برآوردیابی
- آمار و احتمالات مهندسی -

مدرس: مشکانی فراهانی

دانشگاه صنعتی امیرکبیر

۲۵ آذر ۱۳۹۹

جامعه‌ی آماری

در یک بررسی آماری هدف به دست آوردن اطلاعات از نمونه‌ی جمع‌آوری شده از جامعه و قضاوت از روی آن‌ها در مورد خصیصه‌های جامعه است.

استنباط آماری:

استنباط آماری روشی است که به وسیله‌ی آن بر اساس نتایج حاصل از نمونه‌ی انتخابی از جامعه، در مورد کل جامعه یا پارامترهای مجهول جامعه نتیجه‌گیری کنیم.

در مبحث استنباط آماری، خصیصه‌ی مورد نظر در جامعه را با یک متغیر تصادفی X نشان می‌دهند. فرض می‌شود توزیع X مشخص و تنها پارامتر θ از این توزیع نامعلوم است.

با استفاده از نمونه‌ی تصادفی $X_1, \dots, X_n$ و محاسبه‌ی آماره‌ی $T = T(X_1, \dots, X_n)$ سعی در استنباط روی پارامتر نامعلوم θ داریم.

استنباط آماری

استنباط آماری دو شاخه‌ی مهم دارد:

۱- **براورد پارامتر نامعلوم جامعه**: که خود به دو روش براورد نقطه‌ای و براورد فاصله‌ای قابل انجام است.

- **براورد نقطه‌ای:** در این روش از روی مقدار مشاهده شده‌ی آماره یعنی $t = T(x_1, \ldots, x_n)$ تنها یک مقدار برای تخمین پارامتر نامعلوم ارائه می‌شود.

- **براورد فاصله‌ای:** در این روش با استفاده از مقدار مشاهده شده‌ی آماره، فاصله‌ای را با یک اطمینان بسیار خوب به عنوان تخمین پارامتر نامعلوم ارائه می‌دهیم.

۲- **آزمون فرض‌های آماری:** در این روش بر اساس مشاهدات در مورد صحت یا عدم صحت ادعایی که در مورد جامعه یا پارامترهای آن انجام شده، قضاوت می‌کنیم.

براورد نقطه‌ای پارامتر نامعلوم جامعه

فرض کنید $X_1, \dots, X_n$ نمونه‌ای تصادفی از توزیعی با تابع چگالی احتمال $f_\theta(x)$ باشد، که θ پارامتری نامعلوم است.

هدف از تخمین (براورد) نقطه‌ای یافتن مقدار عددی یک آماره $t = T(x_1, \dots, x_n)$ از روی مقادیر مشاهده شده‌ی نمونه تصادفی یعنی $x_1, \dots, x_n$ است، که به مقدار نامعلوم θ بسیار نزدیک باشد.

توجه کنید که پارامتر θ دارای مقداری ثابت اما نامعلوم است.

ما تابعی از نمونه تصادفی $X_1, \dots, X_n$ به دست می‌آوریم که مقدار مشاهده شده‌ی این تابع به ازای مشاهدات $x_1, \dots, x_n$ به پارامتر مجهول θ بسیار نزدیک باشد.

براورد نقطه‌ای θ را با $\hat{\theta}$ نشان می‌دهند.

دو ویژگی یک براوردگر نقطه‌ای خوب

براوردگر ناریب:
براوردگر $T = T(X_1, \ldots, X_n)$ را براوردگری ناریب برای پارامتر θ گویند، هرگاه $E(T) = \theta$.

براوردگر کارا:
از یک براوردگر خوب انتظار داریم که در مقایسه با براوردگرهای دیگر دقت عملکرد بیشتری داشته باشد.
بدین معنی که متوسط فاصله آن تا پارامتر مجهول θ کمتر باشد. در این حالت می‌گوییم این براوردگر کاراتر است.
به عبارتی دیگر، اگر تمام براوردگرهای ناریب ممکن θ را در نظر بگیریم، آن‌که کمترین واریانس را دارد، کاراترین براوردگر است.

در روش براورد فاصله‌ای یک فاصله را به عنوان براوردی از پارامتر نامعلوم جامعه ارائه می‌دهیم، به طوری که این فاصله با اطمینان بالایی پارامتر نامعلوم θ را در بر داشته باشد. این فاصله را فاصله‌ی اطمینان می‌نامند.

تعریف: فرض کنید $X_1, \ldots, X_n$ نمونه‌ای تصادفی از توزیعی با تابع چگالی احتمال $f_\theta(x)$ باشد، که θ پارامتری نامعلوم است. همچنین فرض کنید $L(\mathbf{X}) = L(X_1, \ldots, X_n)$ و $U(\mathbf{X}) = U(X_1, \ldots, X_n)$ دو آماره باشند، به گونه‌ای که

$$P\left(L(\mathbf{X}) \le \theta \le U(\mathbf{X})\right) = 1 - \alpha$$

در این صورت فاصله‌ی تصادفی $(L(\mathbf{X}), U(\mathbf{X}))$ را یک خانواده از فاصله‌های اطمینان سطح $1 - \alpha$ برای θ می‌گویند.

$1 - \alpha$ را ضریب اطمینان فاصله می‌گویند.

$L(\mathbf{X})$ را کران پایین فاصله و $U(\mathbf{X})$ را کران بالای فاصله می‌گویند.

نمادها

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

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

- ○ σ: انحراف استاندارد جامعه
- S: انحراف استاندارد نمونه

- ○ n: حجم نمونه

براورد میانگین جامعه

$$\mu$$

براورد میانگین جامعه

فرض کنید از جامعه X با میانگین μ و واریانس $\sigma^{۲}$ یک نمونه تصادفی به اندازه n انتخاب کرده باشیم.

می‌خواهیم میانگین جامعه یعنی μ را برآورد کنیم.

- **بهترین براوردگر نقطه‌ای برای μ عبارت است از:** $\bar{X}$

- برای به دست آوردن یک فاصله اطمینان برای μ سه حالت را بررسی می‌کنیم:
  الف- جامعه نرمال با واریانس معلوم
  ب- جامعه نرمال با واریانس نامعلوم
  ج- واریانس جامعه نامعلوم و $n \geq ۳۰$.

برای به دست آوردن یک فاصله‌ی اطمینان $(۱-\alpha)۱۰۰\%$ برای μ باید فاصله‌ی (L, U) را طوری تعیین کنیم که $P(L < \mu < U) = ۱ - \alpha$.

اگر جامعه نرمال باشد، آن‌گاه $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(۰, ۱)$

برای به دست آوردن فاصله‌ی اطمینان برای μ اعداد $a < b$ را به گونه‌ای انتخاب می‌کنیم که

$$P\left(a < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < b\right) = ۱ - \alpha$$

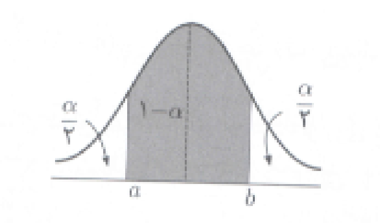
نقطه‌ی b نقطه‌ای روی محور افقی نمودار تابع چگالی نرمال استاندارد است که مساحت زیر منحنی قبل از این نقطه $۱ - \frac{\alpha}{۲}$ است. این نقطه را با $z_{۱-\frac{\alpha}{۲}}$ نشان می‌دهیم. این نقطه را با استفاده از جدول نرمال استاندارد به دست می‌آوریم.

