات دانشگاه علم و صنعت ایران .

۸ - آمار و احتمال مهندسی، تألیف میلر، فروند و جانسون (۱۳۷۵)، ترجمه دکتر غلامحسین یاری، انتشارات دانشگاه علم و صنعت ایران .

۹ - آمار و نظریه احتمال، تألیف علی مدنی (۱۳۷۶)، نشر چاپ بهمن .

۱۰ - آمار و احتمال مقدماتی، تألیف دکتر جواد بهبودیان (۱۳۷۷)، دانشگاه امام رضا (ع) مشهد .

۱۱ - آمار ریاضی، تألیف جان فروند (۱۳۷۸)، ترجمه دکترعلی عمیدی و دکتر محمد قاسم وحیدی اصل، مرکز نشر دانشگاهی .

۱۲ - نظریه احتمالات و نتیجه گیری آماری، تألیف هرولد ج . لارسن (۱۳۷۹)، ترجمه غلام حسین همدانی، انتشارات علمی دانشگاه صنعتی شریف ایران .

۱۳ - مقدمه‌ای براحتمال و آمار ریاضی، تألیف لی بین - ماکس انگهارد (۱۳۸۱)، ترجمه علی شکانی ، دکتر حسنعلی آذرنوش و دکتر ابوالقاسم بزرگنیا، انتشارات دانشگاه فردوسی مشهد .

1. Gupta , S .P . , Statistical Methods , (Sultan chand and Sons , New – Delhi , 1997) .
2. Ingram , J.A. , Introductory Statistics (Cummings Puplishing Company , California , (1974) .
3. Nasiri , P. (1997) Generalizations of the Logarithmic Series Distribution Under Markov Dependence and Its Application to wet – spell Analysis , Unpublished Mphil. Dissertation Submitted to the University of Pune .
4. Rohatgi , V. K. and Ehsanes Saleh , A. K. , Introduction to Probability and Statistics . (John Willey and Sons , Inc , 2001) .
5. Robert , V. H. and Ledolter , J. , Applied Statistics for Engineers and Physical Scientists (Macmillam Published Company , New York , 1987)

واژه نامه انگلیسی – فارسی

Acceptance Region	ناحیه قبول
Alternative Hypothesis	فرض مقابل
Arthmetic Mean	میانگین حسابی
Bar Chart	نمودار میله ای
Bayes Formula	فرمول بیز
Bernoulli Distribution	توزیع برنولی
Beta Distribution	توزیع بتا
Binomial Distribution	توزیع دوجمله ای
Bound	کران
Bounded	کراندار
Central Limit Theorem	قضیه حد مرکزی
Chance	شانس
Chebyshev's Inequality	نامساوی جبیشف
Chi - square Distribution	توزیع کی دو
Class	رده - کلاس
Classification	طبقه بندی
Class Interval	فاصله کلاس – طول کلاس
Coding	کد گذاری
Coefficient of Variation	ضریب تغییر
Combination	ترکیب
Composite Hypothesis	فرض مرکب
Confidence Interval	بازه اطمینان
Consistency	سازگار

Continuous	پیوسته
Correlation Coefficient	ضریب همبستگی
Covariance	کوواریانس
Cumulative Frequency	فراوانی تجمعی
Dependent Events	پیشامدهای وابسته
Descriptive Sttistics	آمار توصیفی
Diagram (Chart)	نمودار
Discrete	گسسته
Dispersion Index	شاخص پراکندگی
Distribution Function	تابع توزیع
Dot Diagram	نمودار نقطه ای
Error Type -I	خطای نوع اول
Error Type - II	خطای نوع دوم
Estimate	برآورد
Estimation	برآورد کردن
Estimator	برآوردگر
Exclusive Events	پیشامدهای ناسازگار
Expected Value	امید ریاضی
Experience	تجربه
Exponential Distribution	توزیع نمایی
Forecast	پیش بینی
Fisher Distribution	توزیع فیشر
Gamma Distribution	توزیع گاما
Geometric Mean	میانگین هندسی
Graphs of Cumulative Frequency Curves (ogives)	منحنی چند ضلعی فراوانی تجمعی(اجایو)

