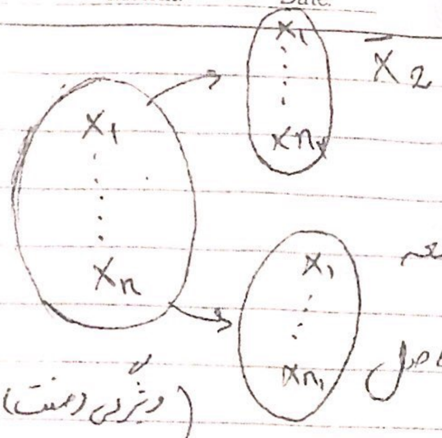
جزوه روش‌های آماری - دکتر حسینی

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 پس 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 بر 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 علاقمند به اطلاع از برخی ویژگی‌های عددی یک جامعه می‌باشیم

معمولاً به ویژگی عددی یک جامعه پارامتر می‌گویند، پارامترها را معمولاً با حروف یونانی مثل θ (تتا)، μ (مو)، η (اتا)، ν (نو) و ... نشان می‌دهند. دقت کنید پارامترها مقادیر ثابت هستند اما نامعلوم که در صورت مطالعه‌ی کل جامعه می‌توان مقدار آن را به طور دقیق به دست آورد. در عمل فقط به نمونه‌ای 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 داده‌ها در تابع، مقدار حاصل به پارامتر جامعه نزدیک باشد.

آماره: تابعی از نمونه تصادفی است: دقت کنید که آماره متغیر تصادفی است، آماره‌ها متغیر تصادفی هستند چون مقدار آماره از نمونه‌ای به نمونه دیگر متفاوت است.

* به عنوان یکی از مطالب مهم این درس انتخاب یکی از این آماره‌ها به عنوان برآورد پارامتر می‌باشد. برآورد دو نوع است: ۱- برآورد نقطه‌ای ۲- برآورد فاصله‌ای

برآورد نقطه‌ای یک پارامتر: هدف از برآورد نقطه‌ای آن است که از روی نمونه عددی به دست آوریم که انتظار داریم به مقدار نامعلوم پارامتر جامعه نزدیک باشد. فرض کنید پارامتر جامعه θ باشد برآورد نقطه‌ای آن را با $\hat{\theta}$ (تتا هت)

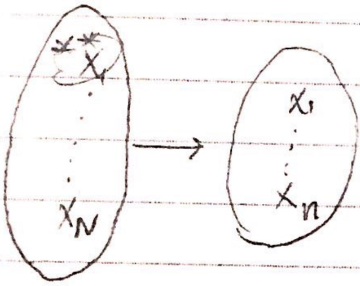
* پارامتر مربوط به جامعه و آماره مربوط به نمونه است *

Subject:
Year: Month: Date:

* هیچوقت x کوچک نمینویسیم چون متغیر تصادفی است

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 هرچه 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.

$$X \sim F(\mu, \sigma^2) \text{ هر توزیعی } \longrightarrow \begin{matrix} X_1 \overset{\text{هم توزیع}}{\equiv} X \\ \vdots \\ X_n \equiv X \end{matrix}$$

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

Subject:
Year: Month: Date:

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

① ناریبی: فرض کنید پارامتر مورد بررسی جامعه θ باشد در اینصورت $\hat{\theta}$ را یک برآورد کننده ناریب می‌گوییم 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 ناریب برای μ است. پس باید

$$E(T) = \mu$$

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

$$E(T) = E\left(\sum a_i x_i\right) = E(a_1 x_1 + \dots + a_n x_n)$$

$$= E(a_1 x_1) + E(a_2 x_2) + \dots + E(a_n x_n)$$

$$= a_1 \underbrace{E(x_1)}_{=\mu} + a_2 E(x_2) + \dots + a_n E(x_n)$$

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

$\sum a_i = 1$ ← * شرط برقراری این ناریبی a_i ها باید ۱ باشند.

Sadra

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

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

$V(\bar{x}) = \dfrac{\sigma^2}{n}$

σ^2 ← واریانس جامعه

$\bar{x}$ ← میانگین نمونه

n ← 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:

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

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

تغییر استفاده شود

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad , \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$ که در آن

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

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

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

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

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

$$= \frac{1}{n-1} E\left((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\right)$$

Sadra

$x_1, \ldots$ تا x_n نمونه تصادفی ← دو به دو مستقلند ← کوواریانسها صفر

Subject
Year: Month: Date:

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

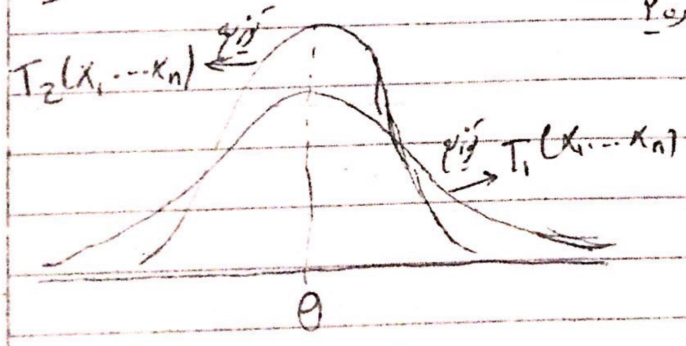
$$+ (\text{کوواریانسها}) = \frac{1}{n^2}\left[\sigma^2 + \cdots + \sigma^2\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

خواص واریانس

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

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

② اکنون سوال این است که بین دو آماره که هر دو ناریب 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 ناریب باشند و توزیعشان 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 ناریبی 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: درصورتیکه بین برآوردهای اریب بخواهیم برآورد خوب انتخاب کنیم

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

زیر تعریف میشود:

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

تمرین: ثابت کنید:

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

«به عنوان سوال در میانترم آید»

«ادامه مثال قبل»

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

خواص امید ریاضی

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

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

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 $\circledast$

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

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

Subject:
Year: Month: Date:

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

۳) 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 $\lim_{n \to \infty} Var(\hat{\theta}_n) = 0$

کارایی یک برآوردکننده: یک معیار برای مقایسه دو برآوردکننده کارایی است. فرض کنید $U(X)$ و $V(X)$ دو برآوردگر پارامتر θ باشد 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 " " U "

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

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

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 $V(U(X)) = V\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}V(X_1+X_2) = \frac{1}{4}\Big[\underbrace{Var(X_1)}_{=1}$

$+ \underbrace{Var(X_2)}_{=1} + 2\,\underbrace{\cancel{Cov(X_1, X_2)}}_{0}\Big] = \frac{1}{2}$

X_1 و X_2 نمونه تصادفی پس مستقلند

$V(T(X)) = V(2X_1 - X_2) = 4V(X_1) + V(X_2) - 4\,\cancel{Cov(X_1, X_2)} = 5$

$\mu(X)$ 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.

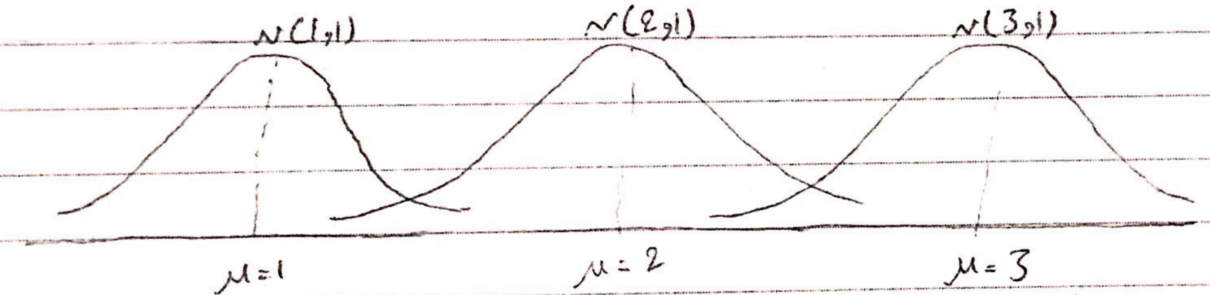
"بخش دوم"

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

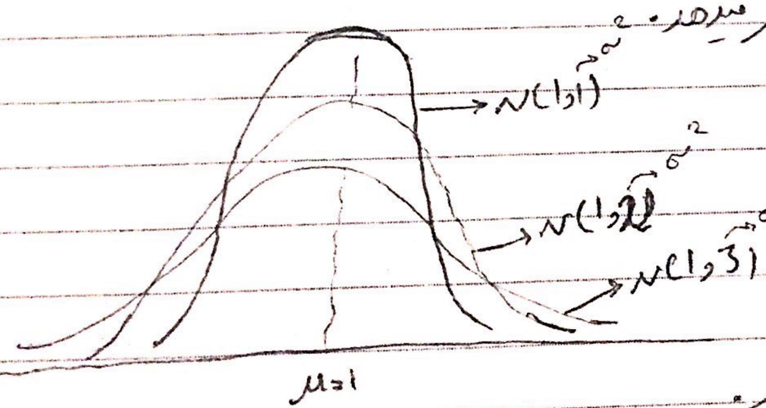
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma}} \qquad -\infty < x < +\infty$$

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

$E(X) = \mu$ و $V(X) = \sigma^2$



* پس μ مکان توزیع را تغییر میدهد.

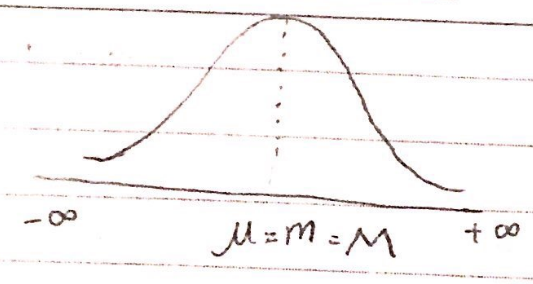


* پس σ^2 شکل توزیع را تغییر میدهد.

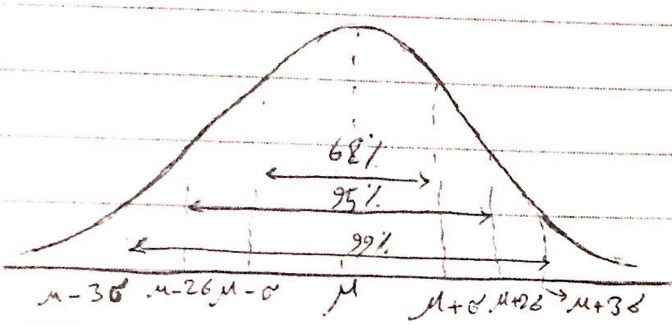
- با توجه به خواص خوبی که توزیع نرمال دارد معمولاً این توزیع مورد توجه است؛ از جمله خواص این توزیع میتوان به موارد زیر اشاره کرد:

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

② توزیع متقارن است و شکل زنگی دارد:



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



④ توزیع نرمال نسبت به ترکیبات خطی بسته است.

دو عدد ثابت

if $X \sim N(\mu, \sigma^2)$; $a, b \Rightarrow aX + b \sim N$

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

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

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

⑤ تحت شرطی کردن بسته است:

$(X_1, X_2) \sim N_2$ ، $X_1 | X_2 \sim N_1$ $\quad X_2 | X_1 \sim N_1$

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 کلا خط بزنه؟

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 مهمی است (حفظ باشید)

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

«مهم» تمرین) نشان دهید اگر $X \sim (\mu, \sigma^2)$ 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.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \dots + x_n)$$

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

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

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. $X \sim F(\mu, \sigma^2) \xrightarrow[X_1 \dots X_n]{\text{نمونه تصادفی}} \bar{X} \approx N(\mu, \frac{\sigma^2}{n})$ هر توزیعی

$\boxed{n \geq 30}$ 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 که 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 که 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.

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

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

$$P(\bar{X} < 25) = P(\underbrace{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}_{Z} < \frac{25-24}{\frac{2}{5}}) = P(Z < 2{,}5)$$

$= 0{,}9938$

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

صدتوی از افراد متوسط تعداد سالهای تحصیل بین ۱۱ تا ۱۲ سال باشد را بیابید.

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

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

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

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

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

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

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

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

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

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:

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

Sadra

* درصورت استفاده از $\bar{x}$ در به جای پارامتر و یا وجود محدودیت در مدل، باید از درجه آزادی به همان تعداد کم کنیم $(n-1)$ *

Subject:
Year: Month: Date:

از طرفی $N(0,1) \sim Z_i = \frac{x_i - \mu}{\sigma}$ اکنون تابع جدید زیر داریم

$i = 1, \dots, n$

توزیع کای دو است :

$$Z_i^2 = \frac{(x_i-\mu)^2}{\sigma^2} \sim \chi^2_1$$

و

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

کای دو (خی دو)

درجه آزادی توزیع

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

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

میانگین جامعه ـــ 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:

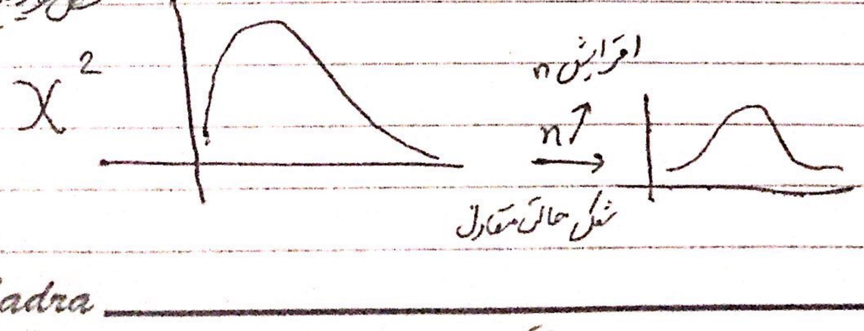
$$\text{if } y \sim \chi^2_n \Rightarrow E(y) = n$$

$$V(y) = 2n$$

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



Scanned with CamScanner

Subject:
Year: 14 Month: Date: $X \sim N(\mu, \sigma^2)$

توزیع S^2 :

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

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

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

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

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 :

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

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

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

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

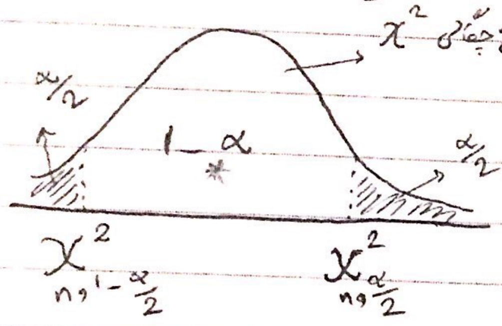
و نشان دهید S_n^2 برای σ^2 اریب است و کارایی S_n^2 را با S^2 مقایسه کنید.

