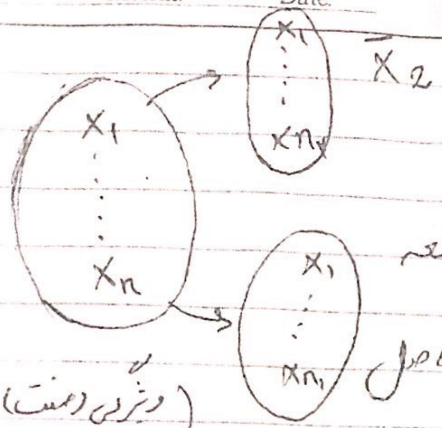
جزوه روش‌های آماری

Subject:
Year: Month: Date:

بخش اول:



$\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ (n : حجم نمونه)

هدف اصلی درس: نمونه‌گیری و سپس تحلیل جامعه و استنباط پیرامون جامعه بر اساس نتایج حاصل از نمونه.

X: ویژگی (صفت)
N: حجم جامعه

جامعه آماری: مجموعه‌ای از تمام عناصر است که به گروه تعریف شده‌ای تعلق دارد. به عبارت دیگر مجموعه‌ی افراد یا چیزهایی که می‌خواهیم یک یا چند ویژگی آنها را مطالعه کنیم.

نمونه: قسمتی از جامعه آماری است که طبق ضوابطی انتخاب می‌شود و مطالعه‌ی آن به جای مطالعه‌ی کل جامعه مقدور است؛ نمونه‌گیری باعث کاهش هزینه و افزایش سرعت می‌شود و همچنین بالا بردن توان و کیفیت کار، عدم صدمه وارد کردن به واحدهای جامعه.

پارامتر: ویژگی عددی جامعه مربوط به صفت مورد بررسی مثل میانگین (μ) جامعه.

σ^2 ← واریانس جامعه p ← نسبت در جامعه که در صورتی که جامعه امکان مطالعه کل پذیر باشد مقدار دقیق آن را می‌توان محاسبه کرد.

آماره ← تابعی از نمونه تصادفی می‌باشد.

پارامتر: بعضی مواقع علاقه‌مند به اطلاع از برخی ویژگی‌های عددی یک جامعه می‌باشیم

Subject:
Year: Month: Date:

معمولاً به ویژگی های عددی یک جامعه پارامتر می گویند، پارامترها را معمولاً با حروف

یونانی مثل θ (تتا)، μ (مو)، η (اتا)، ν (نو) و ... نشان می دهند. دقت کنید پارامترها

مقادیر ثابت هستند اما نامعلوم که در صورت مطالعه ی کل جامعه می توان مقدار

آن را به طور دقیق به دست آورد. در عمل فقط به نمونه ای از جامعه دسترسی

داریم و استنباط ما در مورد پارامترهای جامعه بر نمونه متکی است یعنی هدف یافتن

تابعی از مشاهدات نمونه است که پس از جایگذاری داده ها در تابع، مقدار حاصل

به پارامتر جامعه نزدیک باشد.

آماره: تابعی از نمونه تصادفی است: دقت کنید که آماره متغیر تصادفی است،

آماره ها متغیر تصادفی هستند چون مقدار آماره از نمونه ای به نمونه دیگر متفاوت

است.

* به عنوان یکی از مطالب مهم این درس انتخاب یکی از این آماره ها به عنوان

برآورد پارامتر می باشد. برآورد دو نوع است: ۱- برآورد نقطه ای ۲- برآورد فاصله ای

برآورد نقطه ای یک پارامتر: هدف از برآورد نقطه ای آن است که از روی

نمونه عددی بدست آوریم که انتظار داریم به مقدار نامعلوم پارامتر جامعه نزدیک

باشد. فرض کنید پارامتر جامعه θ باشد برآورد نقطه ای آن را با $\hat{\theta}$ (تتا هت)

Sadra

* پارامتر مربوط به جامعه و آماره مربوط به نمونه است *

* ما هیچ وقت θ را نمیدانیم چون متغیر تصادفی است *

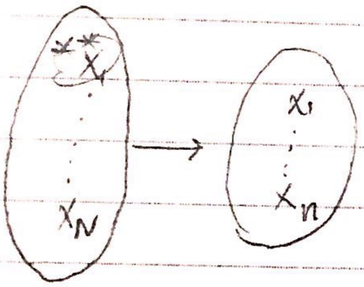
Subject:
Year: Month: Date:

نشان میدهیم. این برآورد یکی از آماره های نمونه است. اصل سوال این

است که کدام آماره به عنوان برآوردکننده انتخاب شود. بنابراین در این

قسمت ویژگی های یک برآوردکننده خوب را بیان میکنیم

کدام را به عنوان $\hat{\theta}$ انتخاب کنیم.



$$\begin{cases} T_2(x_1 \dots x_n) = \bar{X} \\ T_1(x_1 \dots x_n) = \min(x_1 \dots x_n) \\ T_3(x_1 \dots x_n) = \text{median}(x_1 \dots x_n) \end{cases}$$

پارامتر جامعه $\theta \longrightarrow \hat{\theta}$ برآورده $\longrightarrow$ یکی از آماره

۱. برآوردکننده نااریب باشد یعنی مقدار واقعی θ هر چه باشد داشته باشیم:

$E(\hat{\theta}) = \theta \rightarrow$ عدد ثابت

* برای پارامترها امید ریاضی نداریم.

نمونه تصادفی: فرض کنید $X_1, \dots, X_n$ را از جامعه آماری با صفت X (میانگین) نمونه

تصادفی می گوییم به طوری که $X_1 \dots X_n$ دو به دو از هم مستقلند و هم توزیع با

X هستند.

$X \sim F(\mu, \sigma^2)$ هر توزیعی $\longrightarrow$ $X_1 \equiv X$ (هم توزیع) $\vdots$ $X_n \equiv X$

$$X_1 \sim F(\mu, \sigma^2)$$
$$\vdots$$
$$X_n \sim F(\mu, \sigma^2)$$

Subject:
Year: Month: Date:

نکته: از بین آماره هایی که بهتر از بقیه است به عنوان برآورد کننده پارامتر جامعه انتخاب می کنیم. به بهتر بودن تعریف دارد: برآوردی بهتر است که خواص زیر را داشته باشد.

① نااریبی: فرض کنید پارامتر مورد بررسی جامعه θ باشد در اینصورت $\hat{\theta}$ را یک برآورد کننده نااریب می گوییم هرگاه: $E(\hat{\theta})=\theta$

مثال) فرض کنید $x_1, \dots, x_n$ یک نمونه تصادفی از جامعه ای با میانگین μ و واریانس σ^2 باشد. تحت چه شرایطی برآورد کننده (برآوردگر)

$T=\sum_{i=1}^{n} a_i x_i$ (a_i ثابت) یک برآورد کننده نااریب برای μ است. پس باید $E(T)=\mu$

$$E(x)=\sum x \cdot f(x)$$

$$E(T)=E\left(\sum a_i x_i\right)=E(a_1x_1+\dots+a_nx_n)$$

$$=E(a_1x_1)+E(a_2x_2)+\dots+E(a_nx_n)$$

$$=a_1\underbrace{E(x_1)}_{=\mu}+a_2E(x_2)+\dots+a_nE(x_n)$$

$$=a_1\mu+a_2\mu+\dots+a_n\mu=\mu(a_1+\dots+a_n)$$

$\Rightarrow \sum a_i=1$ * شرط برقرار است a_i ها باید ۱ باشند.

Sadra

$E(x^2) = Var(x) + E^2(x)$

$V(x) = E(x)^2 - E^2 x$

میانگین نمونه $V(\bar{x}) = \frac{\sigma^2}{n}$ واریانس جامعه ← حجم نمونه

Subject:
Year: Month: Date:

مثال) فرض کنید $X_1 \dots X_n$ نمونه تصادفی از جامعه با صفت X و میانگین μ و واریانس σ^2 باشد نشان دهید:

$E(\bar{X}) = \mu$

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

تغیرات نمونه $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ و $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$ که در آن

$E(\bar{X}) = \mu$

$E\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right) = \frac{1}{n} E\left(\sum x_i\right) = \frac{1}{n} E(x_1 \dots x_n)$

$= \frac{1}{n} E(x_1) + E(x_2) + \dots + E(x_n) = \frac{1}{\cancel{n}} \times \cancel{n}(\mu) = \mu$

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

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

Sadra

حل:

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i^2+\bar{x}^2-2x_i\bar{x})\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left[E(x_i^2)+E(\bar{x})^2-2E(x_i\bar{x})\right] \quad (*)$$

خواص امید ریاضی

از طرفی میدانیم $V(x_i) = E(x_i)^2 - E^2(x_i)$

$$① \; E(x_i)^2 = \sigma^2 + E^2(x_i) = \sigma^2 + \mu^2 \qquad ② \; E(\bar{x})^2 = V(\bar{x}) = \frac{\sigma^2}{n} + \mu^2$$