Graphs of Frequency Distribution	نمودار توزیعهای فراوانی
Graphs of Frequency Polygon	نمودار چند ضلعی فراوانی
Harmonic Mean	میانگین هارمونیک
Histogram Diagram	نمودار مستطیلی
Homogeneity	همگنی
Hypergeometric Distribution	توزیع فوق هندسی
Hypothesis Distribution	آزمون فرض
Independent Events	پیشامدهای مستقل
Idicator Function	تابع اشاره (نشانگر)
Inferential Statistics	آمار استنباطی
Infinite Population	جامعه نامتناهی
Johnt Distribution	توزیع توام (مشترک)
Kurtosis	کشیدگی
Limit	حد
Linear Regression	خط رگرسیون (برگشت)
Logarithmic Series Distribution	توزیع سری لگاریتمی
Lower Bound	کران پایین
Lower Limit	حد پایین
Maximum	ماکزیمم
Markov Logarithmic Series Distribution	توزیع سری لگاریتمی مارکف
Markov's Inequality	نامساوی مارکف
Mean	میانگین
Mean Square Error	میانگین مربعات خطا
Median	میانه
Method of Maximum Likelihood	روش درستنمایی ماکزیمم
Method of Moments	روش گشتاوره ا

Mid Point	حد متوسط
Minimum	می نیمم
Mode	مد - نما
Moments Generating Function	تابع مولد گشتاورها
Negative Binomial Distribution	توزیع دو جمله ای منفی
Normal Curve	منحنی نرمال
Normal Distribution	توزیع نرمال
Null Hypothesis	فرض صفر
Operator	عملگر
Order	مرتب
Out Come	برآمد
Paire	جفت - زوج
Parameter Space	فضای پارامتر
Permutation	جایگشت
Pivotal Quantity	کمیت محوری
Point Estimation	برآورد نقطه ای
Poisson Distribution	توزیع پواسن
Population	جمعیت - جامعه
Primary Data	داده های دسته اول (خام)
Probability	احتمال
Probability Conditional	احتمال شرطی
Probability Generating Function	تابع مولد احتمال
P- Value	مقدار
Quantitative Variable	متغیر کیفی
Quantitative Variable	متغیر کمی
Quartitle	چارک

Range	دامنه
Rejection Region	ناحیه بحرانی (رد)
Relative Cumulative Frequency	فراوانی تجمعی نسبی
Relative Frequency	فراوانی نسبی
Sample Size	حجم نمونه
Sample Space	فضای نمونه
Sampling Distribution	توزیع نمونه گیری
Scatter Plot	نمودار پراکنش
Secondary Data	داده های دسته دوم
Signification Level	سطح اطمینان (معنی دار)
Simple Hypotheis	فرض ساده
Skewenss	چولگی
Standard Deviation	انحراف معیار
Standard Variable	متغیر استاندارد
Statistic	آماره
Statistics	آمار
Symmetric	متقارن
Tree Diagram	نمودار درختی
Trimmed Mean	میانگین پیراسته
Unbiased	نا اریب
Upper Bound	کران بالا
Upper Limit	حد بالا
Variance	واریانس
Uniform Distribution	توزیع یکنواخت
Weighted Mean	میانگین موزون

خوانندهٔ محترم

این پرسشنامه به منظور ارتقای کیفیت کتابهای درسی و رفع نواقص آنها تهیه شده است. دقت شما در پاسخگویی به این پرسشنامه در پایان هر نیمسال ما را در تحقق این هدف یاری خواهد کرد.

نام کتاب .... نام مؤلف/مترجم .... سال انتشار ....

وضعیت پاسخگو: عضوعلمی پیام نور ☐ عضوعلمی سایر دانشگاهها ☐ رشتهٔ تخصصی .... سابقهٔ تدریس ....

دانشجوی پیام نور ☐ دانشجوی سایر دانشگاهها ☐ رشتهٔ تحصیلی .... ورودی سال ....

