توزیع‌های نمونه‌گیری

فردوس گرجی

جامعه‌ی آماری

در هر مطالعه‌ی آماری با مجموعه‌ای از افراد یا اشیاء که در یک یا چند صفت با یکدیگر مشترک هستند، سروکار داریم و هدف از مطالعه کسب اطلاعات درباره‌ی آن‌ها است. این مجموعه را جامعه‌ی آماری یا به اختصار جامعه می‌گویند.

متغیر تصادفی X نمایان‌گر یک جامعه است؛ به طوری که این متغیر تصادفی دارای توزیع احتمال $f_X(x)$ است.

در مطالعه یک جامعه، می‌خواهیم ویژگی‌های جامعه، رفتار، توزیع آماری و پارامترهای (میانگین، واریانس و ...) آن را بدانیم که همگی ثابت هستند ولی ما لزوما آن‌ها را نمی‌دانیم.

مثال: می‌خواهیم توزیع آماری جرم ذرات معلق ناخالصی در یک محلول تولید شده در پالایشگاه را بدانیم. جرم این ذرات توزیعی مانند $f(x)$ دارد که ثابت است ولی ما لزوما آن را نمی‌دانیم و می‌خواهیم تحقیق کنیم.

مثال: می‌خواهیم از میان افرادی که در یک کشور قهوه مصرف می‌کنند، نسبت کسانی که نوع خاصی از قهوه را ترجیح می‌دهند، به دست آوریم.

مثال: می‌خواهیم میانگین برد نوعی از موشک و واریانس آن را بدانیم (که از پارامترهای توزیع جامعه است). هر موشک بردی دارد که میانگین همه آن‌ها یک عدد ثابت است که در واقع پارامتر میانگین در جامعه برد موشک ها است.

میانگین بر موشک‌ها
درصد افرادی که نوع خاصی از قهوه را مصرف می‌کنند
دامنه تغییرات حقوق کارمندان بخش خصوصی
میانگین قد جوانان ایرانی در سال ۹۹
نسبت یک نوع ماده شیمیایی در یک محلول تولیدشده
توزیع آماری جرم ذرات معلق ناخالصی در یک محلول (توزیع آماری مجهول است)

پس یک جامعه آماری داریم با توزیع احتمال $f_X(x,\theta)$ که θ پارامترهای مربوط به آن، مثل میانگین، واریانس، دامنه، میانه، ... بوده و مقادیر ثابتی هستند؛ ما به دنبال یافتن و یا تخمین آن ها هستیم تا بتوانیم در حوزه کاربردی خود تصمیم‌گیری و برنامه‌ریزی و تحلیل انجام دهیم. دو راه داریم:

۱- همه داده‌ها و اعضای جامعه را بررسی کنیم.

- گاهی غیر ممکن (محاسبه توزیع طول عمر یک قطعه، محاسبه میانگین برد موشک‌ها)
- هزینه‌بر (اندازه‌گیری قد همه جوانان، نیروی کار، آموزش نیرو و فرهنگ‌سازی لازم دارد؛ تست کردن همه محلول‌های شیمیایی زمان زیادی می‌برد.)
- گاهی ممکن و حتی لازم (سرشماری‌های دوره‌ای، برخی از انواع کنترل کیفی محصولات که در آن همه قطعات تولیدشده بررسی می‌شوند.)

۲- بخشی از اعضای جامعه $f_X(x)$ را به عنوان نمونه بررسی کنیم و نتیجه را به کل جامعه تعمیم دهیم.

- یعنی $X_1, X_2, \ldots, X_n$ را از جامعه $f(x)$ را انتخاب کرده و مقادیر آن‌ها یعنی $x_1, x_2, \ldots, x_n$، را مشاهده می‌کنیم. سپس پارامتر مورد نظر در جامعه را به کمک این مشاهدات محاسبه می‌کنیم و مقدار به دست آمده را به کل جامعه نسبت می‌دهیم.

مثال: قد عده‌ای از جوانان را اندازه‌گیری کرده و میانگین و واریانس آن را محاسبه می‌کنیم و مقادیر به دست آمده را به میانگین و واریانس قد همه جوانان نسبت می‌دهیم.

نمونه تصادفی

نمونه‌گیری یکی از موضوعات بسیار مهم در تحلیل داده‌ها و تصمیم‌گیری‌ها در حوزه‌های مهندسی، مدیریت، جامعه‌شناسی، پزشکی و صنعت است.
هدف ما از انتخاب نمونه‌ی تصادفی دست‌یابی به اطلاعاتی درباره‌ی پارامترهای مجهول جامعه‌ی آماری است.
مثال: برای تصمیم‌گیری درباره نحوه آموزش ریاضی به دانش‌آموزان نیاز به مجموعه‌ای از اطلاعات راجع به آن‌ها داریم تا طبق آن برنامه‌ریزی کنیم. معمولا نمی‌توانیم همه دانش‌آموزان را بررسی کنیم، بنابراین بخشی از آن‌ها را به عنوان نمونه بررسی کنیم.
مثال: می‌خواهیم از میان افرادی که در یک کشور قهوه مصرف می‌کنند، نسبت کسانی که نوع خاصی از قهوه را ترجیح می‌دهند، به دست آوریم. غیرممکن است که هر آمریکایی که قهوه می‌نوشد را برای محاسبه‌ی پارامتر p مورد پرسش قرار دهیم. به جای آن نمونه‌ی تصادفی بزرگی انتخاب کرده و $\hat{p}$ نسبت مصرف‌کنندگان قهوه‌ی مورد نظر در این نمونه محاسبه می‌شود. حال برای استنباط درباره‌ی p از مقدار $\hat{p}$ استفاده می‌کنیم.
مثال: چند لامپ به عنوان نمونه انتخاب می‌کنیم. طول عمر آن‌ها را حساب کرده و توزیع احتمال آن‌ها ($\hat{f}$) را پیدا می کنیم. مثلا آیا نرمال است؟ آیا نمایی است؟ سپس نتیجه به دست آمده را به توزیع آماری کل لامپ‌های تولیدشده (f) نسبت می‌دهیم.