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

Sadra

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 (μ) معلوم باشد 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 $X \sim N(\mu,\sigma^2)$ و $\overbrace{X_1,\dots,X_n}^{\text{نمونه تصادفی}} \sim N(\mu,\sigma^2)$

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:

و داریم:

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

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

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.

$X \sim N(\mu, \sigma^2) \quad X_1 \dots X_n \xrightarrow{\text{پس}} Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ ①

$\xrightarrow{\text{قبلاً دیدیم}} \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

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

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

درجه آزادی ← 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.

$\mu = 15$

$n = 20$

$S = 4{,}28$

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

t_{n-1}

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

$= 0{,}025$

توزیع فیشر:

if $u_1 \sim \chi^2_{n_1}$, $u_2 \sim \chi^2_{n_2}$

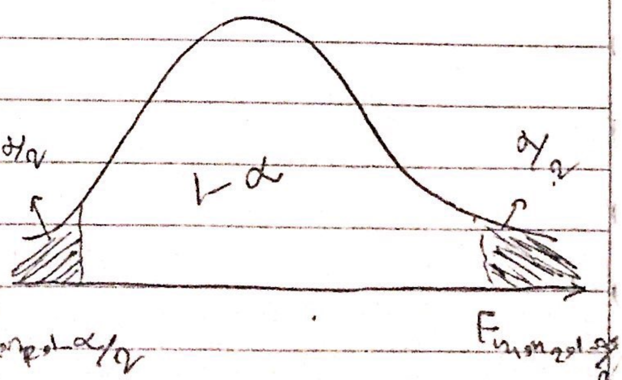
درجه آزادی صورت

$$\Longrightarrow F = \frac{u_1/n_1}{u_2/n_2} \sim F_{n_1, n_2}$$

← درجه آزادی مخرج

توزیع فیشر

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



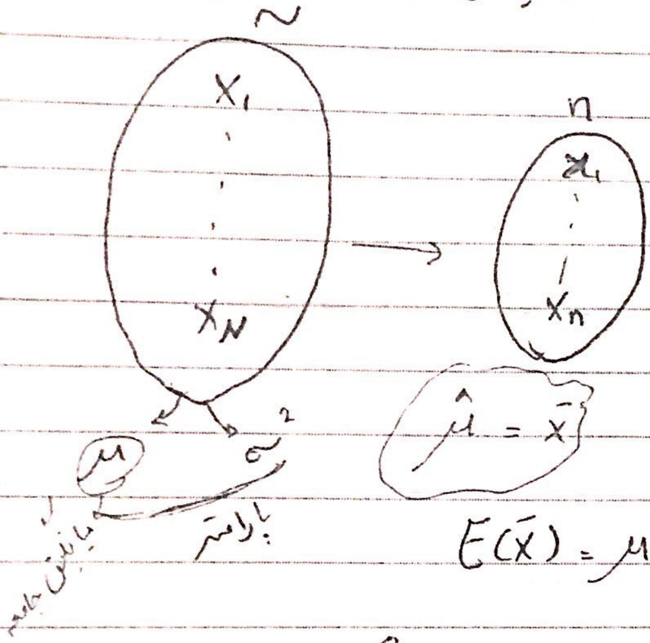
Sadra

"بخش سوم"

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

نقطه‌ی 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.

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

(p نسبت همان احتمال موفقیت است)

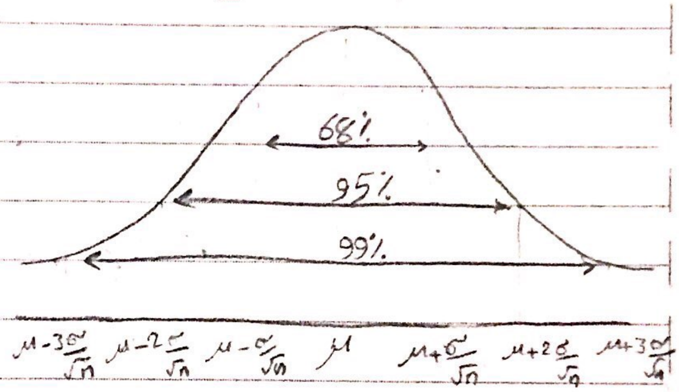
Subject:
Year: Month: Date:

نکته: اگر $X \sim F$ (جامعه هر توزیعی داشته باشد) به طوری که نمونه استخراج شده به اندازه‌ی کافی بزرگ (هر توزیعی) باشد آنگاه طبق قضیه حد مرکزی:

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

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 1 000 1 1 1 0 1 0 1 0 1 $\rightarrow \boxed{P = \frac{X}{N}}$

تعداد کل جامعه — احتمال موفقیت

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

X تعداد افرادی که در نمونه شاغل هستند.

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

① $E(X) = np$ ② $V(X) = npq$ $\boxed{q = 1 - p}$

چون X تعداد افرادی هستند که در نمونه صفت خاص را دارند (شاغل بودن) پس

$\hat{p} = \frac{X}{n}$ (نمونه)

Sadra

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

برآورد نقطه ای خوبی برای نسبت در جامعه است.

$$\hat{\underline{p}} = p$$

$\underline{p}$ نسبت جامعه p نسبت نمونه

از ① و ②

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

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

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

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

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

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

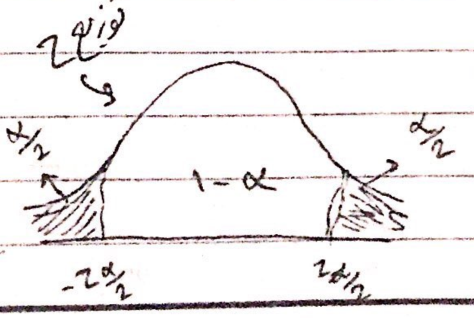
نکته: اگر n به اندازه کافی بزرگ باشد میتوان نشان داد $\hat{\underline{p}}$ دارای توزیع

تقریبی نرمال می باشد:

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

یا

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



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

با جایگذاری Z :

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

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

نکته: چون $\frac{pq}{n}$ مجهول است معمولاً از برآوردن 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.
کران خطا

مثال ۱) شرکتی که از طریق سفارش پستی فروش دارد به طور ماهانه پیشنهادهای ویژه‌ای را از طریق پست برای افراد 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 خطا برآورد ۹۵٪ را بدست آورید ($\sigma=1$)

۱ ۲ ۴ ۳ ۵

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

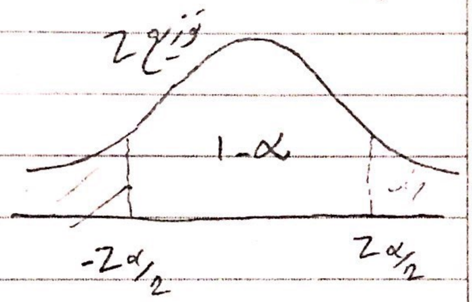
$$\hat{\mu}=\bar{x}=\frac{1+2+4+3+5}{15}=3$$

$1-\alpha=0.95$, $\alpha=0.05 \rightarrow \frac{\alpha}{2}=0.025$

$Z_{0.025}=1.96$

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

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



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

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

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

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

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

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

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

Sadra

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

Subject:
Year: Month: / Date:

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

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

است.

* دقت کنید که نااریبی تحت جذر گرفتن حفظ نمی‌شود

$E(S^2) = \sigma^2 \cancel{\Rightarrow} \cancel{E(S) = \sigma}$ 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 پارامتر 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

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

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

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

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

یا نمونه اولیه استخراج می‌کنیم و p را حساب می‌کنیم.

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

مثال) یک زیست شناس می‌خواهد میانگین نطفات موجود در هر واحد حجم آب دریاچه‌ای را محاسبه کند. در محاسبات سال گذشته معلوم شده است که انحراف معیار نطفات 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:

$n = (1.64)^2 \times \frac{16}{(0.8)^2} = 67.75 \Rightarrow n \approx 68$

مثال) یک بررسی به منظور برآورد نسبت کسانیکه دید کافی ندارند در یک برنامه بهداشت عمومی انجام می‌شود 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 0/3 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/ $n = (2.33)^2 \times \frac{(0.3)(0.7)}{(0.05)^2} = 456$

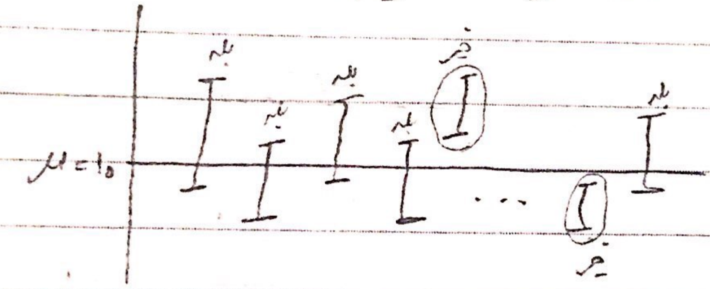
(یکی از حالت‌های بدست آوردن فاصله اطمینان در امتحان می‌آید)

(برآورد فاصله‌ای (فاصله اطمینان، بازه اطمینان): چون با احتمال 1، برآورد نقطه‌ای برابر با پارامتر نیست معمولاً علاقمند به ارائه یک فاصله برای پارامتر می‌باشیم که با اطمینان $100(1-\alpha)\%$ (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 بر اساس برآورد نقطه‌ای محاسبه می‌کنند و به صورت زیر بیان می‌شود.

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

ضریب اطمینان

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: 〃 〃 نمی‌شود

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

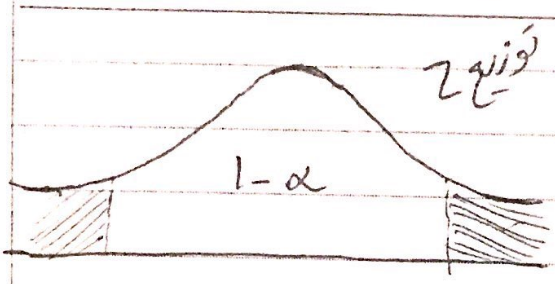
حالت اول: (فاصله اطمینان برای μ وقتی $\boxed{\text{جامعه نرمال}}$ است)، n بزرگ یا کوچک

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 $\bar{X}$ ترکیب خطی از $X_1 \dots X_n$ است نیز نرمال است:

نمونه تصادفی

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

Subject:
Year: Month: Date:



$-Z_{\frac{\alpha}{2}}$ $\qquad$ $Z_{\frac{\alpha}{2}}$

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

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:

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

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.

Year: Month: Date:

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

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

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

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

به جای σ استفاده کرد: ②

$$\boxed{\mu : (\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}})}$$

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

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

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

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

با جایگذاری Z:

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

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 $(1-\alpha)100\%$ برای μ وقتی جامعه نرمال نیست،

n بزرگ و σ^2 معلوم:

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

σ^2 معلوم و n بزرگ

واگر σ^2 مجهول و n بزرگ

$$\boxed{\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)} \quad ④$$

← 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:

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

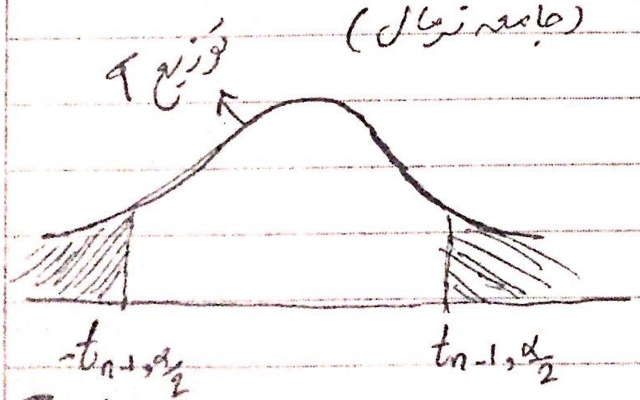
ودقیقاً مشابه مطالب حالت 1 و 2 تکرار می‌شود داریم:

$$\boxed{\mu : \bar{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}$$

← 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 از اریبی چشم‌پوشی شود 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 فاصله ای 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 *
    - σ^2 معلوم (ق.ح.م): $\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$
    - σ^2 مجهول (ق.ح.م): $\mu : \left(\bar{X} \pm Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$
  - n کوچک باشد بستگی به توزیع جامعه دارد

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

نامعلوم

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

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

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

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

Subject:
Year: Month: Date:

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:

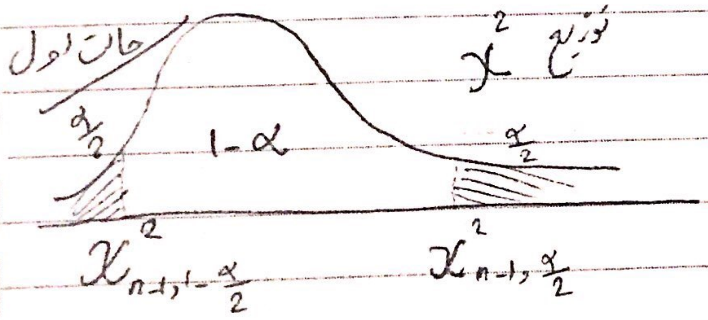
$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

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:

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

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

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

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

سمت اول کوچیکه بعد بزرگه

چرا دو فاصله برای وقتی μ مجهول است.