از طرفی $\sum_{i=1}^{n} 2E(x_i\bar{x})$ را میتوان به صورت زیر نوشت:

$$③ \; 2E\left(\sum_{i=1}^{n} x_i\bar{x}\right) = 2nE(\bar{x})^2 = 2n\frac{\sigma^2}{n} + 2n\mu^2$$

با جایگذاری ①, ②, ③ در (*)

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) + \sum_{i=1}^{n}\left(\frac{\sigma^2}{n}+\mu^2\right) - 2n\frac{\sigma^2}{n} - 2n\mu^2\right]$$

$$= \frac{1}{n-1}\left[-\sigma^2 + n\sigma^2\right] = \sigma^2$$

x_1 تا x_n نمونه تصادفی است دو به دو مستقل‌اند کوواریانس‌ها صفر

Subject
Year: Month: Date:

محاسبه واریانس میانگین

$$V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} V(x_1+\cdots+x_n) = \frac{1}{n^2}\left[V(x_1)+\cdots+V(x_n)\right.$$

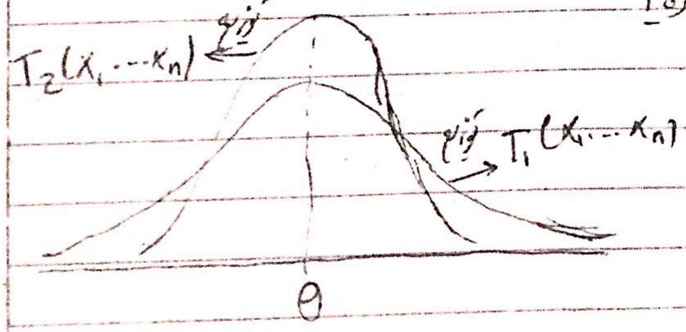
$$\left. + \text{(کوواریانس‌ها)}\right] = \frac{1}{n^2}\left[\sigma^2+\cdots+\sigma^2\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

خواص واریانس

$$\begin{cases} V(x+y) = V(x) + V(y) + 2\operatorname{cov}(x,y) \\ V(ax) = a^2 V(x) \\ V(a) = 0 \quad a \text{ ثابت} \end{cases}$$

$$X \sim F(\mu, \sigma^2) \quad \frac{x_1 \ldots x_n}{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i} \quad \begin{array}{l} \nearrow E(\bar{x}) = \mu \\ \searrow V(\bar{x}) = \frac{\sigma^2}{n} \end{array}$$

(۲) اکنون سوال این است که بین دو آماره که هر دو نااریب هستند برای پارامتر مورد نظر جامعه (مثلاً θ) کدام را به عنوان برآوردکننده انتخاب کنیم. فرض کنید X متغیر تصادفی از هر توزیعی با میانگین μ و واریانس σ^2 باشد و $x_1 \ldots x_n$ نمونه‌های تصادفی از X باشند و $\underbrace{T_1(x_1 \ldots x_n)}_{\text{آماره ۱}}$ و $\underbrace{T_2(x_1 \ldots x_n)}_{\text{آماره ۲}}$ هر دو برای θ نااریب باشند و توزیعی به صورت شکل زیر داشته باشند



$$E(T_1(x_1 \ldots x_n)) = \theta$$

$$E(T_2(x_1 \ldots x_n)) = \theta$$

با توجه به شکل ترجیح میدهیم برآوردکننده نااریبی را انتخاب کنیم که پراکندگی کمتری

Sadra

*mean square error

« میانگین مربع خطا »

Subject:
Year: Month: Date:

به عبارت دیگر واریانس کوچکتری داشته باشد برای این مثال چون T_2

واریانس کوچکتری دارد پس T_2 را بر T_1 ترجیح میدهیم.

نکته: در صورتیکه بین برآوردهای اریب میخواهیم برآورد خوب انتخاب کنیم

معمولاً برآوردی را انتخاب میکنیم که MSE^* کوچکتری داشته باشد که به صورت

زیر تعریف میشود:

$$\begin{cases} MSE = \underbrace{var(\hat{\theta})}_{\text{مقدار اریبی}} + \overbrace{bias^2(\hat{\theta})}^{\text{بایاس}} \\ E(\hat{\theta}) \neq \theta \quad \text{یا} \quad E(\hat{\theta}) = \theta + \underbrace{bias}_{\text{اریبی}} \end{cases}$$

تمرین: ثابت کنید:

$$MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = var(\hat{\theta}) + bias^2(\hat{\theta})$$

« به توان برسون و میانگین بگیر »

« ادامه مثال قبل »

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i^2+\bar{x}^2 - 2x_i\bar{x})\right) \underset{\text{خواص امید ریاضی}}{=} \frac{1}{n-1}\sum_{i=1}^{n}\left[E(x_i^2)+E(\bar{x})^2 - 2E(x_i\bar{x})\right] \circledast$$

از طرفی میدانیم $V(x_i) = E(x_i)^2 - E^2(x_i)$

$$① \; E(x_i)^2 = \sigma^2 + E^2(x_i) = \sigma^2 + \mu^2 \qquad ② \; E(\bar{x})^2 = V(\bar{x}) = \frac{\sigma^2}{n} + \mu^2$$

از طرفی $\sum_{i=1}^{n} 2E(x_i\bar{x})$ را میتوان به صورت زیر نوشت:

$$③ \; 2E\left(\sum_{i=1}^{n} x_i\bar{x}\right) = 2nE(\bar{x})^2 = 2n\frac{\sigma^2}{n} + 2n\mu^2$$

با جایگذاری ①، ②، ③ در $\circledast$

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) + \sum_{i=1}^{n}\left(\frac{\sigma^2}{n}+\mu^2\right) - 2n\frac{\sigma^2}{n} - 2n\mu^2\right]$$

Sadra

$$= \frac{1}{n-1}\left[-\sigma^2 + n\sigma^2\right] = \sigma^2$$

Subject:
Year: Month: Date:

نکته: برآوردکننده‌ای که ناریب باشد و دارای کمترین واریانس و البته در فضای پارامتر تعریف شده باشد یک برآوردگر خوب یا برآوردگر ناریب با واریانس مینیمم نامیده میشود. خصوصیات دیگری برای برآوردکننده تعریف میشود که به شرح زیر است:

③ سازگاری: سازگاری یعنی اگر حجم نمونه را افزایش دهیم $\hat{\theta}$ به θ میل کند احتمال اختلاف آنها به اندازه ϵ هم صفر باشد یعنی

$$\forall \epsilon > 0 \quad \lim_{n \to \infty} P(|\hat{\theta} - \theta| \geq \epsilon) = 0$$

تمرین: از نامساوی چبیشف ثابت کنید عبارت بالا معادل است با $\lim_{n \to \infty} Var(\hat{\theta}_n) = 0$

Sadra

Subject:
Year: Month: Date:

کارایی یک برآوردکننده: یک معیار برای مقایسه دو برآوردکننده کارایی است. فرض کنید $u(X)$ و $v(X)$ دو برآورد پارامتر θ باشند درانصورت کارایی این دو برآوردکننده به صورت زیر تعریف می‌شود:

$$e = \frac{Var(u)}{Var(v)}$$

$e<1$ u کاراتر از v است

$e>1$ v کاراتر از u است

مثال) برای یک نمونه تصادفی دوتایی X_1, X_2 از توزیع $X \sim N(\mu, 1)$ داریم: ($E(X)=\mu$، $Var(X)=1$، پارامتر)

$$u(X) = \frac{X_1+X_2}{2} \quad , \quad T(X) = 2X_1 - X_2$$

۱) آیا $u(X)$ و $T(X)$ برای μ نااریبند؟

۲) کدام برآوردکننده بهتر است؟

۳) کارایی این دو برآوردکننده را مقایسه کنید.

① $E(u(x)) = E\left(\frac{X_1+X_2}{2}\right)$

$= \frac{1}{2}\left[\underbrace{E(X_1)}_{\mu} + \underbrace{E(X_2)}_{\mu}\right]$

$= \frac{1}{2}[2\mu] = \mu$

$E(T(x)) = E(2X_1 - X_2) = 2E(X_1 - X_2) = 2\underbrace{E(X_1)}_{\mu} - \underbrace{E(X_2)}_{\mu}$

$= 2\mu - \mu = \mu$

② $v(u(x)) = v\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}v(X_1+X_2) = \frac{1}{4}\left[\underbrace{var(X_1)}_{=1}\right.$

$\left. + \underbrace{var(X_2)}_{=1} + 2\underbrace{cov(X_1, X_2)}_{0}\right] = \frac{1}{2}$

نمونه تصادفی X_1, X_2 مستقلند

$v(T(x)) = v(2X_1 - X_2) = 4v(X_1) + v(X_2) - 4cov(X_1, X_2) = 5$

Sadra

$\mu(X)$ نسبت به $T(X)$ واریانس کوچکتری دارد پس بهتر است.