نقطه‌ی a قرینه‌ی نقطه‌ی b است: $a = z_{\frac{\alpha}{۲}} = -z_{۱-\frac{\alpha}{۲}}$

براورد میانگین جامعه
حالت الف- واریانس جامعه معلوم

با قرار دادن این دو مقدار a و b به دست می‌آوریم:

$$P\left(-z_{۱-\frac{\alpha}{۲}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{۱-\frac{\alpha}{۲}}\right) = ۱-\alpha$$
$$P\left(\bar{X} - z_{۱-\frac{\alpha}{۲}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{۱-\frac{\alpha}{۲}}\frac{\sigma}{\sqrt{n}}\right) = ۱-\alpha$$



یک فاصله‌ی اطمینان $(۱-\alpha)۱۰۰\%$ برای میانگین جامعه‌ی نرمال μ زمانی که واریانس $\sigma^۲$ معلوم است، عبارت است از

$$\mu \in \left(\bar{x} - z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}} \right)$$

که در آن $\bar{x}$ میانگین نمونه‌ی تصادفی و $z_{۱-\frac{\alpha}{۲}}$ مقدار متغیر نرمال استاندارد است که سطح زیر منحنی در سمت چپ آن $۱-\frac{\alpha}{۲}$ باشد.

از یک جامعه نرمال با واریانس ۴ یک نمونه تصادفی به اندازه ۲۵ انتخاب کرده‌ایم و میانگین این نمونه ۲۰ شده است. یک فاصله اطمینان ۹۰ درصدی برای میانگین این جامعه پیدا کنید.

راه‌حل:

$$1-\alpha = 0/90 \quad \Rightarrow \quad \alpha = 0/1 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/1}{2} = 0/95$$

$$\Rightarrow \quad z_{0/95} = 1/645$$

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$\mu \in \left(20 - 1/645 \times \frac{2}{\sqrt{25}} \ , \ 20 + 1/645 \times \frac{2}{\sqrt{25}}\right)$$

$$\mu \in (19/342 \ , \ 20/658)$$

مثال ۲

یک نوع خازن الکترونیکی به وسیله یک شرکت ساخته می‌شود و در طی سال‌ها شرکت دریافته است که طول عمر این خازن‌ها دارای توزیع نرمال با انحراف استاندارد ۲۲۵ ساعت است. میانگین یک نمونه ۳۰ تایی از این خازن‌ها برابر ۱۴۰۷/۶۵ ساعت است. یک فاصله‌ی اطمینان ۹۸ درصدی برای میانگین طول عمر خازن‌های این شرکت به دست آورید.

راه‌حل:

$$1-\alpha = 0/98 \quad \Rightarrow \quad \alpha = 0/02 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/02}{2} = 0/99$$
$$\Rightarrow \quad z_{0/99} = 2/33$$

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \;,\; \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$
$$\mu \in \left(1407/65 - 2/33 \times \frac{225}{\sqrt{30}} \;,\; 1407/65 + 2/33 \times \frac{225}{\sqrt{30}}\right)$$
$$\mu \in (1311/93 \;,\; 1503/37)$$

اگر جامعه نرمال باشد و واریانس آن $\sigma^۲$ نامعلوم باشد، آن‌گاه $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-۱)}$

برای به دست آوردن فاصله‌ی اطمینان برای μ اعداد $a < b$ را به گونه‌ای انتخاب می‌کنیم که

$$P\left(a < \frac{\bar{X}-\mu}{s/\sqrt{n}} < b\right) = ۱ - \alpha$$

نقطه‌ی b نقطه‌ای روی محور افقی نمودار تابع چگالی t-استیودنت است که مساحت زیر منحنی قبل از این نقطه $۱ - \frac{\alpha}{۲}$ است. این نقطه را با $t_{۱-\frac{\alpha}{۲}}$ نشان می‌دهیم. این نقطه را با استفاده از جدول t به دست می‌آوریم.

نقطه‌ی a قرینه‌ی نقطه‌ی b است: $a = t_{\frac{\alpha}{۲}} = -t_{۱-\frac{\alpha}{۲}}$

با قرار دادن این دو مقدار a و b به دست می‌آوریم:

$$P\left(-t_{۱-\frac{\alpha}{۲},(n-۱)} < \frac{\bar{X}-\mu}{s/\sqrt{n}} < t_{۱-\frac{\alpha}{۲},(n-۱)}\right) = ۱-\alpha$$

$$P\left(\bar{X} - t_{۱-\frac{\alpha}{۲},(n-۱)}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{۱-\frac{\alpha}{۲},(n-۱)}\frac{s}{\sqrt{n}}\right) = ۱-\alpha$$

یک فاصله‌ی اطمینان $(1-\alpha)100\%$ برای میانگین جامعه‌ی نرمال μ زمانی که واریانس σ^2 نامعلوم است، عبارت است از

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} \;\; , \;\; \bar{x} + t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}}\right)$$

که در آن $\bar{x}$ میانگین نمونه‌ی تصادفی و $t_{1-\frac{\alpha}{2},(n-1)}$ مقدار متغیر t-استیودنت است که سطح زیر منحنی در سمت چپ آن $1-\frac{\alpha}{2}$ باشد.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - n(\bar{x})^2\right]$$

مثال ۳

یک نمونه تصادفی ۸-تایی از یک نوع سیگار به طور متوسط دارای ۱۸/۶ میلی‌گرم نیکوتین با انحراف استاندارد ۲/۴ میلی‌گرم است. با فرض نرمال بودن میزان نیکوتین در این نوع سیگار یک فاصله اطمینان ۹۹ درصدی برای حد متوسط واقعی نیکوتین این نوع سیگار پیدا کنید.

راه‌حل:

$$1-\alpha = 0/99 \quad \Rightarrow \quad \alpha = 0/01 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/01}{2} = 0/995$$

$$\Rightarrow \quad t_{0/995,(8-1)} = 3/5$$

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} \ , \ \bar{x} + t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}}\right)$$

$$\mu \in \left(18/6 - 3/5 \times \frac{2/4}{\sqrt{8}} \ , \ 18/6 + 3/5 \times \frac{2/4}{\sqrt{8}}\right)$$

$$\mu \in (15/63 \ , \ 21/57)$$

مثال ۴

اگر طول قد کارمندان یک اداره دارای توزیع نرمال باشد، یک فاصله اطمینان ۹۵ درصدی برای میانگین طول قد کارمندان این اداره پیدا کنید؛ در حالی که یک نمونه ۵ تایی از بین کارمندان انتخاب شده باشد و مقدار ۱۶۰، ۱۷۰، ۱۶۵، ۱۷۵ و ۱۸۰ به دست آمده باشد.

راه‌حل:

$$1-\alpha = 0/95 \Rightarrow \alpha = 0/05 \Rightarrow 1-\frac{\alpha}{2} = 1-\frac{0/05}{2} = 0/975 \Rightarrow t_{0/975,(5-1)} = 2/78$$

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{850}{5} = 170$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{4}\left\{(-10)^2+0+(-5)^2+5^2+10^2\right\} = 62/5$$

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} \ , \ \bar{x} + t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}}\right)$$

$$\mu \in \left(170 - 2/78 \times \sqrt{\frac{62/5}{5}} \ , \ 170 + 2/78 \times \sqrt{\frac{62/5}{5}}\right)$$

$$\mu \in (160/17 \ , \ 179/83)$$

براورد میانگین جامعه
حالت ج- واریانس جامعه نامعلوم و حجم نمونه زیاد

اگر $n > ۳۰$ باشد، آنگاه $z_{۱-\frac{\alpha}{۲}} = t_{۱-\frac{\alpha}{۲},(n-۱)}$ است.

بنابراین در حالتی که σ نامعلوم و $n > ۳۰$ باشد، می‌توان از فاصله اطمینان حالت (الف) با قرار دادن s به جای σ استفاده کرد:

$$\mu \in \left(\bar{x} - z_{۱-\frac{\alpha}{۲}} \frac{s}{\sqrt{n}} \ , \ \bar{x} + z_{۱-\frac{\alpha}{۲}} \frac{s}{\sqrt{n}}\right)$$

مثال ۵

قطر ۲۰۰ بلبرینگ ساخت یک کارخانه اندازه‌گیری شده و میانگین ۰/۸۲۴ اینچ و انحراف استاندارد ۰/۴۲ اینچ بوده است. یک فاصله‌ی اطمینان ۹۹ درصدی برای میانگین قطر بلبرینگ‌های ساخت این کارخانه بیابید.