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

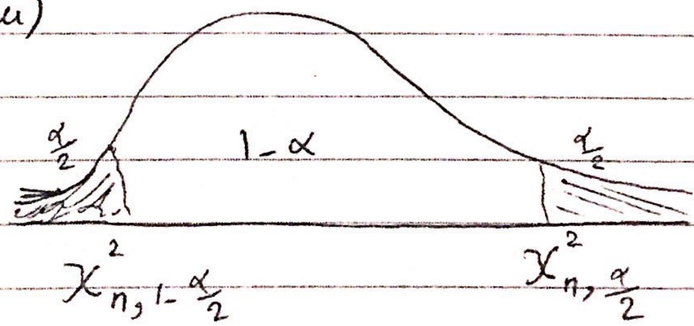
از جذر رابطه بالا:

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

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:

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

→ 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:

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

$$\begin{cases} 1-\alpha = 0.95 \\ \alpha = 0.05 \\ \frac{\alpha}{2} = 0.025 \end{cases} \rightarrow t_{9,0.025} = 2.26$$ جدول

$$\bar{X} = \frac{230 + \cdots + 226}{10} = 227.6$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{9}\left[(230-227.6)^2 + \cdots + (226-227.6)^2\right]$$

Sadra

Subject:
Year: Month: Date:

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

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

– 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:

$\hat{P} = p = \frac{X}{n}$ → تعداد افرادی که در نمونه صفت خاص را دارند

n → نمونه

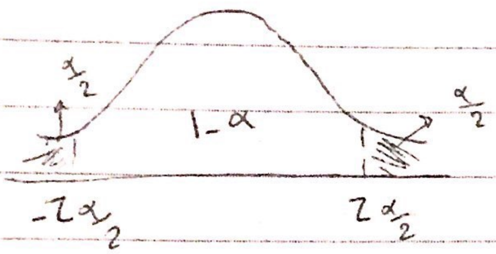
Sadra

از طرفی دیدیم اگر n بزرگ باشد:

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

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

توزیع



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

با جایگذاری Z:

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

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

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 برآورد آن در فرمول بالا استفاده می‌شود.

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

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

مورد ~~[illegible]~~ مصاحبه قرار گرفتند و معلوم شده است که 165 نفر بیکار بوده اند. $n=2000$

$X=165$

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

$$\hat{P}=\frac{165}{2000}=0{,}0825$$

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

$$P:\left(\hat{p}\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)=\left(0{,}085\pm 1{,}96\sqrt{\frac{0{,}0825(1-0{,}0825)}{2000}}\right)$$

$$(1-\alpha)=0{,}95\rightarrow\alpha=0{,}025\rightarrow\frac{\alpha}{2}=0{,}025\rightarrow Z_{0{,}025}=1{,}96$$

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 از آنها 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 که نسبت صاحبان اتومبیل که اتومبیل خود را بیش از مبلغ معین بیمه کرده

Subject:
Year: Month: Date:

$$n = \frac{\hat{p}\hat{q}Z^2_{\alpha/2}}{d^2} = \frac{(0{,}14)(1-0{,}14)(1{,}96)^2}{(0{,}008)^2} \simeq 7228$$

تعریف آزمون فرض: هر وقت بخواهیم یک ادعا یا فرضیه‌ای را از روی داده‌های نمونه بررسی کنیم مسئله آزمون فرض نامیده می‌شود. برای آزمون فرض در هر مسئله اولین گام تشکیل فرضیات است که به طور قراردادی به این صورت است که ادعا را در فرض مقابل قرار می‌گیرد و آنچه به تجربه در مورد مسئله وجود دارد در فرض صفر به صورت زیر

فرض صفر — مثال

$H_0: \mu = 1200$ → طول عمر لامپ‌های تولید شده کارخانه

$H_1: \mu > 1200$ → مهندسی فرآیند تولید پیشنهاد می‌دهد که طول عمر لامپ‌ها بزرگتر شود

فرض مقابل H_1

$$\equiv \begin{cases} H_0: \mu \leq 1200 \\ H_1: \mu > 1200 \end{cases}$$

پس از تشکیل فرضیه برای آزمون کردن نمونه‌ای تصادفی استخراج می‌شود، سطح آزمون تعیین می‌کنیم (چون فرضیه‌ای که در H_0 قرار می‌دهیم معمولاً به تجربه برقرار بوده و اکنون براساس نمونه H_0 رد شود به غلط این خطا، خطای بدی است و همین خطا که خطای نوع اول نامیده می‌شود به عنوان سطح آزمون در نظر گرفته می‌شود و قبل از نتیجه‌گیری کنترل می‌شود که با α نشان می‌دهند و معمولاً عدد کوچکی مثل $\frac{1}{100}$، 0.01، 0.05، 0.1، ... فرض می‌شود) پس احتمال رد H_0 (به شرطی که H_0 درست باشد) سطح آزمون می‌گویند:

سطح آزمون $\alpha = P(\underset{\text{reject}}{R} H_0 \mid H_0 \text{ درست باشد})$

* در تحلیل دنبال جزهای هستیم که توان بالایی دارند *

Subject:
Year: Month: Date:

و در مرحله آخر با فرض درست بودن H_0 و بر اساس نمونه استخراج شده و با توجه به مسئله آماره آزمونی تشکیل می شود و تصمیم گیری انجام می شود.

خطای نوع دوم را معمولاً با β نشان میدهیم و به صورت زیر تعریف می‌شود:

$$\beta = P(RH_1 \mid H_1 \text{ درست باشد})$$

توان آزمون = $1 - \beta$

وضعیت / تصمیم	AH0 (Accept) H_0 پذیرفته می شود	RH0 (Reject) H_0 رد می شود
اگر H_0 درست باشد	تصمیم صحیح است. احتمال $= 1 - \alpha$	خطای نوع اول احتمال $= \alpha$
اگر H_0 غلط باشد	خطای نوع دوم: احتمال $= \beta$	تصمیم صحیح است. توان آزمون احتمال $= 1 - \beta$

مقدار احتمال (p_value): یک روش کلاسیک برای آزمون کردن فرضیه ها محاسبه مقدار احتمال است که به صورت زیر تعریف می شود:

کوچکترین مقداری که برای α میتوان در نظر گرفت = p_value => مقدار احتمال

Sadra تا H_0 رد شود. مثلاً فرض کنید T_0 آماره آزمون تحت H_0 و t_0 مقدار مشاهده شده از آماره

Subject:
Year: Month: Date:

آنگاه:

$$\begin{cases} H_0 = \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \qquad p\text{-value} = p(T_0 > t)$$

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \qquad p\text{-value} = p(T_0 < t)$$

برای تصمیم گیری:

$$\text{if} \quad p\text{-value} \leq \alpha \Rightarrow RH_0$$

اگر فرضیه به صورت $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$ باشد $p\text{-value}$ عبارت است از:

$$p\text{-value} = 2\min\{p(T_0 > t), p(T_0 < t)\}$$

یا

$$p(|T_0| > t)$$

نکته: مساوی همیشه در H_0 قرار میگیرد.

- آزمون فرض برای میانگین (μ) جامعه: برای میانگین جامعه به فرضیه زیر ممکن است مورد نظر باشد.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \equiv \begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

فرضیه‌های یک طرفه

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \equiv \begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

* معادل! if $|Z_0| > Z_{\frac{\alpha}{2}}$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \longrightarrow$$ فرضیه دو طرفه

سطح آزمون را α در نظر میگیریم نمونه ای استخراج می شود $X_1 \dots X_n$ و اکنون

برای حالت های مختلف آماره های آزمون بدست آورده می شود

حالت اول: اگر جامعه نرمال باشد و واریانس جامعه معلوم باشد، دیدیم:

(یا n بزرگ باشد) و واریانس معلوم

$$X \sim N(\mu, \sigma^2) \xrightarrow{x_1 \dots x_n} \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

به $Z_0 = \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ آماره آزمون تحت H_0 می گوییم. پس از محاسبه آماره در مورد فرضیه تصمیم گیری می کنیم:

1 گام اول تعیین فرضیه

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \alpha$$

2 گام دوم تعیین سطح آزمون

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \alpha$$

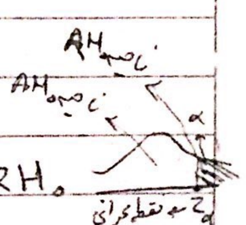
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad \alpha$$

3 گام سوم نمونه گیری و محاسبه آماره

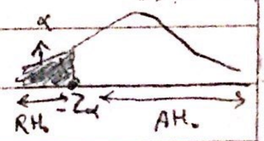
$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4 گام چهارم تصمیم گیری

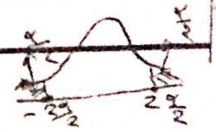
if $Z_0 > Z_\alpha \Rightarrow RH_0$



if $Z_0 < -Z_\alpha \Rightarrow RH_0$



*
$$\begin{cases} \text{if } Z_0 \geq Z_{\frac{\alpha}{2}} \\ \text{if } Z_0 \leq -Z_{\frac{\alpha}{2}} \end{cases} \Rightarrow RH_0$$



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

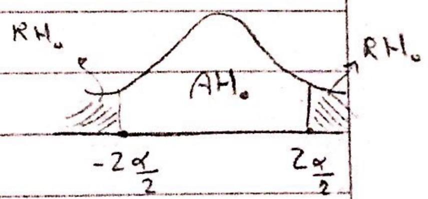
Subject:
Year: Month: Date:

حالت دوم: n بزرگ باشد (جامعه نرمال یا غیرنرمال) و واریانس مجهول باشد در اینصورت

قبلاً دیدیم:

$$Z = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim N(0,1) \xrightarrow[\text{آزمون}]{\text{آماره}} Z_0 = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$$

1	2	3	4
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	α	$Z_0 = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	if $Z_0 \geq Z_\alpha \Rightarrow RH_0$ 
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	α	"	if $Z_0 \leq -Z_\alpha \Rightarrow RH_0$ 
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	α	"	if $\lvert Z_0 \rvert \geq Z_{\frac{\alpha}{2}} \Rightarrow RH_0$ $\begin{cases} Z_0 \geq Z_{\frac{\alpha}{2}} \\ Z_0 \leq -Z_{\frac{\alpha}{2}} \end{cases}$



حالت سوم: وقتی جامعه نرمال است و n کوچک و واریانس مجهول:

برای این نیز فرض کنید نمونه تصادفی $x_1 \dots x_n$ در اختیار است.

(1)

$$\boxed{T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}}$$

قبلاً دیدیم:

$$\begin{cases} \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i & \text{میانگین نمونه} \\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{X})^2 & \text{تغییرات نمونه} \end{cases}$$

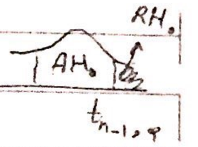
اکنون فرض کنید فرضیه ای $\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu>\mu_0 \end{cases}$ ، $\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu<\mu_0 \end{cases}$ ، $\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu\neq\mu_0 \end{cases}$

مورد نظر هستند. پس آماره ی آزمون با توجه به رابطه (1)

$$\underset{H_0 \text{ تحت}}{T_0} = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$$

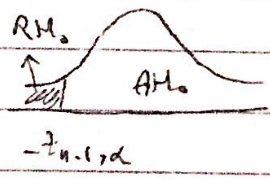
اکنون برای فرضیه اول چون $\bar{X}$ باید به اندازه کافی از μ_0 بزرگتر باشد تا H_0 را رد کنیم پس قاعده ی

فرضیه	سطح آزمون	آماره ی آزمون	قاعده تصمیم گیری
$\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu>\mu_0 \end{cases}$	α	$T_0=\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	if $T_0 > t_{n-1,\frac{\alpha}{2}} \rightarrow RH_0$



برای فرضیه دوم: برای این فرضیه باید $\bar{X}$ به اندازه کافی از μ_0 کوچکتر باشد که H_0 را رد کنیم.

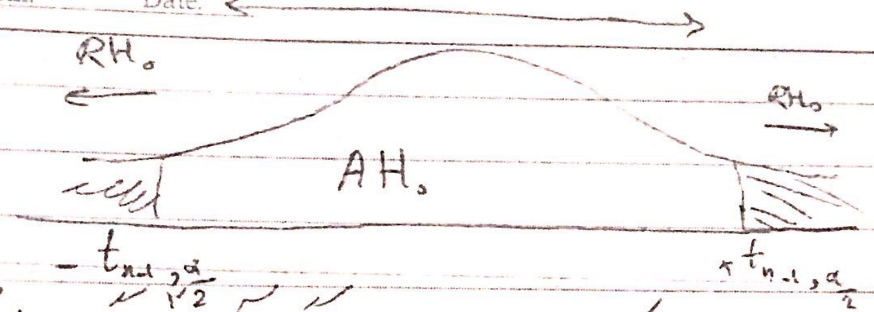
فرضیه	سطح آزمون	آماره آزمون	قاعده تصمیم
$\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu<\mu_0 \end{cases}$	α	$T_0=\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	if $T_0 < -t_{n-1,\alpha} \Rightarrow RH_0$



فرضیه سوم: برای این فرضیه باید $\bar{X}$ به اندازه کافی از μ_0 کوچکتر یا بزرگتر باشد پس:

فرضیه	سطح آزمون	آماره آزمون	قاعده تصمیم
$\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu\neq\mu_0 \end{cases}$	α	$T_0=\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	if $\left.\begin{cases} T_0 > t_{n-1,\frac{\alpha}{2}} \\ T_0 < -t_{n-1,\frac{\alpha}{2}} \end{cases}\right\} \Rightarrow RH_0$

Subject:
Year: Month: Date:



مثال) تعداد زیادی از بیماران مبتلا به یک بیماری بخصوص را گرد آورده و گزارش کرده اند که مدت

درمان بیماری به روش استاندارد دارای میانگین 15 روز و انحراف معیار 3 روز است. ادعا

شده که یک روش جدید میتواند مدت درمان را کوتاه‌تر کند. برای روش جدید درمان

نمونه تصادفی به حجم 70 انتخاب شده است. و پس از اعمال روش جدید درمان بر روی

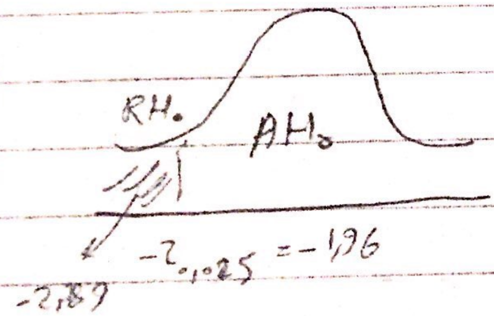
این 70 نفر میانگین مدت درمان 14 روز شده است. آیا در سطح معنی داری (سطح آزمون)

۰/۰25 روش جدید بهتر است؟ $\bar{X}=14$ $\quad \mu_0=15$, $\sigma=3$, $n=70$

انحراف معیار معلوم — حجم نمونه بزرگ

$\alpha=0,025$

$$\begin{cases} H_0: \mu=15 \\ H_1: \mu<15 \end{cases} \longrightarrow Z_0=\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\frac{14-15}{\frac{3}{\sqrt{70}}}=-2,789$$



$-2,789<-1,96 \Rightarrow RH_0$

مثال۲) نمونه تصادفی از پرونده‌های فراوان شرکتی نشان میدهد که سفارشات برای قطعات

معینی از ماشین‌ها به ترتیب روز بایگانی شده است. اگر تعداد روزهای بایگانی از توزیع نرمال

Sadra

چه بگویند آیا در سطح آزمون ۰/۰۱ میتوان ادعا کرد که میانگین زمان بایگانی از 10,5 روز بیشتر است.