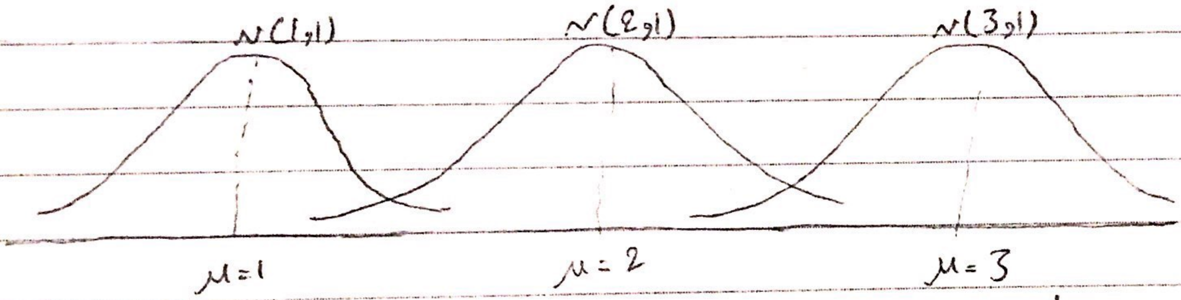
③ $e = \frac{Var(U)}{Var(T)} = \frac{\frac{1}{2}}{5} = \frac{1}{10} < 1$ U، ده برابر کاراتر از T است.

"بخش دوم"

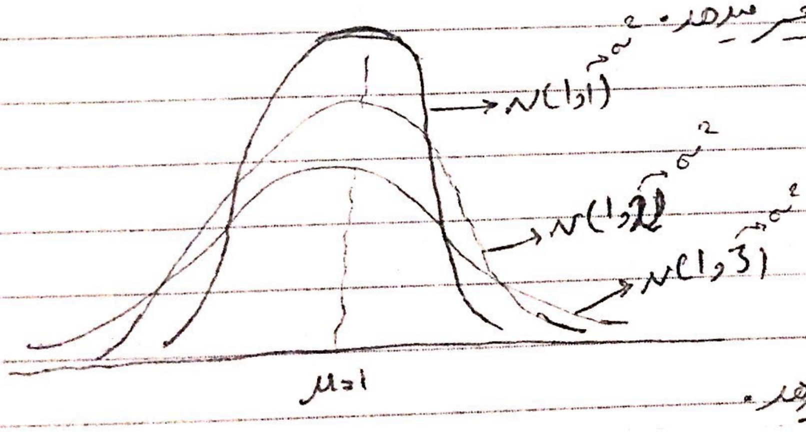
- توزیع نرمال: فرض کنید X یک متغیر تصادفی پیوسته با تابع چگالی زیر باشد:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty$$

و با نماد $N(\mu,\sigma^2)$ نشان میدهیم که به μ میانگین یا پارامتر مکان و به σ^2 واریانس یا پارامتر شکل (مقیاس) می گویند. $E(X) = \mu$ و $V(X) = \sigma^2$



* پس μ مکان توزیع را تغییر میدهد.

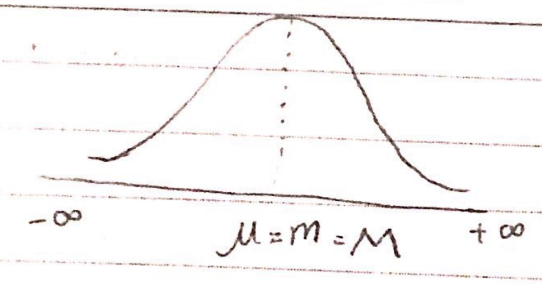


* پس σ^2 شکل توزیع را تغییر میدهد.

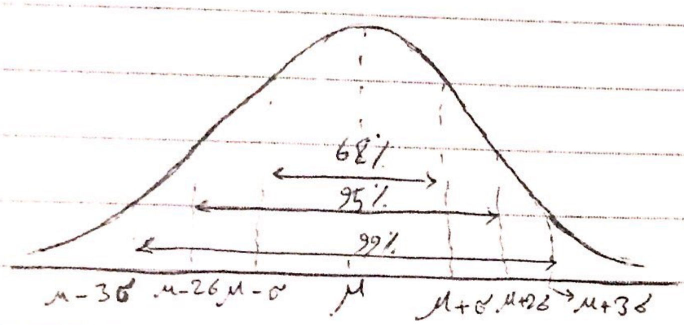
- با توجه به خواص خوبی که توزیع نرمال دارد معمولاً این توزیع مورد توجه است؛ از جمله خواص این توزیع میتوان به موارد زیر اشاره کرد:

① دامنه تعریف X، کل اعداد حقیقی است $(-\infty, +\infty)$

② توزیع متقارن است و شکل زنگی دارد:



③ بازه بندی توزیع مشخص است یعنی اگر $X \sim N(\mu, \sigma^2)$ داریم:



④ توزیع نرمال نسبت به ترکیبات خطی بسته است.

if $X \sim N(\mu, \sigma^2)$; $\underset{\text{دو عدد ثابت}}{a, b} \Rightarrow aX + b \sim N$

$$\begin{cases} E(aX+b) = aE(X) + b = a\mu + b \\ V(aX+b) = a^2 V(X) = a^2 \sigma^2 \end{cases}$$

if $X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$ $\underset{\text{مستقل از } X_2 \perp X_1}{\Longrightarrow}$ $X_1 , X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

⑤ تحت شرطی بودن بسته است:

$(X_1, X_2) \sim N_2$ و $X_1 | X_2 \sim N_1$ $X_2 | X_1 \sim N_1$

⑥ تحت حاشیه سازی هم بسته است:

$(X_1, X_2) \sim N_2 \Rightarrow X_1 \sim N_1$

$X_2 \sim N_1$

Subject:
Year: Month: Date:

یعنی، [illegible]؟

مثال) نشان دهید $E(S_n^2) \neq \sigma^2$

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

نکته: بعد از ساده کردن S^2، می‌توان از فرمول زیر هم استفاده کرد:

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

فرمول مهم است (حتماً یاد بگیر)

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i^2 + \underbrace{\sum_{i=1}^{n} \bar{x}^2}_{n\bar{x}^2} - \underbrace{2\sum_{i=1}^{n} x_i \bar{x}}_{2n\bar{x}^2} \right]$$

«مهم» تمرین) نشان دهید اگر $X \sim N(\mu, \sigma^2)$ باشد و $X_1, \dots, X_n$ نمونه تصادفی از X باشند آنگاه $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

چون

$X \sim N(\mu, \sigma^2)$

$$\begin{cases} X_1 \sim N(\mu, \sigma^2) \\ \vdots \\ X_n \sim N(\mu, \sigma^2) \end{cases}$$

طبق خاصیت ۴ نرمال تحت ترکیبات خطی بسته است پس براین ترکیب نرمال خطی است.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} (X_1 + \dots + X_n)$$

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$\mu \to E(\bar{X})$

$\frac{\sigma^2}{n} \to V(\bar{X})$

$$X \sim N(\mu, \sigma^2) \xrightarrow[\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i]{\text{نمونه تصادفی } X_1, \dots, X_n} \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Sadra

قضیه حد مرکزی : اگر X متغیر تصادفی از هر توزیعی با میانگین μ و واریانس σ^2 باشد و $X_1 \dots X_n$ نمونه تصادفی "به حد کافی بزرگ" از X باشد آنگاه $\bar{X}$ دارای توزیع نرمال است.

$$\underset{\text{هر توزیعی}}{X \sim F(\mu, \sigma^2)} \xrightarrow[\text{نمونه تصادفی}]{x_1 \dots x_n} \bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

$\boxed{n \geq 30}$ — معمولاً n

توزیع نرمال استاندارد: هر توزیع نرمالی را میتوان به توزیع نرمال استاندارد تبدیل کرد، فرض کنید X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، آنگاه متغیر تصادفی جدید Z به صورت :

$$Z = \frac{X-\mu}{\sigma}$$

تعریف میشود دارای توزیع $N(0,1)$ است که به توزیع نرمال استاندارد معروف است و مقادیر $P(Z < z)$ به ازای z های مختلف با استفاده از نرم افزار محاسبه شده است و در انتهای اکثر کتب آماری موجود است.

مثال) یک شرکت تولیدی لاستیک اتومبیل، لاستیک‌های تولید میکند، طول عمر این لاستیک‌ها دارای توزیع نرمال با میانگین 24 ماه و انحراف معیار 2 ماه است! احتمال اینکه در یک نمونه تصادفی 25 تایی از لاستیک‌ها میانگین طول عمر کمتر از 25 ماه باشد را بیابید.

$X \sim N(24, 4)$ $\quad \mu = 24, \ \sigma = 2, \ n = 25$

$\Rightarrow \bar{X} \sim N\left(24, \frac{4}{25}\right)$

$$P(\bar{x} < 25) = P\Bigg(\underbrace{\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}_{Z} < \underbrace{\frac{25-24}{\frac{2}{5}}}_{z}\Bigg) = P(Z < 2.5) = 0.9938$$

Subject:
Year: Month: Date:

مثال) فرض کنید مقدار سالهای تحصیل در بین افراد بالغ در کشوری دارای میانگین ۱۱٫۱ سال و انحراف معیار سه سال باشد، احتمال اینکه در یک نمونه تصادفی صدتایی از افراد متوسط تعداد سالهای تحصیل بین ۱۱ تا ۱۲ سال باشد را بیابید.

