Subject ..... Date .....

تحلیل آماری ، امیرزاده نماینده کلاس: ۰۹۳۶۹۰۶۵۱۹۰

آمار
- توصیفی ← هر آنچه در جامعه مشاهده می‌شود.
- استنباطی ← با کمک نمونه‌گیری به نتیجه می‌رسیم.

شاخص یا معیارهای آماری
- ۱- مرکزی ←
  - ۱- میانگین ← ساده، هندسی، موزون، همساز
  - ۲- میانه
  - ۳- مد
- ۲- پراکندگی
  - ۱- انحراف از معیار
  - ۲- واریانس
  - ۳- R یا دامنه تغییرات
  - ۴- چارک‌ها و صدک‌ها
  - ۵- CV یا ضریب تغییرات

## آمار استنباطی

آماره و پارامتر: برای پیدا نمودن پارامترهای مجهول جامعه از نمونه تصادفی استفاده می‌کنند یعنی در یک نمونه تصادفی اطلاعات لازم را بدست می‌آورند و به کمک روش‌های آماری اطلاعات مجهول جامعه را پیدا می‌کنند که به این عمل استنباط یا استنتاج آماری می‌گویند.

پارامتر: اصطلاحی است که در مورد جامعه به کار می‌رود. و خصوصیتی از یک جامعه و واریانس جامعه پارامتر می‌باشد، پارامترها معمولاً ثابت ولی

Subject ..... Date .....

میانگین = μ پارامتر

آماره = $\bar{x}$

Subject ..... Date .....

مجهول هستند.

آماره: اصطلاحی است که در مورد جامعه به کار می‌رود و خصوصیت از نمونه را بیان می‌کند مثلاً نسبت نمونه یا میانگین نمونه یک آماره می‌باشد. آماره‌ها معمولاً به سهولت به دست می‌آیند ولی متغیر می‌باشند یعنی از یک نمونه به نمونه دیگر تغییر می‌کنند و به کمک آماره‌ها می‌توان پارامترها را پیدا نمود که به این عمل استنباط آماری می‌گویند و چون آماره‌ها متغیر می‌باشند بنابراین دانستن نحوه‌ی تغییر متغیرهای آماری در نمونه‌های مختلف حائز اهمیت است.

میانگین و انحراف معیار و واریانس:

جامعه‌ای به حجم N میانگین $\mu$ و $\sigma^2$ واریانس را در نظر می‌گیریم از این جامعه یک نمونه تصادفی به حجم n میانگین $\bar{x}$ و واریانس $s^2$ در نظر می‌گیریم.

اگر نمونه را با حجم ثابت n تغییر دهیم $\bar{x}$ که یک آماره می‌باشد از نمونه‌ای به نمونه دیگر تغییر خواهد نمود بنابراین $\bar{x}$ یک (متغیر تصادفی) است. که دارای یک توزیع خواهد بود. برای یک متغیر تصادفی یعنی $\bar{x}$ میانگین و واریانس و انحراف معیار به شرح زیر می‌باشد.



Subject ..... Date .....

میانگین و انحراف معیار و واریانس

جامعه‌ای به حجم نمونه N میانگین $\mu$ و $\sigma^2$ در نظر می‌گیریم:

جامعه $\mu = \frac{\sum x_i}{n}$ $\qquad \sigma^2 = \frac{\sum (x_i - \mu)}{N}$

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

$$\bar{x} \begin{cases} E(\bar{x}) = \mu \\ \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \end{cases} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

قضیه حد مرکزی

یکی از اهمیت است نحوه‌ی توزیع $\bar{x}$ را در نمونه‌های مختلف بیان می‌کند اگر جامعه اصلی نرمال باشد توزیع $\bar{x}$ همواره نرمال خواهد بود و استاندارد شده آن