10 12 19 14 15 18 11 13 $\rightarrow n=8$

$$\begin{cases} H_0: \mu = 10{,}5 = \mu \leq 10{,}5 \\ H_1: \mu > 10{,}5 \end{cases}$$

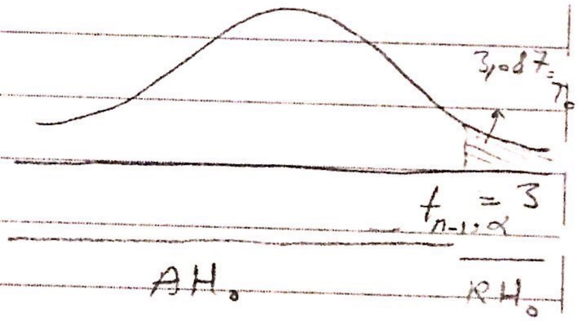
جامعه نرمال و واریانس جامعه مجهول

$$\bar{X} = \frac{10+\dots+13}{8} = 14 \quad , \quad S^2 = \frac{1}{7}\left((10-14)^2+\dots+(13-14)^2\right) =$$

10,286

$$T_0 = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} = \frac{14-10{,}5}{\frac{\sqrt{10{,}286}}{\sqrt{8}}} = 3{,}087$$

$$\alpha = 0{,}01 \quad t_{n-1,0{,}01} = t_{8-1,0{,}01} = 3$$



$$\text{if } T_0 > t_{n-1,\alpha} \Rightarrow RH_0$$

$3{,}087 > 3 \Rightarrow RH_0$ / یعنی میانگین زمان بایگانی قطعات بیشتر از 10,5 روز است.

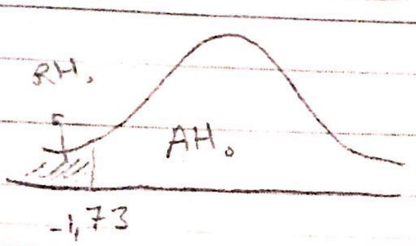
مثال) نمونه شامل 20 رسید فروش کتاب یک فروشگاه کتاب با دلار میانگین 121 دلار و انحراف معیار 10,2 دلار است. از این داده ها برای آزمون این فرض که میانگین فروش کتاب کمتر از 125 دلار است استفاده کنید. سطح آزمون را 0,05 در نظر بگیرید. (فرض کنید جامعه نرمال است.)

$n = 20 < 30 \quad , \quad \bar{x} = 121 \quad , \quad S = 10{,}2$

$$T_0 = \frac{121-125}{\frac{10{,}2}{\sqrt{20}}} = -1{,}75$$

«اگر α معلوم نبود 0,05 در نظر بگیریم»

Subject:
Year: Month: Date:



$$-t_{n-1,\,0,05} = -t_{19,\,0,05} = -1,73$$

$$-1,75 < -1,73 \Rightarrow RH_0$$

مثال) از یک نمونه تصادفی شامل 36 مشاهده نتایج $\bar{X} = 80,4$ و $S = 16,2$ بدست آمده‌اند در مورد فرضیه‌های زیر تصمیم بگیرید. از روش آماره آزمون (مستقیم) و از روش P_Value

مقدار احتمال

الف)

$$\begin{cases} H_0 : \mu = 74 \\ H_1 : \mu > 74 \end{cases} \qquad \alpha = 0,05$$

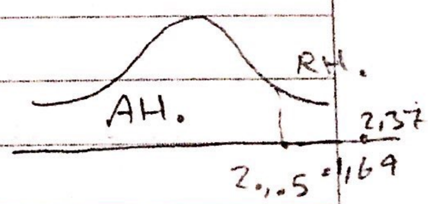
ب)

$$\begin{cases} H_0 : \mu = 76 \\ H_1 : \mu \neq 76 \end{cases}$$

الف)

$$n = 36 \geqslant 30 \ , \ \bar{X} = 80,4 \ , \ S = 16,2$$

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{80,4 - 74}{\frac{16,2}{\sqrt{36}}} = 2,37$$



$$\text{if } \underbrace{Z_0}_{2,37} > Z_\alpha = Z_{0,05} = 1,64 \Rightarrow RH_0$$

جدول نرمال $P(Z \leqslant z) = \Phi(z)$, $P\Big(\underbrace{\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}}_{Z_0} \geqslant 2,37\Big) = 1 - \Phi(2,37)$

$$= 0,0089$$

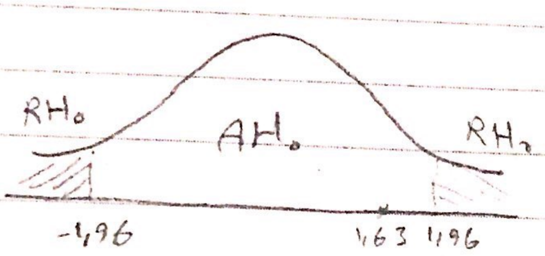
$$\text{if } p\text{-value} < \alpha \Rightarrow RH_0$$

$$0,0089 < 0,05 \Rightarrow RH_0$$

ب/

$$Z_0 = \frac{80,4 - 76}{\frac{16,2}{\sqrt{36}}} = 1,63$$

$$\text{if } Z_0 > Z_{\frac{\alpha}{2}} \text{ or } Z < -Z_{\frac{\alpha}{2}} \Rightarrow RH_0$$



$$\cancel{1,63 > 1,96} \Rightarrow AH_0$$

$$\longrightarrow 2\min\{p(Z_0 > Z_0), p(Z_0 < -Z_0)\}$$

$$P\text{-value} = p(|Z_0| > 1,63) = p(Z_0 > 1,63) + p(Z_0 < -1,63)$$

$$= (1 - \Phi(1,63)) + (\Phi(-1,63)) = 0,1032 < 0,05 \Rightarrow AH_0$$

پیرامون

- آزمون فرض برای نسبت در جامعه: فرض کنید هدف مطالعه نسبت در جامعه است و در خصوص نسبت جامعه ادعایی شده است که مطالعه‌ی کل جامعه امکان‌پذیر نیست و از روی نمونه می‌خواهیم ادعا را بررسی کنیم سه فرضیه زیر ممکن است مورد نظر باشد:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases} \qquad \begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases} \qquad \begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

برای این منظور نمونه‌ای استخراج می‌شود و قبلاً دیدیم:

$$\hat{p} = p \underset{\text{بزرگ}}{\overset{n \text{ اگر}}{\approx}} N\left(p, \frac{pq}{n}\right)$$

پس تحت H_0:

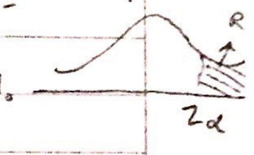
$$\hat{p} \approx N\left(p_0, \frac{p_0 q_0}{n}\right)$$

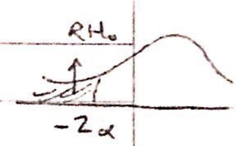
Subject:
Year: Month: Date:

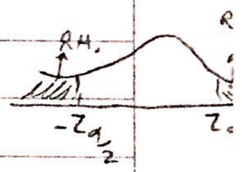
$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0,1) \longrightarrow \text{آماره‌ی آزمون}$$

پس داریم:

فرضیه آزمون	سطح آزمون	آماره	
$H_0: p = p_0$ $H_1: p > p_0$	α	$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$	if $Z_0 > Z_\alpha \Rightarrow RH_0$
$H_0: p = p_0$ $H_1: p < p_0$	α	"	if $Z_0 < -Z_\alpha \Rightarrow RH_0$
$H_0: p = p_0$ $H_1: p \neq p_0$	α	"	if $Z_0 > Z_{\alpha/2}$ or $Z_0 < -Z_{\alpha/2} \Rightarrow RH_0$







مثال) بر مبنای یک سرشماری در 5 سال پیش 20% خانواده های یک شهر زیر خط فقر زندگی می کردند. برای معلوم شدن این موضوع که آیا این درصد تغییری کرده است یا خیر نمونه ای تصادفی به حجم 400 خانواده مطالعه شد و معلوم شد ۷۰ نفر زیر خط فقرند آیا این یافته ها دلالت بر این دارد که درصد خانواده های زیر خط فقر با 5 سال پیش فرق کرده است؟ این فرض H_1 را نشان میدهد $\Rightarrow p \neq 0/2$

همه‌ی خانواده‌ها

نمونه خانواده‌ها

p

آزمون را از روش مستقیم و از روش مقدار احتمال انجام دهید.

$$\begin{cases} H_0: p = 0/2 \\ H_1: p \neq 0/2 \end{cases}$$

$n = 400$ $\quad \alpha = 0/05$

$X = 70$ تعداد افرادی که صفت را دارند $\longrightarrow \hat{p} = \frac{70}{400} = 0/175$

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0/175 - 0/2}{\sqrt{\frac{0/2 \times 0/8}{400}}} = -1/25$$

Sadra

قدر مطلق برای آزمون دو طرفه

$\alpha = 0{,}025 \rightarrow Z_{0{,}025} = \boxed{1{,}96}$

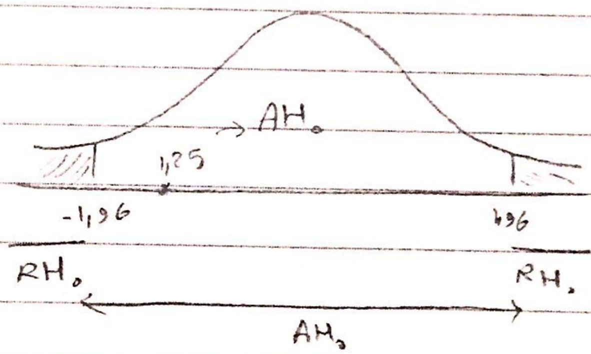
Conclusion $\rightarrow$ if $|Z_0| > Z_{\frac{\alpha}{2}} = 1{,}96$

$\Rightarrow 1{,}25 \ngtr 1{,}96 \Rightarrow AH_0$

$p\text{-value} = p(|Z_0| > -1{,}25) = p(Z_0 > 1{,}25) + p(Z_0 < -1{,}25)$

$= 1 - p(Z_0 < 1{,}25) + p(Z_0 < -1{,}25) = 1 - \phi(1{,}25) + \phi(-1{,}25) = \boxed{0{,}21}$

$p\text{-value} < 0{,}05 \Rightarrow AH_0$

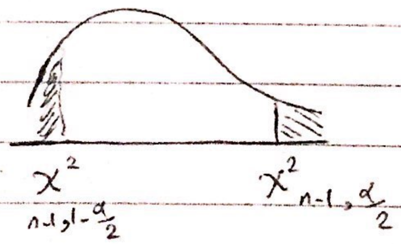


۰ آزمون برای واریانس جامعه : دیدیم برآورد نقطه‌ای خوب برای واریانس جامعه وقتی

μ مجهول است عبارت است از: $\hat{\sigma}^2 = S^2$ $\quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

از طرفی دیدیم وقتی جامعه نرمال است و یا n بزرگ:

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



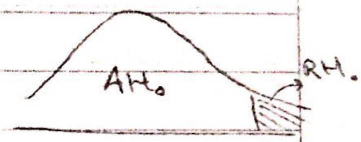
برای آزمون فرض، فرضهای مختلف داریم:

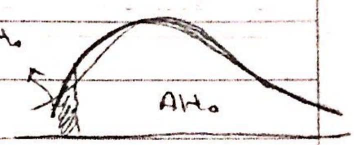
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \rightarrow \mu \text{ مجهول}$$

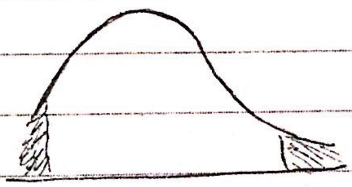
Subject:
Year: Month: Date:

$$S_n^2 = \frac{1}{n}\sum(x_i-\mu)^2 \rightarrow \mu \text{ معلوم.}$$

تشکیل فرضیه	سطح آزمون	نمونه‌گیری و محاسبه آماره	نتیجه‌گیری
$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases}$	α	$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$	if $\chi_0^2 > \chi^2_{n-1,\alpha} \Rightarrow RH_0$
$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases}$	α	$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$	if $\chi_0^2 < \chi^2_{n-1,1-\alpha}$
$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$	α	$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$	if $\chi_0^2 > \chi^2_{n-1,\frac{\alpha}{2}}$ or $\chi_0^2 < \chi^2_{n-1,1-\frac{\alpha}{2}}$







نکته: اگر μ معلوم باشد می‌دانیم برآوردگر خوب برای σ^2، S_n^2 است یعنی:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2, \quad \frac{nS_n^2}{\sigma^2} \sim \chi^2_n$$

و در روابط بالا باید از آماره‌ی آزمون زیر استفاده کنید و از توزیع کای‌دو با درجه آزادی n.

$$\chi_0^2 = \frac{nS_n^2}{\sigma_0^2}$$

یعنی هرجایی که χ^2_{n-1} است جای (n-1)، n بذاریم، چون μ معلوم است.