$\mu = 11{,}1 \qquad \sigma = 3 \qquad n = 100$

$P(11 < \bar{X} < 12)$

چون جامعه نرمال نیست اما n به اندازه کافی بزرگ است پس از قضیه‌ی حد مرکزی استفاده میکنیم.

$X \sim f(11{,}1, 9) \xrightarrow{n \geq 30} \bar{X} \sim N(11{,}1, 9/100)$

$E(\bar{X}) = \mu = 11{,}1 \qquad V(\bar{X}) = \frac{\sigma^2}{n} = \frac{9}{100}$

$P\left(\frac{11-11{,}1}{0{,}3} < \underbrace{Z}_{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}} < \frac{12-11{,}1}{0{,}3}\right) = P(-0{,}33 < Z < 3)$ $P(Z<z$

$= P(Z<3) - P(Z<-0{,}33)$

$= \Phi(3) - \Phi(-0{,}33)$

$= 0{,}9987 - 0{,}3707 = 0{,}628$

توزیع کای دو: فرض کنید جامعه‌ای دارای توزیع نرمال است و X متغیر تصادفی مورد بررسی:

$X \sim N(\mu, \sigma^2)$

و $X_1 \dots X_n$ نمونه‌های تصادفی از این جامعه، طبق تعریف نمونه تصادفی:

$X_1 \dots X_n \sim N(\mu, \sigma^2) \rightsquigarrow X_i \sim N(\mu, \sigma^2)$
$i = 1, \dots, n$

Sadra

* درصورت استفاده از برآورد میانگین به جای پارامتر و با وجود محدودیت در مدل باید از درجه آزادی به همان تعداد کم کنیم *

Subject:
Year: Month: Date:

از طرفی (از قبل) می‌دانیم $Z_i = \dfrac{x_i - \mu}{\sigma}$ اکنون با یک متغیر جدید زیر داریم

$i = 1, \dots, n$

توزیع کای دو است: $Z_i^2 = \dfrac{(x_i-\mu)^2}{\sigma^2} \sim \chi^2_1$

$$\sum_{i=1}^{n} z_i^2 \sim \chi^2_{n}$$

درجه آزادی توزیع

مجموع کای دو (خی دو)

نکته: اگر در جامعه μ مجهول باشد و در رابطه: $\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \dfrac{(x_i-\mu)^2}{\sigma^2}$

به جای μ از $\bar{X}$ استفاده می‌شود از درجه آزادی، یک کم می‌شود

میانگین جامعه — میانگین نمونه

$$* \quad \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} \sim \chi^2_{n-1}$$

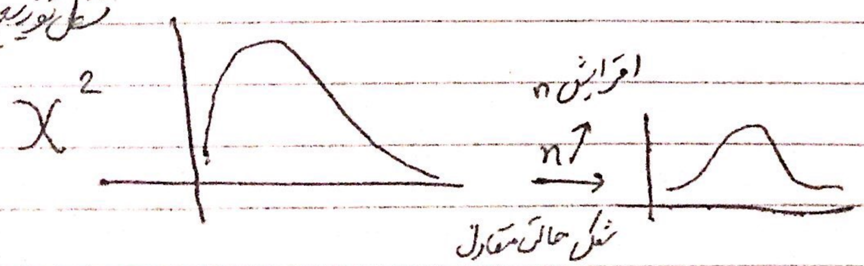
نکته: همچنین ثابت می‌شود که:

$$\text{if } \overset{y}{\uparrow}{y} \sim \chi^2_n \Longrightarrow E(y) = n$$

$$V(y) = 2n$$

نکته: توزیع کای دو به درجه آزادی حساس است و با افزایش درجه آزادی توزیع کای دو به نرمال همگرا می‌شود $(n \geq 30)$.

شکل توزیع



Subject:
Year: ۱۴ Month: Date: $X \sim N(\mu, \sigma^2)$

توزیع S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \xrightarrow{\sigma^2/\sigma^2} \frac{\sigma^2}{n-1}\sum_{i=1}^{n}\underbrace{\left(\frac{X_i-\bar{X}}{\sigma}\right)^2}$$

$X_i \sim N(\mu, \sigma^2)$

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

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

طبق * در صفحه قبل

اکنون داریم:

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1$$

$$\frac{(n-1)}{\sigma^2}E(S^2) = n-1 \Rightarrow E(S^2) = \sigma^2$$

به همین ترتیب S^2 برای σ^2 نااریب است.

به همین ترتیب:

$$V\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma^4}V(S^2) = 2(n-1) \Rightarrow V(S^2) = \frac{2\sigma^4}{(n-1)}$$

$$\lim_{n\to\infty} V(S^2) = 0 \Rightarrow$$ S^2 برای σ^2 سازگار است.

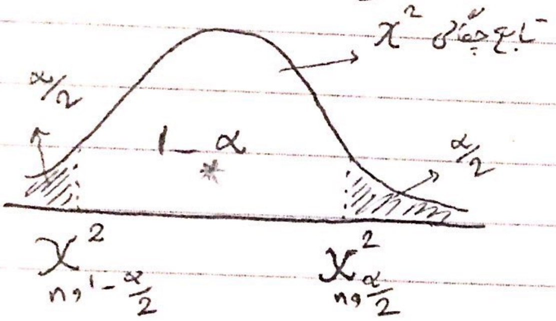
تمرین) برای S_n^2 تمام مراحل بالا را طی کن. (توزیع، *) $(X_1 \cdots X_n,\ X \sim N(\mu, \sigma^2))$

و نشان دهید S_n^2 برای σ^2 نااریب است و کارایی S_n^2 را با S^2 مقایسه کنید.

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Subject:
Year: Month: Date:

نکته: توزیع کای دو به صورت چوله به راست است بنابراین:



بازای مقادیر مختلف α، مقادیر $\chi^2_{\frac{\alpha}{2}}$ و $\chi^2_{1-\frac{\alpha}{2}}$ را می‌توانید از جدول انتهای کتاب استخراج کنید.

نکته: دقت کنید اگر در مسئله‌ای میانگین معلوم باشد (μ) آنگاه در فرمولهای S^2 و S_n^2 به جای $\bar{X}$ از μ استفاده می‌شود آنگاه:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2 \quad , \quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_n \quad , \quad \Rightarrow \frac{nS_n^2}{\sigma^2} \sim \chi^2_n$$

تمرین) به عنوان تمرین در حالتی که μ معلوم است، $E(S^2)$ و $E(S_n^2)$ را بیابید و واریانس آنها را مقایسه کنید.

توزیع t استودنت: فرض کنید $X \sim N(\mu, \sigma^2)$ ($X_1, \dots, X_n \sim N(\mu, \sigma^2)$) نمونه تصادفی

آنگاه می‌دانیم: $Z = Z_i = \frac{X_i-\mu}{\sigma} \sim N(0,1)$

و

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

$$\frac{N(0,1)}{\sqrt{\chi^2_{df}/df}} = t_{df}$$

Subject:
Year: Month: Date:

وداریم:

$$T = \frac{\overset{\to N(0,1)}{Z}}{\sqrt{U/n}} \sim t_n$$

$$\sqrt{\chi^2_{df}/df} = \sqrt{\chi^2_n/n}$$

degree of freedom

درجه آزادی

* متقارن ولی پخش‌تر از نرمال

اگر درجه آزادی بالا برود مثل نرمال کشیده شده و بالا می‌رود.

درجه آزادی

$n \nearrow \quad t \approx N$

مثال) از جامعه‌ای با توزیع نرمال نمونه‌ای به حجم n انتخاب می‌شود، نشان دهید $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ دارای توزیع t استودنت با n-1 درجه آزادی است.

$X \sim N(\mu, \sigma^2) \quad X_1 \dots X_n$

$\xrightarrow{\text{قبلاً دیدیم}} \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\xrightarrow{\text{پس}} Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ ①

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

$$①, ② \Longrightarrow T = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ n-1}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

درجه آزادی ← صورت

درجه آزادی مخرج

Subject: درجه آزادی مخرج
Year: Month: Date:

مثال) نمرات یک کلاس از دانشجویان دارای توزیع نرمال با میانگین 15 است. اگر از این کلاس یک نمونه 20 تایی انتخاب کنیم و مشاهده کنیم که انحراف استاندارد نمرات (انحراف معیار (جذر واریانس)) 4,28 است. احتمال اینکه میانگین نمرات این افراد از 17 بیشتر باشد را بیابید.

