جزوه تحلیل آماری

خانم دکتر عاطفه فرخی

این جزوه به وسیله دانشجویان نوشته شده است لذا ممکن است دارای اشتباهاتی نیز باشد.

داده هامون پیوسته باشد ۳۰ تا بالاتر

پواسن ← راجب بدی ها - رگ و سرد - نرخ شکست و ...

آمار ← داده ها را جمع آوری و خلاصه و تجزیه و تحلیل و نتیجه

۱- خلاصه سازی (نمودار - جدول فراوانی) در قسمت پایان نامه

۲- خلاصه سازی (شاخص مرکزی - شاخص پراکندگی)

جمع آوری اطلاعات را از جامعه آماری بدست می آوریم

جامعه آماری: مجموعه ای از افراد و اشیاء که ویژگی مشخصه ای دارند.

ویژگی مشترک ← متغیر
- کیفی ← ویژگی های بالای پرسشنامه
- کمی ← تعداد سؤالات
  - گسسته
  - پیوسته: بین آنها عدد قرار می گیرند

روی جامعه کار می کنیم و روی نمونه کار می کنیم
- هم از نظر زمان ← نمونه آماری
- هم از نظر مالی ← جامعه آماری

جامعه: $\mu = \frac{1}{N}\sum f_i x_i$

نمونه: $\bar{x} = \frac{1}{n}\sum f_i x_i$

در جامعه: $P = \frac{x}{N}$

$\hat{P} = \frac{x}{n}$

جامعه: $\sigma^2 = \frac{1}{N}\sum (x_i - \bar{x})^2$

نمونه: $s^2 = \frac{1}{n-1}\sum f_i (x_i - \bar{x})^2$

جامعه:

میانگین	μ
واریانس	σ^2
نسبت	P

نقطه‌ای   ص ۱۶

براورد

فاصله‌ای   ۱۷_۱۶/۵ فاصله ← بین آن بشود عددی انتخاب کرد.

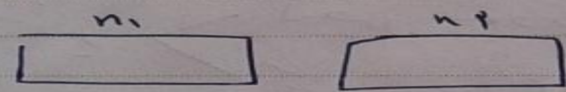
ویژگی محصول جامعه ← پارامتر   پارامتر ← p σ^2 μ

آماره ← برآورد پارامتر به کمک نمونه   آماره $\hat{p}$ s^2 $\bar{x}$

$\bar{x}$ براوردی است برای μ

میزان انتظار   Expect $E(\bar{x}) = \mu$

یعنی انتظار داریم که نمونه اونقدر نمونه خوبی باشد که میانگین نمونه با جامعه برابر باشد.



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

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

$E(\hat{p}) = p$

نمونه هرچه بیشتر بهتر ← به واقعیت نزدیکتر

براوردی بهتر است که دو ویژگی را داشته باشد:

۱- نااریب باشد (کج نباشد) دقیقاً خودش باشد. $E(\bar{x}) = \mu$

۲- میزان واریانس کمتر باشد. $var(\bar{x})$

براوردگری را انتخاب می‌کنیم که واریانسش کمتر باشد.

x_1 → اریب → var = 4 → این بهتر است

x_2 → نااریب → var = 9

$E(\bar{x}) = \text{عددی} + \mu$ اریب

$E(x) = \mu$ نا اریب

روش های نمونه گیری :

۱- تصادفی ساده : راحت ترین و بهترین برای حالتی که محدودیت نداریم و

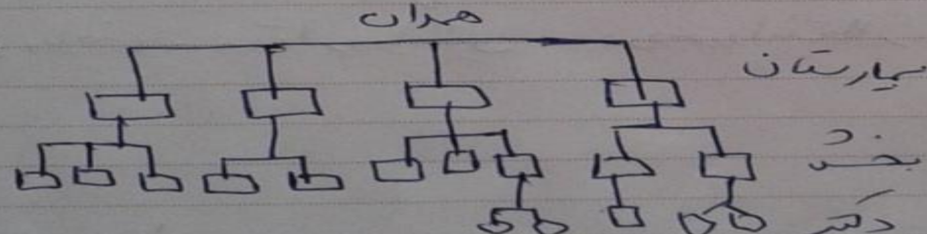
با جایگذاری ^ بدون جایگذاری ← بهترین حالت

۲- طبقه بندی : از قشرهای مختلف می خواهیم نمونه بگیریم.

۳- خوشه ای : مقالاتش خوب و سریع اکسپت می گیرد. (به صورت سلسله ای جلو می رود)

۴- سیستماتیک (منظم) : خیلی در واقعیت با آن کار نمی کنیم. نقطه شروع را خودتان باید انتخاب کنید مثلاً ۲ تا ۲ یا ۸ تا ۸ تا جلو برود

شکل خوشه ای



جامعه هدف ← نمونه ← ویژگی ها را روی آن اجرا و قضاوت می کنیم.

برآورد نقطه ای و فاصله ای

۱- برنولی ← شکست و موفقیت

۱- برنولی ← شکست و موفقیت

۲- دوجمله‌ای ← بیشتر از یک بار ← توزیع برنولی را n بار به طور مستقل انجام دهیم.

$P(X=\bar{x}) = \binom{k}{n}$ $\quad x \sim Bin(n, p)$

$P(X=\bar{x}) = \binom{n}{x} p^x q^{n-x}$ $\quad x=0,\dots,n$

Expect انتظار $\quad E(X)=np$

برنولی $\quad E(X) = p$ انتظار داشتیم که این یک بار نتیجش موفقیت باشد.

دوجمله‌ای $\quad E(X) = np$ انتظار داشتیم که ... کارهای که انجام با موفقیت

دوجمله‌ای $\quad var(X) = npq$

واریانس یعنی ریسک و خطا

$var(X) = p(1-p) = pq$

امید ریاضی - مقدار میانگین - سود مورد انتظار $\quad E(X) = \bar{x}$

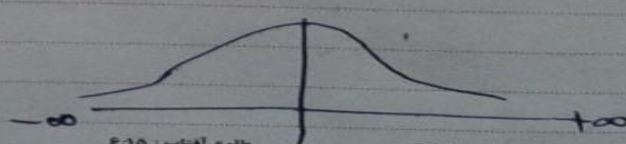
$E(X_i) = \sum x_i p_i$

$\bar{x} = \frac{1}{n}\sum f_i x_i$

توزیع نرمال:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$x \in R$ بهترین توزیع در آمار



احتمال در حالت پیوسته ← انتگرال

نرمال ← تبدیل

چرا از استاندارد استفاده می‌کنیم؟

$x \sim N$ (میانگین , واریانس)

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

$z \sim N(0,1)$ و $z = \frac{x-\mu}{\sigma}$ z نرمال استاندارد

z	٪۰۰	٪۰۱	...	٪۰۹
-۳/۵				
-۳/۴				
⋮				
۰				

z	٪۰۰	٪۰۱	...	٪۰۹
۰				
۰/۱				
۰/۲				
⋮				
۳/۵				۰/۹۹۹۹

نمونه‌گیری مرحله‌ای

مثال: اگر وزن شیشه‌های آبلیمو دارای توزیع نرمال با میانگین ۵۰۰ و واریانس ۲۵ باشد مطلوب است محاسبهٔ هر یک از موارد زیر:

الف) $P(x > 495)$

ب) $P(500 < x < 502)$

ج) $P(x < 482) = P(z < \frac{482-500}{5}) = P(z < -\frac{18}{5}) =$

الف) $P(x > 495) = 1 - P(z \le \frac{495-500}{5}) = 1 - P(z \le -1) =$

ب) $P(x < 502) - P(x < 500) = P(z < \frac{502-500}{5}) - P(z < \frac{500-500}{5}) =$

$P(\frac{2}{5}) - P(z < 0) = 0/6554 - 0/5 =$

هرچه به ۳,۵- نزدیک به صفر نزدیک و هرچه به ۳,۵+ نزدیک شویم به ۱ نزدیک می‌شویم.