نمونه تصادفی

کدام بخش از داده‌ها را به عنوان نمونه در نظر بگیریم؟
- آن بخش که راحت‌تر است وبیشتر در دسترس می‌باشد؟! آن بخش که مطابق سلیقه شخصی ماست؟!
- مقدار واقعی پارامتر را به دست نمی‌دهد، به برآوردهای کمتر یا بیشتر از مقدار واقعی منجر می‌شود که اصطلاحا می‌گوییم اریب است.

- آن بخش که کاملا تصادفی و بدون پیش‌فرض و دخالت شخصی انتخاب می‌شود؟ $\surd$
- به تعداد مناسب (n) متغیر تصادفی از جامعه $f(x)$ را که از هم مستقل هستند انتخاب می کنیم.

این متغیر ها X_1، ...، X_n هستند که هر کدام دارای توزیع $f(x)$ می‌باشند. X_i درواقع iامین اندازه‌گیری یا iامین مقدار نمونه ای (مثلا طول عمر iامین لامپ انتخاب شده، یا حقوق iامین کارمند انتخاب شده) را نشان می‌دهد که مقدار عددی x_i را اختیار می‌کند. توزیع توام آن‌ها چیست؟

تعاریف

جامعه: جامعه مشاهداتی است که با آن‌ها سر و کار داریم.

نمونه: نمونه یک زیرمجموعه از جامعه است.

نمونه تصادفی: فرض کنید $X_1, X_2 \ldots, X_n$ متغیرهای تصادفی مستقل با توزیع احتمال یکسان $f_X(x)$ باشند. (این تابع به پارامتر مجهوا θ بستگی دارد.) در این صورت $X_1, X_2 \ldots, X_n$ را نمونه تصادفی از اندازه n از جامعه $f_X(x)$ گوییم که توزیع احتمال توام آن به صورت زیر است:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$$

آماره‌ها

برای تخمین زدن پارامتر مورد نظر جامعه، آن پارامتر را در نمونه محاسبه می کنیم. هر ویژگی یک جامعه را پارامتر و ویژگی متناظر آن در نمونه را آماره گویند. یک آماره تابعی از نمونه تصادفی است که به پارامتر مجهول بستگی ندارد.

مثلا برای تخمین میانگین طول عمر همه لامپ‌ها، μ، میانگین طول عمر لامپ‌ها را در نمونه اندازه گیری می کنیم. برای این کار مقادیر طول عمرها را جمع زده و بر تعداد نمونه تقسیم می کنیم. $\frac{X_۱+\ldots+X_n}{n}$

یا دامنه حقوق کارمندان نمونه ای را حساب می کنیم. برای این کار بیشترین حقوق را منهای کمترین حقوق می‌کنیم. $\max\{X_i\}-\min\{X_i\}$

این‌ها در واقع توابعی از نمونه تصادفی هستند که پارامتر مجهول در آن‌ها وجود ندارد. به این توابع آماره (Statistic) گویند.

تعریف

آماره: به هر تابعی مانند $U=g(X_۱,X_۲,\ldots,X_n)$ از متغیرهای تصادفی حاصل از یک نمونه تصادفی یک آماره گویند.

نکته: مقدار به دست آماده برای یک آمار در نمونه تصادفی، از نمونه‌ای به نمونه دیگر ممکن است تغییر کند، ولی پارامتر مربوط به جامعه مقدارش ثابت است.

نکته: آماره تابعی از نمونه تصادفی بوده و خود نیز یک متغیر تصادفی است. توزیع احتمال آماره را **توزیع نمونه‌ای** گویند.

آماره های گرایش به مرکز

میانگین نمونه: اگر $X_n, \dots, X_1$ یک نمونه تصادفی از اندازه n باشد، آنگاه میانگین نمونه‌ای به صورت زیر تعریف می‌شود:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

میانه نمونه: اگر $X_n, \dots, X_1$ نمونه تصادفی از اندازه n باشد که به ترتیب بزرگی به صورت $\underbrace{X_{(1)}, X_{(2)}, \dots, X_{(n)}}_{\text{آماره‌های مرتب یا ترتیبی}}$ مرتب شده‌اند، آنگاه میانه نمونه با آماره زیر تعریف می‌شود:

$$\tilde{X} = \begin{cases} X_{(\frac{n+1}{2})} & \text{; اگر n فرد باشد} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} & \text{; اگر n زوج باشد} \end{cases}$$

مد نمونه: اگر $X_n, \dots, X_1$ نمونه تصادفی از اندازه n باشد (که ممکن است لزوما متفاوت از هم نباشند)، مد نمونه، M، مقداری از نمونه تصادفی است که بیش از همه واقع می‌شود یا بیشترین فراوانی را دارد. مد ممکن است وجود نداشته باشد و یا در صورت وجود، می‌تواند منحصر به فرد نباشد.