$\mu = 15$

$n = 20$

$S = 4{,}28$

$$P(\bar{X} > 17) = P\left(\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{17-15}{\frac{4{,}28}{\sqrt{20}}}\right) \quad t_{n-1}$$

$$= P(T > 2{,}09) = 1 - \underbrace{P(T < 2{,}09)}_{0{,}975 \text{ از جدول}}$$

$$= 0{,}025$$

توزیع فیشر:

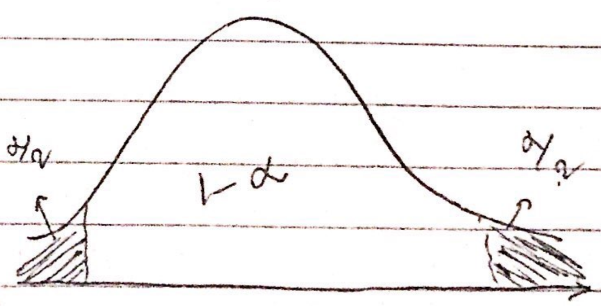
$$\text{if} \quad U_1 \sim \chi^2_{n_1} \;,\; U_2 \sim \chi^2_{n_2}$$

درجه آزادی صورت

$$\Longrightarrow \quad F = \frac{U_1/n_1}{U_2/n_2} \sim F_{n_1, n_2} \rightarrow \text{درجه آزادی مخرج}$$

توزیع فیشر

$$F_{n_1, n_2, \alpha} = \frac{1}{F_{n_2, n_1, 1-\alpha}}$$



$F_{n_1, n_2, 1-\alpha/2}$ $\qquad$ $F_{n_1, n_2, \alpha/2}$

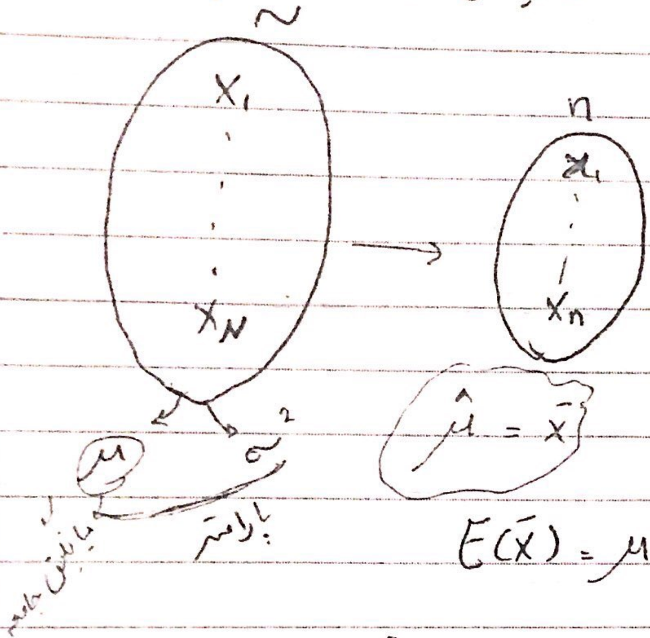
Sadra

«بخش سوم»

برآورد نقطه‌ای : همانطور که قبلاً اشاره شد یکی از اهداف اصلی این درس برآورد

نقطه‌ای سه پارامتر «میانگین جامعه μ»، «واریانس جامعه» و «نسبت p»

جامعه است.



میانگین نمونه

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

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

همچنین قبلاً دیدیم

$$E(\bar{x}) = \mu \,,\, V(\bar{x}) = \frac{\sigma^2}{n}$$

$$\lim_{n\to\infty} V(\bar{x}) = 0$$

و ثابت می‌شود که $V(\bar{x})$ بین برآوردکننده‌های نااریب کمترین واریانس است

پس $\bar{x}$ را به عنوان برآورد نقطه‌ای خوب μ می‌شناسیم.

نکته: توزیع $\bar{x}$ بستگی به توزیع نمونه دارد مثلاً اگر جامعه نرمال باشد

دیدیم چون $\bar{x}$ تابعی خطی از نمونه تصادفی است و توزیع نرمال تحت ترکیبات خطی

بسته است $\bar{x}$ هم دارای توزیع نرمال است.

$$X \sim N(\mu, \sigma^2) \xrightarrow[\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i]{x_1 \cdots x_n} \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

(نیز همان احتمال موفقیت است p)

Subject:
Year: Month: Date:

نکته: اگر $X \sim F$ (جامعه هر توزیعی داشته باشد) به طوری که نمونه استخراج

هر توزیعی

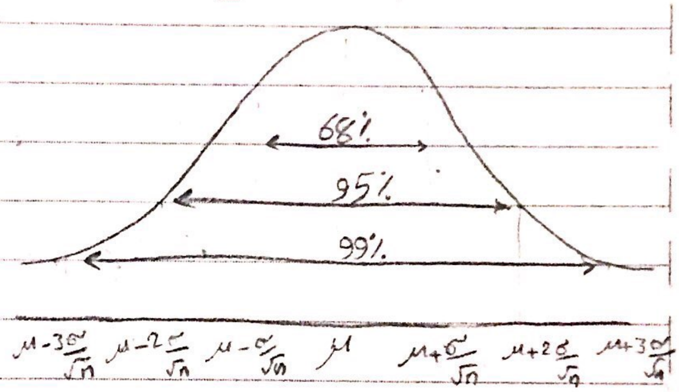
شده به اندازه‌ی کافی بزرگ باشد آنگاه طبق قضیه حد مرکزی:

$$X \sim F(\mu, \sigma^2) \xrightarrow{x_1, \dots, x_n} \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

هر توزیعی

$n \nearrow$

$n \geq 30$



p بزرگ

برآورد نقطه‌ای نسبت جامعه p: گاهی در عمل هدف برآورد نسبت جامعه است،

برای مثال فرض کنید ویژگی شاغل بودن یا نبودن افراد جامعه مورد نظر است که مطالعه‌ی

کل جامعه امکان پذیر نیست و می‌خواهیم از روی یک نمونه تصادفی p را برآورد کنیم:

شاغل — تعداد شاغل (تعداد افراد دارای صفت خاص)

$N \leftarrow$ بیکار 0 1 0 0 1 0 0 1 0 0 0 1 1 1 0 1 0 1 0 1 $\rightarrow \boxed{p = \frac{X}{N}}$

تعداد کل جامعه — احتمال موفقیت

$\underbrace{10 \quad 10010}_{N} \longrightarrow X \sim bin(n, p)$

X تعداد افرادی که در نمونه شاغل هستند

از طرفی میدانیم در توزیع دوجمله‌ای $X \sim bin(n, p)$.

① $E(X) = np$ ② $V(X) = npq$ $\quad q = 1 - p$

چون X تعداد افرادی هستند که در نمونه صفت خاص را دارند پس

شاغل بودن

$\hat{p} = \frac{X}{n}$

Sadra

نسبت در نمونه است (نسبت شاملین نمونه) اکنون میتوان نشان داد که نسبت نمونه :

برآورد نقطه‌ای خوبی برای نسبت در جامعه است.

$$\hat{p} = p$$

p نسبت جامعه p نسبت نمونه

از ① و ②

$$\boxed{E(\hat{p})} = E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) \overset{①}{=} \frac{1}{n}\cdot np = p \quad ③$$

یعنی $\hat{p} = p$ برآورد نااریب برای p است.

$$\boxed{V(\hat{p})} = V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{1}{\cancel{n^2}}\cdot \cancel{n}pq = \frac{pq}{n} \quad ④$$

$$\lim_{n\to\infty} V(\hat{p}) = 0$$

و ثابت میشود $\hat{p}$ دارای کمترین واریانس بین برآوردهای نااریب p است

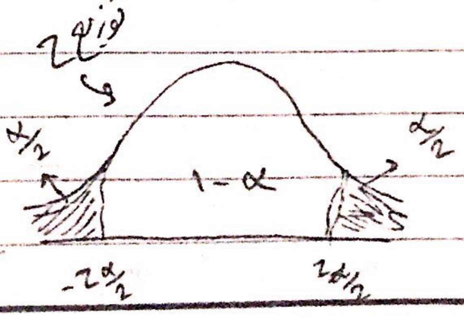
پس $\hat{p}=p$ به عنوان "برآورد خوب" p شناخته میشود.

نکته: اگر n به اندازه کافی بزرگ باشد میتوان نشان داد $\hat{p}$ دارای توزیع تقریبی نرمال می باشد:

$$\hat{p} \underset{③,④}{\approx} N\left(p, \frac{pq}{n}\right)$$

یا

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$



$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

Subject:
Year: Month: Date:

با جایگذاری Z:

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\hat{P}-P}{\sqrt{\frac{pq}{n}}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow P\left(|\hat{P}-P| < Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1-\alpha$$

نکته: چون $\frac{pq}{n}$ مجهول است معمولاً از برآورد آن استفاده می‌شود یعنی

حاشیه خطا برابر است با $Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ و معمولاً با "S.E" نشان می‌دهند
یران خطا

مثال ۱) شرکتی که از طریق سفارش پستی فروش دارد به طور ماهانه پیشنهادهای ویژه‌ای را از طریق پست برای افراد عضو می‌فرستد یک پیشنهاد آزمایشی برای نمونه تصادفی شامل 250 نفر انتخاب شده و به وسیله پست برای افراد نمونه ارسال می‌شود بر مبنای این نمونه 70 نفر تصمیم به خرید می‌گیرند یک برآورد نقطه‌ای برای نسبت اعضائی که انتظار می‌رود از طریق پست خرید کنند بدست آورید و حاشیه‌ی خطای ٪95 را برای این برآورد بدست آورید.