۱) $P(x<a) = P(z \le \frac{a-\mu}{\sigma})$

۲) $P(x>b) = 1 - P(x<b) = 1 - P(z < \frac{b-\mu}{\sigma})$

۳) $P(a<x<b) = P(x<b) - P(z \le \frac{a-\mu}{\sigma})$

آماره → برآورد پارامتر

برآورد < نقطه‌ای / فاصله‌ای

df درجه آزادی	۰.۰۱	...	۰.۹۷۵
۱			
۲			
۳۰			

$Z\alpha$ $\begin{cases} P(z \le 2) \\ P(t < \dots)^{3} \end{cases}$

$T\alpha\ df$

df = درجه آزادی

قضیه حد مرکزی: مهمترین قضیه در آمار

هر توزیعی که باشد وقتی تعداد اعضای نمونه بیشتر از ۳۰ باشد به جای آن از توزیع نرمال استفاده می‌کنیم.

هر توزیعی که داشته وقتی $n \to \infty$ (اندازه نمونه بزرگ $n \geq ۳۰$) به سمت توزیع نرمال میل می‌کند. (در کتاب لاشین $n \geq ۲۵$ است).

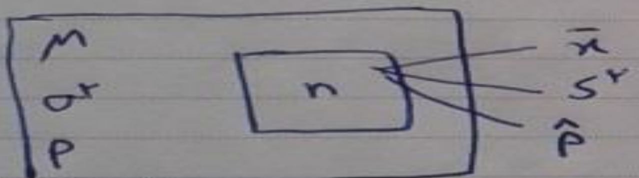
فاصله اطمینان ← یک فاصله را مشخص که چقدر به فاصله، اطمینان داریم. یا خطا را می‌دهد یا میزان اطمینان

در نرم افزار spss مقدار خطا را ۰/۰۵ می‌گذاریم.

همیشه ۰/۰۵ = خطا   ۱ > خطا > ۰

هرچه خطا کمتر باشد بهتر است یا می‌گویند خطا یا سطح معنی داری.

فاصله اطمینان برای میانگین جامعه



فاصله اطمینان برای میانگین جامعه
- واریانس جامعه معلوم ← در کتاب‌ها فقط هست در واقعیت نیست.
- واریانس جامعه مجهول

مراحل فاصله اطمینان

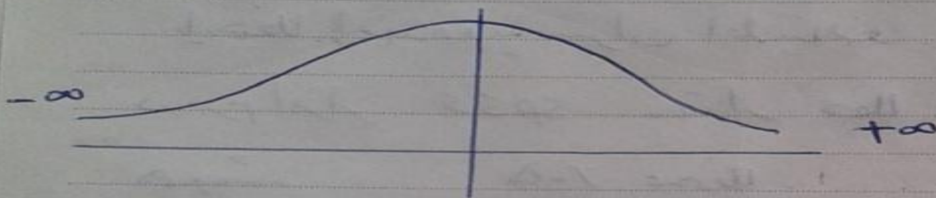
① واریانس جامعه معلوم ← نرمال z $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

واریانس جامعه مجهول ← T Student ← t ← تست مجهوری

کران بالا > ∞ > فاصله اطمینان یک طرفه

∞ > > کران پایین

کران بالا > > کران پایین فاصله اطمینان دو طرفه



① $z = \dfrac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$

② $\alpha \,,\, z_{\alpha/2}$

③ $-z_{\alpha/2} < z < z_{\alpha/2}$

$$-z_{\alpha/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2} \rightarrow \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\mu \in \left(\bar{x} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

مثال: با توجه به اینکه واریانس هوش کودکان برابر ۴ واحد است به کمک نمونه‌ای به اندازه ۱۶ متوسط هوش آنها برابر ۱۰۰ واحد گزارش شده است در سطح معنی داری ۰۵٪ یک فاصله اطمینان مناسب برای متوسط هوش جامعه کودکان بدست آوریم

اول واریانس جامعه   بعد واریانس نمونه ← پس واریانس به نمونه ربط ندارد

پس این واریانس می‌شود واریانس جامعه

$\alpha = \%05 \rightarrow Z_{\alpha/2} = Z_{\%.025} = -1.96$

$Z(1-\alpha/2) = Z_{1-\%.025} = Z_{\%.975} = 1.96$

$$\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}$$

$$100 - \frac{2}{4}\, 1.96 < \mu < 100 + \frac{2}{4}\, 1.96$$

فاصله اطمینان برای میانگین یک جامعه با واریانس مجهول

گام ① $T = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

درجه آزادی

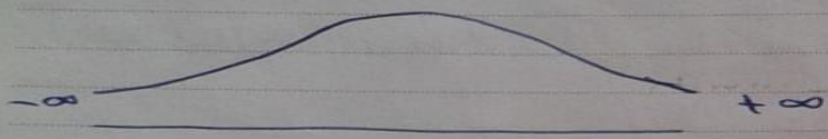
② $\alpha ,\ t(\alpha/2 ,\ df)$

③

$$\sigma^2 = \frac{1}{N} \sum f_i (x_i - \mu)^2$$

$$S = \frac{1}{n-1} \sum f_i (x_i - \bar{x})^2$$

به این خاطر $\frac{1}{n-1}$ می‌گذارد که از حالت اریب خارج می‌شود.



درجه آزادی به خاطر اینکه از n به $n-1$ است.

(۳) $-t(\alpha/2, n-1) < T < t(\alpha/2, n-1) \rightarrow$

$$-t(\alpha/2, n-1) < \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} < t(\alpha/2, n-1)$$

$$\bar{x} - t(\alpha/2, n-1)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t(\alpha/2, n-1)\frac{s}{\sqrt{n}}$$

باتوجه به یک نمونه ۶۴ تایی میانگین و واریانس وزن قوطی های رب به ترتیب ۴۸۰ و ۹ بدست آمده است. در سطح معنی داری ۰۵٪ برای متوسط وزن قوطی‌ها یک فاصله اطمینان بدست آورید؟

چون نمونه زیاد است به جای t از z استفاده می‌کنیم.

داده های زیر معدل ۱۰ نفر از دانشجویان رشته کامپیوتر را نشان می دهد در سطح معنی داری ۰۵٪ برای معدل کل دانشجویان یک فاصله اطمینان بدست آورید

۱۶ ، ۱۸ و ۱۶٫۵ و ۱۷ ، ۱۳ و ۱۵ و ۱۴ و ۱۸ و ۱۶ و ۱۴

$t(\alpha/2, n-1) = T(\%0.025, 9) = 2.262$

$\bar{x} = 15.75 \rightarrow \bar{x} = \frac{157.5}{10} = 15.75$ $\quad n-1 = 10-1 = 9$

$S^2 = \frac{1}{n-1}\sum f_i(x_i-\bar{x})^2 = \frac{1}{9}\left((14-15.75)^2 + \dots + (16-15.75)^2\right) =$

$= \frac{1}{9}(22.5) = 2.411 \rightarrow S^2 = 2.411 \rightarrow S = \sqrt{2.411} = 1.615$

$\bar{x} - T(\alpha/2, n-1)\frac{S}{\sqrt{n}} < \mu < \bar{x} + t(\alpha/2, n-1)\frac{S}{\sqrt{n}}$

$3.825 < \mu < 6.136$

می توان میانگین آن ۴ هم باشد در فاصله اطمینان بررسی می کنم که در فاصله قرار می گیرد یا نه