مثال ۱

تعداد زندگی‌ها در هر متر مربع از یک نوع پارچه تولید شده، برای ۱۵ قطعه یک مترمربعی که به تصادف از یک توپ پارچه انتخاب شده‌اند به صورت زیر به دست آمده است. میانگین، میانه و مد نمونه گرفته شده را حساب کنید. ۲ ,۴ ,۲ ,۲ ,۷ ,۶ ,۱ ,۳ ,۳ ,۲ ,۳ ,۵ ,۱ ,۲ ,۱

میانگین نمونه عبارت است از:

$$\bar{X} = \frac{۱+۲+۱+۵+۳+۲+۳+۳+۱+۶+۷+۲+۲+۴+۲}{۱۵} = ۲/۹۳$$

برای میانه ابتدا داده‌ها را مرتب می کنیم:

۷ ,۶ ,۵ ,۴ ,۳ ,۳ ,۳ ,۲ ,۲ ,۲ ,۲ ,۲ ,۱ ,۱ ,۱

میانه نمونه در نمونه تصادفی ۱۵ تایی برابر با داده وسطی، یعنی داده هشتم است که می‌شود: ۲

مد نمونه داده‌ای است که بیشترین تکرار را داشته باشد که در این نمونه مقدار مد برابر با ۲ است.

نکته

تغییرات میانگین از نمونه‌ای به نمونه دیگر نسبتا کم است. در حالی که میانه از نمونه‌ای به نمونه دیگر بیشتر تغییر می‌کند. میانگین تحت تاثیر اعداد خیلی بزرگ یا خیلی کوچک قرار می‌گیر در حالی که میانه کمتر تحت تاثیر این اعداد قرار دارد. مد در نمونه های کوچک در صورت وجود هم عملا معنی ندارد، در عوض اولا نیاز به محاسبه ندارد، ثانیا هم برای داده‌های کیفی و هم داده‌های کمی قابل استفاده است.

آماره‌های پراکندگی

دامنه نمونه: دامنه نمونه تصادفی $X_۱, \ldots, X_n$ برابر با آماره زیر است:

$$\max_{۱\leq i\leq n} X_i - \min_{۱\leq i\leq n} X_i = X_{(n)} - X_{(۱)}$$

واریانس نمونه: اگر $X_۱, \ldots, X_n$ یک نمونه تصادفی از اندازه n باشد، واریانس نمونه عبارت است از آماره

$$S^۲ = \frac{\sum_{i=۱}^{n}(X_i - \bar{X})^۲}{n-۱}$$

انحراف معیار نمونه: اگر $X_۱, \ldots, X_n$ نمونه‌ای تصادفی از اندازه n باشد، انحراف معیار نمونه عبارت است از $S = \sqrt{S^۲}$

قضیه ۱

اگر $S^۲$ واریانس یک نمونه تصادفی از اندازه n باشد، می‌توان نوشت:

$$S^۲ = \frac{n\sum_{i=۱}^{n} X_i^۲ - (\sum_{i=۱}^{n} X_i)^۲}{n(n-۱)}$$

مثال ۲:

در اندازه‌گیری حجم دو نمونه تصادفی از آب پرتغال‌های تولید شده در دو شرکت الف و ب، اعداد زیر به دست آمده‌اند. دامنه و واریانس و انحراف معیار دو نمونه را محاسبه کرده و دو شرکت را با یکدیگر مقایسه کنید.

نمونه الف	۰/۹۷	۱/۰۰	۰/۹۴	۱/۰۳	۱/۱۱
نمونه ب	۱/۰۶	۱/۰۱	۰/۸۸	۰/۹۱	۱/۱۴

میانگین هر دو نمونه برابر یا ۱/۰۰ است. دامنه نمونه الف برابر است با:

$$۱/۱۱ - ۰/۹۴ = ۰/۱۷$$

و دامنه نمونه ب برابر است با:

$$۱/۱۴ - ۰/۸۸ = ۰/۲۶$$

دامنه نمونه الف کمتر است که نشان می‌دهد پراکندگی داده‌ها در آن کمتر است. یعنی اگر از شرکت ب خرید کنیم، اطمینان بیشتری داریم که حجم آب پرتغال به میانگین اعلام شده نزدیک تر باشد.
واریانس و انحراف معیار شرکت الف به ترتیب ۰/۰۰۳۵ و ۰/۰۶ است و واریانس و انحراف معیار شرکت ب به ترتیب ۰/۰۰۹۲ و ۰/۱ است که نتیجه‌گیری قبلی را تایید می‌کند.

نکته

محاسبه دامنه نمونه بسیار راحت است ولی این معیار در نمونه های بزرگ کارایی ندارد و فقط کمترین داده و بیشترین داده بررسی می‌شوند و داده‌های میانی دیده نمی‌شوند.

تعریف

توزیع نمونه‌ای: آماره تابعی از نمونه تصادفی بوده و خود نیز یک متغیر تصادفی است. توزیع احتمال آماره را توزیع نمونه‌ای گویند.

نمادها

- ○ μ: میانگین جامعه
- ● $\bar{X}$: میانگین نمونه

- ○ $\sigma^۲$: واریانس جامعه
- ● $S^۲$: واریانس نمونه

- ●○ n: حجم نمونه

توزیع نمونه‌ای میانگین نمونه

قضیه ۲

اگر $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی مستقل باشند و $X_i \sim N(\mu_i, \sigma_i^2), i = 1, \ldots, n$ و قرار دهیم

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

آنگاه

$$Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

توزیع نمونه‌ای میانگین نمونه $\bar{X}$

فرض کنید از جامعه‌ای با میانگین μ و واریانس σ^2 نمونه تصادفی $X_1, X_2, \ldots, X_n$ به اندازه n انتخاب کرده باشیم. به علت مستقل و هم‌توزیع بودن $X_1, X_2, \ldots, X_n$ داریم:

$$E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$$
$$Var(X_1) = Var(X_2) = \cdots = Var(X_n) = \sigma^2$$

می‌خواهیم توزیع نمونه‌ای $\bar{X} = \frac{1}{n}\sum X_i$ میانگین نمونه را به دست آوریم:

$$E\left(\bar{X}\right) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}E\left(X_1 + \ldots + X_n\right) = \frac{1}{n} \times n\mu = \mu$$

$$Var\left(\bar{X}\right) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2}Var\left(X_1 + \ldots + X_n\right)$$
$$= \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$$

حال می‌خواهیم بررسی کنیم که متغیر تصادفی $\bar{X}$ از چه تابع چگالی تبعیت می‌کند.
دو حالت را در نظر می‌گیریم (۱- جامعه با توزیع نرمال و ۲- جامعه با توزیع غیر نرمال):

۱- اگر جامعه دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد:
چون $\bar{X}$ ترکیب خطی از متغیرهای تصادفی نرمال مستقل است، طبق قضیه ۲ داریم:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Longrightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

نکته

با افزایش حجم نمونه واریانس $\bar{X}$ کاهش می‌یابد.