تعداد افرادی که خرید می‌کنند $X = 70$ ، $n = 250$ تعداد نمونه

$$P = \hat{P} = \frac{70}{250} = 28\%$$

چون $1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow \frac{\alpha}{2} = 0{,}025$

$\boxed{Z_{0,025} = 1{,}96 \simeq 2}$

$$S.E(\hat{P}) = Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}\hat{q}}{n}} = 2\sqrt{\frac{0{,}28(1-0{,}28)}{250}} = 0{,}056$$

Subject:
Year: Month: Date:

مثال) فرض کنید داده‌های زیر از جامعه‌ی نرمال استخراج شده است. میانگین جامعه را برآورد کنید و حاشیه‌ی خطای برآورد را بدست آورید ($\sigma=1$) ٪۹۵

۱ ۲ ۴ ۳ ۵

حاشیه خطا $: Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 1.96\times\frac{1}{\sqrt{5}}$

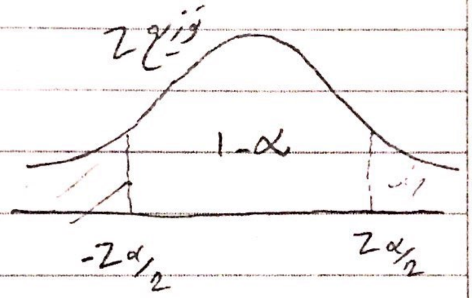
$$\hat{\mu}=\bar{x}=\frac{1+2+4+3+5}{15}=3$$

$1-\alpha=0.95$

$\alpha=0.05 \rightarrow \frac{\alpha}{2}=0.025$

$Z_{0.025}=1.96$

میدانیم اگر جامعه نرمال باشد $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$ پس:



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

$$P(-Z_{\frac{\alpha}{2}}<Z<Z_{\frac{\alpha}{2}})=1-\alpha$$

$$P(-Z_{\frac{\alpha}{2}}<\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<Z_{\frac{\alpha}{2}})=1-\alpha$$

$$\Rightarrow P(|\bar{X}-\mu|<Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})=1-\alpha$$

برآورد نقطه‌ای واریانس جامعه (σ^2): یکی دیگر از پارامترهای مهم برای استنباط پیرامون جامعه σ^2 است که قبلاً دیدیم:

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

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

اگر جامعه نرمال یا $n\nearrow$ $\quad \frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1}$

Sadra

(وقتی از Z استفاده می‌کنیم یعنی از توزیع نرمال استفاده می‌کنیم)

Subject:
Year: Month: / Date:

ثابت میشود S^2 "برآورد خوب" σ^2 بهترین برآوردکننده‌های نااریب

$\hat{\sigma}^2 = \sigma^2$

است.

* دقت کنید به نااریبی تحت جذر گرفتن حفظ نمی‌شود

$E(S^2) = \sigma^2 \not\Rightarrow \cancel{E(S) = \sigma}$ و $E(S) \neq \sigma$ و

$E(S) < \sigma$ و مقدار اریبی به n بستگی دارد بطوریکه اریبی با افزایش n به سمت صفر می‌رود.

$E(S) \underset{n \uparrow}{\approx} \sigma$

تعیین حجم نمونه: برای تعیین حجم نمونه معمولاً بستگی به مطالعه مقدار خطا (d) مشخص می‌شود و براساس d، حجم نمونه مشخص می‌شود:

① اگر مسئله برآورد میانگین جامعه باشد دیدیم: $P(\underbrace{|\bar{x} - \mu|}_{\text{خطا}} < Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1-\alpha$

② $|\bar{x} - \mu| \leq d$ از طرفی

①, ② $\Rightarrow d = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow d^2 = Z^2_{\frac{\alpha}{2}} \frac{\sigma^2}{n} \Rightarrow n = Z^2_{\frac{\alpha}{2}} \frac{\sigma^2}{d^2}$

نکته: اگر σ^2 مجهول باشد مجبوریم یک نمونه اولیه استخراج کنیم $\hat{\sigma}^2 = S^2$ را محاسبه کنیم پس $\hat{n} = Z^2_{\frac{\alpha}{2}} \frac{\hat{\sigma}^2}{d^2}$

③ اگر مسئله "نسبت جامعه" باشد دیدیم $n \uparrow$

$P(|\hat{p} - p| < Z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}) = 1-\alpha$

Sadra

Subject:
Year: Month: Date:

$$\rightarrow \hat{p} - p \le d \Rightarrow d = Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$\Rightarrow d^2 = Z^2_{\alpha/2}\,\frac{pq}{n} \Rightarrow n = Z^2_{\alpha/2}\,\frac{pq}{d^2}$$

مشکل

p و q مجهولند و در نتیجه pq مجهول است و باید از برآورد آن استفاده کرد، در بدترین حالت میتوان $p = q = \frac{1}{2}$

$$n = Z^2_{\alpha/2} \cdot \frac{1}{4d^2}$$

یا نمونه اولیه استخراج میکنیم و p را حساب میکنیم.

$$\hat{n} = Z^2_{\alpha/2}\,\frac{\hat{p}\hat{q}}{d^2}$$

مثال) یک زیست شناس میخواهد میانگین نفعات موجود در هر واحد حجم آب دریاچه‌ای را محاسبه کند در محاسبات سال گذشته معلوم شده است که انحراف معیار نفعات موجود در دریاچه ۴ است، چه تعداد از نمونه این زیست شناس برای برآورد میانگین باید بگیرد تا ۹۰٪ مطمئن باشد که خطای برآورد از ۰/۸ بیشتر نخواهد بود.

$$n = ?\ ,\ 1-\alpha = 0.9\ ,\ d = 0.8\ ,\ \sigma = 4$$

$$1-\alpha = 0.9 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow Z_{0.05} = 1.64$$

Subject:
Year: Month: Date:

$$n = (1{,}64)^2 \times \frac{16}{(0{,}8)^2} = 67{,}75 \Rightarrow n \simeq 68$$

مثال) یک بررسی به منظور برآورد نسبت کسانیکه دید کافی ندارند در یک برنامه بهداشت عمومی انجام شود چند نفر را باید معاینه کرد؟ اگر وزارت بهداشت بخواهد با ٪98 اطمینان خطای برآورد کمتر از 0/05 باشد وقتیکه:

$d = 0{,}05$

الف) هیچ اطلاعی از P نداریم.

ب) P بر اساس نمونه اولیه ٪30 تخمین زده شده است.

$1 - \alpha = 0{,}98$

$\rightarrow \alpha = 0{,}02$

$\rightarrow \frac{\alpha}{2} = 0{,}01$

$\rightarrow Z_{0{,}01} = 2{,}33$

الف/ $n = ? \Rightarrow n = (2{,}33)^2 \times \frac{1}{4(0{,}05)} = 543$

ب/ $n = (2{,}33)^2 \times \frac{(0{,}3)(0{,}7)}{(0{,}05)^2} = 456$

(یکی از حالت های بدست آوردن فاصله اطمینان در امتحان میاد)

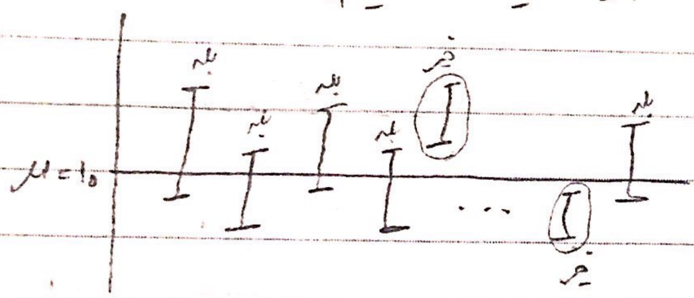
(برآورد فاصله ای (فاصله اطمینان، بازه اطمینان): چون با احتمال ! ، برآورد نقطه ای برابر با پارامتر نیست معمولاً علاقمندیم به ارائه یک فاصله برای پارامتر می باشیم که با اطمینان $\%(1-\alpha)100$ (مثلاً $\alpha = 95$ ، $1-\alpha = 0{,}95$ ، اطمینان ٪95) مطمئن باشیم که فاصله مذکور پارامتر را دربردارد. برای مثال اگر پارامتر مورد نظر θ باشد، معمولاً یک فاصله اطمینان براساس برآورد نقطه ای محاسبه می کنند و به صورت زیر بیان می شود.

Sadra $P(L < \theta < U) = 1 - \alpha$; (L, U) فاصله اطمینان

ضریب اطمینان

14

تفسیر فاصله اطمینان: منظور این است که اگر مثلاً فاصله اطمینان 95% برای پارامتر میانگین جامعه محاسبه شده است، انتظار داریم اگر 100 بار نمونه‌گیری انجام دهیم و 100 فاصله اطمینان برای μ بدست آوریم، 95 بار فاصله μ را شامل شوند (فرض کنید μ واقعاً جامعه‌ای 10 باشد از این جامعه 100 بار نمونه‌گیری کرده‌ایم)



بله: فاصله μ را شامل می‌شود
خیر: 〃 〃 نفی شود

از 100 فاصله اطمینان، 95 فاصله، μ را شامل می‌شوند

برآورد فاصله‌ای برای پارامتر میانگین جامعه: در این بخش در حالت‌های مختلف برآورد فاصله‌ای را برای پارامتر μ که مجهول است بدست می‌آوریم.