Sadra

Subject:
Year: Month: Date:

مثال: فرض کنید بر مبنای نمونه تصادفی به حجم $n=25$ از توزیع نرمال، فاصله اطمینان

برای میانگین جامعه به صورت $(106,8 \,,\, 115,2)$ باشد. از این نتیجه استفاده شده و

الف) مقدار انحراف معیار نمونه را بدست آورید.

ب) در سطح آزمون $0,05$ فرض انحراف معیار جامعه برابر 3 هست یا نه؟

الف) $n=25$ — جامعه نرمال

$\mu = (106,8 \,,\, 115,2)$ — حجم نمونه کوچک

$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma \neq 3 \end{cases}$ — واریانس مجهول

$$\mu : \left(\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}}\right) \rightsquigarrow \begin{cases} \bar{X} - t_{n-1,\frac{\alpha}{2}} \dfrac{S}{\sqrt{n}} = 106,8 \\ \bar{X} + t_{n-1,\frac{\alpha}{2}} \dfrac{S}{\sqrt{n}} = 115,2 \end{cases}$$

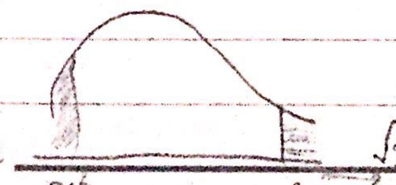
$$t_{24,0,025} = 2,064 \rightsquigarrow \begin{cases} \bar{X} - 2,064 \dfrac{S}{5} = 106,8 \\ \bar{X} + 2,064 \dfrac{S}{5} = 115,2 \end{cases}$$

$$2\bar{X} = 222 \Rightarrow \bar{X} = 111 \,,\, \bar{x} = 111 \rightsquigarrow 111 + 2,064 \frac{S}{5} = 115,2$$

$$\Rightarrow \frac{2,064\,S}{5} = 4,2 \Rightarrow 2,064\,S = 21 \Rightarrow S = \boxed{\frac{21}{2,064} = 10,17}$$

ب)

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \xrightarrow{\text{انحراف معیار}} \sqrt{\chi_0^2} = \frac{\sqrt{(n-1)}\,S}{\sigma_0} = \frac{\sqrt{24} \times 10,17}{3} = 16,61$$



$\sqrt{39,36} = 6,27$

Sadra

$3,52 = \sqrt{12,40} = \chi^2_{24,0,975} \quad \chi^2_{24,0,025}$

اگر دو جامعه نسبت میزان بارندگی یا نوع ... دارو - کود - بارش و ...) *

Subject:
Year: Month: Date:

مبحث پایان ترم ۲

استنباط پیرامون دو جامعه ۲

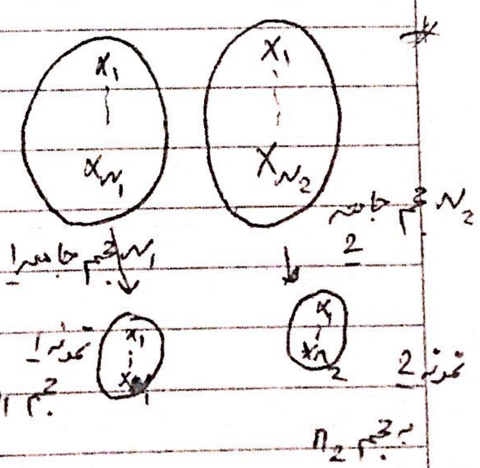
در این قسمت روش‌های استنباط آماری برای مقایسه دو جامعه بر مبنای نمونه‌های مستقل را مورد بحث قرار میدهیم دقت کنید حتماً نباید دو جامعه داشته باشیم گاهی منظور دو تیمار (روش درمان) مثل دو نوع لاستیک یا دو نوع دارو مقایسه می‌شوند.

مثلاً فرض کنید 8 بیمار برای مطالعه‌ی اثر نوع دارو در اختیار یک محقق قرار میدهند که هدف مقایسه اثر داروی شماره یک با اثر داروی شماره 2 در مداوای بیماران است.

ابتدا به طور تصادفی 8 بیمار به دو گروه تقسیم میشوند و به گروه ۱ داروی شماره ۱ و به گروه 2 داروی شماره 2 داده میشود مشاهدات حاصل از داروی شماره یک هیچ ارتباطی با مشاهدات حاصل از داروی شماره 2 ندارد و تقسیم بیمارها به دو گروه کاملاً تصادفی انجام شده است.

پس در این بخش میخواهیم دو تیمار یا دو جامعه را بر مبنای نمونه‌های مستقل مقایسه کنیم.

برای استنباط پیرامون دو جامعه با فرض اینکه به هر دلیلی سرشماری امکان پذیر نیست ابتدا دو نمونه از هر جامعه استخراج میکنیم. استخراج دو نمونه کاملاً مستقل است.



از جامعه اول ← $\underbrace{X_1 \dots X_{n_1}}_{\text{نمونه از هم مستقل}}$

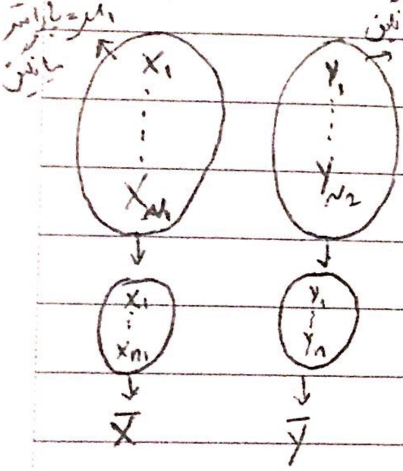
از هم مستقل

از جامعه دوم ← $\underbrace{Y_1 \dots Y_{n_1}}_{\text{نمونه از هم مستقل}}$

Sadra

ما با استنباط پیرامون یک جامعه می‌خواهیم از روی نمونه‌های تصادفی دو جامعه را باهم مقایسه کنیم و پارامترهای مهم میانگین و واریانس نسبت بودن را مورد بررسی قرار میدهیم.

مقایسه میانگین دو جامعه ($\mu_2-\mu_1$) یا ($\mu_1-\mu_2$ فرقی ندارد) : ابتدا دو نمونه مستقل از دو جامعه به ترتیب با میانگین μ_1 و μ_2 استخراج می‌شود دیده‌ایم میانگین نمونه برآورد خوبی برای میانگین‌های جامعه است.



$$\hat{\mu}_1 = \bar{X} \to E(\bar{X}) = \mu_1$$

$$\hat{\mu}_2 = \bar{Y} \to E(\bar{Y}) = \mu_2$$

از خواص امید ریاضی:

$$E(\bar{Y}-\bar{X}) = E(\bar{Y}) - E(\bar{X}) = \mu_2-\mu_1$$

پس $\bar{Y}-\bar{X}$ یک برآوردکننده نااریب برای $\mu_2-\mu_1$ است.

پس یک برآوردکننده برای $\mu_2-\mu_1$ عبارت است از $\bar{Y}-\bar{X}$.

$$\boxed{\widehat{\mu_2-\mu_1} = \bar{Y}-\bar{X}} \rightarrow$$ «برآورد نقطه‌ای»

برآورد فاصله‌ای و فاصله اطمینان برای $\mu_2-\mu_1$:

برای بدست آوردن برآورد فاصله‌ای حالت‌های مختلف را بررسی می‌کنیم:

حالت اول: نمونه‌ها به اندازه کافی بزرگ باشند و واریانس‌ها معلوم یا جامعه نرمال باشد

$$V(ax+by) = a^2V(X) + b^2V(Y) - 2ab\,cov(X,Y)$$

واریانس ها معلوم (یا جامعه نرمال باشد و واریانس معلوم).

از طرفی $E(\bar{Y}-\bar{X}) = \mu_2-\mu_1$ داریم

چون از هم مستقل اند

$$V(\bar{Y}-\bar{X}) = V(\bar{Y}) + V(\bar{X}) - 2\,\overset{0}{cov}(\;) = V(\bar{Y}) + V(\bar{X})$$

بنابراین در این حالت چون نمونه به اندازه کافی بزرگ هستند (یا جامعه نرمال):

طبق خواص نرمال

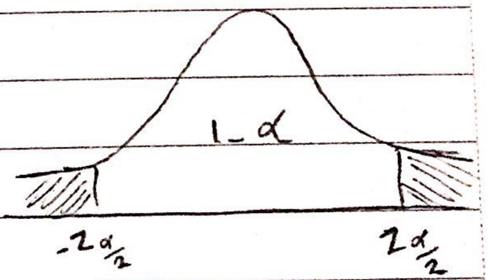
$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\longrightarrow \boxed{\bar{Y}-\bar{X} \sim N\left(\mu_2-\mu_1, \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}\right)}$$

$$\Longrightarrow Z = \frac{(\bar{Y}-\bar{X}) - (\mu_2-\mu_1)}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} \sim N(0,1)$$

فاصله اطمینان با ضریب اطمینان $1-\alpha$ مورد نظر



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

با جایگذاری Z :

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{(\bar{Y}-\bar{X}) - (\mu_2-\mu_1)}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

با نگه داشتن $\mu_2-\mu_1$ در وسط و انتقال سایر جمله ها به طرفین داریم:

$$P\left((\bar{y}-\bar{x}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_1}} < \mu_2-\mu_1 < (\bar{y}-\bar{x}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_1}}\right)$$

بنابراین فاصله اطمینان $\%\ 100(1-\alpha)$ برای $\mu_2-\mu_1$ وقتی n_1 و n_2 بزرگ باشند و واریانس ها معلوم هست به صورت زیر است:

$$\boxed{\mu_2-\mu_1 : \left((\bar{y}-\bar{x}) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_1}}\right)}$$

نکته: فقط در حالتیکه n_1 و n_2 هر دو بزرگ هستند و واریانس ها مجهول در این صورت میتوان در روابط بالا به جای σ_1^2 و σ_2^2 از S_1^2 و S_2^2 استفاده کرد:

$$\boxed{\mu_2-\mu_1 : \left((\bar{y}-\bar{x}) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{S_2^2}{n_2}+\frac{S_1^2}{n_1}}\right)}$$

$$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i-\bar{y})^2 \quad , \quad S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2$$

مثال) برای مقایسه سن ازدواج زنان در دو ایل A و B از هر ایل نمونه ای شامل 100 زن استخراج شده است و سن ازدواج آنها را ثبت کرده ایم و خلاصه آماره ها به شرح زیر بدست آمده است.

	A	B
حجم نمونه	$n_1=100$	$n_2=100$
میانگین	$\bar{x}=18{,}5$	$\bar{y}=20{,}7$
واریانس	$S_A=6{,}2$	$S_B=5{,}8$

$\mu_B-\mu_A=$

برآورد نقطه ای $\Leftarrow \bar{y}-\bar{x}=20{,}7-18{,}5=2{,}2$

سن ازدواج در ایل A کمتر از ایل B است.

برآورد فاصله ای: $\alpha=0{,}05 \Rightarrow 1-\alpha=\%95$ و $Z_{0{,}025}=1{,}96$

$$\mu_B - \mu_A : \left((\bar{y} - \bar{x}) \pm 1{,}96 \sqrt{\frac{S_B^2}{n_2} + \frac{S_A^2}{n_1}} \right)$$

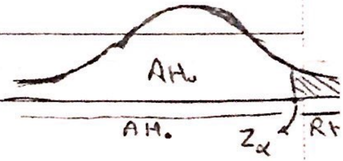
$$= \left((20{,}7 - 18{,}5) \pm 1{,}96 \sqrt{\frac{(5{,}8)^2}{100} + \frac{(6{,}2)^2}{100}} \right)$$

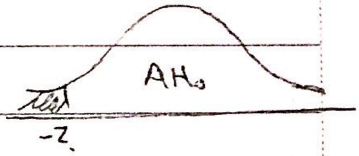
چون کران پایین و کران بالا $\rightarrow (0{,}52 \,,\, 3{,}88) =$
فاصله اطمینان مثبت است پس با اطمینان ۰٫95 سن ازدواج زنان ایل B از ایل A بیشتر است.