توزیع نمونه‌ای میانگین نمونه $\bar{X}$

۲- اگر جامعه دارای توزیع نرمال نباشد:
اگر چه توزیع $\bar{X}$ به توزیع جامعه نمونه‌گیری شده وابسته است، ولی طبق قضیه‌ی حد مرکزی با افزایش n توزیع نمونه‌ای $\bar{X}$ به توزیع نرمال نزدیک می‌شود.

قضیه حد مرکزی (۳)

اگر $\bar{X}$ میانگین نمونه‌ای تصادفی با اندازه n انتخاب شده از جامعه‌ای با میانگین μ و واریانس $\sigma^۲$ باشد، آن‌گاه شکل حدی توزیع

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

وقتی $n \to \infty$، توزیع نرمال استاندارد $n(z; ۰, ۱)$ است.

بنابراین طبق قضیه حد مرکزی وقتی اندازه نمونه n افزایش یابد، توزیع میانگین نمونه $\bar{X}$ یک نمونه تصادفی که از هر جامعه‌ای گرفته شده باشد، توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^۲}{n}$ است.

نکته

تقریب نرمال برای توزیع نمونه‌ای $\bar{X}$ معمولاً زمانی که $n \geq ۳۰$ باشد یک تقریب مناسب است.

مثال ۳

یک شرکت تولیدی لاستیک اتومبیل، لاستیک‌هایی تولید می‌کند که طول عمر این لاستیک‌ها دارای توزیع نرمال با میانگین ۲۴ ماه و انحراف معیار ۲ ماه است. احتمال اینکه در یک نمونه‌ی ۲۵ تایی از لاستیک‌ها، میانگین طول عمر کمتر از ۲۵ ماه باشد، چه‌قدر است؟

جامعه نرمال است، پس داریم:

$$\bar{X} \sim N\left(\mu = ۲۴ \ , \ \frac{\sigma^۲}{n} = \frac{۲^۲}{۲۵}\right)$$

$$P(\bar{X} < ۲۵) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{۲۵-۲۴}{\frac{۲}{۵}}\right) = P(Z < ۲/۵) = ۰/۹۹۳۸$$

مثال ۴

یک آسانسور طوری طراحی شده که حد ظرفیت بار آن ۵۰۰۰ کیلوگرم باشد. ادعا می‌شود که این آسانسور گنجایش ۵۰ نفر را دارد. اگر وزن تمام کسانی که از این آسانسور استفاده می‌کنند دارای میانگین ۹۵ کیلوگرم و انحراف معیار ۱۲ کیلوگرم باشد، احتمال اینکه وزن یک گروه تصادفی ۵۰ نفری از حد ظرفیت آسانسور تجاوز کند چه‌قدر است؟

چون حجم نمونه بیشتر از ۳۰ است، پس طبق قضیه‌ی حد مرکزی داریم:

$$\bar{X} \sim N\left(\mu = ۹۵ \;,\; \frac{\sigma^۲}{n} = \frac{۱۲^۲}{۵۰}\right)$$

$$P\left(\sum_{i=۱}^{۵۰} X_i > ۵۰۰۰\right) = P(\bar{X} > ۱۰۰) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{۱۰۰ - ۹۵}{\frac{۱۲}{\sqrt{۵۰}}}\right)$$

$$= P(Z > ۲/۹۵) = ۱ - P(Z < ۲/۹۵) = ۱ - ۰/۹۹۸۴ = ۰/۰۰۱۶$$

مثال ۵

عرض یک شکاف که بر یک قطعه از آلیاژ آلومینیوم که با ریخته‌گری تولید می‌شود، توزیع نرمال با میانگین ۰/۹ و انحراف معیار ۰/۰۲ است. حدود مشخصات طراحی عبارتند از $۰/۹ \pm a$ اینچ. هر ساعت نمونه‌هایی ۵ تایی از آلیاژ ریخته‌گری گرفته شده و میانگین آن محاسبه می‌شود. حدود را طوری تعیین کنید که درصد میانگین‌های نمونه که خارج از حدود قرار می‌گیرند، معادل ۰/۲۷ درصد باشد.

راه‌حل:

$$\bar{X} \sim N\left(\mu = ۰/۹ \ , \ \frac{\sigma^۲}{n} = \frac{(۰/۰۲)^۲}{۵}\right)$$

$$P(۰/۹ - a < \bar{X} < ۰/۹ + a) = ۱ - ۰/۰۰۲۷ = ۰/۹۹۷۳$$

$$۰/۹۹۷۳ = P\left(\frac{-a}{\frac{۰/۰۲}{\sqrt{۵}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{a}{\frac{۰/۰۲}{\sqrt{۵}}}\right) = P(-۱۱۱/۸a < Z < ۱۱۱/۸a)$$

$$= ۲P(Z < ۱۱۱/۸a) - ۱$$

$$\Rightarrow P(Z < ۱۱۱/۸a) = ۰/۹۹۸۶۵ \quad \Rightarrow \quad ۱۱۱/۸a = ۲/۹۹۵ \quad \Rightarrow \quad a = ۰/۰۲۷$$

توزیع نمونه‌ای واریانس نمونه

نکته

فرض کنید Z متغیری تصادفی با توزیع نرمال استاندارد $Z \sim N(0, 1)$ باشد و قرار دهیم $Y = Z^2$.
در این صورت تابع توزیع Y عبارتست از:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(Z^2 \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}} f_Z(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_Z(-\sqrt{y}) = \frac{1}{\sqrt{y}} f_Z(\sqrt{y}) \\ &= \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} = \frac{1}{2^{\frac{1}{2}} \Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y} \qquad y > 0 \end{aligned}$$

بنابراین $Y = Z^2 \sim \chi^2_{(1)}$.