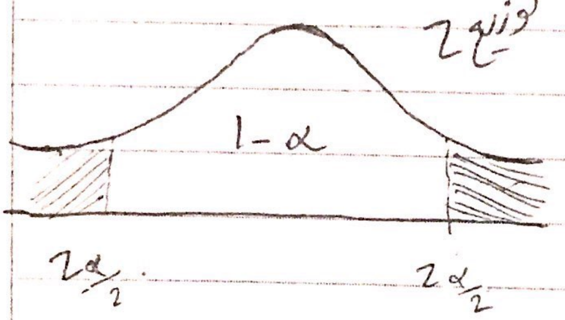
و واریانس جامعه معلوم

حالت اول: (فاصله اطمینان برای μ وقتی جامعه نرمال است)، n بزرگ یا کوچک

فرض کنید $X_1 \dots X_n$ نمونه‌ای تصادفی از جامعه نرمال با میانگین μ و واریانس σ^2 باشد. قبلاً دیدیم $\bar{X}$ برآورد خوب μ است و وقتی جامعه نرمال است $\bar{X}$ که ترکیب خطی از $X_1 \dots X_n$ است نیز نرمال است:

نمونه تصادفی

$$X \sim N(\mu, \sigma^2) \xrightarrow{X_1 \dots X_n} \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Longrightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$



$$P(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}) = 1-\alpha$$

با جایگذاری Z:

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

با نگه داشتن μ در وسط و انتقال بقیه جملات به طرفین نامساوی داریم:

$$P\Big(\underbrace{\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{L \text{ کران پایین فاصله اطمینان}} < \mu < \underbrace{\bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{U = \text{ کران بالا فاصله اطمینان}}\Big) = \underbrace{1-\alpha}_{\text{ضریب فاصله اطمینان}}$$

$$\boxed{U - L = \text{طول فاصله اطمینان}}$$

پس یک فاصله اطمینان ٪ $100(1-\alpha)$ برای μ

وقتی جامعه نرمال است و σ^2 معلوم:

① $$\boxed{\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)}$$

σ^2 معلوم و جامعه نرمال

مثال) برمبنای یک نمونه تصادفی شامل 25 مشاهده از جامعه نرمال با میانگین μ و انحراف معیار 8 میانگین جامعه 42,7 بدست آمده است یک فاصله اطمینان ٪95 برای میانگین جامعه بسازید.

$n=25$, $\bar{X}=42{,}7$, $\sigma=8$, $1-\alpha=0{,}95 \Rightarrow \alpha=0{,}05 \Rightarrow \frac{\alpha}{2}=0{,}025$

$Z_{0{,}025}=1{,}96$ $\qquad \mu : \left(42{,}7 \pm 1{,}96\,\frac{8}{\sqrt{25}}\right)=(39{,}6 , 45{,}8)$

نکته ۱: اگر واریانس σ^2 جامعه معلوم نباشد فقط در صورتیکه n بزرگ باشد چون میتوان

از واریانس S نسبت به σ چشم پوشی کرد در رابطه فاصله اطمینان ① میتوان از S

به جای σ استفاده کرد: $\boxed{\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\,\frac{S}{\sqrt{n}}\right)}$ ②

جامعه نرمال و σ^2 مجهول

حالت دوم: اگر جامعه نرمال نباشد اما n بزرگ باشد ثابت می شود در حالتیکه

واریانس معلوم است از رابطه ① و اگر واریانس مجهول است میتوان از رابطه ②

استفاده کرد طبق قضیه حد مرکزی

فرض کنید $X_1 \dots X_n$ نمونه ای از جامعه ای با هر توزیع باشد چون فرض

کرده ایم n بزرگ است طبق قضیه حد مرکزی $\bar{X} \sim N\left(\mu , \frac{\sigma^2}{n}\right)$

پس داریم: $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$$P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right)=1-\alpha$$

با جایگذاری Z:

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right)=1-\alpha$$

با نگه داشتن μ در وسط و انتقال بقیه جملات به طرفین نامساوی داریم

Sadra

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

پس یک فاصله اطمینان $\%\,100(1-\alpha)$ برای μ وقتی جامعه نرمال نیست،

n بزرگ و σ^2 معلوم:

$$\boxed{\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)}$$ ③

σ^2 معلوم و n بزرگ

و اگر σ^2 مجهول و n بزرگ

$$\boxed{\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)}$$ ④ ← n بزرگ و σ^2 مجهول

حالت سوم: (n کوچک باشد و توزیع جامعه نرمال باشد). فرض کنید $X_1 \dots X_n$

نمونه‌ای تصادفی از $X \sim N(\mu, \sigma^2)$ باشد اگر σ^2 معلوم باشد:

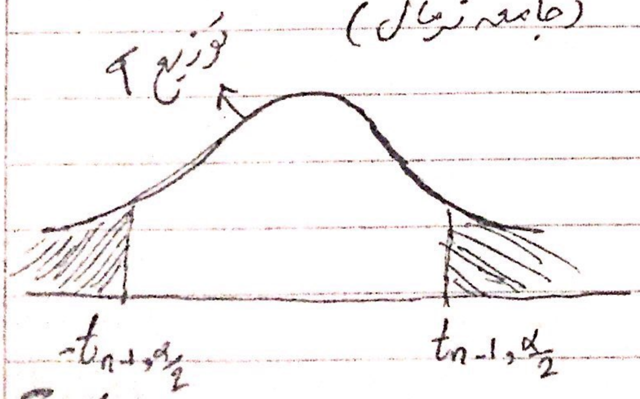
$$\bar{X} \sim N(\mu, \sigma^2) \longrightarrow Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

و دقیقاً مشابه مطالب حالت ① و ② تکرار می‌شود داریم:

$$\boxed{\mu : \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$ ← جامعه نرمال، σ^2 معلوم، n کوچک

اما اگر σ^2 معلوم نباشد:

قبلاً دیدیم: $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$ (جامعه نرمال)



* n بزرگ و واریانس چشم‌پوشی ز و در هر دو حالت از Z استفاده می‌شود *

Subject:
Year: Month: Date:

$$P\left(-t_{n-1,\frac{\alpha}{2}} < T < t_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha$$

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

با نگه داشتن μ در وسط و انتقال سایر جملات به طرفین نامساوی داریم:

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

پس یک فاصله اطمینان $\%100(1-\alpha)$ برای μ وقتی جامعه نرمال است، n کوچک و σ^2 مجهول:

$$\boxed{\mu : \left(\bar{X} \pm t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)}$$

σ^2 مجهول و n کوچک، جامعه نرمال

خلاصه مطالب:

برآورد فاصله‌ای μ:

- جامعه نرمال باشد
  - σ^2 معلوم باشد: $\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$
  - σ^2 مجهول باشد
    - n بزرگ: $\mu : \left(\bar{X} + Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$
    - n کوچک: $\mu : \left(\bar{X} \pm t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$
- جامعه نرمال نباشد
  - * n بزرگ — ق.ح.م
    - σ^2 معلوم: $\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$
    - σ^2 مجهول: $\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$ ق.ح.م
  - n کوچک باشد [illegible]

Sadra

$n>30$ بزرگ ، $n<30$ کوچک

Subject:
Year: Month: Date:

مثال) اداره بهداشت یک شهر ساحلی در نظر دارد در مورد میزان باکتری در واحد حجم آب ساحل یک دریاچه تحقیق کند. برای این منظور ۱۰ نمونه از واحد حجم آب را آزمایش و میزان باکتری موجود در آب را به شرح زیر گزارش میکند

175 190 215 198 184 207 310 193 196 180

یک برآورد نقطه ای برای میانگین جامعه و یک فاصله اطمینان ٪۹۵ برای میانگین میزان باکتری بدست آورید، فرض کنید جامعه نرمال است.

جامعه نرمال

$n=10$ کوچک

σ مجهول

$$\mu : \left(\bar{X} \pm t_{n-1, \frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)$$

شر $= \bar{X} = \dfrac{175 + \dots + 180}{10} = 194{,}8 \quad , S^2 = \dfrac{1}{9}\left[(175 - 194{,}8)^2 + \dots + (180 - 194{,}8)^2\right] = (13{,}14)^2$

$t_{9, 0{,}025} = 2{,}262 \leftarrow 1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow \frac{\alpha}{2} = 0{,}025$

$$\mu : \left(194{,}8 \pm 2{,}262 \frac{13{,}14}{\sqrt{10}}\right)$$

Sadra

Scanned with CamScanner

برآورد فاصله‌ای برای واریانس جامعه (σ^2): در عمل معمولاً علاوه بر پارامتر μ
پارامتر σ^2 نیز مورد نظر است. قبلاً دیدیم برآورد نقطه‌ای خوب برای σ^2 وقتی μ
مجهول است عبارت است از تغییرات نمونه یعنی:

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

وقتی μ معلوم است:

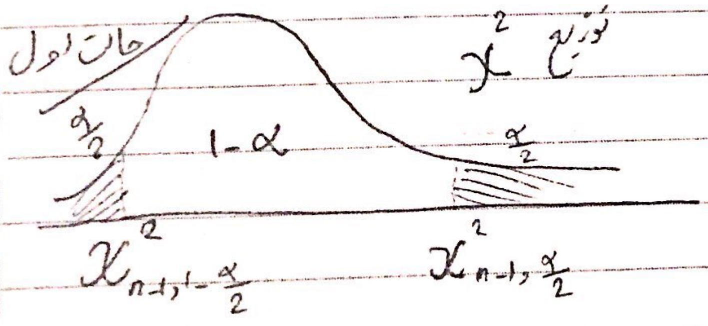
$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

همچنین دیدیم: (تحت فرض نرمال بودن یا $n>30$)

μ مجهول ← حالت اول

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

حالت دوم ← μ معلوم



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

با جایگذاری χ^2:

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

$$\Rightarrow P\Big(\underbrace{\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}}_{L} < \sigma^2 < \underbrace{\frac{(n-1)\,\sigma^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}}_{U}\Big) = 1-\alpha$$

پس یک فاصله اطمینان ٪ $100(1-\alpha)$ برای σ^2 وقتی μ مجهول است:

$$\boxed{\sigma^2 : \left(\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \,,\, \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right)}$$

→ چرا دو فاصله برای وقتی μ مجهول است.