$15.75 - 2.262 \cdot \frac{1.615}{3.162} < \mu < 15.75 + 2.262 \cdot \frac{1.615}{3.162}$

خطا که زیاد باشد کمیت محوری چه تغییری می کند ← کمتر می شود

اندازه نمونه زیاد شود کمیت محوری چگونه می شود ← بیشتر می شود

$T = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{x}-\mu)}{S}$

هرچقدر n بیشتر به جامعه نزدیک تر و T بیشتر شود

نسبت

فاصله اطمینان برای نسبت

① کمیت محوری $z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ مجهول

$p = \frac{\text{تعداد حالات مطلوب}}{\text{تعداد کل حالات}}$

برای نسبت همیشه از z استفاده می‌کنیم قبلاً می‌گفتیم که اگر واریانس معلوم از z و مجهول از t

② $z_{\alpha/2}$

③ $-z_{\alpha/2} < z < +z_{\alpha/2} \rightarrow -z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < z_{\alpha/2}$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

مثال: از بین ۱۰ تیر پرتاب شده توسط یک تیرانداز شش تیر به هدف اصابت می‌کند در سطح معنی داری ۰٫۵٪ برای نسبت کل تیرهایی که به هدف اصابت می‌کنند یک فاصله اطمینان مناسب بدست آورید.

$n = 10$, $x = 6 \rightarrow \hat{p} = .6$

$\alpha = .05 \rightarrow \alpha/2 = .025$

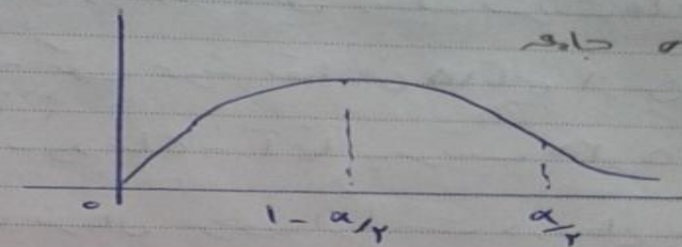
$z_{\alpha/2} = z_{.025} = 1.96$

$$.6 - 1.96\sqrt{\frac{.6(1-.6)}{10}} < p < .6 + 1.96\sqrt{\frac{.6(1-.6)}{10}}$$

برخلاف توان دوم که ... نمی تواند ...
مثل ریشه دوم که نمی تواند منفی باشد

فاصله اطمینان برای واریانس یک جامعه:

(۱) $x^2 = \frac{(n-1)s^2}{\sigma^2}$



(۲) $x^2(\alpha/2, n-1)$ و $x^2(1-\alpha/2, n-1)$

واریانس مجهول باید یک درجه آزادی داشته باشد

(۳) $x^2(1-\alpha/2, n-1) < x^2 < x^2(\alpha/2, n-1)$

$$x^2(1-\alpha/2, n-1) < \frac{(n-1)s^2}{\sigma^2} < x^2(\alpha/2, n-1)$$

$$\frac{(n-1)s^2}{x^2(\alpha/2, n-1)} < \sigma^2 < \frac{(n-1)s^2}{x^2(1-\alpha/2, n-1)}$$

df درجه آزادی	٪.۲۵	٪.۵	٪۱	٪.۹۹۷ → $\alpha/2$
۱				
۲				
۳				
۱۲۰				

$\alpha/2$ را بدست می آوریم چون نصف آن می افتد کران پایین و نصف آن می افتد در کران بالا در قسمت طرف.

مثال: یکی از کارگزاران بازار بورس تهران در مورد تعیین ریسک شرکت هایی است که سهام خود را عرضه می‌کنند در این زمینه یک نمونه تصادفی ۱۵ تایی از بین شرکت‌ها انتخاب کرده که میانگین و واریانس سود سالانه آنها به ترتیب ۲۵ و ۱۲۲۵ است توزیع سود سالانه شرکت ها از تقریب نرمال برخوردار است در سطح معنی داری ۰.۵٪ حدود اطمینان ریسک شرکت ها را بدست آورید؟

$n = 15 \qquad n-1 = 14$

$\bar{x} = 25 \qquad \sigma^2 = 1225 \rightarrow \sigma = 35 \qquad \alpha = 0.05$

$\alpha/2 \quad \chi^2 (0.025 , 14) = 26.1190 \qquad \alpha/2 = 0.025$

$\chi^2 (0.975 , 14) = 5.62872 \qquad (1 - \alpha/2) = 0.975$

$1 - \alpha/2$

$$\frac{(15-1)1225}{26.1190} < \sigma^2 < \frac{(15-1)1225}{5.62872}$$

$$\sqrt{\frac{14 \times 1225}{26.1190}} < \sigma < \sqrt{\frac{14 \times 1225}{5.62872}}$$

$$\sigma^2 = \frac{1}{N} \sum f_i (x_i - \mu)^2$$

نااریب با کمترین واریانس

$$S^2 = \frac{1}{n-1} \sum f_i (x_i - \bar{x})^2 \qquad \frac{(n-1)S^2}{\chi^2} < \quad < \frac{(n-1)S^2}{\chi^2}$$

$$S^2 = \frac{1}{n} (f_i (x_i - \mu)^2) \longrightarrow \frac{nS^2}{\chi^2} < \quad < \frac{nS^2}{\chi^2}$$

میانگین دو جامعه

فاصله اطمینان برای تفاضل دو جامعه که آیا واریانس آنها معلوم است یا مجهول.

فاصله اطمینان برای تفاضل میانگین دو جامعه

۱- واریانس معلوم باشد:

① $$Z = \frac{\overbrace{(\mu_1 - \mu_2)}^{\text{مجهول}} - (\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

② $Z_{\alpha/2}$

③ $$-Z_{\alpha/2} < \frac{(\mu_1 - \mu_2) - (\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < Z_{\alpha/2}$$

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

مثال: هدف از تحقیق مقایسه عملکرد کارمندان در ۲ سازمان است و ملاک مقایسه میانگین دو جامعه آماری است البته به دلیل در دسترس نبودن عملکرد کلیه کارمندان محققان ناچارند به نمونه هایی از هر جامعه اکتفا کنند اطلاعات بدست آمده از نمونه ها عبارت است از جامعه اول، جامعه دوم:

جامعه اول: $n_1 = 25$ و $\bar{x}_1 = 60$ و $\sigma_1^2 = 100$

جامعه دوم: $n_2 = 20$ و $\bar{x}_2 = 55$ و $\sigma_2^2 = 144$

و هر یک از نمونه‌ها از جامعه‌ای نرمال بدست آمده‌اند در سطح معنی‌داری ۰٫۰۱٪ برای تفاضل متوسط عملکرد دو جامعه یک فاصله اطمینان بدست آورده و آن را تفسیر نمایید.

$\alpha = \%1$, $Z_{\%0.5} = 2.58$

$$(60-55) - 2.58\sqrt{\frac{100}{25}+\frac{144}{25}} < \mu_1-\mu_2 < (60-55)+2.58\sqrt{\frac{100}{25}+\frac{144}{25}}$$

$$-10 < \mu_1-\mu_2 < 5 \rightarrow \mu_1-\mu_2 = 0 \rightarrow \mu_1 = \mu_2$$

نکته:

(1) $- < \mu_1-\mu_2 < +$ $\qquad \mu_1 = \mu_2$

(2) $+ < \mu_1-\mu_2 < +$ $\qquad \mu_1 > \mu_2$

(3) $- < \mu_1-\mu_2 < -$ $\qquad \mu_1 < \mu_2$

فاصله اطمینان وقتی واریانس دو تا جامعه معلوم باشد چنین چیزی اتفاق می‌افتد.