راه‌حل:

$$1-\alpha = 0/99 \quad \Rightarrow \quad \alpha = 0/01 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/01}{2} = 0/995$$
$$\Rightarrow \quad z_{0/995} = 2/575$$

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \;,\; \bar{x} + z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$
$$\mu \in \left(0/824 - 2/575 \times \frac{0/42}{\sqrt{200}} \;,\; 0/824 + 2/575 \times \frac{0/42}{\sqrt{200}}\right)$$
$$\mu \in (0/747 \;,\; 0/901)$$

مثال ۶

یک تولیدکننده لامپ‌های روشنایی، لامپ‌هایی تولید می‌کند که انحراف معیار طول عمر آنها ۴۰ ساعت است. اگر یک نمونه تصادفی ۳۶ تایی دارای حد متوسط عمر ۸۷۰ ساعت باشند، یک فاصله اطمینان ۹۵ درصدی برای حد متوسط عمر تمام لامپ‌های تولیدی این کارخانه به دست آورید.

راه‌حل:

$$1-\alpha=0/95 \quad \Rightarrow \quad \alpha=0/05 \quad \Rightarrow \quad 1-\frac{\alpha}{2}=1-\frac{0/05}{2}=0/975$$
$$\Rightarrow \quad z_{0/975}=1/96$$

$$\mu \in \left(\bar{x}-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \;,\; \bar{x}+z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$
$$\mu \in \left(870-1/96\times\frac{40}{\sqrt{36}} \;,\; 870+1/96\times\frac{40}{\sqrt{36}}\right)$$
$$\mu \in (856/93 \;,\; 883/07)$$

خطای برآورد میانگین

چون اغلب مقدار برآورد نقطه‌ای $\bar{x}$ دقیقاً مساوی μ نیست؛ بنابراین برآورد نقطه‌ای دارای خطا است.

با استفاده از حدود فاصله اطمینان می‌توان میزان این خطا یعنی $|\bar{x} - \mu|$ را در دو حالت زیر تعیین کرد.

الف– اگر واریانس $\sigma^۲$ معلوم باشد، آنگاه

$$-z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad |\bar{x} - \mu| < z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}}$$

ب– اگر واریانس $\sigma^۲$ نامعلوم باشد، آنگاه

$$-t_{۱-\frac{\alpha}{۲},(n-۱)} \frac{s}{\sqrt{n}} < \bar{x} - \mu < t_{۱-\frac{\alpha}{۲},(n-۱)} \frac{s}{\sqrt{n}} \quad \Rightarrow \quad |\bar{x} - \mu| < t_{۱-\frac{\alpha}{۲},(n-۱)} \frac{s}{\sqrt{n}}$$

یک جایگاه سوخت‌گیری برای کامیون‌ها یادداشت‌های همه‌ی دادوستدها با مشتریانش را نگهداری می‌کند. اگر نمونه‌ای از ۱۸ یادداشت میانگین فروش‌ها را ۶۳/۸۴ گالن سوخت دیزل با انحراف استاندارد ۲/۷۵ گالن نشان دهد و مقدار ۶۳/۸۴ به عنوان براورد متوسط فروش سوخت دیزل نسبت به هر مشتری در جایگاه به کار رود، با اطمینان ۹۰ درصد حداکثر مقدار خطا را براورد کنید.

راه‌حل:

$$1 - \alpha = 0/90 \quad \Rightarrow \quad \alpha = 0/1 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 1 - \frac{0/1}{2} = 0/95$$

$$\Rightarrow \quad t_{0/95,(17)} = 1/74$$

$$|\bar{x} - \mu| < t_{1-\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}} = 1/74 \times \frac{2/75}{\sqrt{18}} = 1/13$$

در یک بررسی آماری یکی از مهم‌ترین مراحل، تعیین اندازه نمونه قبل از عمل نمونه‌گیری است.

اگر یک حداکثر مقدار خطای e برای برآورد میانگین μ برای نمونه‌گیر قابل تحمل باشد، آنگاه به وسیله‌ی خطای برآورد می‌توان اندازه نمونه n را در دو حالت زیر تعیین کرد.

اگر $\bar{x}$ را به عنوان برآورد نقطه‌ای μ به کار ببریم، آنگاه $(1-\alpha)100\%$ مطمئن هستیم که خطای براورد از مقدار مشخص e کمتر است زمانی که حجم نمونه از رابطه‌ی زیر حساب شود:

الف– اگر واریانس σ^2 معلوم باشد:
$$n = \left(z_{1-\frac{\alpha}{2}}\frac{\sigma}{e}\right)^2$$

ب– اگر واریانس σ^2 نامعلوم باشد:
$$n = \left(t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{e}\right)^2$$

مثال ۸

متوسط روی تغلیظ شده در نمونه‌ای از اندازه‌گیری‌های روی در ۳۶ مکان مختلف رودخانه‌ای، ۲/۶ گرم در هر میلی‌لیتر است. اگر بخواهیم ۹۰ درصد مطمئن باشیم که خطای براورد μ کمتر از ۰/۰۵ است، چه تعداد نمونه باید انتخاب کنیم؟ فرض کنید انحراف معیار جامعه ۰/۳ است.

راه‌حل:

$$1-\alpha = 0/9 \quad \Rightarrow \quad \alpha = 0/1 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/1}{2} = 0/95$$

$$\Rightarrow \quad z_{0/95} = 1/645$$

$$n = \left(z_{1-\frac{\alpha}{2}}\frac{\sigma}{e}\right)^2 \qquad \Rightarrow \qquad n = \left(1/645 \times \frac{0/3}{0/05}\right)^2$$

$$\Rightarrow \qquad n = 97/42 \qquad \Rightarrow \qquad n = 98$$

مثال ۹

در مثال طول قد کارمندان (مثال ۴) اگر بخواهیم با اطمینان ۹۵ درصد خطای براورد میانگین طول قد کارمندان اداره از ۵ سانتی‌متر کمتر باشد، حجم نمونه را تعیین کنید.

راه‌حل:

$$1-\alpha = 0/95 \quad \Rightarrow \quad \alpha = 0/05 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/05}{2} = 0/975$$
$$\Rightarrow \quad t_{0/975,(5-1)} = 2/78$$

$$n = \left(t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{e}\right)^2 \qquad \Rightarrow \qquad n = \left(2/78 \times \frac{7/9}{5}\right)^2$$
$$\Rightarrow \qquad n = 19/29 \qquad \Rightarrow \qquad n = 20$$

مثال ۱۰

اگر $\bar{x}$ میانگین نمونه یک نمونه‌ی تصادفی به اندازه‌ی n از یک جامعه‌ی نرمال با انحراف معیار ۰/۳ باشد، مقدار n را چنان تعیین کنید که فاصله‌ی اطمینان ۹۹ درصدی برای میانگین جامعه به صورت $(\bar{x} - ۰/۱ \ , \ \bar{x} + ۰/۱)$ باشد.

راه‌حل:

$$\sigma = ۰/۳, \qquad e = ۰/۱, \qquad ۱ - \alpha = ۰/۹۹$$

$$\Rightarrow \quad \alpha = ۰/۰۱ \quad \Rightarrow ۱ - \frac{\alpha}{۲} = ۰/۹۹۵ \quad \Rightarrow \quad z_{۰/۹۹۵} = ۲/۵۷۵$$

$$n = \left(z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{e}\right)^۲ \quad \Rightarrow \quad n = \left(۲/۵۷۵ \times \frac{۰/۳}{۰/۱}\right)^۲$$

$$\Rightarrow \quad n = ۵۹/۶۸ \quad \Rightarrow n = ۶۰$$

اگر اندازه‌ی نمونه‌ای به $\frac{۱}{۴}$ تقلیل یابد، طول فاصله‌ی اطمینان $(۱-\alpha)۱۰۰\%$ برای میانگین چه تغییری می‌کند؟

راه‌حل:

طول بازه = انتهای بازه - ابتدای بازه

$$D = \left[\bar{x} + z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}}\right] - \left[\bar{x} - z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}}\right] = ۲\, z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}}$$

$$D' = ۲z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{\frac{n}{۴}}} = ۲z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\frac{۱}{۲}\sqrt{n}} = ۲ \times ۲z_{۱-\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}} = ۲D$$

براورد تفاضل میانگین دو جامعه

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 $X_1, \dots, X_n$ 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 $Y_1, \dots, Y_m$ 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 اختلاف میانگین دو جامعه یعنی $\mu_1 - \mu_2$ را برآورد کنیم.