قضیه ۴

اگر $X_1, \ldots, X_n$ متغیرهای تصادفی مستقل با توزیع خی-۲ (χ^2) به ترتیب با $\nu_n, \ldots, \nu_1$ درجه آزادی باشند، آنگاه متغیر تصادفی $Y = X_1 + \cdots + X_n$ دارای توزیع خی-۲ با $\nu = \nu_1 + \cdots + \nu_n$ درجه آزادی است.

نتیجه: اگر $Z_1, Z_2, \ldots, Z_n$ یک نمونه تصادفی n تایی از جامعه نرمال استاندارد باشند آنگاه $Z_1^2 + Z_2^2 + \ldots + Z_n^2$ دارای توزیع خی-دو با n درجه آزادی، $\chi^2_{(n)}$، است.

یک معیار پراکندگی مناسب واریانس نمونه است: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

این معیار را زمانی به کار می‌بریم که میانگین جامعه یعنی μ شناخته شده نباشد.

دلیل انتخاب آن عبارت است از $E(S^2) = \sigma^2$

اثبات

$$
\begin{aligned}
E(S^۲) &= E\left[\frac{۱}{n-۱}\sum(X_i-\bar{X})^۲\right] = \frac{۱}{n-۱}E\left[\sum(X_i-\bar{X})^۲\right]\\
&= \frac{۱}{n-۱}E\left[\sum_i\left[(X_i-\mu)-(\bar{X}-\mu)\right]^۲\right]\\
&= \frac{۱}{n-۱}E\left[\sum_i\left[(X_i-\mu)^۲+(\bar{X}-\mu)^۲-۲(X_i-\mu)(\bar{X}-\mu)\right]\right]\\
&= \frac{۱}{n-۱}E\left[\sum_i(X_i-\mu)^۲+n(\bar{X}-\mu)^۲-۲(\bar{X}-\mu)\underbrace{\sum_i(X_i-\mu)}_{\sum X_i-\sum\mu=n\bar{X}-n\mu}\right]\\
&= \frac{۱}{n-۱}E\left[\sum_i(X_i-\mu)^۲-n(\bar{X}-\mu)^۲\right]\\
&= \frac{۱}{n-۱}\left[\sum_i E(X_i-\mu)^۲-nE(\bar{X}-\mu)^۲\right]\\
&= \frac{۱}{n-۱}\left[\sum_i\sigma^۲-nVar(\bar{X})\right]\\
&= \frac{۱}{n-۱}\left[n\sigma^۲-n\times\frac{\sigma^۲}{n}\right]=\sigma^۲
\end{aligned}
$$

توزیع نمونه‌ای واریانس نمونه $S^۲$

اگر $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از جامعه نرمال با میانگین μ و واریانس $\sigma^۲$ باشند، آنگاه

$$\frac{(n-۱)S^۲}{\sigma^۲} \sim \chi^۲_{(n-۱)}$$

اثبات:

$$\frac{\sum_i (X_i-\mu)^۲}{\sigma^۲} = \sum_i \left(\frac{X_i-\mu}{\sigma}\right)^۲ = \sum_i Z_i^۲ \sim \chi^۲_{(n)}$$

$$\begin{aligned}
\frac{\sum_i (X_i-\mu)^۲}{\sigma^۲} &= \frac{\sum_i \left[(X_i-\bar{X})+(\bar{X}-\mu)\right]^۲}{\sigma^۲} \\
&= \frac{\sum_i (X_i-\bar{X})^۲}{\sigma^۲} + \frac{n(\bar{X}-\mu)^۲}{\sigma^۲} \\
&= \frac{(n-۱)S^۲}{\sigma^۲} + \underbrace{\frac{n(\bar{X}-\mu)^۲}{\sigma^۲}}_{Z^۲\sim\chi^۲_{(۱)}}
\end{aligned}$$

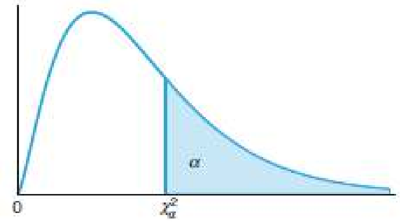


Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.80**	**0.75**	**0.70**	**0.50**
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337

Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619

مثال ۶

یک جامعه‌ی نرمال واریانس ۶ دارد. اگر نمونه‌ی تصادفی ۲۵ تایی از این جامعه انتخاب شود، احتمال این‌که واریانس نمونه بین ۳/۴۵ و ۱۰/۷۵ باشد، چه‌قدر است؟

راه‌حل:

$$\begin{aligned} P(۳/۴۵ < S^۲ < ۱۰/۷۵) &= P\left(\frac{۲۴ \times ۳/۴۵}{۶} < \frac{(n-۱)S^۲}{\sigma^۲} < \frac{۲۴ \times ۱۰/۷۵}{۶}\right) \\ &= P\left(۱۳/۸ < \chi^۲_{(۲۴)} < ۴۳\right) \\ &= P\left(\chi^۲_{(۲۴)} < ۴۳\right) - P\left(\chi^۲_{(۲۴)} \leq ۱۳/۸\right) \\ &= ۰/۹۹ - ۰/۰۵ = ۰/۹۴ \end{aligned}$$