تفسیر فاصله اطمینان: هرگاه کران پایین در فاصله اطمینان عددی منفی و کران بالا عددی مثبت باشد به دلیل اینکه این فاصله عدد صفر را دربر می‌گیرد می‌توان

گفت میانگین دو جامعه در سطح خطای α برابر است

۲- هرگاه کران پایین وکران بالا در یک فاصله اطمینان اعدادی مثبت باشند می‌توان گفت میانگین جامعه اول همواره از میانگین جامعه دوم بزرگتر است.

۳- هرگاه کران پایین وکران بالا هردو اعدادی منفی باشند براین معناست که میانگین جامعه اول در سطح معنی داری α کوچکتر از میانگین جامعه دوم است.

فاصله اطمینان برای تفاضل دوجامعه واریانس مجهول باشد:

$$① \quad T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2 p}{n_1} + \frac{s^2 p}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$② \quad t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)$ (درجه آزادی) ~~مجهول~~

$$\underset{\text{آمیخته}}{s^2 p} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$df: (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

$$③ \quad -t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right) < T < t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)$$

$$-t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right) < \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)$$

$$(\bar{x}_1-\bar{x}_2) - t(\alpha/2, n_1+n_2-2)\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t(\alpha/2, n_1+n_2-2)\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

سؤال: هدف تحقیق مقایسه هزینه سوخت دو منطقه تهران با اطمینان ۹۵ درصد در فصل زمستان است توزیع نمونه سوخت در هر دو منطقه مورد نظر نرمال است در این تحقیق از منطقه اول ۵ خانوار و از منطقه دوم ۸ خانوار به طور تصادفی انتخاب شده و هزینه سوخت آنها به صورت زیر است؟ با فرض تساوی واریانس دو منطقه یک فاصله اطمینان مناسب برای تفاضل میانگین را بدست آورده و آن را تفسیر نمایید.

اول	۱۰۰۰	۱۱۰	۹۰۰	۸۱۰	۸۵۰			
دوم	۵۵۰	۴۵۰	۱۰۰۰	۸۵۰	۸۶۰	۳۵۰	۴۰۰	۴۵۰

واریانس های مساوی: میزان تغییرات دو منطقه یکسان

واریانس یکسان $\sigma_1^2 = \sigma_2^2$

$n_1 = 5 \quad \bar{x}_1 = \frac{1}{n_1}\sum x_i = \frac{1000+\cdots}{5} =$

$n_2 = 8 \quad \bar{x}_2 = \frac{1}{n_2}\sum x_i = \frac{550+\cdots}{8} =$

$S_1^2 = \frac{1}{n-1}\sum (x_{i1}-\bar{x}_1)^2 =$

$S_2^2 = \frac{1}{n_2-1}\sum (x_{i2}-\bar{x}_2)^2 =$

$\alpha = 5\% \rightarrow \alpha/2 = 2.5\% \quad T(\alpha/2 ,\, n_1+n_2-2) = T(2.5\% ,\, 11) \quad 5+8-2=11$

$$S_p^2 = \frac{(5-1)S_1^2+(8-1)S_2^2}{5+8-2} =$$

$$S_p = \sqrt{S_p^2}$$

$$(\bar{x}_1-\bar{x}_2) - T(\alpha/2 ,\, 11)\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t(\alpha/2 ,\, 11)\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$- < \mu_1-\mu_2 < + \longrightarrow \mu_1=\mu_2$

$- < \mu_1-\mu_2 < - \longrightarrow \mu_1<\mu_2$

$+ < \mu_1-\mu_2 < + \longrightarrow \mu_1>\mu_2$

واریانس دو جامعه چون کاربرد ندارد نمی‌نویسیم آن را چون در پایان نامه و واقعیت با آن سر و کار نداریم.

هر مطلبی که خوندیم راجبش مقاله بخوانیم و توضیح بدهیم.

فاصله اطمینان برای تفاضل نسبت دو جامعه:

مجهول

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

(۲) $Z_{\alpha/2}$ یا $Z(1-\alpha/2)$

$$-Z_{\alpha/2} < Z < Z_{\alpha/2}$$

$$-Z_{\alpha/2} < \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} < Z_{\alpha/2}$$

$$(\hat{p}_1-\hat{p}_2)-Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < p_1-p_2 < (\hat{p}_1-\hat{p}_2)+Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$- < p_1-p_2 < + \rightarrow p_1=p_2$

$- < p_1-p_2 < - \rightarrow p_1<p_2$

$+ < p_1-p_2 < + \rightarrow p_1>p_2$

$\hat{p}_1 = \frac{۲۰۰}{۲۰۰۰} = ۰/۱$ $\qquad \hat{p}_2 = \frac{۱۵۰}{۱۰۰۰} = ۰/۱۵$ $\qquad z_{۰/۰۲۵} = 1{,}96$

$(۰/۱ - ۰/۱۵) - 1{,}96\sqrt{\frac{۰/۱(۰/۹)}{۲۰۰۰} + \frac{۰/۱۵ \times ۰/۸۵}{۱۰۰۰}} < P_1 - P_2 < (۰/۱ - ۰/۱۵) +$

$1{,}96 \times \sqrt{\frac{۰/۱ \times ۰/۹}{۲۰۰۰} + \frac{۰/۱۵ \times ۰/۸۵}{۱۰۰۰}}$

آزمون فرض (فرضیه):

آزمون ادعا

فرضیه

$$\begin{cases} H_0 \rightarrow & \text{فرض صفر یا اولیه} \\ H_1 \rightarrow & \text{فرض مقابل یا فرض یک} \end{cases}$$

فرض صفر ← فرض مسئله

فرض یک ← فرض خودما

$$\begin{cases} H_0 = \\ H_1 \neq \end{cases} \qquad \begin{cases} H_0 \geq \\ H_1 < \end{cases} \qquad \begin{cases} H_0 \leq \\ H_1 > \end{cases}$$

آزمون فرض میانگین یک جامعه

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

واریانس معلوم ($\mu < a$ یا $\mu > a$)

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

آزمون مرکب :

یک عدد بدست می‌آوریم

① آماره آزمون $Z = \dfrac{\bar{x} - a}{\frac{\sigma}{\sqrt{n}}}$

$Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ → مجهول

چون در قبلی یک مجهول بود می‌گوییم کمیت محوری

تابعی که در آن یک مجهول وجود دارد کمیت محوری

کمیت محوری : تابعی است که در آن پارامتر مجهول وجود دارد

آماره آزمون : تابعی است که در آن هیچ پارامتر مجهولی وجود ندارد.

آزمون مرکب ($\alpha/2$) : آزمونی که فرض مقابل نامساوی باشد ($\neq$)

آزمون ساده (α) : آزمونی که فرض مقابل جهت داشته باشد. ($>$ و $<$)

② $Z_{\alpha/2}$

③ مرحله تصمیم گیری $|Z| > Z_{\alpha/2}$

فرض صفر را رد می‌کنیم.

مثال : ادعا شده است که متوسط طول عمر ۹ لامپ برابر با ۱۴ روز می‌باشد. اگر بدانیم میزان تغییرات کل لامپ ها برابر ۴ است در سطح معنی داری

از صحت این ادعا را بیازمایید. میانگین نمونه برابر ۱۲ بوسته آمده است.

$$\begin{cases} H_0: \mu = 14 \\ H_1: \mu \neq 14 \end{cases}$$

① $z = \dfrac{\bar{x} - a}{\frac{\sigma}{\sqrt{n}}} = \dfrac{12 - 14}{2/3} = -3$

② $z_{\frac{0.1}{2}} = z_{0.05} = 1.645$

[illegible]

③ $|z| > z_{\alpha/2} \to |-3| = 3 > 1.645$

فرض صفر را رد می‌کنیم.