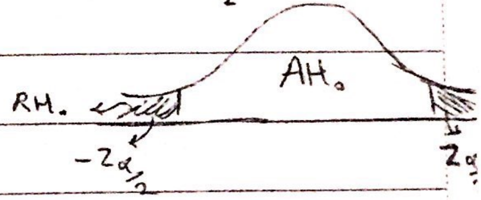
- آزمون فرض در این حالت: (n_1 و n_2 بزرگند و واریانس‌ها مجهول (معلوم)) سه فرضیه مورد توجه است به شرح زیر:

تصمیم‌گیری (قاعده تصمیم)	آماره‌ی آزمون	سطح آزمون	تشکیل فرضیه
if $Z_0 > Z_\alpha \Rightarrow RH_0$	$Z_0 = \dfrac{(\bar{y}-\bar{x}) - \overbrace{(\mu_2-\mu_1)}^{\delta_0}}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}}$	α	$\begin{cases} H_0: \mu_2 - \mu_1 = \delta_0 \\ H_1: \mu_2 - \mu_1 > \delta_0 \end{cases}$
if $Z_0 < -Z_\alpha \Rightarrow RH_0$	〃	α	$\begin{cases} H_0: \mu_2 - \mu_1 = \delta_0 \\ H_1: \mu_2 - \mu_1 < \delta_0 \end{cases}$
* if $\lvert Z_0 \rvert > Z_{\frac{\alpha}{2}} \Rightarrow RH_0$	〃	α	$\begin{cases} H_0: \mu_2 - \mu_1 = \delta_0 \\ H_1: \mu_2 - \mu_1 \neq \delta_0 \end{cases}$

(معادل)







$\Rightarrow Z_0 > Z_{\frac{\alpha}{2}}$ or $Z_0 < -Z_{\frac{\alpha}{2}} \Rightarrow RH_0$

مثال) آیا مثال قبل شواهد قوی برای قبول این ادعا که متوسط سن ازدواج در دو ایل متفاوت است وجود دارد؟ سطح آزمون را ۰٫02 در نظر بگیرید.

$$\begin{cases} H_0: \mu_B - \mu_A = 0 \\ H_1: \mu_B - \mu_A \neq 0 \end{cases}$$

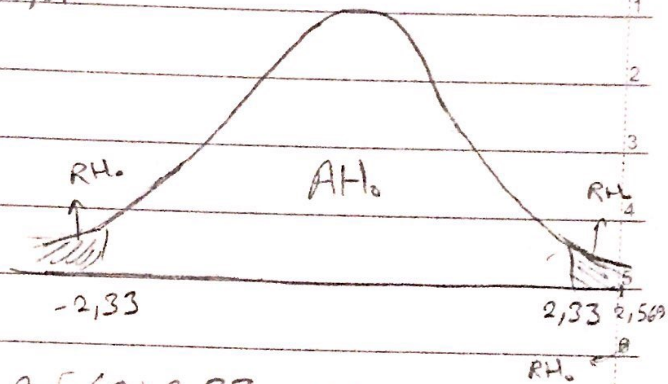
③

Year: Month: Date:

Sa Su Mo Tu We Th Fr

$\alpha = 0.02 \to 1-\alpha = 0.98 \Rightarrow \frac{\alpha}{2} = 0.01$

$$Z_0 = \frac{(20.7-18.5)-(0)}{\sqrt{\frac{(5.8)^2}{100}+\frac{(6.3)^2}{100}}} = 2.569$$



$Z_{0.01} = 2.33$

$2.569 > 2.33 \Rightarrow RH_0$

تفسیر: چون H_0 رد می‌شود ⇐ متوسط سن ازدواج دو ایل متفاوت است.

حالت دوم: اگر n_1 یا n_2 کوچک باشند و جوامع نرمال و واریانس ها مجهول، در این حالت

دو زیر حالت در نظر میگیریم ① زیر حالت اول: دو جامعه از نظر پراکندگی یکسان هستند

$\sigma_1^2 = \sigma_2^2 = \sigma^2$

② زیر حالت دوم: دو جامعه از نظر پراکندگی یکسان نیستند.

$\sigma_1^2 \neq \sigma_2^2$

برای زیر حالت اول داریم:

چون جامعه نرمال $X_1 \dots X_n \longrightarrow \bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1} = \frac{\sigma^2}{n_1}\right)$

نمونه

از جامعه نرمال اول

$Y_1 \dots Y_n \longrightarrow \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2} = \frac{\sigma^2}{n_2}\right)$

نمونه‌ای تصادفی از جامعه نرمال دوم

$S_1^2 = \frac{1}{n_1-1} \sum (x_i - \bar{x})^2$

$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$

$S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2$

$\frac{(n_2-1)s^2}{\sigma^2} \sim \chi^2_{n_2-1}$

(جمع دو توزیع کای دو، کای دو می‌شود (فقط درجه آزادی‌ها با هم جمع می‌شوند

چون $\sigma_1^2=\sigma_2^2=\sigma^2$ پس یک برآورد ترکیبی از دو نمونه به صورت زیر برای آن می‌توان بدست آورد.

$$\left.\begin{aligned} E(S_1^2)&=\sigma_1^2=\sigma^2\\ E(S_2^2)&=\sigma_2^2=\sigma^2\end{aligned}\right\}\rightarrow S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

puld
ترکیب شده

$$E(S_p^2)=\frac{(n_1-1)\sigma^2+(n_2-1)\sigma^2}{n_1+n_2-2}=\frac{\sigma^2(n_1+n_2-2)}{n_1+n_2-2}=\sigma^2$$

از طرفی طبق خواص توزیع نرمال:

$$\bar{y}-\bar{x}\sim N\left(\mu_2-\mu_1,\ \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$$

$$\begin{cases} E(\bar{y}-\bar{x})=\mu_2-\mu_1\\ V(\bar{y}-\bar{x})=V(\bar{y})+V(\bar{x})-2\operatorname{cov}(\bar{x},\bar{y})=\dfrac{\sigma^2}{n_2}+\dfrac{\sigma^2}{n_1}=\sigma^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)\end{cases}$$

با استاندارد کردن:

$$Z=\frac{(\bar{y}-\bar{x})-(\mu_2-\mu_1)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1) \qquad (1)$$

همچنین چون:

$$\frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2_{n_2-1},\quad \frac{(n_1-1)S_1^2}{\sigma^2}\sim\chi^2_{n_1-1}$$

بنابراین:

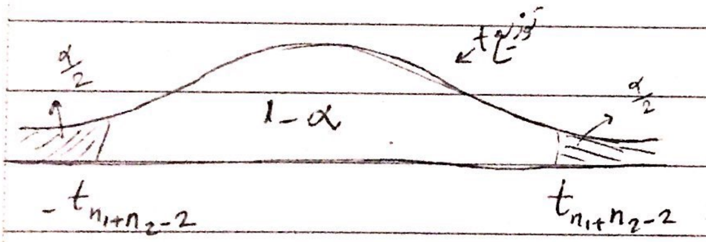
$$\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}\sim\chi^2_{\underbrace{(n_2-1)+(n_1-1)}_{n_1+n_2-2}} \qquad (2)$$

اکنون یک توزیع نرمال استاندارد (1) و یک توزیع کای دو (2) داریم و می‌دانیم:

$$T = \frac{\dfrac{(\bar{Y}-\bar{X})-(\mu_2-\mu_1)}{\cancel{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\dfrac{\left(\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\cancel{\sigma^2}}\right)}{n_1+n_2-2}}} = t_{n_1+n_2-2}$$

$$\Rightarrow \boxed{T = \frac{(\bar{Y}-\bar{X})-(\mu_2-\mu_1)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}} \quad \star$$

برای بدست آوردن فاصله اطمینان $100(1-\alpha)\%$ برای $\mu_2-\mu_1$ داریم:



$$P\left(-t_{n_1+n_2-2,\frac{\alpha}{2}} < T < T_{n_1+n_2-2,\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow P\left(-t_{n_1+n_2-2,\frac{\alpha}{2}} < \frac{(\bar{Y}-\bar{X})-(\mu_2-\mu_1)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < t_{n_1+n_2-2,\frac{\alpha}{2}}\right) = 1-\alpha$$

پس:

$$\Rightarrow P\left((\bar{Y}-\bar{X}) - t_{n_1+n_2-2,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_2-\mu_1 < (\bar{Y}-\bar{X}) + t_{n_1+n_2-2,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha$$

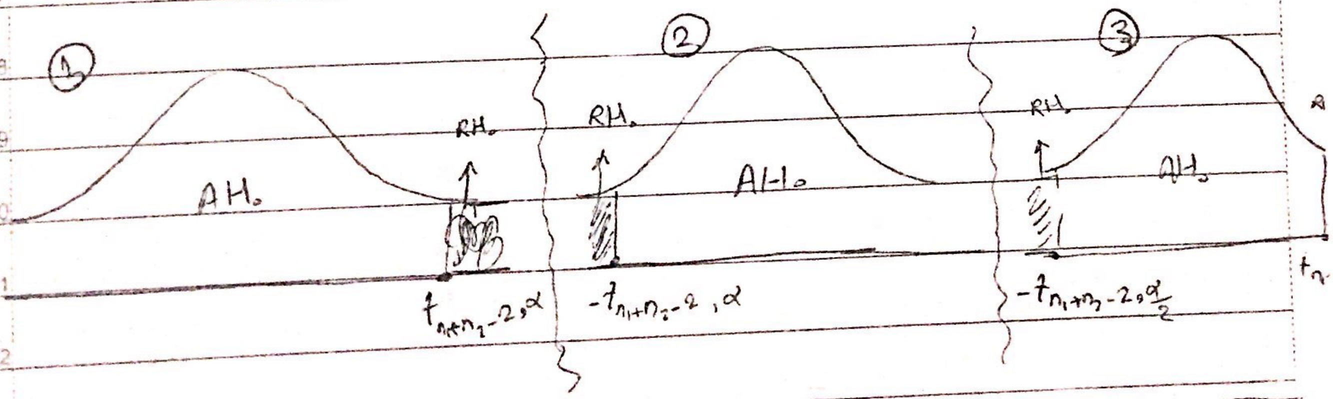
پس فاصله اطمینان $100(1-\alpha)\%$ برای حالتی که: جوامع نرمال‌اند، σ_1^2 و σ_2^2 مجهول‌اند ولی برابر، n_1 و n_2 یا هر دو کوچک باشند:

$\delta_0 = 0$

$$\mu_2 - \mu_1 : (\bar{y} - \bar{x}) \pm t_{n_1+n_2-2,\frac{\alpha}{2}} \, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

برای آزمون فرضیه های زیر هم از آماره ی (*) استفاده می کنیم:

فرضیه ها	سطح آزمون	آماره آزمون	قاعده تصمیم
① $\begin{cases} H_0 : \mu_2 - \mu_1 = \delta_0 \quad (\equiv H_0 : \mu_2 - \mu_1 \leq \delta_0) \\ H_1 : \mu_2 - \mu_1 > \delta_0 \end{cases}$	α	$T_0 = \dfrac{(\bar{y} - \bar{x}) - \delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	if $T_0 > t_{n_1+n_2-2,\alpha} \Rightarrow RH_0$
② $\begin{cases} H_0 : \mu_2 - \mu_1 = \delta_0 \quad (\equiv H_0 : \mu_2 - \mu_1 \geq \delta_0) \\ H_1 : \mu_2 - \mu_1 < \delta_0 \end{cases}$	α	//	if $T_0 < t_{n_1+n_2-2,\alpha} \Rightarrow RH_0$
③ $\begin{cases} H_0 : \mu_2 - \mu_1 = \delta_0 \\ H_1 : \mu_2 - \mu_1 \neq \delta_0 \end{cases}$	α	//	if $\lvert T_0 \rvert < t_{n_1+n_2-2,\frac{\alpha}{2}} \Rightarrow R$



برای زیر حالت دوم داریم:

جوامع نرمال و نمونه کوچک و واریانس ها مجهول و نابرابر:

$$\sigma_1^2 \neq \sigma_2^2$$

در این زیر حالت از آماره‌ی زیر استفاده می‌شود:

$$T^* = \frac{(\bar{Y}-\bar{X}) - (\mu_2-\mu_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t^*_{\underbrace{\min(n_1-1, n_2-1)}_{df}}$$

یا df را به صورت زیر بدست می‌آوریم که دقیق‌تر است:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$

و برای آزمون کردن فرضیه‌های بالا از آماره آزمون:

$$T_0^* = \frac{(\bar{Y}-\bar{X}) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (**)$$

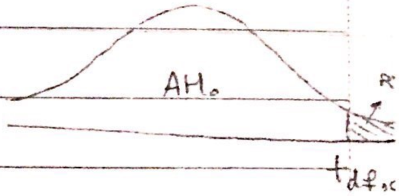
استفاده می‌کنیم و فاصله اطمینان را در این حالت:

$$\mu_2 - \mu_1 : (\bar{Y}-\bar{X}) \pm t^*_{df, \frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

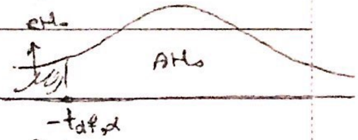
و برای آزمون فرضیه‌های دیگر هم از آماره‌ی ** استفاده می‌کنیم:

فرضیه‌ها — سطح آزمون — آماره آزمون — قاعده تصمیم

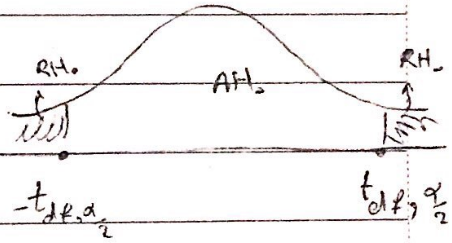
$$\begin{cases} H_0: \mu_2-\mu_1=\delta_0 \\ \quad = \mu_2-\mu_1 \leq \delta_0 \\ H_1: \mu_2-\mu_1 > \delta_0 \end{cases} \quad \alpha \quad T_0^* = \frac{(\bar{y}-\bar{x})-\delta_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \quad \text{if } T_0^* > t^*_{df,\alpha} \Rightarrow RH_0$$



$$\begin{cases} H_0: \mu_2-\mu_1=\delta_0 \\ \quad = \mu_2-\mu_1 \geq \delta_0 \\ H_1: \mu_2-\mu_1 < \delta_0 \end{cases} \quad \alpha \quad '' \quad \text{if } T_0^* < -t_{df,\alpha} \Rightarrow RH_0$$



$$\begin{cases} H_0: \mu_2-\mu_1=\delta_0 \\ \\ H_1: \mu_2-\mu_1 \neq \delta_0 \end{cases} \quad \alpha \quad '' \quad \text{if } |T_0^*| < t^*_{df,\frac{\alpha}{2}} \Rightarrow RH_0$$



مقایسه واریانس‌های دو جامعه: هدف مقایسه σ_1^2 و σ_2^2 یعنی واریانس‌های دو جامعه

است. قبلاً دیدیم: (با فرض نرمال بودن دو جامعه یا n_1 و n_2 بزرگ):

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$$

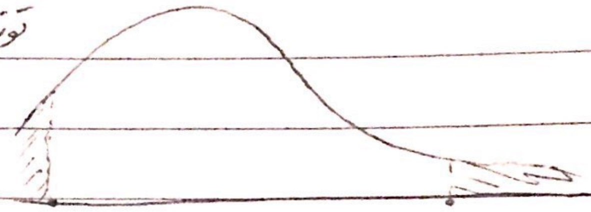
$$\Longrightarrow F = \frac{\left(\frac{(n_1-1)S_1^2}{\sigma_1^2}\right)\Big/ n_1-1}{\left(\frac{(n_2-1)S_2^2}{\sigma_2^2}\right)\Big/ n_2-1}$$

$$\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$$

(توزیع F و کای دو هر دو مثبت هستند و مقدار منفی معنا ندارد)