بهترین براوردگر نقطه‌ای برای $\mu_1 - \mu_2$ عبارت است از: $\bar{X} - \bar{Y}$

برای به دست آوردن فاصله‌ی اطمینان برای $\mu_1 - \mu_2$ دو حالت را در نظر می‌گیریم:

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

در فصل قبل دیدیم:

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

برای به دست آوردن فاصله‌ی اطمینان برای $\mu_1 - \mu_2$، مقادیر $a < b$ را با توجه به توزیع نرمال استاندارد تعیین می‌کنیم:

$$P\left(a < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} < b\right) = 1 - \alpha$$

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

براورد تفاضل میانگین دو جامعه
حالت اول: واریانس دو جامعه معلوم

یک فاصله اطمینان $(1-\alpha)100\%$ برای تفاضل میانگین دو جامعه نرمال که واریانس‌های آنها معلوم هستند عبارتست از:

$$\mu_1 - \mu_2 \in \left([\bar{x} - \bar{y}] - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \ , \ \ [\bar{x} - \bar{y}] + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

که در آن $\bar{x}$ و $\bar{y}$ به ترتیب میانگین نمونه‌های تصادفی n و m تایی از دو جامعه هستند.

تفسیر فاصله‌ی اطمینان برای $\mu_1 - \mu_2$

$$
\begin{array}{llllcl}
if & L_+ & \& & U_+ & \Rightarrow & \mu_1 > \mu_2 \\
if & L_- & \& & U_- & \Rightarrow & \mu_1 < \mu_2 \\
if & L_- & \& & U_+ & \Rightarrow & \mu_1 \simeq \mu_2
\end{array}
$$

مثال ۱۲

دو شرکت A و B لامپ‌های روشنایی تولید می‌کنند که به ترتیب دارای انحراف معیارهای ۲۷ و ۳۱ ساعت است. یک نمونه‌ی تصادفی ۴۰ تایی از لامپ‌های تولیدی هر شرکت انتخاب می‌کنیم و به ترتیب متوسط طول عمر لامپ‌ها برای این دو نمونه را ۶۴۹ و ۶۳۵ به دست می‌آوریم. با ساختن یک فاصله‌ی اطمینان ۹۶ درصدی برای اختلاف متوسط طول عمر لامپ‌های دو شرکت، چه نتیجه‌ای به دست می‌آورید؟

راه‌حل:

$$1-\alpha=0/96 \quad\Rightarrow\quad 1-\frac{\alpha}{2}=1-\frac{0/04}{2}=0/98 \quad\Rightarrow\quad z_{0/98}=2/05$$

$$\mu_1-\mu_2\in\left([\bar{x}-\bar{y}]-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\ ,\ [\bar{x}-\bar{y}]+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

$$\mu_1-\mu_2\in\left([649-635]-2/05\sqrt{\frac{27^2}{40}+\frac{31^2}{40}}\ ,\ [649-635]+2/05\sqrt{\frac{27^2}{40}+\frac{31^2}{40}}\right)$$

$$\mu_1-\mu_2\in(0/675\ ,\ 27/325) \qquad\Rightarrow\qquad \mu_1>\mu_2$$

مثال ۱۳

دو ماشین A و B جعبه‌های ۸ گرمی از یک ماده را بسته‌بندی می‌کنند. با استفاده از تجربیات گذشته با این ماشین‌ها می‌پذیریم که انحراف معیار وزن بسته‌های پر شده به وسیله ماشین A و B به ترتیب ۰۴/۰ و ۰۵/۰ گرم است. از جعبه‌های پر شده هر یک از ماشین‌ها صد جعبه به تصادف انتخاب شده و نتایج زیر به دست آمده است:

ماشین A: $n_A = ۱۰۰$ و $\bar{x}_A = ۸/۱۸$

ماشین B: $n_B = ۱۰۰$ و $\bar{x}_B = ۸/۱۵$

برای تفاضل میانگین وزن‌های پر شده با دو ماشین A و B یک فاصله اطمینان ۹۹ درصدی به دست آورید.

راه‌حل:

$$1 - \alpha = ۰/۹۹ \quad \Rightarrow \quad 1 - \frac{\alpha}{۲} = 1 - \frac{۰/۰۱}{۲} = ۰/۹۹۵ \quad \Rightarrow \quad z_{۰/۹۹۵} = ۲/۵۷۵$$

$$\mu_A - \mu_B \in \left([\bar{x}_A - \bar{x}_B] - z_{1-\frac{\alpha}{۲}} \sqrt{\frac{\sigma_۱^۲}{n} + \frac{\sigma_۲^۲}{m}} \ , \ [\bar{x}_A - \bar{x}_B] + z_{1-\frac{\alpha}{۲}} \sqrt{\frac{\sigma_۱^۲}{n} + \frac{\sigma_۲^۲}{m}} \right)$$

$$\mu_A - \mu_B \in \left([۸/۱۸ - ۸/۱۵] - ۲/۵۷۵\sqrt{\frac{(۰/۰۴)^۲}{۱۰۰} + \frac{(۰/۰۵)^۲}{۱۰۰}} \ , \ [۸/۱۸ - ۸/۱۵] + ۲/۵۷۵\sqrt{\frac{(۰/۰۴)^۲}{۱۰۰} + \frac{(۰/۰۵)^۲}{۱۰۰}} \right)$$

$$\mu_A - \mu_B \in (۰/۰۱۳ \ , \ ۰/۰۴۷) \qquad \Rightarrow \qquad \mu_A > \mu_B$$

حالت دوم: واریانس دو جامعه نامعلوم امّا مساوی باشند
در این حالت $\sigma_1^2 = \sigma_2^2 = \sigma^2$ است.

ابتدا واریانس مشترک دو جامعه را با واریانس مشترک دو نمونه به صورت زیر براورد می‌کنیم:

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

اگر دو جامعه نرمال باشند، در فصل قبل اثبات شد که:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{(n+m-2)}$$

با عملیات مشابه اسلایدهای ۱۵ و ۱۶ به دست می‌آوریم:

براورد تفاضل میانگین دو جامعه
حالت دوم: واریانس دو جامعه نامعلوم امّا مساوی

یک فاصله اطمینان $(1-\alpha)100\%$ برای تفاضل میانگین دو جامعه نرمال که واریانس‌های آنها نامعلوم و مساوی هستند عبارتست از:

$$\mu_1 - \mu_2 \in \left([\bar{x} - \bar{y}] \pm t_{1-\frac{\alpha}{2},(n+m-2)} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

که در آن

- $\bar{x}$ و $\bar{y}$ به ترتیب میانگین نمونه‌های تصادفی n و m تایی از دو جامعه هستند.
- s_p^2 واریانس مشترک دو نمونه‌ی تصادفی n و m تایی است.