آزمون ساده:

$$\begin{cases} H_0: \mu \geq a \\ H_1: \mu < a \end{cases}$$

① $z = \dfrac{\bar{x} - a}{\frac{\sigma}{\sqrt{n}}}$

② z_α

③ $z < -z_\alpha$

فرض صفر را رد می‌کنیم

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

① $z = \dfrac{\bar{x} - a}{\frac{\sigma}{\sqrt{n}}}$

② z_α

③ $z > z_\alpha$

فرض صفر را رد می‌کنیم

$$|z| > z_{\alpha/2} \to \begin{cases} z > z_{\alpha/2} \\ z < -z_{\alpha/2} \end{cases}$$

مثال: مدیر گروهی ادعا می‌کند معدل دانشجویان گروهش بیشتر از ۱۷ است
اگر بدانیم میزان تغییر پذیری معدل دانشجویان برابر ۱ واحد است و به کمک
نمونه ۲۵ تایی معدل دانشجویان برابر ۱۶ بدست آورده باشد صحت ادعای
این مدیر گروه را بیازمائید.

$\alpha = \%5$

$\bar{\mu} > 17$

$n = 25$

$\mu = 16$

$$\begin{cases} H_0 : \mu \geq 17 \\ H_1 : \mu < 17 \end{cases}$$

① $z = \dfrac{16 - 17}{\frac{1}{5}} = -5$

② $z_{\%5} = 1{,}645$

③ $-5 < -1{,}645 \rightarrow$ فرض صفر را رد می‌کنیم

آزمون فرض میانگین: (واریانس مجهول)

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

① $T = \dfrac{\bar{x} - a}{\frac{s}{\sqrt{n}}}$

② $t(\alpha/2, n-1)$

③ $|T| > t(\alpha/2, n-1)$

فرض صفر را رد می‌کنیم

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

① $T =$

② $T(\alpha, n-1)$

③ $t > t(\alpha, n-1)$

فرض صفر رد می‌شود

$$\begin{cases} H_0 : \mu \geq a \\ H_1 : \mu < a \end{cases}$$

① $T =$

② $t(\alpha, n-1)$

③ $t < -t(\alpha, n-1)$

فرض صفر رد می‌شود.

مثال: ادعا شده است که میانگین برق مصرفی فروردین ماه یک ناحیه بزرگ تهران دست کم ۱۳۰۰ کیلووات ساعت بوده است. برای این منظور یک نمونه تصادفی به اندازه ۴۰۰ خانوار از منطقه انتخاب شود که میانگین وانحراف معیار برق مصرفی آنها به ترتیب ۱۲۵۲ و ۲۵۷ کیلووات ساعت است. سطح خطای ۱۰٪ را در نظر گرفته و صحت ادعا را بررسی کنید.

$$\begin{cases} H_0: \mu \geq ۱۳۰۰ \\ H_1: \mu < ۱۳۰۰ \end{cases}$$

اندازه نمونه بیشتر از ۲۵ یا مساوی به جای T، Z را در نظر بگیریم.

① $$z = \frac{\bar{x} - a}{\frac{s}{\sqrt{n}}} = \frac{۱۲۵۲ - ۱۳۰۰}{\frac{۲۵۷}{۲۰}} = \frac{-۴۸ \times ۲۰}{۲۵۷} = ۳,۷۳۵$$

② $$z_a = z_{\%.۱} = ۲,۳۵$$

③ $z < -z_a$ $\quad -۳,۷۳۵ < -۲,۳۵$ فرض صفر رد می شود.

تمرین: ادعا شده است که میانگین زمان انتظار هواپیماهای مسافربری فرودگاه مهرآباد برای فرود روی باند ۳ دقیقه است در این زمینه ۷ پرواز به طور تصادفی انتخاب شود که انتظار آن ها برای فرود به این ترتیب است توزیع

زمان انتظار برای فرود روی باند نرمال است. ادعای فوق را در سطح خطای ۱۰٪ محاسبه کنید.

$$\begin{cases} H_0: \mu = ۳ \\ H_1: \mu \neq ۳ \end{cases}$$

$$\bar{x} = \frac{۱}{۷}(۱,۵ + \cdots) = ۳,۲۱$$

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

$$s^2 = \frac{۱}{۷-۱}\left[(۱,۵ - ۳,۲۱)^2 + \cdots + (۴,۵ - \ )^2\right] =$$

$$t\left(\frac{٪۰۱}{۲}, ۷-۱\right)$$

$|t| > t(٪۰۰۵, ۶)$ فرض صفر را رد می‌کنیم.

آزمون فرض برای نسبت در یک جامعه

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

① $z = \dfrac{\hat{p} - a}{\sqrt{\dfrac{a(1-a)}{n}}} \approx \dfrac{\hat{p} - a}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}}$

② $z_{\alpha/۲}$

③ $|z| > z_{\alpha/۲}$

فرض صفر رد می‌شود

① z

② z_α

③ $z > z_\alpha$

فرض صفر را رد

① z

② z_α

③ $z < -z_\alpha$

فرض صفر را رد می‌کنیم.

تیراندازی ادعا می‌کند ۶/٪ از تیرهایش به هدف اصابت می‌کند اگر از بین ده تیر پرتاب شده توسط این تیرانداز پنج تیر به هدف اصابت کند صحت ادعای این تیرانداز را در سطح معنی داری ۰۵/٪ بیازمایید؟

$$\begin{cases} H_0: p = 0.6 \\ H_1: p \neq 0.6 \end{cases} \qquad \hat{p} = \frac{x}{n} = \frac{5}{10} = 0.5$$

① $$z = \frac{0.5 - 0.6}{\sqrt{\frac{0.6(1-0.6)}{10}}} = 0.1$$

② $$z_{\alpha/2} = z_{\frac{0.05}{2}} = 1.96$$

③ $|z| > 1.96$ فرض صفر را رد می‌کنیم

و اگر این رابطه برقرار نشود فرض صفر را می‌پذیریم و می‌گوییم ادعاشون ادعای درستی بوده است.

$|-0.51| = 0.5 > 1.96 \rightarrow$ صفر را می‌پذیریم

آزمون فرض برای واریانس یک جامعه:

آزمون فرض برای واریانس یک جامعه:

$$\begin{cases} H_0: \sigma^2 \geq a \\ H_1: \sigma^2 < a \end{cases}$$

① χ^2:

② $\chi^2(1-\alpha, n-1)$

③ $\chi^2 < \chi^2(1-\alpha, n-1)$

فرض صفر رد می‌شود

$$\begin{cases} H_0: \sigma^2 \leq a \\ H_1: \sigma^2 > a \end{cases}$$

① $\chi^2 =$

② $\chi^2(\alpha, n-1)$

③ $\chi^2 > \chi^2(\alpha, n-1)$

فرض صفر رد

$$\begin{cases} H_0: \sigma^2 = a \\ H_1: \sigma \neq a \end{cases}$$

① $\chi^2 = \frac{(n-1)S^2}{a}$

② $\chi^2(\alpha/2, n-1)$ و $\chi^2(1-\alpha/2, n-1)$

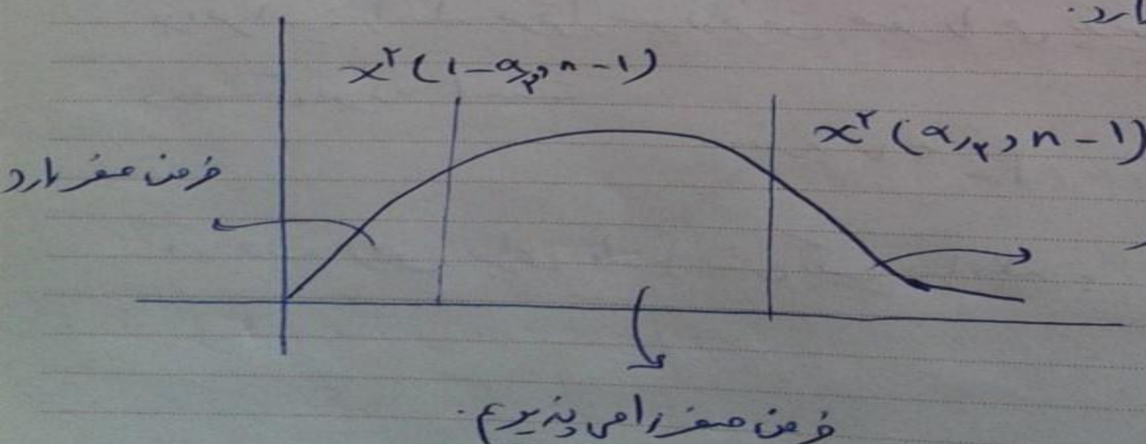
③ $\chi^2 < \chi^2(1-\alpha/2, n-1)$

یا فرض صفر را رد می‌کنیم

$\chi^2 > \chi^2(\alpha/2, n-1)$

فرض صفر رد می‌شود.

همیشه مثبت قدر مطلق معنی ندارد.



مدیر عامل بازار بورس تهران ادعا کرده است که ریسک بازده (انحراف معیار) سهام شرکت های عرضه کننده در بازار بورس کمتر از ۵ تومان است بدین منظور یکی از کارگزاران ۲۵ شرکت را به طور تصادفی از بین شرکت های عرضه کننده سهام در بازار بورس انتخاب کرده که میانگین بازده آن ها ۱۴ و انحراف معیارشان ۴ تومان است اگر بازده شرکت های بازار بورس از توزیع نرمال برخوردار باشد صحت این ادعا را در سطح ۰۵٪ بیازمائید.

$$\begin{cases} H_0 : \sigma^2 \leq 25 \\ H_1 : \sigma^2 > 25 \end{cases}$$

① $\chi^2 = \frac{(n-1)s^2}{a} = \frac{24 \times 16}{25} = 15,36$

② $\chi^2(\%0.05, 24) = 36,14151$

③ $15,36 < 36,14151$ فرض صفر را می پذیریم.

آزمون فرض برای میانگین ۲ جامعه:

۱- واریانس معلوم باشد.

واریانس معلوم

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$$

① Z

② Z_α

③ $Z < -Z_\alpha$

فرض صفر را رد می کنم.

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

① $Z =$

② Z_α

③ $Z > Z_\alpha$

فرض صفر رد می شود

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

① $Z = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

② $Z_{\alpha/2}$

③ $|Z| > Z_{\alpha/2}$

فرض صفر رد می شود

واریانس مجهول:

① T

② $t(\alpha, df)$

③ $t < -t(\alpha, df)$

فرض صفر را رد می کنم

① $T =$

② $T(\alpha, n_1 + n_2 - 2)$

③ $T > t(\alpha, df)$

فرض صفر رد می شود

① $T = \dfrac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

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

② $t(\alpha/2, n_1 + n_2 - 2)$

③ $|T| > T(\alpha/2, n_1 + n_2 - 2)$

فرض صفر رد می شود

مثال : اطلاعات زیر مربوط به دو شرکت الف و ب می باشد با فرض نرمال بودن توزیع x در دو شرکت تساوی میانگین دو جامعه با فرضیه تساوی واریانس در دو شرکت را بیازمائید :

$n_2 = 15 \qquad n_1 = 10$

$\bar{x}_2 = 120 \qquad \bar{x}_1 = 100$

$S_2 = 10 \qquad S_1 = 8$

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

(1) $$T = \frac{\bar{x}_1 - \bar{x}_2}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{100 - 120}{9.2\sqrt{\frac{1}{10} + \frac{1}{15}}} = \frac{-20}{3.68} = -5.43$$

$$s_p^2 = \frac{(n_1 - 1)\,s_1^2 + (n_2 - 1)\,s_2^2}{n_1 + n_2 - 2} = \frac{(10-1)\,64 + 14 \times 100}{10 + 15 - 2} = 85.91$$

(2) $t(\alpha/2, n_1 + n_2 - 2) = t(0.025, 23)$

④ $|T| > t(\alpha/2, n_1+n_2-2)$

$|-5.43| = 5.4 > t(\%0.025, 23) \Rightarrow 5.4 > 2.69$

فرض صفر رد می شود.

آزمون زوجی :

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \qquad T = \frac{\bar{x} - a}{\frac{s}{\sqrt{n}}}$$

هر نفر که می رود آزمایش انتظار ا جامعه است

	1	2	...	n
قبل				
بعد				

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

$x_d =$ بعد − قبل

نفر	①	②
x_d:	4	5

① $T = \dfrac{\bar{x}_d - a}{\frac{s_d}{\sqrt{n}}}$

② $t(\alpha/2, n-1)$

③ $|T| > t(\alpha/2, n-1)$ فرض صفر رد می شود

$$\begin{cases} H_0: \mu_1 - \mu_2 \geq a \\ H_1: \mu_1 - \mu_2 < a \end{cases}$$

① T

② $t(\alpha, n-1)$

③ $T < -t(\alpha, n-1)$

فرض صفر رد می‌شود.

$$\begin{cases} H_0: \mu_1 - \mu_2 \leq a \\ H_1: \mu_1 - \mu_2 > a \end{cases}$$

① T

② $t(\alpha, n-1)$

③ $T > t(\alpha, n-1)$

فرض صفر رد می‌شود

آزمون زوجی عبارت است از آزمونی که به مقایسه دو جامعه می‌پردازد با توجه به اینکه جامعه اول و جامعه دوم عوض ثابتی مورد بررسی قرار می‌گیرد به عبارت دیگر هر شخص در این نوع آزمون دارای یک زوج داده می‌باشد.

سؤال: آزمایشی درباره زمان تشخیص محصول برحسب استقرار آگهی تبلیغات روی ۱۶ نفر به طور تصادفی انجام گرفته و نتایج حاصل برحسب ثانیه در این جدول آمده است. زمان تشخیص از توزیع نرمال برخوردار است آیا می‌توان گفت بین تشخیص برحسب نوع استقرار دارای اختلاف معنی داری می‌باشد.

استقرار نوع اول	۱	۳	۲	۱	۲	۱	۳	۲
استقرار نوع دوم	۴	۲	۳	۳	۱	۲	۳	۳

$x_d = x_2 - x_1 = \quad 3 \quad -1 \quad 1 \quad 2 \quad -1 \quad 1 \quad 0 \quad 1$

$$\bar{x}_d = \frac{\sum x_d}{n} = \frac{6}{8}$$

$$S_d^2 = \frac{\sum (x_d - \bar{x}_d)^2}{n-1}$$

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \qquad \mu_d = 0 \\ H_1 : \mu_d \neq 0 \end{cases}$$

① $T = \dfrac{\bar{x}_d}{\frac{S_d}{\sqrt{n}}}$

② $t(\alpha/2 , n-1)$

③ $|T| > t(\alpha/2 , 7)$ فرض صفر رد می‌شود

آزمون استقلال :

در انجام آزمون استقلال ابتدا جدول توافقی را رسم می نماییم که در این جدول حداقل یکی از متغیرها از نوع کیفی می باشد و حالت کلی جدول عبارت است از :

جدول تقاطعی - توافقی

① / ②	متغیر ①			
متغیر ②	o_{11}	o_{12}	$\cdots$	o_{1n}
	o_{21}	o_{22}	$\cdots$	o_{2n}
	$\vdots$			
	o_{m1}	o_{m2}	$\cdots$	o_{mn}