$$* \quad \boxed{Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

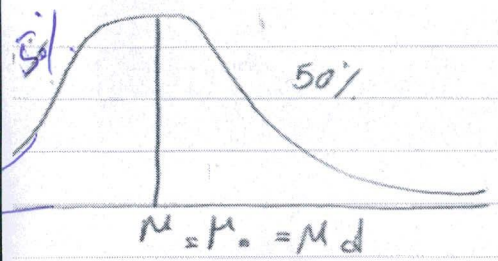
مهم

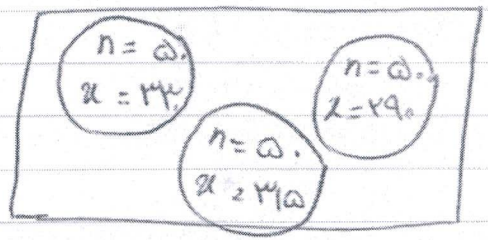
و بستگی به حجم نمونه ندارد همواره نرمال و استاندارد خواهد بود.

ولی اگر توزیع جامعه اصلی نرمال باشد. فقط وقتی توزیع $\bar{x}$ تقریباً نرمال است که حجم نمونه

بزرگ باشد ( $n > 30$ ) و اگر حجم نمونه کوچک باشد ( $n \leq 30$ ) توزیع نرمال نیست

توزیع نرمال $\rightarrow n > 30$




$$\bar{x} = \frac{290 + 330 + 315}{3} = \simeq 300$$

میانگین نمونه‌ها تقریباً برابر با میانگین کل جامعه

| | | میانگین | واریانس | انحراف معیار | حجم نمونه | نسبت ٤ | |
|---|---|---|---|---|---|---|---|
| جامعه | پارامتر | $\mu$ | $\sigma^2$ | $\sigma$ | $N$ | $\pi$ | $\theta$ |
| نمونه | آماره | $\bar{x}$ | $S^2$ | $S$ | $n$ | $p$ | $\hat{\theta}$ |



واریانس $\mu$ جامعه $= \frac{\sum x_i}{N}$

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

$\sigma^2 = \frac{\sum (x_i - \mu)}{N}$ انحراف از معیار (جامعه)

$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ ✓ انحراف معیار (نمونه) *

$$\bar{x} = \begin{cases} E(\bar{x}) = \mu \\ \sigma_{\bar{x}} = \frac{\sigma^2}{n} \end{cases}$$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

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

$S_{\bar{x}} = \frac{S}{\sqrt{n}}$

برآوردکننده‌ها: نقطه‌ای / فاصله‌ای

آزمون فرضیه

## استنباط آماری:

منظور از استنباط آماری این است که با کمک آمارهای معلوم نمونه پارامترهای مجهول جامعه پیدا کنیم. در استنباط آماری انتخاب نمونه و در نظر گرفتن احتمال P الزامی است که این عمل و استنباط آماری به دو روش زیر صورت می‌گیرد.

**برآورد آماری:** در این صورت با انتخاب نمونه تصادفی و پیدا نمودن آماره‌ها در نمونه و با استفاده از فرمولها و روشهای آماری برای پارامتر مجهول جامعه مقدار مشخصی بدست می‌آید. برآورد آماری بر دو نوع است:

الف) برآورد نقطه‌ای: یکی از آماره‌های نمونه را به جای پارامتر جامعه قرار می‌دهیم که به آن آماره [برآوردکننده و یا تخمین‌زن] می‌گویند.

برآورد نقطه‌ای مناسب نمی‌باشد چون نمی‌دانیم که از مقدار واقعی پارامتر جامعه کمتر است و یا بیشتر و حتی خطای آنها را هم نمی‌دانیم.

برای پارامتر مجهول جامعه چندین برآوردکننده را می‌توان از نمونه انتخاب نمود ولی آن برآوردکننده‌ای بهتر است که به آن پارامتر مجهول جامعه نزدیکتر باشد و برآوردکننده باید دارای خصوصیت باشد که مهمترین آن خاصیت نااریبی است



بودن است. به آماره‌ای برآوردکننده نااریب می‌گویند که امید ریاضی آن پارامتر جامعه‌ی نمونه‌های مختلف دقیقاً برابر پارامتر مجهول جامعه گردد و به عبارت دیگر $\hat{\theta}$ (تا حد) در صورتی برآوردکننده نااریب برای پارامتر $\theta$ می‌باشد که داشته باشیم:

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

پارامتر مجهول = امید ریاضی

مثلاً $\bar{x}$ یعنی میانگین نمونه یک برآوردکننده‌ی نااریب برای $\mu$ یعنی میانگین جامعه می‌باشد چون داریم:

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

امید ریاضی

و واریانس نمونه یعنی $S^2$ یک برآورد یکبرآوردکننده نااریب برای واریانس جامعه یعنی $\sigma^2$ می‌باشد:

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

ب) برآورد فاصله‌ای: برای این منظور دو مقدار حداقل و حداکثر به صورت یک فاصله در نظر می‌گیرند که با احتمال مشخص P احتمال یک جامعه در آن فاصله قرار می‌گیرد. معمولاً برآوردکننده نقطه‌ای پارامتر جامعه در وسط فاصله قرار می‌گیرد. هر قدر احتمال بیشتر شود فاصله اعتماد بیشتر می‌شود و بالعکس.

Subject ..... Date .....

| | | |
|---|---|---|
| 10 | | |
| (95 | 1.5) | $p = 90\%$ |
| (80 | 40 12.0) | $p = 99\%$ |
| (90 | 25 11.5) | $p = 95\%$ |

* هر چقدر احتمال بیشتر شود فاصله اعتماد بیشتر می‌گردد.

* برای آزمون فرضیه: در این صورت ابتدا پارامتر مجهول جامعه عدد مناسبی فرض می‌کنیم و آن را فرضیه صفر ($H_0$) در نظر می‌گیریم و فرضیه مخالف آن را $H_1$ می‌نامیم. الکول با انتخاب یک نمونه تصادفی و با در نظر گرفتن احتمال $P$ صحت فرضیه را آزمون می‌کنیم. اگر در نمونه دلایلی بر رد $H_0$ پیدا شود فرضیه $H_1$ را می‌پذیریم و اگر دلیلی بر رد $H_0$ پیدا نشود آن را می‌پذیریم و در این آزمون دو نوع خطا بوجود می‌آید:

* خطای نوع اول: که احتمال آن را $\alpha$ می‌گویند و آن عبارتست از احتمال رد کردن فرضیه $H_0$ در حالیکه واقعاً $H_0$ درست باشد.

* خطای نوع دوم: که احتمال آن را $\beta$ می‌گوییم و آن احتمال اینست که فرضیه $H_0$ را بپذیریم در حالیکه واقعاً $H_1$ درست باشد، هر قدر یکی از این دو کمتر شود دیگری بیشتر می‌شود و بالعکس.

تنها راه برای کاهش هر دو نوع خطا اینست که حجم نمونه را بزرگ انتخاب کنیم



Subject ..... Date .....

تصمیم مهم: در انتخاب فرضیه $H_0$ باید عباراتی را انتخاب نمائیم که شامل = باشد.

$$\alpha \begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases} \qquad \begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases} \qquad \begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta \end{cases}$$

مثال: از جامعه‌ای نمونه‌ای به حجم $n = 100$ انتخاب کرده‌ایم $\bar{x} = 20$ و $s = 10$ شده است، انحراف معیار توزیع میانگین نمونه‌ها یعنی $\sigma_{\bar{x}}$ چقدر است؟

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{100}} = \boxed{1}$$ فرمول جلسه قبل

مثال: اگر $\bar{x}$ و $s_p$ به ترتیب میانگین و واریانس یک نمونه تصادفی ۲۵ تایی باشد برآورد نقطه‌ای انحراف معیار میانگین نمونه‌ها چقدر است؟

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{s}{\sqrt{25}} = \boxed{\frac{s}{5}}$$

مثال: ادعا شده است که نسبت ضایعات در کارخانه حداقل ۶۰٪ است فرض $H_0$ را بسازید.