مثال ۱۴

میزان تقاضا بر حسب کیلوگرم برای ۲ محصول x و y دارای توزیع نرمال با واریانس‌های یکسان می‌باشند. میزان تقاضا برای ۶ روز این دو محصول را جمع‌آوری و در جدول زیر ثبت کرده‌ایم. با تشکیل یک فاصله اطمینان ۹۰ درصد برای تفاضل میانگین تقاضای دو محصول بیان کنید که میزان تقاضا برای کدام محصول بیشتر است.

x	۳۰	۲۶	۳۰	۱۹	۲۶	۳۱
y	۴۰	۳۴	۳۰	۲۹	۳۵	۴۲

راه‌حل:

$$1-\alpha = 0/90 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/1}{2} = 0/95 \quad \Rightarrow \quad t_{0/95\,,\,(6+6-2)} = 1/81$$

$$\bar{x} = \frac{30+\cdots+31}{6} = 27 \qquad s_X^2 = \frac{1}{6-1}\left[(30-27)^2+\cdots+(31-27)^2\right] = 20$$

$$\bar{y} = \frac{40+\cdots+42}{6} = 35 \qquad s_Y^2 = \frac{1}{6-1}\left[(40-35)^2+\cdots+(42-35)^2\right] = 27/2$$

$$s_p^2 = \frac{(6-1)\times 20 + (6-1)\times 27/2}{6+6-2} = 23/6 \qquad \Rightarrow \qquad s_p = 4/86$$

ادامه‌ی راه‌حل:

$$\mu_X - \mu_Y \in \left(\bar{x} - \bar{y} - t_{۱-\frac{\alpha}{۲},(n+m-۲)} s_p \sqrt{\frac{۱}{n} + \frac{۱}{m}} \ , \ \bar{x} - \bar{y} + t_{۱-\frac{\alpha}{۲},(n+m-۲)} s_p \sqrt{\frac{۱}{n} + \frac{۱}{m}}\right)$$

$$\mu_X - \mu_Y \in \left([۲۷ - ۳۵] - ۱/۸۱ \times ۴/۸۶\sqrt{\frac{۱}{۶} + \frac{۱}{۶}} \ , \ [۲۷ - ۳۵] + ۱/۸۱ \times ۴/۸۶\sqrt{\frac{۱}{۶} + \frac{۱}{۶}}\right)$$

$$\mu_X - \mu_Y \in (-۱۳/۰۸ \ , \ -۲/۹۲)$$

$$\Rightarrow \qquad \mu_X < \mu_Y$$

مثال ۱۵

دو آزمایشگاه به طور مستقل برای اندازه‌گیری چربی موجود در شیرهای پاستوریزه اقدام می‌کنند. هر یک تعدادی نمونه انتخاب کرده و نتایج در جدول زیر ثبت شده است. با فرض نرمال بودن دو جامعه و مساوی بودن واریانس‌ها، یک فاصله‌ی اطمینان ۹۵ درصدی برای اختلاف میانگین چربی موجود در شیرهای دو کارخانه به دست آورید. چه نتیجه‌ای می‌گیرید؟

آزمایشگاه ۱	۳	۵	۷	۳	۸	۶	۸	۹	۴	۷
آزمایشگاه ۲	۹	۸	۸	۴	۷	۶	۸	۶		

راه‌حل:

$$۱-\alpha=۰/۹۵ \quad\Rightarrow\quad ۱-\frac{\alpha}{۲}=۱-\frac{۰/۰۵}{۲}=۰/۹۷۵ \quad\Rightarrow\quad t_{۰/۹۷۵\,,\,(۱۰+۸-۲)}=۲/۱۲$$

$$\bar{x}=\frac{۳+\cdots+۷}{۱۰}=۶ \qquad\qquad s_۱^۲=\frac{۱}{۱۰-۱}\left[(۳-۶)^۲+\cdots+(۷-۶)^۲\right]=۴/۶۷$$

$$\bar{y}=\frac{۹+\cdots+۶}{۸}=۷ \qquad\qquad s_۲^۲=\frac{۱}{۸-۱}\left[(۹-۷)^۲+\cdots+(۶-۷)^۲\right]=۲/۵۷$$

$$s_p^۲=\frac{(۱۰-۱)\times ۴/۶۷+(۸-۱)\times ۲/۵۷}{۱۰+۸-۲}=۳/۷۵ \qquad\Rightarrow\qquad s_p=۱/۹۳۷$$

مثال ۱۵

ادامه‌ی راه‌حل:

$$\mu_۱ - \mu_۲ \in \left((\bar{x} - \bar{y}) - t_{۱-\frac{\alpha}{۲},(n+m-۲)} s_p \sqrt{\frac{۱}{n} + \frac{۱}{m}} \; , \; (\bar{x} - \bar{y}) + t_{۱-\frac{\alpha}{۲},(n+m-۲)} s_p \sqrt{\frac{۱}{n} + \frac{۱}{m}} \right)$$

$$\mu_۱ - \mu_۲ \in \left((۶ - ۷) - ۲/۱۲ \times ۱/۹۳۷ \sqrt{\frac{۱}{۱۰} + \frac{۱}{۸}} \; , \; (۶ - ۷) + ۲/۱۲ \times ۱/۹۳۷ \sqrt{\frac{۱}{۱۰} + \frac{۱}{۸}} \right)$$

$$\mu_۱ - \mu_۲ \in (-۲/۹۵ \; , \; ۰/۹۵)$$

$$\Rightarrow \qquad \mu_۱ \simeq \mu_۲$$

براورد تفاضل میانگین مشاهدات زوجی

براورد تفاضل میانگین مشاهدات زوجی

در برخی از آزمایش‌های آماری می‌خواهیم تأثیر یک روش یا آزمایش را روی اعضای جامعه بررسی کنیم.

به عنوان مثال:
می‌خواهیم تأثیر یک نوع رژیم غذایی را در کاهش وزن افراد جامعه بررسی کنیم.
می‌خواهیم تأثیر یک دارو جدید را در کاهش یا افزایش فشار خون بیماران بررسی کنیم.

ابتدا یک نمونه از اعضای جامعه انتخاب کرده و خصوصیت مورد نظر را روی اعضای نمونه اندازه‌گیری می‌کنیم.
سپس بعد از انجام روش یا آزمایش همان خصوصیت را مجدداً روی همان اعضای نمونه اندازه‌گیری می‌کنیم.

فرض کنید X و Y به ترتیب اندازه‌گیری خصوصیت مد نظر قبل و بعد از انجام آزمایش باشد.

نمونه تصادفی حاصل به صورت $(X_۱, Y_۱), \ldots, (X_n, Y_n)$ است.

در این حالت زوج‌های $(X_۱, Y_۱), \ldots, (X_n, Y_n)$ از یکدیگر مستقل هستند ولی X_i و Y_i برای $i = ۱, ۲, \ldots, n$ که مربوط به اندازه‌گیری قبل و بعد از انجام آزمایش عضو i-ام است، به یکدیگر وابسته هستند.

براورد تفاضل میانگین مشاهدات زوجی

اگر $(\mu_X \ , \ \mu_Y)$ میانگین مقدار خصوصیت مورد نظر قبل و بعد از انجام آزمایش باشد، می‌خواهیم تفاضل میانگین‌ها یعنی $\mu_D = \mu_X - \mu_Y$ را براورد کنیم.

برای براورد μ_D قرار می‌دهیم:

$$D_i = X_i - Y_i \qquad \bar{D} = \frac{۱}{n}\sum_{i=۱}^{n} D_i \qquad S_D^۲ = \frac{۱}{n-۱}\sum_{i=۱}^{n}(D_i - \bar{D})^۲$$

در این صورت $D_۱ , \dots , D_n$ نمونه‌ای تصادفی از جامعه هستند که فرض می‌کنیم دارای توزیع نرمال با میانگین μ_D و واریانس نامعلوم $\sigma_D^۲$ است.

براوردگر نقطه‌ای برای μ_D برابر است با $\bar{D}$ و براوردگر نقطه‌ای برای $\sigma_D^۲$ برابر است با $S_D^۲$.

