توزیع‌های نمونه‌ای
- آمار و احتمالات مهندسی -

مدرس: مشکانی فراهانی

دانشگاه صنعتی امیرکبیر

۹ آذر ۱۳۹۹

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.

نمونه: نمونه زیر مجموعه‌ای از جامعه است.

هدف ما از انتخاب نمونه‌ی تصادفی دست‌یابی به اطلاعاتی درباره‌ی پارامترهای مجهول جامعه‌ی آماری است.

برای مثال، فرض کنید بخواهیم از میان افرادی که در آمریکا 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.

چند تعریف

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

نمونه تصادفی: یک نمونه تصادفی به اندازه‌ی n از این جامعه عبارتست از جمع‌آوری n متغیر تصادفی مستقل $X_1, X_2, \ldots, X_n$ که هر کدام دارای توزیع احتمال $f_X(x)$ هستند. این تابع به پارامتر مجهول θ بستگی دارد.

آماره: 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.

نکته: مقدار آماره از یک نمونه به نمونه دیگر تغییر می‌کند، اما مقدار پارامتر جامعه همواره ثابت است.

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

- ○ σ: انحراف معیار جامعه
- ● S: انحراف معیار نمونه

- ●○ n: حجم نمونه

توزیع نمونه‌ای میانگین نمونه

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:

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

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

یادآوری یک قضیه

اگر $X_۱, X_۲, \ldots, X_n$ متغیرهای تصادفی مستقل باشند و $X_i \sim N(\mu_i, \sigma_i^۲)$ برای $i = ۱, \ldots, n$ و قرار دهیم

$$Y = a_۱ X_۱ + a_۲ X_۲ + \cdots + a_n X_n = \sum_{i=۱}^{n} a_i X_i$$

در این صورت

$$Y = \sum_{i=۱}^{n} a_i X_i \sim N\left(\sum_{i=۱}^{n} a_i \mu_i, \sum_{i=۱}^{n} a_i^۲ \sigma_i^۲\right)$$

توزیع نمونه‌ای میانگین نمونه $\bar{X}$

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 صفحه قبل 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.

قضیه حد مرکزی

فرض کنید $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی هم‌توزیع با میانگین μ و واریانس σ^2 متناهی باشند به قسمی که هر تعداد متناهی از این متغیرهای تصادفی از یکدیگر مستقل باشند. در این صورت توزیع حدی $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ موقعی که $n \to \infty$ توزیع نرمال استاندارد است.

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

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.

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.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^۲}{n}\right) \Longrightarrow Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(۰,۱)$$

مثال ۱

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؟

راه‌حل:

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

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

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؟

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

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

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

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

مثال ۳

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:

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

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

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

$$= ۲P(Z \leq ۱۱۱/۸a) - ۱$$

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

توزیع نمونه‌ای واریانس نمونه

توزیع کای-دو

فرض کنید Z متغیری تصادفی با توزیع نرمال استاندارد $Z \sim N(۰,۱)$ باشد و قرار دهیم $Y = Z^۲$.
در این صورت تابع چگالی Y عبارتست از:

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(Z^۲ \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\
&= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \\
f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{۱}{۲\sqrt{y}} f_Z(\sqrt{y}) + \frac{۱}{۲\sqrt{y}} f_Z(-\sqrt{y}) = \frac{۱}{\sqrt{y}} f_Z(\sqrt{y}) \\
&= \frac{۱}{\sqrt{y}} \frac{۱}{\sqrt{۲\pi}} e^{-\frac{۱}{۲}(\sqrt{y})^۲} = \frac{۱}{۲^{\frac{۱}{۲}} \Gamma(\frac{۱}{۲})} y^{\frac{۱}{۲}-۱} e^{-\frac{۱}{۲}y} \qquad y > ۰
\end{aligned}
$$

- بنابراین $Y = Z^۲ \sim \chi^۲_{(۱)}$.
- حال اگر $Z_۱, Z_۲, \ldots, Z_n$ یک نمونه تصادفی n تایی از جامعه نرمال استاندارد باشند آنگاه $Z_۱^۲ + Z_۲^۲ + \ldots + Z_n^۲$ دارای توزیع کای-دو با n درجه آزادی $\chi^۲_{(n)}$ است.
- جدول مربوط به توزیع کای-دو در پیوست کتاب آمده است.
- **یادآوری:** توزیع $\chi^۲_{(n)}$ حالت خاصی از توزیع گاما با پارامترهای $r = \frac{n}{۲}$ و $\beta = ۲$ است.

مثال ۴

اگر $X \sim \chi^{۲}_{(۱۸)}$، مطلوبست:
الف- محاسبه‌ی $P(X \leq ۷/۰۱)$
ب- اگر $P(X > x) = ۰/۰۵$، مقدار x را به دست آورید.