صورت اول کوچکتر یعنی بعدی بزرگتره

نکته: فاصله اطمینان ٪ $100(1-\alpha)$ برای انحراف معیار عبارت است

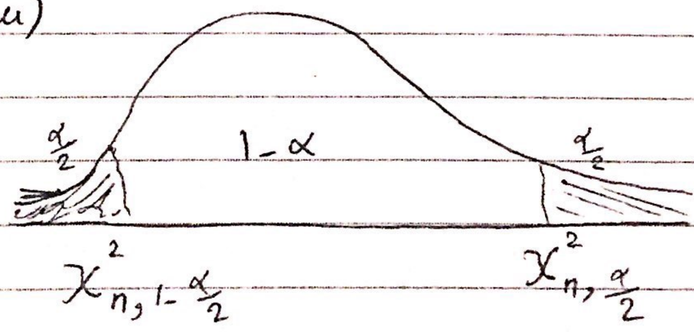
از جذر رابطه بالا:

$$\sigma : \left(\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}} \,,\, \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}}\right)$$

حالت دوم

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\chi^2 = \frac{nS_n^2}{\sigma^2} \sim \chi^2_n$$



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

با جایگذاری χ^2:

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

Subject:
Year: Month: Date:

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

پس یک فاصله اطمینان $\%100(1-\alpha)$ برای σ^2 وقتی μ معلوم است:

$$\boxed{\sigma^2 : \left(\frac{nS_n^2}{\chi^2_{n,\frac{\alpha}{2}}} , \frac{nS_n^2}{\chi^2_{n,1-\frac{\alpha}{2}}}\right)} \longrightarrow$$ برآورد فاصله‌ای برای σ^2 وقتی μ معلوم

نکته: فاصله اطمینان $\%100(1-\alpha)$ برای انحراف معیار وقتی μ معلوم است از جذر رابطه بالا:

$$\sigma : \left(\sqrt{\frac{nS_n^2}{\chi^2_{n,\frac{\alpha}{2}}}} , \sqrt{\frac{nS_n^2}{\chi^2_{n,1-\frac{\alpha}{2}}}}\right)$$

مثال) ده نمونه زیر از جامعه‌ای در دسترس می‌باشد فاصله اطمینان $\%95$ برای میانگین جامعه و واریانس جامعه بدست آورید: (با فرض نرمال بودن)

230 225 229 227 228
232 225 226 228 226

- واریانس جامعه مجهول، حجم نمونه کم $n < 30$، جامعه نرمال:

$$\mu : \left(\bar{X} \pm t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) \qquad \begin{cases} 1-\alpha = 95 \\ \alpha = 0{,}05 \\ \frac{\alpha}{2} = 0{,}025 \end{cases} \rightarrow t_{9,0{,}025} = 2{,}26$$ جدول

$$\bar{X} = \frac{230 + \cdots + 226}{10} = 227{,}6$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{9}\left[(230-227{,}6)^2 + \cdots + (226-227{,}6)^2\right]$$

$\simeq 5,26 \rightarrow S = \sqrt{5,26} = 2,27$

$$\mu : \left(227,6 \pm 2,262 \frac{2,27}{\sqrt{10}}\right) = (226 , 229,2)$$

- برای بدست آوردن فاصله اطمینان برای σ^2:

$$\sigma^2 : \left(\underbrace{\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}} , \underbrace{\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}}\right) = \left(\frac{9 \times 5,26}{19,022} , \frac{9 \times 5,26}{2,7}\right)$$

که این مقادیر را از جدول بدست آمد

$$= (1,56 , 4,14)$$

مثال) مقدار آهن موجود در یک کیلوگرم سنگ آهن متغیر تصادفی با توزیع نرمال است، اگر برای یک نمونه 26 تایی داشته باشیم $S^2 = 93,24$ یک فاصله اطمینان با ضریب اطمینان ٪۹۰ برای σ^2 بدست آورید.

$$n = 26 , S^2 = 93,24 , \begin{cases} 1-\alpha = 0,9 \\ \alpha = 0,1 \\ \frac{\alpha}{2} = 0,05 \end{cases}$$

توزیع نرمال، σ^2 مجهول

$$\chi^2_{25,0,05} = 37,65 \qquad \chi^2_{25,0,95} = 14,6$$

$$\sigma^2 : \left(\frac{25 \times 93,24}{37,65} , \frac{25 \times 93,24}{14,6}\right) = (61,91 , 159,65)$$

فاصله اطمینان برای نسبت جامعه: قبلاً دیدیم برآورد خوب برای نسبت جامعه نسبت نمونه است:

$$\hat{p} = p = \frac{X}{n}$$

$X \rightarrow$ تعداد افرادی که در نمونه صفت خاص را دارند

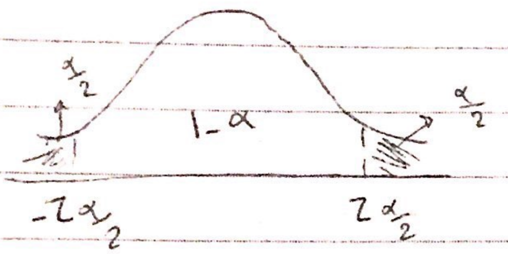
$n \rightarrow$ نمونه

از طرفی دیدیم اگر n بزرگ باشد:

$$\hat{p} \approx N\left(p, \frac{pq}{n}\right)$$

$$E(\hat{p}) = p \quad var(\hat{p}) = \frac{pq}{n} \Rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

توزیع



$$P(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}) = 1-\alpha$$

با جایگذاری Z:

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow P\Big(\underbrace{\boxed{\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}}}_{L} < p < \underbrace{\boxed{\hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}}}_{U}\Big) = 1-\alpha$$

پس فاصله اطمینان $\%100(1-\alpha)$ برای p:

$$\boxed{p: \left(\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right)}$$

نکته: چون $var(\hat{p}) = \frac{pq}{n}$ معمولاً مجهول است از برآوردگر آن در فرمول بالا استفاده می‌شود.

$$\widehat{var(\hat{p})} = \frac{\hat{p}\hat{q}}{n}, \quad p: \left(\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

Subject:
Year: Month: Date:

مثال) از نیروی کار در شهر برای نمونه تصادفی مرکب از دو هزار نفر انتخاب شده اند، افراد نمونه

مورد ~~...~~ مصاحبه قرار گرفتند و معلوم شده است که 165 نفر بیکار بودند. $n = 2000$

$x = 165$

الف) نرخ بیکاری را بر مبنای این داده ها برآورد کنید.

الف

$$\hat{P} = \frac{165}{2000} = 0.0825$$

ب) فاصله اطمینان ٪۹۵ برای نرخ بیکاری بدست آورید.

ب

$$P : \left(\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \left(0.085 \pm 1.96 \sqrt{\frac{0.0825(1-0.0825)}{2000}}\right)$$

$$(1-\alpha) = 0.95 \rightarrow \alpha = 0.025 \rightarrow \frac{\alpha}{2} = 0.025 \rightarrow Z_{0.025} = 1.96$$

مثال) برای بدست آوردن نسبت صاحبان اتومبیل که ماشین خود را به مبلغ بیش از یک مقدار

معین بیمه کرده اند، نمونه ای تصادفی مرکب از 400 صاحب اتومبیل برگزیده شده است،

اگر حق بیمه اتومبیل 56 نفر آن ها بیش از این مقدار معین باشد یک فاصله اطمینان

٪۹۵ برای نسبت این افراد بیابید.

ب) اگر بخواهیم ٪۹۵ مطمئن باشیم $d = 0.08$ (خطای برآورد) چه حجم نمونه ای باید اختیار کنیم؟

$$\hat{P} = \frac{56}{400} = 0.14$$

$$P : \hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(0.14 \pm 1.96 \sqrt{\frac{0.14(1-0.14)}{400}}\right)$$

$$= (0.106, 0.174)$$

پس ٪۹۵ اطمینان دارم نسبت صاحبان اتومبیل که اتومبیل خود را بیش از مبلغ معین بیمه کرده