مطابق با اثبات روابط قبلی:

$$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \sim t_{(n-۱)}$$

براورد تفاضل میانگین مشاهدات زوجی

یک فاصله‌ی اطمینان $(1-\alpha)100\%$ برای تفاضل میانگین مشاهدات زوجی عبارت است از:

$$\mu_D = \mu_X - \mu_Y \in \left(\bar{d} - t_{1-\frac{\alpha}{2},(n-1)} \frac{s_d}{\sqrt{n}} \;\; , \;\; \bar{d} + t_{1-\frac{\alpha}{2},(n-1)} \frac{s_d}{\sqrt{n}} \right)$$

که در آن:
$t_{1-\frac{\alpha}{2},(n-1)}$ مقدار متغیر t-استیودنت است که سطح زیر منحنی در سمت چپ آن $1-\frac{\alpha}{2}$ باشد.

$$d_i = x_i - y_i \qquad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

مثال ۱۶

داده‌های زیر مربوط به نظرخواهی از ۵ نفر قبل و بعد از دیدن یک فیلم تبلیغاتی است. با فرض نرمال بودن در سطح اطمینان ۹۶ درصد داده‌ها بیان کنید که آیا نمایش فیلم موجب افزایش میانگین سطح فکر افراد شده است؟

قبل	۶۰	۲۹	۳۶	۵۳	۲۵
بعد	۶۷	۳۳	۳۶	۵۰	۴۲
قبل - بعد	-۷	-۴	۰	۳	-۱۷

$$\bar{d} = \frac{-7 - 4 + 0 + 3 - 17}{5} = -5$$

$$s_d^2 = \frac{1}{5-1}\left\{(-7+5)^2 + (-4+5)^2 + (0+5)^2 + (3+5)^2 + (-17+5)^2\right\} = 59/5 \quad \Rightarrow \quad s_d = 7/71$$

$$1-\alpha = 0/96 \quad \Rightarrow \quad \alpha = 0/04 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 1 - \frac{0/04}{2} = 0/98 \quad \Rightarrow \quad t_{0/98\,,\,(4)} = 2/999$$

$$\mu_D \in \left(\bar{d} - t_{1-\frac{\alpha}{2},(n-1)}\frac{s_d}{\sqrt{n}} \;,\; \bar{d} + t_{1-\frac{\alpha}{2},(n-1)}\frac{s_d}{\sqrt{n}}\right)$$

$$\mu_D \in \left(-5 - 2/999 \times \frac{7/71}{\sqrt{5}} \;,\; -5 + 2/999 \times \frac{7/71}{\sqrt{5}}\right) \quad \Rightarrow \quad \mu_D \in (-15/34\,,\, 5/34) \quad \Rightarrow \quad \mu_{\text{قبل}} \simeq \mu_{\text{بعد}}$$

مثال ۱۷

فرض کنید میزان مصرف بنزین ۶ اتومبیل قبل و بعد از تعویض فیلتر بنزین اندازه‌گیری شده و نتایج زیر به دست آمده است. با ساختن یک فاصله اطمینان ۹۸ درصد بیان کنید که آیا میانگین مصرف بنزین بعد از تعویض فیلتر کاهش یافته است؟

میزان مصرف قبل از تعویض فیلتر	۶۴	۶۶	۸۹	۷۷	۷۰	۶۳
میزان مصرف بعد از تعویض فیلتر	۵۴	۵۴	۷۰	۶۲	۷۸	۶۳
قبل - بعد	۱۰	۱۲	۱۹	۱۵	-۸	۰

$$\bar{d} = \frac{10 + 12 + 19 + 15 - 8 + 0}{6} = 8$$

$$s_d^2 = \frac{1}{6-1}\left\{(10-8)^2 + (12-8)^2 + (19-8)^2 + (15-8)^2 + (-8-8)^2 + (0-8)^2\right\} = 102 \quad \Rightarrow \quad s_d = 10/1$$

$$1-\alpha = 0/98 \quad \Rightarrow \quad \alpha = 0/02 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/02}{2} = 0/99 \quad \Rightarrow \quad t_{0/99\,,\,(5)} = 3/365$$

$$\mu_D \in \left(\bar{d} - t_{1-\frac{\alpha}{2},(n-1)}\frac{s_d}{\sqrt{n}} \;,\; \bar{d} + t_{1-\frac{\alpha}{2},(n-1)}\frac{s_d}{\sqrt{n}}\right)$$

$$\mu_D \in \left(8 - 3/365 \times \frac{10/1}{\sqrt{6}} \;,\; 8 + 3/365 \times \frac{10/1}{\sqrt{6}}\right) \quad \Rightarrow \quad \mu_D \in (-5/875\,,\, 21/875) \quad \Rightarrow \quad \mu_{\text{قبل}} \simeq \mu_{\text{بعد}}$$

براورد نسبت در جامعه

براورد نسبت در جامعه

در برخی از بررسی‌های آماری نسبتی از اعضای جامعه p که دارای خصوصیت معینی هستند، مد نظر است.

این مطالعات مشاهده‌ای بر روی یک متغیر دو ارزشی (مثل: زن و مرد - شهری و روستایی - موافق و مخالف) انجام می‌شود.

از آن‌جا که متغیر مورد مطالعه تنها دو ارزش دارد، هر یک از اعضای جامعه در یکی از دو ارزش قرار می‌گیرند.

نسبت موفقیت‌ها (افراد/اشیاء مورد نظر و سوال شده) را با p و نسبت افراد/اشیائی که در موفقیت قرار نمی‌گیرند را با q نشان می‌دهیم.

برای بررسی این نسبت p نمونه تصادفی $X_1, \ldots, X_n$ را از جامعه جمع‌آوری می‌کنیم:

$$X_i = \begin{cases} 1, & \text{اگر عضو } i\text{-ام نمونه دارای خصوصیت معین باشد} \\ 0, & \text{در غیر این‌صورت} \end{cases}$$

قرار می‌دهیم: $X = X_1 + X_2 + \cdots + X_n$؛ که در آن X تعدادی اعضای نمونه است که دارای خصوصیت معین هستند؛ پس $X \sim Bin(n, p)$

بهترین براوردگر نقطه‌ای برای p عبارت است از: $\frac{X}{n}$

از آن‌جا که X دارای توزیع دوجمله‌ای است پس امید ریاضی $\frac{X}{n}$ برابر است با p و واریانس آن برابر است با $\frac{pq}{n}$.

اگر حجم نمونه n بزرگ باشد، آن‌گاه با توجه به قضیه حد مرکزی توزیع $\frac{X-np}{\sqrt{npq}} = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$ به توزیع نرمال استاندارد میل می‌کند.

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \sim N(۰,۱)$$

پس داریم:

$$P\left(-z_{۱-\frac{\alpha}{۲}} < \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} < z_{۱-\frac{\alpha}{۲}}\right) = ۱-\alpha$$

یک فاصله‌ی اطمینان $(1-\alpha)100\%$ برای نسبت p از اعضای جامعه که خصوصیت معینی دارند، عبارتست از:

$$p \in \left(\hat{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

که در آن $\hat{p} = \frac{X}{n}$ و $\hat{q} = 1 - \hat{p}$ است.
هم‌چنین، X تعداد اعضای نمونه n تایی است که دارای خصوصیت مد نظر است.

مثال ۱۸

یک سیستم موشکی جدید به منظور توسعه‌ی سیستم‌های موشکی برد کوتاه بررسی می‌شود. سیستم موجود دارای احتمال ۰/۸ موفقیت در هر پرتاب است. یک نمونه ۴۰ تایی از پرتاب‌های آزمایش سیستم جدید منجر به ۳۴ موفقیت شده است. یک فاصله اطمینان ۹۰ درصد برای p بسازید.