راه‌حل:

$$\text{الف-}\quad \chi^{۲}_{۰/۰۱,(۱۸)} = ۷/۰۱ \qquad \Rightarrow \qquad P(X \leq ۷/۰۱) = ۰/۰۱$$

$$\text{ب-}\quad P(X > x) = ۰/۰۵ \qquad \Rightarrow \qquad P(X \leq x) = ۰/۹۵$$

$$\Rightarrow \qquad x = \chi^{۲}_{۰/۹۵,(۱۸)} = ۲۸/۹$$

توزیع نمونه‌ای واریانس نمونه S^2

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 $E(S^2) = \sigma^2$

اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از جامعه نرمال با میانگین μ و واریانس σ^2 باشند، آنگاه

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

اثبات:

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)} \quad \Rightarrow \quad \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \sum_{i=1}^{n} Z_i^2$$

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 ۶

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 $\frac{\bar{X}-\mu}{S/\sqrt{n}}$

توزیع t-استیودنت

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.
جدول مربوط به توزیع t در پیوست کتاب آمده است.

مثال ۷

اگر $T \sim t_{(۱۶)}$، مطلوب است
الف- محاسبه‌ی احتمال $P(T > ۱/۳۴)$
ب- اگر $P(T < t) = ۰/۸$ باشد، مقدار t را به دست آورید.

راه‌حل:

الف- $t_{۰/۹,(۱۶)} = ۱/۳۴ \quad \Rightarrow \quad P(T > ۱/۳۴) = ۱ - P(T \leq ۱/۳۴) = ۱ - ۰/۹ = ۰/۱$
ب- $P(T < t) = ۰/۸ \quad \Rightarrow \quad t = t_{۰/۸,(۱۶)} = ۰/۸۶۵$

قضیه: 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:

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

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

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

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 ۸

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

توزیع نمونه‌ای اختلاف میانگین‌ها $\mu_۱ - \mu_۲$

فرض کنید دو جامعه داشته باشیم.
جامعه‌ی اول دارای میانگین $\mu_۱$ و واریانس $\sigma_۱^۲$ باشد.
جامعه‌ی دوم دارای میانگین $\mu_۲$ و واریانس $\sigma_۲^۲$ باشد.
یک نمونه‌ی تصادفی n تایی $X_۱, \ldots, X_n$ از جامعه‌ی اول انتخاب کرده و میانگین این نمونه را با $\bar{X}$ و واریانس آن را با $S_۱^۲$ نمایش می‌دهیم.
یک نمونه‌ی تصادفی m تایی $Y_۱, \ldots, Y_m$ از جامعه‌ی دوم انتخاب کرده و میانگین این نمونه را با $\bar{Y}$ و واریانس آن را با $S_۲^۲$ 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).

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:

$$\bar{X}_A - \bar{X}_B \sim N\left(۴۰۰۰ - ۴۵۰۰\,,\ \frac{۳۰۰^{۲}}{۱۰۰} + \frac{۲۰۰^{۲}}{۵۰}\right) \quad \Rightarrow \quad \bar{X}_A - \bar{X}_B \sim N\left(-۵۰۰\,,\ ۱۷۰۰\right)$$

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

مثال ۱۰

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:

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

$$P\left(|\bar{X}-\bar{Y}|<۱۲\right)=P\left(-۱۲<\bar{X}-\bar{Y}<۱۲\right)$$

$$=P\left(\frac{-۱۲-۲۵}{\sqrt{۱۸}}<\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_۱-\mu_۲\right)}{\sqrt{\frac{\sigma_۱^۲}{n}+\frac{\sigma_۲^۲}{m}}}<\frac{۱۲-۲۵}{\sqrt{۱۸}}\right)$$

$$=P\left(-۸/۷۲<Z<-۳/۰۶\right)=۰/۰۰۱۱-۰=۰/۰۰۱۱$$

مثال ۱۱

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.

راه‌حل:

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

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

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

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:

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

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

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

مثال ۱۲

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؟

راه‌حل:

$$
\begin{aligned}
S_p^{۲} &= \frac{(n-۱)S_۱^{۲}+(m-۱)S_۲^{۲}}{n+m-۲} = \frac{(۸\times ۹)+(۹\times ۱۶)}{۹+۱۰-۲} = ۱۲/۷ \\
P\left(\bar{X}_۱ \geq \bar{X}_۲ + ۱۰/۷۵\right) &= P\left(\bar{X}_۱ - \bar{X}_۲ \geq ۱۰/۷۵\right) \\
&= P\left(\frac{(\bar{X}_۱-\bar{X}_۲)-(\mu_۱-\mu_۲)}{S_p\sqrt{\frac{۱}{n}+\frac{۱}{m}}} \geq \frac{۱۰/۷۵-۶}{\sqrt{۱۲/۷}\sqrt{\frac{۱}{۹}+\frac{۱}{۱۰}}}\right) \\
&= P\left(T_{(۱۷)} \geq ۲/۹\right) = ۱ - P\left(T_{(۱۷)} < ۲/۹\right) \\
&= ۱ - ۰/۹۹۵ = ۰/۰۰۵
\end{aligned}
$$