مشاهدات : observation

انتظار : expext

H_0 : دو متغیر ① و ② مستقلند

H_1 : دو متغیر ① و ② مستقل نیستند

① $x^2 = \dfrac{\sum_i \sum_j (o_{ij} - e_{ij})}{e_{ij}}$ $\qquad e_{ij} = \dfrac{n_{i.}\, n_{.j}}{n}$

② $x^2(\alpha,$ درجه آزادی$)$ ← df

(تعداد ستون - ۱) (تعداد سطر - ۱) : درجه آزادی

مرحله تصمیم گیری

③ $x^2 > x^2(\alpha, df)$ فرض صفر رد می شود

جدول تقابلی :

تعداد فرزند

متغیر ۲ \ متغیر ۱	۰-۲	۳-۴	۵ و بالاتر
زیر دیپلم	۴ o_{11} e_{11}	۳ o_{12} e_{12}	۰ o_{13} e_{13}
دیپلم و کارشناسی	۶	۱	۴
ارشد و بالاتر	۸	۲	۵ o_{33} e_{33}

سطح سواد

سؤال : با توجه به داده های زیر بررسی کنید که آیا بین مصرف سیگار و سرطان ریه در سطح معنی داری ۰.۰۵ رابطه معنی داری وجود دارد یا خیر؟

سرطان ریه \ مصرف سیگار	بله	خیر
بله	۴	۱
خیر	۲	۲

$$\begin{cases} H_0: & \text{دو متغیر مستقل باشند} \\ H_1: & \text{دو متغیر مستقل نباشند} \end{cases}$$

① $$\chi^2 : \frac{(o_{11}-e_{11})^2}{e_{11}} + \frac{(o_{12}-e_{12})^2}{e_{12}} + \frac{(o_{21}-e_{21})^2}{e_{21}} + \frac{(o_{22}-e_{22})^2}{e_{22}} =$$

$$= \frac{(4-3.3)^2}{3.3} + \frac{(1-1.6)^2}{1.6} + \frac{(2-2.6)^2}{2.6} + \frac{(2-1.3)^2}{1.3} = 0.917$$

$$e_{ij} = \frac{n_{i.} \times n_{.j}}{n}$$

$n_{i.} =$ جمع سطر i ام

$n_{.j} =$ جمع ستون j ام

$$e_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{۵ \times ۶}{۹} = ۳,۳$$

$$e_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{۵ \times ۳}{۹} = ۱,۶$$

$$e_{21} = \frac{n_{2.} \times n_{.1}}{n} = \frac{۴ \times ۶}{۹} = ۲,۶$$

$$e_{22} = \frac{n_{2.} \times n_{.2}}{n} = \frac{۴ \times ۳}{۹} = ۱,۳$$

۲) $\chi^2(\alpha = \%۵$ و $($۱- سطر$)($۱- ستون$)) = \chi^2(\%۵$ و $(۲-۱)(۲-۱))$

$= \chi^2(\%۵$ و $۱) = ۳,۸۴۱۴۶$

۳) $\chi^2 > \chi^2(\%۵$ و $۱)$ $\qquad ۰,۹۱۷ \not> ۳,۸۴۱۴۶$

صفر پذیرفته می شود

متغیر کیفی ۹

ضریب همبستگی پیرسون (ρ یا r)

$$\rho = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum (x_i-\bar{x})^2 \sum (y_i-\bar{y})^2}}$$

$-1 < \rho < +1$

تفسیر ρ :

① $\rho = 0 \rightarrow$ رابطه خطی بین x و y وجود ندارد

② $\rho = +1$ رابطه بین x و y خطی کامل و مثبت است

x و y صد درصد در جهت هم هستند

③ $\rho = -1$ رابطه بین x و y خطی کامل در جهت عکس است.

$0 < \rho < 1$ رابطه بین x و y مثبت است

هرچه ρ نزدیک به ۱ بود رابطه قوی است

$-1 < \rho < 0$ رابطه بین x و y منفی است

هرچه به ۱- نزدیک تر بود رابطه قوی تر است.

r ضریب همبستگی $\quad -1 \le r \le +1$

r^2 ضریب تعیین $\quad 0 \le r^2 \le 1$

هرچه ضریب تعیین بالاتر باشد مدل ما مدل مناسب تر است.

ضریب تعیین ← تعیین کننده

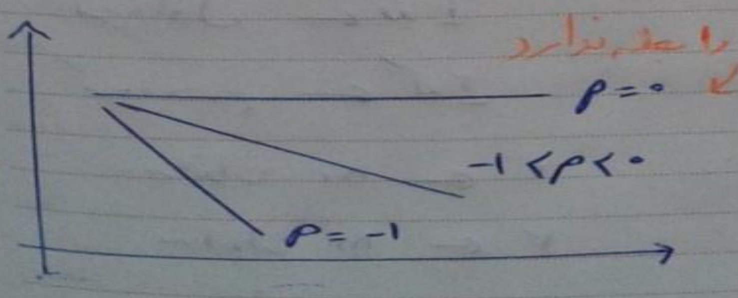
x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
۲	۶	$۲ - ۵ = -۳$	$(۶ - ۵) = +۱$	-۳
۸	۴	$۸ - ۵ = ۳$	$(۴ - ۵) = -۱$	-۳
$\bar{x} = \frac{1}{n}\sum x_i$	$\bar{y} = \frac{1}{n}\sum y_i$			
$\frac{۱}{۲}(۸+۲) = ۵$	$\frac{۱}{۲}(۶+۴) = ۵$			

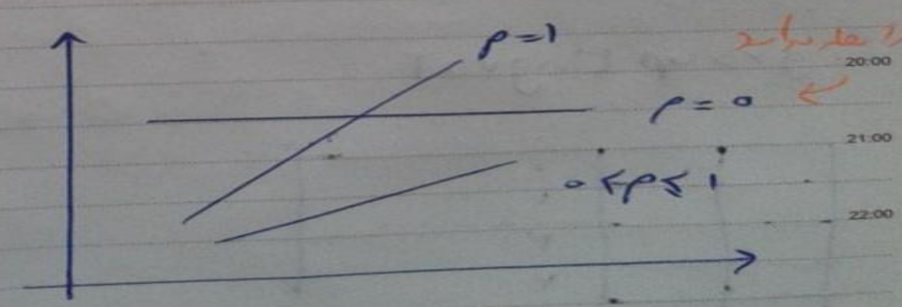
$$\sum (x_i - \bar{x})^2 = (-۳)^2 + (۳)^2 = ۱۸$$

$$\sum (y_i - \bar{y})^2 = (۱)^2 + (-۱)^2 = ۲$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = -۶$$

$$\rho = \frac{-۶}{\sqrt{۱۸ \times ۲}} = -۱$$





تحلیل واریانس:

تحلیل (میزان تغییرات در بیشتر از ۲ تا جامعه):

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_K \rightarrow \text{تعداد گروه} \\ H_1: \text{حداقل دو گروه متفاوت باشند} \end{cases}$$