راه‌حل:

$$\hat{p} = \frac{X}{n} = \frac{۳۴}{۴۰} = ۰/۸۵ \qquad \Rightarrow \qquad \hat{q} = ۱ - ۰/۸۵ = ۰/۱۵$$

$$۱ - \alpha = ۰/۹۰ \qquad \Rightarrow \qquad ۱ - \frac{\alpha}{۲} = ۱ - \frac{۰/۱}{۲} = ۰/۹۵ \qquad \Rightarrow \qquad z_{۰/۹۵} = ۱/۶۴۵$$

$$p \in \left(\hat{p} - z_{۱-\frac{\alpha}{۲}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p} + z_{۱-\frac{\alpha}{۲}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

$$p \in \left(۰/۸۵ - ۱/۶۴۵ \times \sqrt{\frac{۰/۸۵ \times ۰/۱۵}{۴۰}}\ ,\ ۰/۸۵ + ۱/۶۴۵ \times \sqrt{\frac{۰/۸۵ \times ۰/۱۵}{۴۰}}\right)$$

$$p \in (۰/۷۵۷\ ,\ ۰/۹۴۳)$$

مثال ۱۹

یک نمونه تصادفی ۱۰۰ تایی از رأی‌های جمع‌آوری شده در یک انتخابات نشان می‌دهد که ۵۹ نفر به کاندید A رای داده‌اند. یک فاصله اطمینان ۹۵ درصدی برای درصد افرادی که به نفع A رأی داده‌اند، تشکیل دهید.

راه‌حل:

$$\hat{p} = \frac{X}{n} = \frac{۵۹}{۱۰۰} = ۰/۵۹ \qquad \Rightarrow \qquad \hat{q} = ۱ - ۰/۵۹ = ۰/۴۱$$

$$۱ - \alpha = ۰/۹۵ \qquad \Rightarrow \qquad ۱ - \frac{\alpha}{۲} = ۱ - \frac{۰/۰۵}{۲} = ۰/۹۷۵ \qquad \Rightarrow \qquad z_{۰/۹۷۵} = ۱/۹۶$$

$$p \in \left(\hat{p} - z_{۱-\frac{\alpha}{۲}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p} + z_{۱-\frac{\alpha}{۲}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

$$p \in \left(۰/۵۹ - ۱/۹۶ \times \sqrt{\frac{۰/۵۹ \times ۰/۴۱}{۱۰۰}}\ ,\ ۰/۵۹ + ۱/۹۶ \times \sqrt{\frac{۰/۵۹ \times ۰/۴۱}{۱۰۰}}\right)$$

$$p \in (۰/۴۹۴\ ,\ ۰/۶۸۶)$$

اگر $\hat{p}$ به عنوان براوردی از p به کار برده شود، می‌توانیم $\%۱۰۰(۱-\alpha)$ مطمئن باشیم که خطا از $z_{۱-\frac{\alpha}{۲}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ بیشتر نیست:

$$|\hat{p}-p| < z_{۱-\frac{\alpha}{۲}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- در مثال قبل (۱۹)، ۹۵ درصد مطمئن هستیم که نسبت نمونه $\hat{p} = ۰/۵۹$ با نسبت واقعی p حداکثر به اندازه ۰/۰۹۶ اختلاف دارد.

تعیین حجم نمونه

اگر $\hat{p}$ برای براورد p به کار برده شود، می‌توانیم $\%۱۰۰(۱-\alpha)$ مطمئن باشیم وقتی اندازه نمونه تقریبا برابر با

$$n = \frac{z^{۲}_{۱-\frac{\alpha}{۲}}\hat{p}\hat{q}}{e^{۲}}$$

است، خطا کمتر از مقدار مشخص e است.

مثال ۲۰

از ۱۶۰۰ بزرگ‌سالی که با تلفن نظر آن‌ها درباره‌ی برنامه فضایی پرسیده شده، ۳۲ درصد اظهار کرده‌اند که برنامه فضایی سرمایه‌گذاری خوبی برای کشور است و باید ادامه پیدا کند. اگر بخواهیم ۹۸ درصد مطمئن باشیم درصد براورد شده از درصد واقعی ۲٪ کمتر است، از چه تعدادی از افراد باید نظرخواهی کنیم؟

راه‌حل:

$$1-\alpha=0/98 \quad \Rightarrow \quad 1-\frac{\alpha}{2}=1-\frac{0/02}{2}=0/99 \quad \Rightarrow \quad z_{0/99}=2/33$$

$$n=\frac{z^2_{1-\frac{\alpha}{2}}\hat{p}\hat{q}}{e^2}=\frac{(2/33)^2\times 0/32\times 0/68}{(0/02)^2}=2953/3 \quad \Rightarrow \quad n=2954$$

یکی از اعضای هیأت علمی بخش میکروبیولوژی حدس می‌زند که مصرف روزانه‌ی دو فنجان چای فلوراید کافی را برای حفاظت دندان‌ها از پوسیدگی تأمین می‌کند. اگر بخواهیم حداقل ۹۹ درصد مطمئن باشیم که این براورد از درصد واقعی ۱٪ کمتر است، چه تعداد نمونه احتیاج است؟

راه‌حل:

$$1-\alpha = 0/99 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/01}{2} = 0/995 \quad \Rightarrow \quad z_{0/995} = 2/57$$

$$n = \frac{z^2_{1-\frac{\alpha}{2}}\hat{p}\hat{q}}{e^2} = \frac{(2/57)^2 \times 0/5 \times 0/5}{(0/01)^2} = 16512/25 \quad \Rightarrow \quad n = 16513$$

براورد تفاضل نسبت دو جامعه

براورد تفاضل نسبت دو جامعه

در برخی از بررسی‌ها می‌خواهیم نسبت یک خصوصیت را در دو جامعه با یکدیگر مقایسه کنیم.

برای این منظور از دو جامعه نمونه‌های تصادفی مستقل به اندازه‌های n و m انتخاب می‌کنیم.

p_1: نسبت اعضای جامعه اول که دارای خصوصیت معینی هستند
p_2: نسبت اعضای جامعه دوم که دارای خصوصیت معینی هستند
X_1: تعداد اعضای نمونه از جامعه اول که دارای خصوصیت معینی هستند
X_2: تعداد اعضای نمونه از جامعه دوم که دارای خصوصیت معینی هستند

می‌خواهیم $p_1 - p_2$ را براورد کنیم.

یک براورد نقطه‌ای خوب برای $p_1 - p_2$ عبارت است از: $\frac{X_1}{n} - \frac{X_2}{m}$

یک فاصله‌ی اطمینان $(۱-\alpha)۱۰۰\%$ برای تفاضل نسبت دو جامعه یعنی $p_۱ - p_۲$ عبارت است از:

$$p_۱ - p_۲ \in \left(\hat{p}_۱ - \hat{p}_۲ \pm z_{۱-\frac{\alpha}{۲}} \sqrt{\frac{\hat{p}_۱\hat{q}_۱}{n} + \frac{\hat{p}_۲\hat{q}_۲}{m}}\right)$$

که در آن $\hat{p}_۱ = \frac{X_۱}{n}$ و $\hat{p}_۲ = \frac{X_۲}{m}$ است.

مثال ۲۲

مطالعه‌ای به منظور ارزشیابی عوارض جانبی دو دارو طراحی شده است. بدین منظور به ۵۰ حیوان داروی A و به ۵۰ حیوان داروی B داده‌اند. از میان حیواناتی که داروی A را استفاده کرده‌اند، ۱۱ حیوان و از حیواناتی که داروی B را استفاده کرده‌اند، ۸ حیوان عوارض جانبی نامطلوب نشان داده‌اند. برای $p_A - p_B$ حدود اطمینان ۹۶ درصد به دست آورید.