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{n_1-1,\,n_2-1}$$

توزیع



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

$$P\left(F_{(n_1-1),(n_2-1),1-\frac{\alpha}{2}}<\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}<F_{n_1-1,n_2-1,\frac{\alpha}{2}}\right)=1-\alpha$$

بنابراین فاصله اطمینان ٪۱۰۰(1-α) برای نسبت واریانس‌ها:

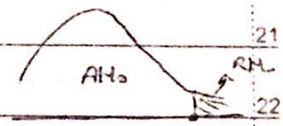
$$\frac{\sigma_2^2}{\sigma_1^2}:\left(\frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\ ,\ \frac{S_2^2}{S_1^2}F_{n_1-1,n_2-1,\frac{\alpha}{2}}\right)$$

$$\boxed{F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}=\frac{1}{F_{n_2-1,n_1-1,\frac{\alpha}{2}}}}$$

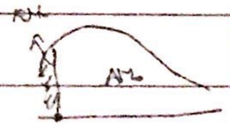
برای آزمون کردن فرضیه‌های مربوط به واریانس دو جامعه داریم:

از *** در آزمون فرضیه‌ها

$\begin{cases}H_0:\sigma_1^2=\sigma_2^2\\ H_1:\sigma_1^2>\sigma_2^2\end{cases}$ $\quad\alpha\quad$ $F_0=\dfrac{S_1^2}{S_2^2}$ $\quad$ if $F_0>F_{n_1-1,n_2-1,\alpha}\Rightarrow RH_0$



$\begin{cases}H_0:\sigma_1^2=\sigma_2^2\\ H_1:\sigma_1^2<\sigma_2^2\end{cases}$ $\quad\alpha\quad$ // $\quad$ if $F_0<F_{(n_1-1),(n_2-1),1-\alpha}$



$\begin{cases}H_0:\sigma_1^2=\sigma_2^2\\ H_1:\sigma_1^2\neq\sigma_2^2\end{cases}$ $\quad\alpha\quad$ // $\quad$ if $\begin{cases}F_0<F_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\\ F_0>F_{n_1-1,n_2-1,\frac{\alpha}{2}}\end{cases}\Rightarrow RH_0$

مقایسه نسبت های دو جامعه: فرض کنید هدف مقایسه نسبت صفت خاص در دو جامعه است.

مثلاً مقایسه نرخ بیکاری در شهرستان دماوند، یا نسبت افراد دارای ماشین ظرفشویی

در منطقه یک و یازده تهران.

با فرض بزرگ بودن نمونه ها (n_1 و n_2) پس دو نمونه با اندازه کافی بزرگ از دو جامعه استخراج

می کنیم. قبلاً دیدیم نسبت نمونه برآورد خوبی برای نسبت جامعه است یعنی:

نسبت نمونه برای جامعه 2

$$\underbrace{\hat{p}_1}_{\text{نسبت نمونه برای جامعه 1}} = \frac{X}{n_1} \quad , \quad \hat{p}_2 = \frac{Y}{n_2}$$

$\Longleftarrow$ نمونه ها مستقل

$$E(\hat{p}_1) = p_1 \quad , \quad E(\hat{p}_2) = p_2$$

طبق خواص امید ریاضی:

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

پس برآورد نقطه ای برای $p_1 - p_2$، $\hat{p}_1 - \hat{p}_2$ است.

و چون نمونه ها به اندازه کافی بزرگ هستند پس می دانیم:

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right) \xrightarrow{\text{پس}} V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2)$$

$$\hat{p}_2 \sim N\left(p_2, \frac{p_2 q_2}{n_2}\right) \qquad = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

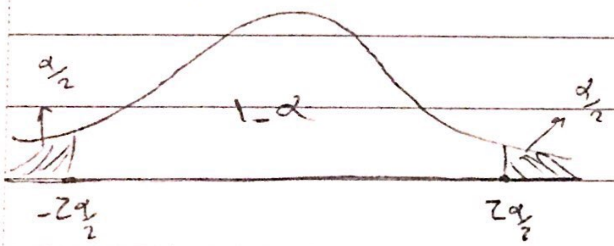
و از خواص توزیع نرمال:

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

با استاندارد کردن:

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} \quad (1)$$

نکته: چون $\frac{p_1q_1}{n_1}$, $\frac{p_2q_2}{n_2}$ مجهولند به‌دلیل از برآوردگرها یعنی $\frac{\hat{p}_1\hat{q}_1}{n_1}$, $\frac{\hat{p}_2\hat{q}_2}{n_2}$ استفاده می‌کنیم. از رابطه (1)



$$P\left(-Z_{\frac{\alpha}{2}}<\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}}<Z_{\frac{\alpha}{2}}\right)=1-\alpha$$

$$\Rightarrow P\left((\hat{p}_1-\hat{p}_2)-Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}<p_1-p_2<(\hat{p}_1-\hat{p}_2)+Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\right)$$
$$=1-\alpha$$

بین فاصله اطمینان ٪۱۰۰(۱-α) برای p_1-p_2:

$$p_1-p_2:(\hat{p}_1-\hat{p}_2)\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$

اگر خودشون نبود از برآوردشون استفاده می‌کنیم یعنی $\frac{\hat{p}_1\hat{q}_1}{n_1}$, $\frac{\hat{p}_2\hat{q}_2}{n_2}$

برای آزمون کردن فرضیه $H_0: p_1-p_2=0$ (آیا دو جامعه دارای نسبت برابر هستند؟)

یک مقدار

$$H_0: p_1-p_2=0 \Rightarrow p_1=p_2=p$$

پس آماره آزمون از رابطه ①

$$Z_0 = \frac{(\hat{p}_1-\hat{q}_1)-0}{\sqrt{\frac{pq}{n_1}+\frac{pq}{n_2}}} \quad ②$$

تحت H_0

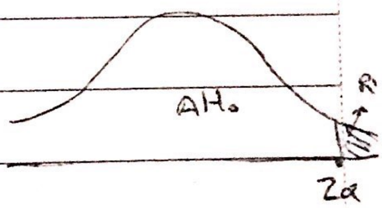
$p_1-p_2=0 \Rightarrow p_1=p_2=p$ این مقدار مجهوله

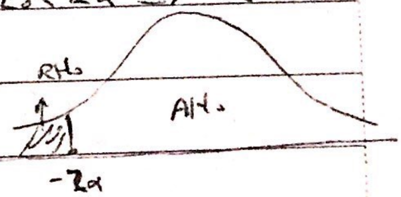
چون معمولاً p و q مجهول‌اند پس یک برآورد ترکیبی از هر دو نمونه برای p در نظر می‌گیریم

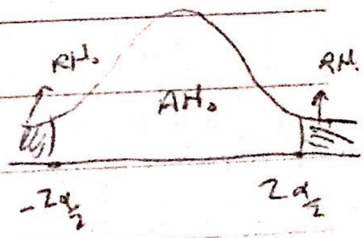
$$\hat{p} = \frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$$

به طور کلی:

فرضیه	سطح آزمون	آماره آزمون ②	قاعده تصمیم
$\begin{cases} H_0: p_1-p_2=0 \\ H_1: p_1-p_2>0 \end{cases}$	α	$Z_0=\dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$	if $Z_0>Z_\alpha \Rightarrow RH_0$
$\begin{cases} H_0: p_1-p_2=0 \\ H_1: p_1-p_2<0 \end{cases}$	α	//	if $Z_0<Z_\alpha \Rightarrow RH_0$
$\begin{cases} H_0: p_1-p_2=0 \\ H_1: p_1-p_2\neq 0 \end{cases}$	α	//	if $\lvert Z_0\rvert>Z_{\alpha/2} \Rightarrow RH_0$







مقایسات زوجی: در این مقایسات نمونه‌ها وابسته هستند. برای مثال فرض کنید هدف مطالعه (رفتنی)
تاثیر دو محیط متفاوت بر روی ظرفیت یادگیری کودکان قبل از مدرسه است. در این گونه
مسائل بهتر است از دوقلوهای یکسان استفاده کرد که از نظر سنی و عوامل ژنتیکی یکسان
هستند یا فرض کنید هدف مقایسه نمرات پاییز و بهار دانشجویان است که از یک گروه
دانشجویی برای هر دو نمونه استفاده می‌شود. پس دو نمونه از دو جامعه مستقل نداریم به عبارتی دو نمونه
وابسته هستند و یک جامعه یا گروه داریم که دو بار مورد آزمایش قرار گرفته اند.

دانشجو	نمره های بهار	نمره های پاییز	
1	X_1	Y_1	$\to (X_1, Y_1)$
$\vdots$	$\vdots$	$\vdots$	
n	X_n	Y_n	$\to (X_n, Y_n)$

برای تحلیل این نوع داده‌ها ابتدا تفاضل‌ها محاسبه می‌شوند.

$$\begin{array}{lll} X_1 & Y_1 & D_1 = X_1 - Y_1 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & D_n = X_n - Y_n \end{array}$$

$X_1 \ldots X_n$

$Y_1 \ldots Y_n$

$, D_i = X_i - Y_i$

بطوریکه $D_1, D_2, \ldots, D_n$ مستقل هستند. فرض کنید δ و σ_D^2 به ترتیب
میانگین تفاضل های جامعه و واریانس آن‌ها باشد.

$(X_1, Y_1) \to D_1 = X_1 - Y_1$

$(X_n, Y_n) \to D_n = X_n - Y_n$

$(X_1, Y_1) \to D_1 = X_1 - Y_1$

$\vdots$

$(X_n, Y_n) \to D_n = X_n - Y_n$

میانگین تفاضل جفتی جامعه

Year: Month: Date:

Sa Su Mo Tu We Th Fr

$$\delta = \frac{1}{N}\sum_{i=1} D_i$$

$$\sigma_D^2 = \frac{1}{N}\sum_{i=1}^{N} (D_i - \delta)^2$$ واریانس تفاضل های جامعه

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$$ $E(\bar{D}) = \delta$ و میانگین تفاضل نمونه

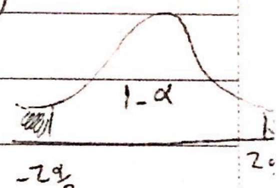
برآورد نقطه‌ای δ

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n} (D_i - \bar{D})$$ $E(S_D^2) = \sigma_D^2$ و واریانس تفاضل نمونه

برآورد نقطه‌ای σ_D^2

با فرض نرمال بودن جامعه یا n به اندازه کافی بزرگ:

$$\bar{D} \sim N\left(\delta, \frac{\sigma_D^2}{n}\right) \longrightarrow Z = \frac{\bar{D} - \delta}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0,1)$$



و مشابه یک جامعه فاصله اطمینان برای میانگین تفاضل جامعه:

$$\underbrace{\delta : \bar{D} \pm Z_{\frac{\alpha}{2}} \frac{\sigma_D}{\sqrt{n}}}_{\text{جامعه نرمال یا n بزرگ و } \sigma_D \text{ معلوم}} \xrightarrow[\text{مجهول باشد یا n بزرگ}]{\sigma_D} \bar{D} \pm Z_{\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

و مشابه یک جامعه اگر جامعه نرمال، n کوچک و σ_D مجهول باشد از رابطه زیر برای محاسبه فاصله اطمینان استفاده می‌کنیم:

$$\delta : \bar{D} \pm t_{n-1, \frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

برای آزمون فرض فرضیه های زیر:

I
$$\begin{cases} H_0 : \delta = \delta_0 \\ H_1 : \delta > \delta_0 \end{cases}$$

II
$$\begin{cases} H_0 : \delta = \delta_0 \\ H_1 : \delta < \delta_0 \end{cases}$$

III
$$\begin{cases} H_0 : \delta = \delta_0 \\ H_1 : \delta \neq \delta_0 \end{cases}$$

! در صورتیکه n بزرگ باشد، جامعه نرمال و σ_D معلوم:

$$Z_0 = \frac{\bar{D} - \delta_0}{\frac{\sigma_D}{\sqrt{n}}}$$ آماره آزمون

TAHERIAN

۲) درصورتیکه n بزرگ باشد و σ_D مجهول:

$$Z_0 = \frac{\bar{D} - \delta_0}{\frac{S_D}{\sqrt{n}}} \longrightarrow$$ آماره‌ی آزمون

۳) اگر n کوچک و جامعه نرمال و σ_D مجهول:

$$T_0 = \frac{\bar{D} - \delta_0}{\frac{S_D}{\sqrt{n}}} \longrightarrow$$ آماره‌ی آزمون

برای حالت 1 و 2 و فرضیه‌های I و II و III:

I فرضیه : قاعده تصمیم $\quad$ if $Z_0 > Z_\alpha \Rightarrow RH_0$

II " : " $\quad$ if $Z_0 < -Z_\alpha \Rightarrow RH_0$

III " : " $\quad$ if $|Z_0| > Z_{\frac{\alpha}{2}} \Rightarrow RH_0$

برای حالت سوم:

I فرضیه : قاعده تصمیم $\quad$ if $T_0 > t_{n-1,\alpha} \Rightarrow RH_0$

II فرضیه : " $\quad$ if $T_0 < -t_{n-1,\alpha} \Rightarrow RH_0$

III فرضیه : " $\quad$ if $|T_0| > t_{n-1,\frac{\alpha}{2}} \Rightarrow RH_0$

مثال) در تحقیقی به منظور مقایسه سیگار طبی با فیلتر، تعداد افراد ترک سیگاری و افرادی که سیگار را ترک نکرده به صورت زیر است.

	ترک سیگار	ادامه سیگار	
سیگار طبی	21	36	57
سیگار فیلتر	11	44	55

صفت خاص

یک فاصله اطمینان ۹۵٪ برای تفاضل بین نسبت افرادی که سیگار را ترک کرده اند در دو گروه به دست آورید.