درجه آزادی		
آماره		

آماره فیشر

① $F = \dfrac{MSb}{MSt \text{ یا } MSE}$

tritment

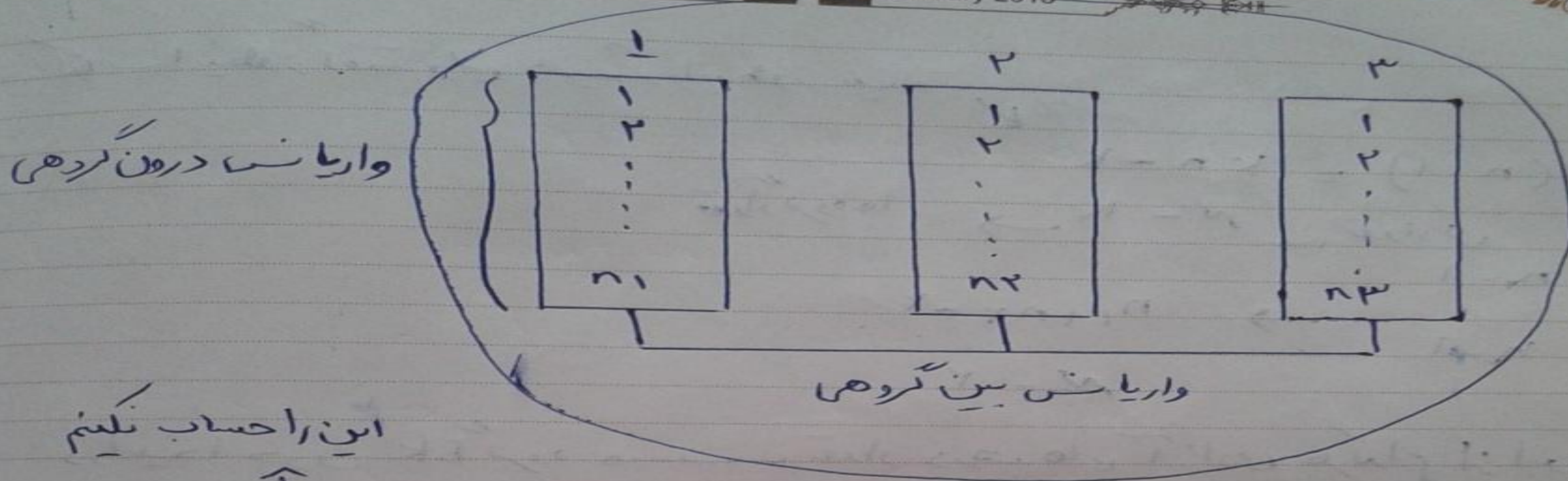
Error

② $F(\alpha$ و $\underbrace{\text{درجه آزادی صورت}}_{df_1}$, $\underbrace{\text{درجه آزادی مخرج}}_{df_2})$

③ $F > F(\alpha, df_1, df_2)$ فرض صفر رد می شود.

$MSb = \dfrac{SSb}{\text{درجه آزادی}} \rightarrow$ واریانس درون گروه

$MSE = \dfrac{SSE}{\text{درجه آزادی}} \rightarrow$ واریانس بین گروه



این را حساب نکنیم

$$SS_T = SS_b + SS_E$$

E و T را حساب کنیم و b را بدست آوریم

درجه آزادی ک $\quad N-1 = \underbrace{k-1} + \underbrace{N-K}$

$\bar{x}$ = میانگین N مشاهده = $\dfrac{\text{همه مشاهدات}}{N}$

$$SS_T = \sum_{i=1}^{K} \sum_{j=1}^{n_K} (x_{ij} - \bar{x})^2$$

$$SS_b = n_1 \sum_j (\bar{x}_{1j} - \bar{x}_{1.})^2 + n_2 \sum_j (x_{2j} - \bar{x}_{2.})^2 + \dots + n_K \sum_j (x_{Kj} - \bar{x}_{K.})^2$$

$$SS_E = n_1(\bar{x}_{1.} - \bar{x})^2 + n_2(\bar{x}_{2.} - \bar{x})^2 + \dots + n_K(\bar{x}_{K.} - \bar{x})^2 =$$

کتاب اینطور گفته ولی همیشه اینطور نیست

$k(n-1) = kn - k$

تعداد گروه ها

تعداد کل ← $N - K$ →

$n_1 - 1$

$\rightarrow n_1 + n_2 - 2$

$n_2 - 1$

$N - 2$

دراینجا چون K تا گروه هست ، تعداد درجه های آزادی هرکدام از این گروه ها یک درجه آزادی دارند

درمثال ۳ تا جامعه ← چون ۳ تا است درجه آزادی می شود $3 - 1 = 2$

مثال ۲_۱۲ فرض کنید ضایعات سه ماشین در ۵ روز ، ۲ روز و ۳ روز بترتیب بر این صورت باشد :

ماشین اول	۷۵	۷۴	۸۳	۸۵	۶۸
ماشین دوم	۶۰	۳۰			
ماشین سوم	۷۱	۷۱	۶۸		

	df_1
df_2	۲
۷	۴٫۷۴

F (α ، df صورت ، df مخرج)

$F(\%0.5 ،\ 3-2=1 ،\ 10-3=7)$

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 \\ H_1 : \text{حداقل دوتا برابر نباشد} \end{cases}$$

$$\bar{x}_1 = \frac{1}{5}(75 + 74 + 83 + 85 + 68) = 77$$

$$\bar{x}_2 = \frac{1}{2}(60 + 30) = 45$$

$$\bar{x}_3 = \frac{1}{3}(71 + 71 + 68) = 70$$

$$\bar{x} = \frac{1}{10}(75 + 74 + 83 + 85 + 68 + 60 + 30 + 71 + 71 + 68) = 68{,}5$$

$$K - 1 = 3 - 1 = 2$$

$$N - K = 10 - 3 = 7$$

① $$F_0 = \frac{MSb}{MSE} = \frac{736{,}25}{92{,}84} = 7{,}93$$

$$MSb = \frac{SSb}{K-1} = \frac{1472{,}5}{2} = 736{,}25$$

$$MSE = \frac{SSE}{N-K} = \frac{650}{7} = 92{,}84$$

$$SST = (75 - 68.5)^2 + (74 - 68.5)^2 + \dots = 2122.5$$

$$SSE = 5(77 - 68.5)^2 + 2(45 - 68.5)^2 + 3(70 - 68.5)^2 = 650$$

$$SSb = SST - SSE = 1472.5$$

(۲) $F(\%05,\ 2,7) = 4.74$

(۳) $7.93 > 4.74$ فرض صفر رد می‌شود

رگرسیون: (برگشت پذیری)

$y = 2x + 1$

رگرسیون یک برگشت پذیری انجام می‌دهد با توجه به این $x \leftarrow y$ هم تغییر می‌کند.

$$\underset{\text{متغیر وابسته}}{y} = \overset{\text{شیب}}{\beta}\ \underset{\text{متغیر مستقل}}{x} + \alpha \longrightarrow \text{عرض از مبدأ}$$

$$\beta = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^۲}$$

$$\alpha = \bar{y} - \beta \bar{x}$$

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
۱۰۰	۹۰	۱۰۰ − ۱۰۰ = ۰	۹۰ − ۱۰۰ = −۱۰	۰
۱۱۰	۱۰۰	۱۱۰ − ۱۰۰ = ۱۰	۱۰۰ − ۱۰۰ = ۰	۰
۹۰	۱۱۰	۹۰ − ۱۰۰ = −۱۰	۱۱۰ − ۱۰۰ = ۱۰	−۱۰۰
$\bar{x}$ = ۱۰۰	$\bar{y}$ = ۱۰۰			

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = ۰ + ۰ + ۱۰۰ = ۱۰۰$$

$$\sum (x_i - \bar{x})^۲ = ۰^۲ + ۱۰^۲ + (-۱۰)^۲ = ۲۰۰$$

$$\beta = \frac{-۱۰۰}{۲۰۰} = \frac{-۱}{۲}$$

$$\alpha = \bar{y} - \beta \bar{x} = ۱۰۰ - (-\tfrac{۱}{۲}) \times ۱۰۰ = +۱۵۰$$

$$y = -\tfrac{۱}{۲} x + ۱۵۰$$

با آرزوی موفقیت برای شما دوستان عزیزم

شاد و موفق باشید.