مثال ۷

طول عمر لامپ‌های تصویر تلویزیون ساخت کارخانه‌ای دارای توزیع نرمال با میانگین ۲۰۰۰ ساعت و انحراف معیار ۶۰ ساعت است. اگر ۱۰ لامپ تصویر تلویزیون ساخت این کارخانه به طور تصادفی انتخاب شود، احتمال این‌که انحراف استاندارد این ۱۰ لامپ بیش از ۵۰ ساعت نباشد، چه‌قدر است؟

راه‌حل:

$$\begin{aligned} P(S \leq ۵۰) &= P(S^۲ \leq ۲۵۰۰) \\ &= P\left(\frac{(n-۱)S^۲}{\sigma^۲} \leq \frac{۹ \times ۲۵۰۰}{۳۶۰۰}\right) \\ &= P\left(\chi^۲_{(۹)} \leq ۶/۲۵\right) \\ &\simeq ۰/۳ \end{aligned}$$

توزیع نمونه‌ای $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$

اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از جامعه نرمال با میانگین μ و واریانس σ^2 باشند، آنگاه $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ دارای توزیع نرمال استاندارد است.

حال اگر σ^2 مجهول باشد، به جای آن می‌توان از واریانس نمونه S^2 استفاده کرد.

اکنون اگر در Z به جای σ مقدار S را قرار دهیم، آنگاه $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ دارای توزیع t است.

تعریف توزیع t-استیودنت:
اگر $Z \sim N(0, 1)$ و $Y \sim \chi^2_{(n)}$ و Z و Y از یکدیگر مستقل باشند، آنگاه متغیر تصادفی $T = \frac{Z}{\sqrt{Y/n}}$ دارای توزیع t با n درجه آزادی $T \sim t_{(n)}$ است.

قضیه ۷

اگر $\bar{X}$ و $S^{۲}$ به ترتیب میانگین و واریانس یک نمونه‌ی تصادفی به اندازه n از یک جامعه نرمال با میانگین μ و واریانس $\sigma^{۲}$ باشند، آنگاه

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-۱)}$$

اثبات:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(۰, ۱) \qquad \perp \qquad Y = \frac{(n-۱)S^{۲}}{\sigma^{۲}} \sim \chi^{۲}_{(n-۱)}$$

$$T = \frac{Z}{\sqrt{\frac{Y}{n-۱}}} \sim t_{(n-۱)}$$

$$T = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-۱)S^{۲}}{\sigma^{۲}(n-۱)}}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-۱)}$$

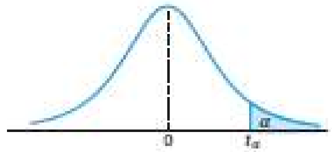
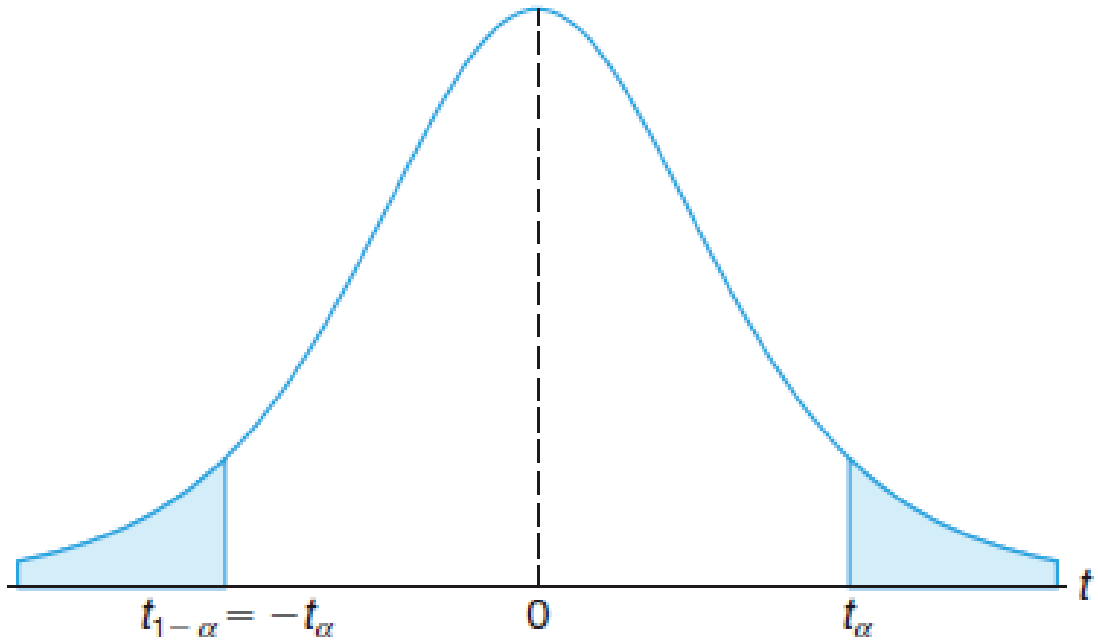


Table A.4 Critical Values of the t-Distribution

v	α = 0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060



نکته

برای $n \geq ۳۰$ توزیع t تقریباً با توزیع نرمال استاندارد برابر می‌شود.
به همین علت در جدول t مقادیر درجه‌ی آزادی بزرگ‌تر از ۳۰ با ∞ نشان داده شده است و مقادیر این ردیف از جدول با جدول توزیع نرمال استاندارد یکی است.