سیگار طبی: گروه اول $\hat{p} = \frac{21}{57}$

سیگار فیلتر: گروه دوم $\hat{p} = \frac{11}{55}$

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$P_1 - P_2 > 0 \Rightarrow P_1 > P_2$

$$\Rightarrow P_1 - P_2 : \left(\frac{21}{57} - \frac{11}{55}\right) \pm 1.96 \sqrt{\frac{\frac{21}{57} \times \frac{36}{57}}{57} + \frac{\frac{11}{55} \times \frac{44}{55}}{55}} \simeq (\underset{>0}{0.010}\ ,\ \underset{>0}{0.33})$$

با توجه به فاصله اطمینان به دست آمده چون صفر در بازه اطمینان نیست نتیجه می‌گیریم موفقیت سیگار طبی از سیگار فیلتر بیشتر است. (چون بازه اطمینان مثبت هم هست)

مثال) آماره‌های خلاصه زیر از دو نمونه تصادفی مستقل از دو جامعه ثبت شده است. با توجه به این اطلاعات فرضیه را آزمون کنید همچنین یک فاصله اطمینان ۹۵٪ برای $\mu_1 - \mu_2$ به دست آورید (با فرض نرمال)

$n_1 = 12 \quad \bar{x} = 249 \quad S_1 = 19$

$n_2 = 15 \quad \bar{y} = 233 \quad S_2 = 45$

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$$

حل: جامعه نرمال است، حجم نمونه کوچک است و واریانس‌ها مجهول‌اند ← $\sigma_1^2 = \sigma_2^2$

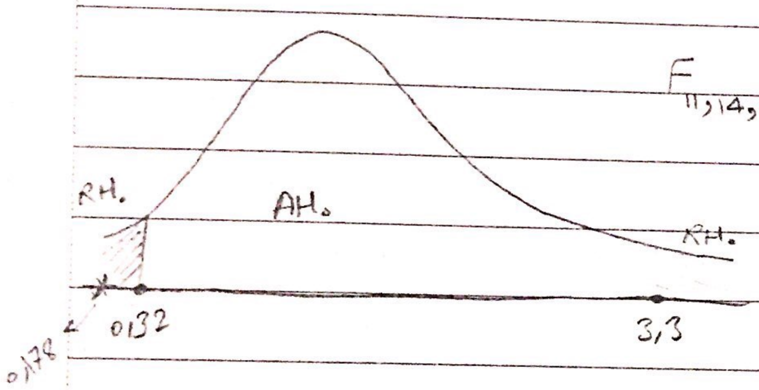
$\sigma_1^2 \neq \sigma_2^2$ ✓

پس اولین گام این است که دو جامعه از نظر پراکندگی بررسی شوند:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \qquad F_0 = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$$

$$F_0 = \frac{361}{2025} = 0{,}178 \qquad F_{11,14,0{,}025} = 3{,}3$$

$$F_{11,14,0{,}975} = \frac{1}{\underbrace{F_{14,11,0{,}025}}_{3{,}1}} = 0{,}32$$



$$0{,}178 < 0{,}32 \Rightarrow RH_0$$

پس دو جامعه از نظر پراکندگی یکسان نیست.

پس از زیر حالت دوم استفاده میشود:

فاصله اطمینان:

$$\mu_1 - \mu_2 : (\bar{X} - \bar{Y}) \pm t^*_{\min\{\overset{11}{(n_1-1)}, \overset{14}{(n_2-1)}\}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$= (249 - 233) \pm t_{11, 0{,}025} \sqrt{\frac{361}{12} + \frac{2025}{15}} \simeq (\underbrace{12{,}28}_{>0}, \underbrace{44{,}28}_{>0})$$

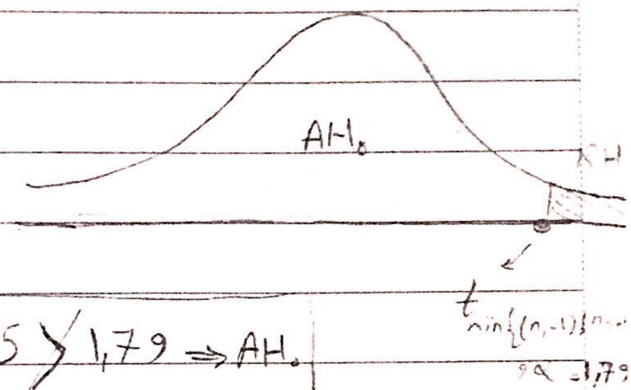
$\mu_1 - \mu_2 > 0$

چون هم کران پایین و هم کران بالای میانگین مثبت است و شامل صفر نیست پس نتیجه میگیریم

با اطمینان 95% میانگین جامعه اول از میانگین جامعه دوم بزرگتر است که از دل این نتیجه را بیرون میکشیم

$$T_0 = \frac{(\bar{X}-\bar{Y})-\delta_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{(249-233)-0}{\sqrt{\frac{19^2}{12}+\frac{45^2}{15}}} \simeq 1{,}25$$

$$\begin{cases} H_0: \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = \overset{\delta_0}{0} \\ H_1: \mu_1 > \mu_2 \rightarrow \mu_1 - \mu_2 > 0 \end{cases}$$



$$t_{11,0{,}05} = 1{,}7959 \qquad \boxed{1{,}25 \not> 1{,}79 \Rightarrow AH_0}$$

در این آزمون یک آزمون یک‌طرفه است و نمی‌توان طبق فاصله اطمینان آن را بررسی کرد.

مثال) آماره‌های خلاصه زیر برای نمونه‌های تصادفی مستقل دو جامعه ثبت شده‌اند با فرض نرمال بودن جوامع فاصله اطمینان ۹۵٪ را برای $\mu_1-\mu_2$ بدست آورید و فرضیه

$$\begin{cases} H_0: \mu_1 \le \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$$

را در سطح ۰٫۰۵ آزمون کنید

$n_1 = 11 \quad \bar{X} = 10{,}7 \quad S_1^2 = 1{,}36$

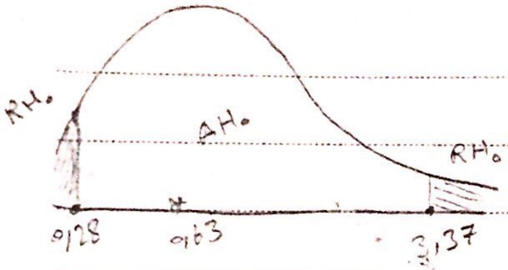
$n_2 = 13 \quad \bar{Y} = 9{,}6 \quad S_2^2 = 2{,}17$

ابتدا فرضیه برابری واریانس‌ها بررسی می‌شود:

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases} \qquad F_0 = \frac{S_1^2}{S_2^2} = \frac{1{,}36}{2{,}17} = 0{,}63$$

$$F_{n_1-1,n_2-1,\frac{\alpha}{2}} = F_{10,12,0{,}025} = 3{,}37 \rightarrow F_{10,12,0{,}975} = \frac{1}{F_{12,10,0{,}025}} = \frac{1}{3{,}62} = 0{,}2$$

Subject:
Date



$3,37 > 0,63 > 0,28 \Rightarrow AH_0$

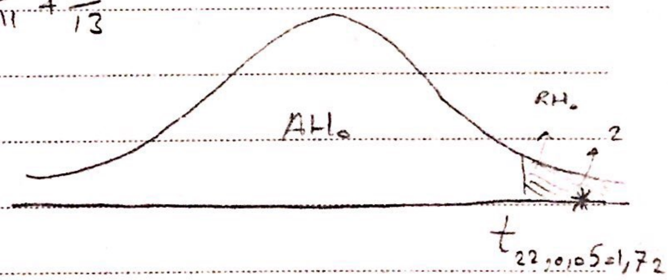
پس از زیر حالت اول استفاده می‌کنیم یعنی (جوامع نرمال، n ها کوچک،
واریانس ها مجهول ولی برابر):

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{10 \times 1,36 + 12 \times 2,17}{11+13-2} \simeq 1,8$$

$$\mu_1-\mu_2 : (\bar{X}-\bar{Y}) \pm \underbrace{t_{n_1+n_2-1,\ \frac{\alpha}{2}}}_{22,\ 0,025} \cdot S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= (10,7 , 9,6) \pm 2,074 \sqrt{1,8}\sqrt{\frac{1}{11}+\frac{1}{13}} = (-0,04 , 2,24)$$

$$T_0 = \frac{(\bar{X}-\bar{Y})-\delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(10,7-9,6)-0}{\sqrt{1,8}\sqrt{\frac{1}{11}+\frac{1}{13}}} \simeq 2$$



$2 > 1,72 \Rightarrow RH_0$

مثال ۱) با توجه به آماره های زیر فرضیه $\begin{cases} H_0 : P_1 = P_2 \\ H_1 : P_1 < P_2 \end{cases}$ و $\begin{cases} H_0 : P_1 = P_2 \\ H_1 : P_1 \neq P_2 \end{cases}$ را آزمون کنید.

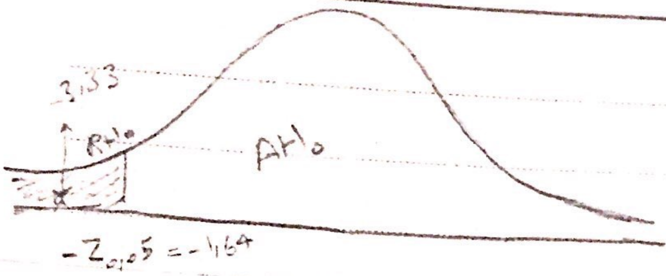
فرض اول را از روش p_value نیز آزمون کنید.

$n_1 = 100 \quad \hat{P}_1 = 0,5$

$n_2 = 200 \quad \hat{P}_2 = 0,7$

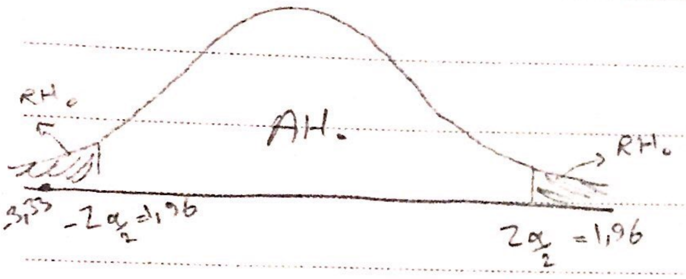
$$\hat{P} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1+n_2} = \frac{(100\times 0,5 + 200 \times 0,7)}{100+200} \simeq 0,63$$

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0,5-0,7}{\sqrt{0,63(1-0,63)\left(\frac{1}{100}+\frac{1}{200}\right)}} \simeq -3,33$$



$-3.33 < -1.64 \Rightarrow RH_0$

برای فرضیه دوم:



پس برای فرضیه $\begin{cases} H_0: P_1 = P_2 \\ H_1: P_1 \neq P_2 \end{cases}$ هم H_0 رد می‌شود

$P\text{-Value} = p(Z < -3.33) = \Phi(-3.33) = 0.0004 < 0.05 \Rightarrow RH_0$

اثرپذیری

مثال) برای تاثیرات تبلیغات ضدچاقی دانشگاه ۵ تن از کارمندان را در نظر گرفت و وزن ایشان را قبل از تبلیغات و بعد از تبلیغات اندازه گرفت، یک فاصله اطمینان ۹۵٪ برای تغییر وزن بیابید. آیا تبلیغات تاثیری روی وزن داشته است؟ چون نمونه تغییری نکرده پس دو نمونه مستقل نداریم پس از استنباط زوجی استفاده می‌کنیم.

قبل از تبلیغات	84	97	78	91	85
بعد " "	80	98	75	90	82

$D_i =$ قبل - بعد $-4 \quad 1 \quad -3 \quad -1 \quad -3 \rightarrow \bar{D} = \frac{1}{5}(-4+1-3-1-3) = -2$

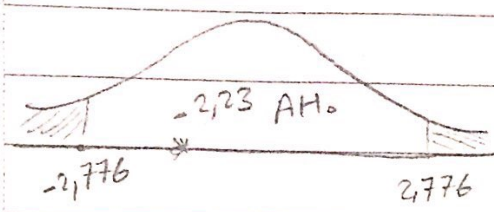
$(D_i - \bar{D})^2 = (-4+2)^2 = 4 \quad 9 \quad 1 \quad 1 \quad 1 \rightarrow S_D^2 = \frac{1}{4}(4+9+1+1+1) = 4$

$\delta : \bar{D} \pm t_{4,0.025} \frac{S_D}{\sqrt{n}} : -2 \pm 2.776 \frac{2}{\sqrt{5}} = (-4.48, 0.48)$

صفر در بازه اطمینان هست پس $\delta = 0$ پذیرفته می‌شود

$$T_0 = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}} = \frac{-2}{\frac{2}{\sqrt{5}}} = -2,23$$

$$\begin{cases} H_0: \delta = 0 \\ H_1: \delta \neq 0 \end{cases}$$



$\rightarrow -2,23 > -2,776 \Rightarrow AH_0$

مثال) باتوجه به اطلاعات زیر فرضیه

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

دو جامعه داریم واریانس ها معلوم

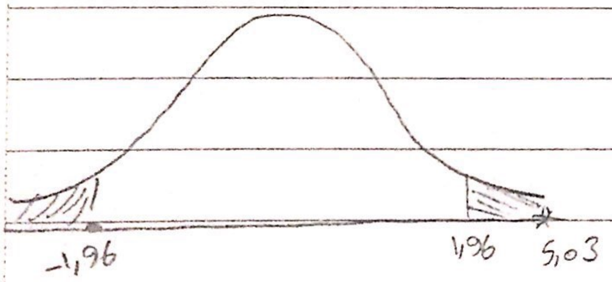
آزمون کنید.

$n_1 = 36 \quad \bar{X} = 81 \quad \sigma_1 = 5,2$

$n_2 = 49 \quad \bar{Y} = 76 \quad \sigma_2 = 3,4$

چون واریانس ها معلوم است از Z استفاده می کنیم:

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{81 - 76 - 0}{\sqrt{\frac{(5,2)^2}{36} + \frac{(3,4)^2}{49}}} = 5,023$$



پس چون $5,03 > 1,96$ فرض صفر را رد می کنیم