راه‌حل:

$$\hat{p}_1 = \frac{X_1}{n} = \frac{11}{50} = 0/22 \qquad \Rightarrow \qquad \hat{q}_1 = 1 - 0/22 = 0/78$$

$$\hat{p}_2 = \frac{X_2}{m} = \frac{8}{50} = 0/16 \qquad \Rightarrow \qquad \hat{q}_2 = 1 - 0/16 = 0/84$$

$$1 - \alpha = 0/96 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 1 - \frac{0/04}{2} = 0/98 \quad \Rightarrow \quad z_{0/98} = 2/05$$

$$p_1 - p_2 \in \left(\hat{p}_1 - \hat{p}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n} + \frac{\hat{p}_2\hat{q}_2}{m}}\right)$$

$$p_1 - p_2 \in \left(0/22 - 0/16 \pm 2/05\sqrt{\frac{0/22 \times 0/78}{50} + \frac{0/16 \times 0/84}{50}}\right)$$

$$-0/1004 \leq p_1 - p_2 \leq 0/2204 \qquad \Rightarrow \qquad p_1 \simeq p_2$$

مثال ۲۳

در تحقیقی پژوهشگران دریافتند که در دمای ۵ درجه‌ی سانتی‌گراد، از ۲۰ دانه کاشته شده ۱۰ دانه به ثمر رسید و در ۱۵ درجه‌ی سانتی‌گراد پس از کاشتن ۲۰ دانه ۱۵ دانه به ثمر رسید. یک فاصله اطمینان ۹۴ درصد برای تفاضل بین نسبت رویش دانه‌ها در این دو درجه حرارت بسازید و در صورتی که اختلاف معنی‌داری وجود داشته باشد، آن را بیان کنید.

راه‌حل:

$$\hat{p}_1 = \frac{X_1}{n} = \frac{10}{20} = 0/5 \quad \Rightarrow \quad \hat{q}_1 = 1 - 0/5 = 0/5$$

$$\hat{p}_2 = \frac{X_2}{m} = \frac{15}{20} = 0/75 \quad \Rightarrow \quad \hat{q}_2 = 1 - 0/75 = 0/25$$

$$1 - \alpha = 0/94 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 1 - \frac{0/06}{2} = 0/97 \quad \Rightarrow \quad z_{0/97} = 1/88$$

$$p_1 - p_2 \in \left(\hat{p}_1 - \hat{p}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}\right)$$

$$p_1 - p_2 \in \left(0/5 - 0/75 \pm 1/88 \sqrt{\frac{0/5 \times 0/5}{20} + \frac{0/75 \times 0/25}{20}}\right)$$

$$-0/528 \leq p_1 - p_2 \leq 0/028 \quad \Rightarrow \quad p_1 \simeq p_2$$

براورد واریانس جامعه
σ^2

فرض کنید از یک جامعه با میانگین μ و واریانس $\sigma^۲$ یک نمونه‌ی تصادفی به اندازه‌ی n انتخاب کرده باشیم و بخواهیم واریانس جامعه یعنی $\sigma^۲$ را براورد کنیم.

بهترین براوردگر نقطه‌ای برای $\sigma^۲$ عبارت است از واریانس نمونه یعنی

$$s^۲ = \frac{۱}{n-۱}\sum_{i=۱}^{n}(x_i-\bar{x})^۲$$

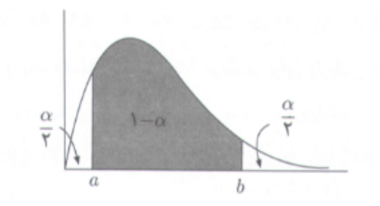
اگر جامعه نرمال باشد، در فصل قبل اثبات کردیم: $\frac{(n-۱)s^۲}{\sigma^۲} \sim \chi^۲_{(n-۱)}$

اعداد a و b را چنان تعیین می‌کنیم که

$$P\left(a<\frac{(n-۱)s^{۲}}{\sigma^{۲}}<b\right)=۱-\alpha$$

$$P\left(\chi^{۲}_{\frac{\alpha}{۲},(n-۱)}<\frac{(n-۱)s^{۲}}{\sigma^{۲}}<\chi^{۲}_{۱-\frac{\alpha}{۲},(n-۱)}\right)=۱-\alpha$$

$$P\left(\frac{(n-۱)s^{۲}}{\chi^{۲}_{۱-\frac{\alpha}{۲},(n-۱)}}<\sigma^{۲}<\frac{(n-۱)s^{۲}}{\chi^{۲}_{\frac{\alpha}{۲},(n-۱)}}\right)=۱-\alpha$$



براورد واریانس جامعه

یک فاصله‌ی اطمینان $(۱-\alpha)۱۰۰\%$ برای واریانس جامعه‌ی نرمال $\sigma^۲$ عبارت است از

$$\sigma^۲ \in \left(\frac{(n-۱)s^۲}{\chi^۲_{۱-\frac{\alpha}{۲},(n-۱)}} \; , \; \frac{(n-۱)s^۲}{\chi^۲_{\frac{\alpha}{۲},(n-۱)}} \right)$$

در رابطه‌ی بالا $s^۲$ واریانس یک نمونه‌ی n-تایی است.

یک نمونه تصادفی ۸-تایی از یک نوع سیگار به طور متوسط دارای ۱۸/۶ میلی‌گرم نیکوتین با انحراف استاندارد ۲/۴ میلی‌گرم است. با فرض نرمال بودن میزان نیکوتین در این نوع سیگار یک فاصله اطمینان ۹۰ درصدی برای واریانس نیکوتین این نوع سیگار پیدا کنید.

راه‌حل:

$$1-\alpha = 0/90 \quad \Rightarrow \quad \alpha = 0/1$$

$$\Rightarrow \quad \frac{\alpha}{2} = \frac{0/1}{2} = 0/05 \quad \Rightarrow \quad \chi^2_{0/05,(8-1)} = 2/17$$

$$\Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/1}{2} = 0/95 \quad \Rightarrow \quad \chi^2_{0/95,(8-1)} = 14/1$$

$$\sigma^2 \in \left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}} \;,\; \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},(n-1)}}\right)$$

$$\sigma^2 \in \left(\frac{(8-1)\times(2/4)^2}{14/1} \;,\; \frac{(8-1)\times(2/4)^2}{2/17}\right)$$

$$\sigma^2 \in (2/86 \;,\; 18/58)$$

مثال ۲۵

صفحه‌های پلاستیکی که توسط یک ماشین تولید می‌شود به طور متناوب مورد بازبینی قرار می‌گیرند تا تفاوت‌های ضخامت آن‌ها بررسی گردد. ناهمگنی در غلظت ماده‌ای که به کار می‌رود، وجود تفاوت‌هایی در ضخامت صفحه‌ها را غیر قابل اجتناب می‌کند. در ۱۰ صفحه‌ی تولید شده در یک نوبت کاری، اندازه‌های ضخامت بر حسب میلی‌متر به صورت مقابل بوده است: ۲۲۵ ۲۳۲ ۲۲۸ ۲۲۷ ۲۲۹ ۲۲۵ ۲۳۰ ۲۲۶ ۲۲۸ ۲۲۶

یک فاصله‌ی اطمینان ۹۵ درصدی برای انحراف معیار واقعی ضخامت صفحه‌های تولید شده در این نوبت کاری بسازید.

راه‌حل:

$$1-\alpha=0/95 \Rightarrow \alpha=0/05 \Rightarrow \begin{cases} \frac{\alpha}{2}=\frac{0/05}{2}=0/025 \Rightarrow \chi^2_{0/025,(10-1)}=2/7 \\ 1-\frac{\alpha}{2}=1-\frac{0/05}{2}=0/975 \Rightarrow \chi^2_{0/975,(10-1)}=19 \end{cases}$$

$$\bar{x}=\frac{226+228+\cdots+225}{10}=227/6$$

$$s^2=\frac{1}{n-1}\sum(x_i-\bar{x})^2=\frac{1}{9}\left[(226-227/6)^2+\cdots+(225-227/6)^2\right]=5/156$$

$$\sigma^2\in\left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}},\ \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},(n-1)}}\right) \Rightarrow \sigma^2\in\left(\frac{(10-1)\times 5/156}{19},\ \frac{(10-1)\times 5/156}{2/7}\right)$$

$$\sigma^2\in(2/442\ ,\ 17/187) \Rightarrow \sigma\in(1/563\ ,\ 4/146)$$