توزیع نمونه‌ای نسبت واریانس‌های دو نمونه

توزیع فیشر

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 از جامعه‌ی اول انتخاب کرده و واریانس آن را با S_1^2 نمایش می‌دهیم.
یک نمونه‌ی تصادفی m تایی از جامعه‌ی دوم انتخاب کرده و واریانس آن را با S_2^2 نشان می‌دهیم.
فرض کنید نمونه‌گیری از دو جامعه مستقل از یکدیگر باشد.
می‌خواهیم توزیع نمونه‌ای نسبت واریانس‌های دو نمونه یعنی $\frac{S_1^2}{S_2^2}$ را به دست آوریم.

تعریف توزیع فیشر F: اگر $U \sim \chi^2_{(n)}$ و $V \sim \chi^2_{(m)}$ و متغیرهای تصادفی U و V از یکدیگر مستقل باشند، آن‌گاه توزیع متغیر تصادفی $\frac{U/n}{V/m}$ را توزیع $F \sim F_{(n,m)}$ می‌نامند.
جدول این توزیع در پیوست کتاب آمده است.

توزیع نمونه‌ای نسبت واریانس‌های دو نمونه

قضیه: اگر S_1^2 و S_2^2 به ترتیب واریانس‌های نمونه‌های تصادفی مستقل به اندازه‌ی n و m از جامعه‌های نرمال با واریانس‌های σ_1^2 و σ_2^2 باشند، آن‌گاه

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n-1\ ,\ m-1)}$$

اثبات:

$$U = \frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2_{(n-1)} \qquad \perp \qquad V = \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2_{(m-1)}$$

$$F = \frac{\frac{U}{n-1}}{\frac{V}{m-1}} \sim F_{(n-1\ ,\ m-1)}$$

$$F = \frac{\frac{(n-1)S_1^2}{\sigma_1^2(n-1)}}{\frac{(m-1)S_2^2}{\sigma_2^2(m-1)}} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{(n-1\ ,\ m-1)}$$

از دو جامعه‌ی نرمال با واریانس‌های ۲۰ و ۳۰ به ترتیب نمونه‌های تصادفی ۸ و ۱۰ تایی انتخاب کرده‌ایم. احتمال اینکه واریانس نمونه‌ی اول بیش از دو برابر واریانس نمونه‌ی دوم باشد، چه‌قدر است؟

راه‌حل:

$$P\left(S_1^2 > 2S_2^2\right) = P\left(\frac{S_1^2}{S_2^2} > 2\right) = P\left(\frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} > 2 \times \frac{30}{20}\right)$$
$$= P\left(F_{(7,9)} > 3\right) = 1 - P\left(F_{(7,9)} \leq 3\right)$$
$$= 1 - 0/95 = 0/05$$

مثال ۱۴

اگر S_1^2 و S_2^2 به ترتیب واریانس‌های دو نمونه‌ی تصادفی مستقل از دو جامعه‌ی نرمال باشند و بدانیم واریانس جامعه‌ی دوم ۳ برابر واریانس جامعه‌ی اول است و به ترتیب نمونه‌هایی به اندازه‌ی ۸ و ۱۲ انتخاب شده باشد، مطلوبست محاسبه‌ی $P(S_1 < \sqrt{1/63}S_2)$.

راه‌حل:

$$\begin{aligned} P\left(S_1 < \sqrt{1/63}S_2\right) &= P\left(S_1^2 < 1/63 S_2^2\right) \\ &= P\left(\frac{S_1^2}{S_2^2} < 1/63\right) \\ &= P\left(\frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} < 1/63 \times \frac{3\sigma_1^2}{\sigma_1^2}\right) \\ &= P\left(F_{(7,11)} < 4/89\right) = 0/99 \end{aligned}$$

نکته

اگر $Z \sim N(۰,۱)$ و a عددی ثابت باشد:

$$\begin{aligned} P\left(-a<Z<a\right) &= P\left(Z<a\right) - P\left(Z\leq -a\right) \\ &= P\left(Z<a\right) - P\left(Z\geq a\right) \\ &= P\left(Z<a\right) - \left[۱ - P\left(Z<a\right)\right] \\ &= ۲P\left(Z<a\right) - ۱ \end{aligned}$$