سؤال	خیلی زیاد	زیاد	کم	خیلی کم
۱. آیا از زمان تحویل و نحوهٔ دسترسی به کتاب راضی بودید؟				
۲. آیا حجم کتاب با توجه به تعداد واحد مناسب بود؟				
۳. آیا راهنماییهایی لازم برای مطالعهٔ کتاب منظور شده بود؟				
۴. آیا در ترتیب مطالب کتاب سلسله مراتب شناختی (آسان به مشکل) رعایت شده بود؟				
۵. آیا تقسیم‌بندی مطالب در فصلها و یا بخشها متناسب و بجا بود؟				
۶. آیا متن کتاب روان و ساده و جمله‌ها قابل‌فهم بود؟				
۷. آیا به روز بودن مطالب و آمارها رعایت شده بود؟				
۸. آیا مطالب تکراری داشت؟				
۹. آیا پیوستگی مطالب با درسهای پیش‌نیاز رعایت شده بود؟				
۱۰. آیا مثالها، شکلها، نمودارها، جدولها و...گویا بودند و در فهم مطلب تأثیر داشتند؟				
۱۱. مطالعهٔ هدفهای کلی، آموزشی/رفتاری تا چه اندازه به درک بهتر شما کمک کرد؟				
۱۲. آیا خودآزماییهای کتاب به گونه‌ای بود که تمام مطالب درسی را شامل شود؟				
۱۳. آیا پاسخ خودآزماییها و تمرینها کامل و گویا بود؟				
۱۴. چقدر با غلطهای املایی و اشکالهای چاپی مواجه شدید؟				
۱۵. کیفیت چاپ و صحافی کتاب چگونه بود؟				
۱۶. آیا طرح روی جلد کتاب مناسب بود؟				
۱۷. چنانچه از وسایل کمک‌آموزشی از قبیل نوار، فیلم، لوح فشرده و... استفاده کرده‌اید، آیا به درک بهتر شما کمک کرده است؟				
۱۸. تا چه اندازه این کتاب شما را از حضور در کلاس بی‌نیاز کرد؟				

لطفاً چنانچه با اشکالهای تایپی یا محتوایی و مطالب تکراری مواجه شده‌اید، فهرستی از آنها را با ذکر شمارهٔ صفحه ضمیمه کنید.

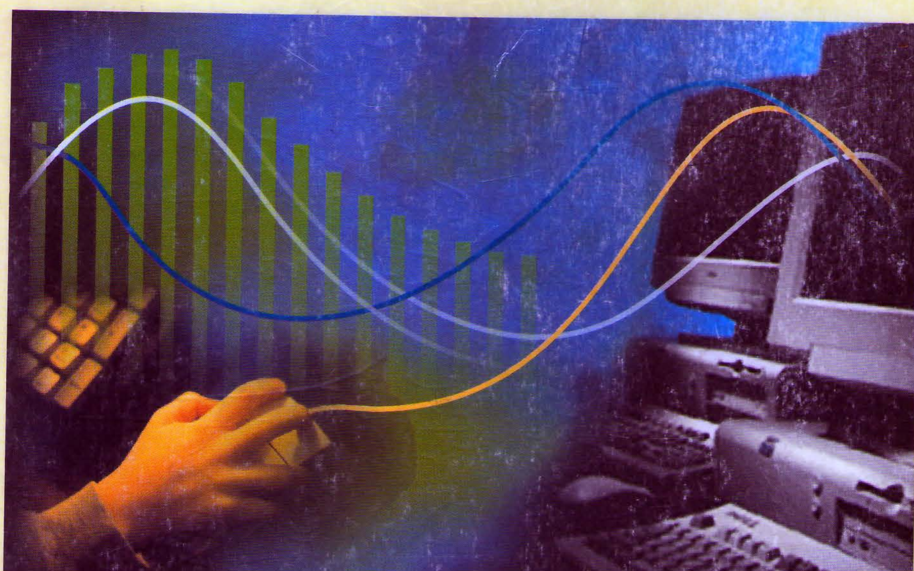
در مجموع کتاب را چگونه ارزیابی می‌کنید؟ عالی ☐ خوب ☐ متوسط ☐ ضعیف ☐

در صورت تمایل سایر پیشنهادهای خود را نیز بنویسید.

این پرسشنامه را پس از تکمیل از کتاب جدا کنید و به قسمت آموزش مرکز تحویل دهید یا مستقیماً به نشانی تهران ۱۹۵۶۹- صندوق پستی ۴۶۹۷-۱۹۳۹۵، مدیریت تدوین ارسال فرمایید.

با تشکر

مدیریت تدوین



دانشگاه پیام نور ۱۱۵۴

گروه آمار (۳۰/د)