برآوردیابی
- آمار و احتمالات مهندسی -

مدرس: مشکانی فراهانی

دانشگاه صنعتی امیرکبیر

۲۴ مرداد ۱۴۰۱

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.

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 $T = T(X_1, \dots, X_n)$ 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

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

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

برآورد میانگین جامعه

μ

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 $P(L < \mu < U) = ۱ - \alpha$.

براورد میانگین جامعه
حالت الف- واریانس جامعه معلوم

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

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

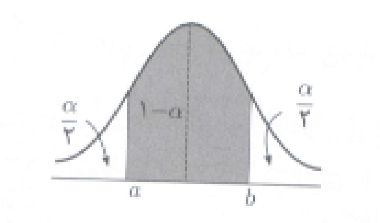
نقطه‌ی b نقطه‌ای روی محور افقی نمودار تابع چگالی نرمال استاندارد است که مساحت زیر منحنی قبل از این نقطه $۱ - \frac{\alpha}{۲}$ است. این نقطه را با $z_{۱-\frac{\alpha}{۲}}$ نشان می‌دهیم. این نقطه را با استفاده از جدول نرمال استاندارد به دست می‌آوریم.

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:

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



براورد میانگین جامعه
حالت الف- واریانس جامعه معلوم

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

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

که در آن $\bar{x}$ میانگین نمونه‌ی تصادفی و $z_{1-\frac{\alpha}{2}}$ مقدار متغیر نرمال استاندارد است که سطح زیر منحنی در سمت چپ آن $1-\frac{\alpha}{2}$ باشد.

مثال ۱

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:

$$1-\alpha = 0/90 \quad \Rightarrow \quad \alpha = 0/1 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/1}{2} = 0/95$$
$$\Rightarrow \quad z_{0/95} = 1/645$$

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$
$$\mu \in \left(20 - 1/645 \times \frac{2}{\sqrt{25}} \ , \ 20 + 1/645 \times \frac{2}{\sqrt{25}}\right)$$
$$\mu \in (19/342 \ , \ 20/658)$$

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:

$$P\left(-t_{1-\frac{\alpha}{2},(n-1)} < \frac{\bar{X}-\mu}{s/\sqrt{n}} < t_{1-\frac{\alpha}{2},(n-1)}\right) = 1-\alpha$$

$$P\left(\bar{X} - t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

براورد میانگین جامعه
حالت ب- واریانس جامعه نامعلوم

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.

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

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:

$$۱ - \alpha = ۰/۹۹ \quad \Rightarrow \quad \alpha = ۰/۰۱ \quad \Rightarrow \quad ۱ - \frac{\alpha}{۲} = ۱ - \frac{۰/۰۱}{۲} = ۰/۹۹۵$$
$$\Rightarrow \quad t_{۰/۹۹۵,(۸-۱)} = ۳/۵$$

$$\mu \in \left(\bar{x} - t_{۱-\frac{\alpha}{۲},(n-۱)} \frac{s}{\sqrt{n}} \ , \ \bar{x} + t_{۱-\frac{\alpha}{۲},(n-۱)} \frac{s}{\sqrt{n}}\right)$$
$$\mu \in \left(۱۸/۶ - ۳/۵ \times \frac{۲/۴}{\sqrt{۸}} \ , \ ۱۸/۶ + ۳/۵ \times \frac{۲/۴}{\sqrt{۸}}\right)$$
$$\mu \in (۱۵/۶۳ \ , \ ۲۱/۵۷)$$

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:

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

$$\Rightarrow \quad z_{0/995} = 2/575$$

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

$$\mu \in \left(0/824 - 2/575 \times \frac{0/42}{\sqrt{200}} \ , \ 0/824 + 2/575 \times \frac{0/42}{\sqrt{200}}\right)$$

$$\mu \in (0/747 \ , \ 0/901)$$

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

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

مثال ۷

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:

$$1-\alpha = ۰/۹۰ \quad \Rightarrow \quad \alpha = ۰/۱ \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{۰/۱}{2} = ۰/۹۵$$

$$\Rightarrow \quad t_{۰/۹۵,(۱۷)} = ۱/۷۴$$

$$|\bar{x}-\mu| < t_{1-\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} = ۱/۷۴ \times \frac{۲/۷۵}{\sqrt{۱۸}} = ۱/۱۳$$

تعیین حجم نمونه

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 براورد 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:

$$1-\alpha = 0/9 \quad \Rightarrow \quad \alpha = 0/1 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 1-\frac{0/1}{2} = 0/95$$
$$\Rightarrow \quad z_{0/95} = 1/645$$

$$n = \left(z_{1-\frac{\alpha}{2}}\frac{\sigma}{e}\right)^2 \qquad \Rightarrow \qquad n = \left(1/645 \times \frac{0/3}{0/05}\right)^2$$
$$\Rightarrow \qquad n = 97/42 \qquad \Rightarrow \qquad n = 98$$

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 نسبت در جامعه
p

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 ۱۲

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 ۱۳

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 ۱۴

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:

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

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

مثال ۱۵

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:

$$۱-\alpha=۰/۹۹ \quad\Rightarrow\quad ۱-\frac{\alpha}{۲}=۱-\frac{۰/۰۱}{۲}=۰/۹۹۵ \quad\Rightarrow\quad z_{۰/۹۹۵}=۲/۵۷$$

$$n=\frac{z^{۲}_{۱-\frac{\alpha}{۲}}\hat{p}\hat{q}}{e^{۲}}=\frac{(۲/۵۷)^{۲}\times ۰/۵\times ۰/۵}{(۰/۰۱)^{۲}}=۱۶۵۱۲/۲۵ \quad\Rightarrow\quad n=۱۶۵۱۳$$

براورد واریانس جامعه
σ^2

براورد واریانس جامعه

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

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

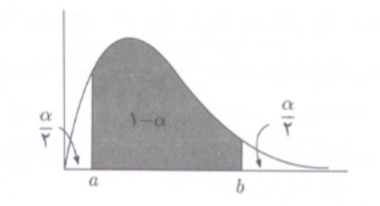
اگر جامعه نرمال باشد، در فصل قبل اثبات کردیم: $\frac{(n-۱)s^۲}{\sigma^۲} \sim \chi^۲_{(n-۱)}$

اعداد a و b را چنان تعیین می‌کنیم که

$$P\left(a < \frac{(n-1)s^2}{\sigma^2} < b\right) = 1-\alpha$$

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

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



براورد واریانس جامعه

یک فاصله‌ی اطمینان $(1-\alpha)100\%$ برای واریانس جامعه‌ی نرمال σ^2 عبارت است از

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

در رابطه‌ی بالا s^2 واریانس یک نمونه‌ی n-تایی است.

مثال ۱۶

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 ۱۷

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:

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

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

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

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

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