مثال ۱۱

نمره‌های یک کلاس از دانشجویان دارای توزیع نرمال با میانگین ۱۵ است. اگر از این کلاس یک نمونه‌ی تصادفی ۲۰ تایی انتخاب کنیم و مشاهده کنیم که انحراف استاندارد نمره‌های آن‌ها ۴/۲۸ است، احتمال این‌که میانگین نمره‌های این افراد از ۱۷ بیشتر باشد، چه‌قدر است؟

راه‌حل:

$$P\left(\bar{X} > ۱۷\right) = P\left(\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{۱۷-۱۵}{\frac{۴/۲۸}{\sqrt{۲۰}}}\right) = P\left(T_{(۱۹)} > ۲/۰۹\right)$$

$$= ۱ - P\left(T_{(۱۹)} \leq ۲/۰۹\right) = ۱ - ۰/۹۷۵ = ۰/۰۲۵$$

توزیع نمونه‌ای اختلاف میانگین‌ها

فرض کنید دو جامعه داشته باشیم.
جامعه‌ی اول دارای میانگین μ_1 و واریانس σ_1^2 باشد.
جامعه‌ی دوم دارای میانگین μ_2 و واریانس σ_2^2 باشد.
یک نمونه‌ی تصادفی n تایی $X_1, \ldots, X_n$ از جامعه‌ی اول انتخاب کرده و میانگین این نمونه را با $\bar{X}$ و واریانس آن را با S_1^2 نمایش می‌دهیم.
یک نمونه‌ی تصادفی m تایی $Y_1, \ldots, Y_m$ از جامعه‌ی دوم انتخاب کرده و میانگین این نمونه را با $\bar{Y}$ و واریانس آن را با S_2^2 نشان می‌دهیم.
فرض کنید نمونه‌گیری از دو جامعه مستقل از یکدیگر باشد.
می‌خواهیم توزیع نمونه‌ای $\bar{X} - \bar{Y}$ را پیدا کنیم.

حالت اول: واریانس دو جامعه σ_1^2 و σ_2^2 معلوم باشد

الف- اگر دو جامعه نرمال باشند، با توجه به این‌که $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n})$ و $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m})$ بوده و از مستقل هستند، پس

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) \Longrightarrow Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

ب- اگر دو جامعه نرمال نباشند، طبق قضیه‌ی حد مرکزی برای حجم نمونه‌ی $n \geq 30$ و $m \geq 30$ از تقریب نرمال استفاده می‌شود (شبیه حالت الف).

مثال ۱۲

دو کارخانه‌ی تولید کابل A و B وجود دارند. کابل‌هایی که کارخانه‌ی A تولید می‌کند، به طور متوسط تحمل ۴۰۰۰ پوند نیروی کششی و انحراف معیار ۳۰۰ پوند را دارند. کابل‌هایی که کارخانه‌ی B تولید می‌کند، به طور متوسط تحمل ۴۵۰۰ پوند نیرو با انحراف معیار ۲۰۰ پوند را دارند. اگر ۱۰۰ کابل نوع A و ۵۰ کابل نوع B آزمایش شوند، احتمال این‌که متوسط تحمل نیروی کششی B حداقل ۶۰۰ پوند بیش از نیروی کششی A باشد، چه‌قدر است؟

راه‌حل:

$$\bar{X}_A - \bar{X}_B \sim N\left(۴۰۰۰ - ۴۵۰۰\ ,\ \frac{۳۰۰^۲}{۱۰۰} + \frac{۲۰۰^۲}{۵۰}\right)$$

$$\Rightarrow \quad \bar{X}_A - \bar{X}_B \sim N\left(-۵۰۰\ ,\ ۱۷۰۰\right)$$

$$\begin{aligned} P\left(\bar{X}_B \geq \bar{X}_A + ۶۰۰\right) &= P\left(\bar{X}_A - \bar{X}_B \leq -۶۰۰\right) \\ &= P\left(\frac{(\bar{X}_A - \bar{X}_B) - (\mu_۱ - \mu_۲)}{\sqrt{\frac{\sigma_۱^۲}{n} + \frac{\sigma_۲^۲}{m}}} \leq \frac{-۶۰۰ + ۵۰۰}{\sqrt{۱۷۰۰}}\right) \\ &= P\left(Z \leq -۲/۴۳\right) = ۰/۰۰۷۵ \end{aligned}$$

مثال ۱۳

فرض کنید در دو جامعه میانگین مصرف روزانه‌ی پروتئین به ترتیب ۱۲۵ و ۱۰۰ گرم باشد. اگر مقادیر مصرف روزانه‌ی پروتئین در دو جامعه دارای توزیع نرمال با انحراف معیار ۱۵ گرم باشد، احتمال این‌که نمونه‌های تصادفی و مستقل ۲۵ نفری از هر جامعه، تفاوت بین میانگین‌هایشان کمتر از ۱۲ گرم باشد را بیابید.

راه‌حل:

$$\bar{X}-\bar{Y}\sim N\left(۱۲۵-۱۰۰\ ,\ \frac{۱۵^{۲}}{۲۵}+\frac{۱۵^{۲}}{۲۵}\right)\quad\Rightarrow\quad \bar{X}-\bar{Y}\sim N\left(۲۵\ ,\ ۱۸\right)$$

$$\begin{aligned}P\left(|\bar{X}-\bar{Y}|<۱۲\right)&=P\left(-۱۲<\bar{X}-\bar{Y}<۱۲\right)\\&=P\left(\frac{-۱۲-۲۵}{\sqrt{۱۸}}<\frac{\left(\bar{X}-\bar{Y}\right)-(\mu_{۱}-\mu_{۲})}{\sqrt{\frac{\sigma_{۱}^{۲}}{n}+\frac{\sigma_{۲}^{۲}}{m}}}<\frac{۱۲-۲۵}{\sqrt{۱۸}}\right)\\&=P\left(-۸/۷۲<Z<-۳/۰۶\right)=۰/۰۰۱۱-۰=۰/۰۰۱۱\end{aligned}$$

مثال ۱۴

فرض کنید $\bar{X}$ و $\bar{Y}$ میانگین‌های دو نمونه‌ی مستقل به اندازه‌ی n از جامعه‌ای نرمال با واریانس $\sigma^۲$ باشد. مقدار n را چنان تعیین کنید تا احتمال این‌که میانگین این دو نمونه بیشتر از σ اختلاف داشته باشند، تقریباً برابر ۰/۰۱ باشد.

داریم:

$$\bar{X}-\bar{Y}\sim N\left(\mu-\mu\,,\,\frac{\sigma^۲}{n}+\frac{\sigma^۲}{n}\right)\quad\Rightarrow\quad \bar{X}-\bar{Y}\sim N\left(۰\,,\,\frac{۲\sigma^۲}{n}\right)$$

$$\begin{aligned}۰/۰۱=P\left(|\bar{X}-\bar{Y}|>\sigma\right)&=۱-P\left(-\sigma\leq\bar{X}-\bar{Y}\leq\sigma\right)\\&=۱-P\left(\frac{-\sigma-۰}{\sigma\sqrt{\frac{۲}{n}}}<\frac{(\bar{X}-\bar{Y})-(\mu-\mu)}{\sqrt{\frac{\sigma^۲}{n}+\frac{\sigma^۲}{n}}}<\frac{\sigma-۰}{\sigma\sqrt{\frac{۲}{n}}}\right)\\&=۱-P\left(-\sqrt{\frac{n}{۲}}\leq Z\leq\sqrt{\frac{n}{۲}}\right)=۲-۲P\left(Z\leq\sqrt{\frac{n}{۲}}\right)\end{aligned}$$

$$\Rightarrow\quad P\left(Z\leq\sqrt{\frac{n}{۲}}\right)=۰/۹۹۵\quad\Rightarrow\quad\sqrt{\frac{n}{۲}}=۲/۵۷۵\quad\Rightarrow\quad n=۱۳/۲۶\simeq ۱۳$$

حالت دوم: واریانس دو جامعه σ_1^2 و σ_2^2 نامعلوم اما مساوی باشد

در این حالت واریانس دو جامعه یعنی σ_1^2 و σ_2^2 در رابطه‌ی $\sigma_1^2 = \sigma_2^2 = \sigma^2$ صدق می‌کنند که σ^2 واریانس مشترک دو جامعه و مقداری نامعلوم است.

در جامعه‌ی اول می‌توان از S_1^2 و در جامعه‌ی دوم می‌توان از S_2^2 به عنوان یک براورد برای σ^2 استفاده کرد.

اما بهتر است که از اطلاعات دو نمونه برای براورد σ^2 استفاده کنیم. بدین منظور از میانگین وزنی S_1^2 و S_2^2 استفاده می‌کنیم:

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

اگر دو جامعه نرمال باشند، آن‌گاه

$$Y = \frac{(n+m-2)S_p^2}{\sigma^2} = \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2_{(n+m-2)}$$

با استفاده از تعریف توزیع t داریم:

$$T = \frac{Z}{\sqrt{\frac{Y}{n+m-۲}}} \sim t_{(n+m-۲)}$$

$$T = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_۱-\mu_۲)}{\sqrt{\frac{\sigma^۲}{n}+\frac{\sigma^۲}{m}}}}{\sqrt{\frac{\frac{(n+m-۲)S_p^۲}{\sigma^۲}}{n+m-۲}}} = \frac{(\bar{X}-\bar{Y})-(\mu_۱-\mu_۲)}{S_p\sqrt{\frac{۱}{n}+\frac{۱}{m}}}$$

$$\Rightarrow \qquad T = \frac{\left(\bar{X}-\bar{Y}\right)-(\mu_۱-\mu_۲)}{S_p\sqrt{\frac{۱}{n}+\frac{۱}{m}}} \sim t_{(n+m-۲)}$$

مثال ۱۵

میانگین نمره‌ی هوش دانشجویان سال اول و دوم یک دانشگاه به ترتیب ۹۱ و ۸۵ است. در یک نمونه‌گیری از ۹ دانشجوی سال اول و ۱۰ دانشجوی سال دوم، انحراف استاندارد نمره‌ی هوش به ترتیب ۳ و ۴ به دست آمده است. با فرض نرمال بودن دو جامعه و برابری واریانس‌های آن‌ها، احتمال این‌که میانگین هوشی دانشجویان سال اول در نمونه حداقل ۱۰/۷۵ نمره بیشتر از میانگین هوشی دانشجویان سال دوم در نمونه باشد، چه‌قدر است؟

داریم:

$$S_p^{۲} = \frac{(n-۱)S_۱^{۲} + (m-۱)S_۲^{۲}}{n+m-۲} = \frac{(۸\times۹)+(۹\times۱۶)}{۹+۱۰-۲} = ۱۲/۷$$

$$P\left(\bar{X}_۱ \geq \bar{X}_۲ + ۱۰/۷۵\right) = P\left(\bar{X}_۱ - \bar{X}_۲ \geq ۱۰/۷۵\right)$$

$$= P\left(\frac{(\bar{X}_۱ - \bar{X}_۲) - (\mu_۱ - \mu_۲)}{S_p\sqrt{\frac{۱}{n}+\frac{۱}{m}}} \geq \frac{۱۰/۷۵ - ۶}{\sqrt{۱۲/۷}\sqrt{\frac{۱}{۹}+\frac{۱}{۱۰}}}\right)$$

$$= P\left(T_{(۱۷)} \geq ۲/۹\right) = ۱ - P\left(T_{(۱۷)} < ۲/۹\right)$$

$$= ۱ - ۰/۹۹۵ = ۰/۰۰۵$$