* سوال:

$$\begin{cases} H_0: p \geq 60\% \\ H_1: p \leq 60\% \end{cases}$$

چرا گفتیم اول $H_1$ بعد $H_0$ اینجا چرا عکس؟

حداقل $\longrightarrow \bar{x} < p$

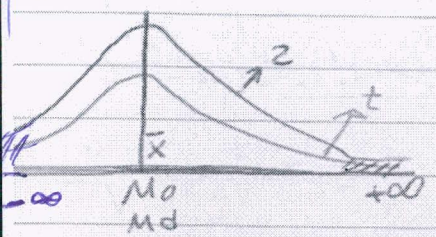
حداکثر $\longrightarrow \bar{x} > p$

مثال: ادعا شده است که نسبت مدیران وظیفه مدار در وزارت خانه A کمتر از وزارتخانه B است. فرض $H_0$ را بسازید.

$H_0: P_A \geq P_B$

اول $H_1$ که صورت مسئله است رو می‌نویسیم و برعکس می‌نویسیم.

$H_1: P_A < P_B$




توزیع tstudent

طبق قضیه حد مرکزی

که قبلاً بیان نمودیم اگر توزیع جامعه اصلی نرمال باشد توزیع $\bar{x}$ همواره نرمال و توزیع استاندارد شده‌ی آن یعنی $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ همواره نرمال استاندارد خواهد بود و بستگی به حجم نمونه ندارد ولی اگر جامعه نرمال نباشد فقط در صورتی توزیع $\bar{x}$ نرمال و توزیع Z نرمال استاندارد است که حجم نمونه بزرگ باشد $(n \geq 30)$

اکنون توزیع جامعه را نرمال فرض می‌کنیم. اگر واریانس جامعه معلوم نباشد به ناچار در فرمول استاندارد شده‌ی Z به جای $\sigma$ از برآورد کننده آن انحراف معیار نمونه یعنی S استفاده می‌کنیم در این صورت متغیر جدیدی پیدا می‌شود که به آن t می‌گویند.

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$



اگر حجم نمونه زیاد شود توزیع t student به نرمال استاندارد نزدیک می‌شود. دلیل درجه آزادی $d_f$ مساوی ۳۰ شود توزیع t تقریباً نرمال می‌شود ولی هر قدر درجه آزادی بیشتر شود توزیع t به نرمال استاندارد نزدیکتر می‌شود و برای $d_f = \infty$ مساوی نهایت توزیع t دقیقاً نرمال می‌شود. ولی اگر درجه آزادی از ۳۰ کمتر شود توزیع t نسبت به نرمال استاندارد خوابیده می‌شود.

تقریب اهمیتی که توزیع tstudent دارد جداولی تنظیم شده است که با آن حرف $d_f$ آزادی یعنی $d_f$ و مساحت زیر منحنی یعنی $\alpha$ می‌توان مقدار t را پیدا نمود که نمونه آن در کتب آمار موجود هست.

نحوه‌ی محاسبه برآورد فاصله‌ای برای میانگین جامعه و نسبت جامعه:

$x - d \qquad x + d$

الف: فاصله اعتماد برای میانگین جامعه:

$\bar{x} \pm Z \frac{\sigma}{\sqrt{n}}$ ← $n \geq 30$ حجم نمونه بزرگ است و $\sigma$ معلوم است.

$\bar{x} \pm Z \frac{S}{\sqrt{n}}$ ← $n \geq 30$ حجم نمونه بزرگ است و $\sigma$ معلوم نیست

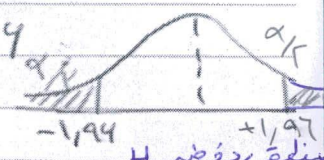
$\bar{x} \pm t \frac{\sigma}{\sqrt{n}}$ ← $n \leq 30$ حجم نمونه کوچک است و $\sigma$ معلوم است

فاصله اعتماد برای نسبت جامعه:

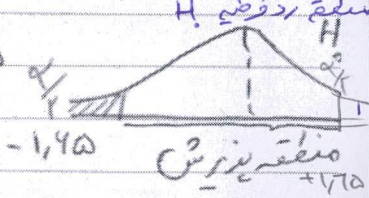
$\bar{x} \pm t \frac{S}{\sqrt{n}}$ ← $n \leq 30$ حجم نمونه کوچک است و $\sigma$ معلوم نیست.

مثال: در یک نمونه ۹۰۰ تایی از بازدیدکنندگان یک فروشگاه تنها ۱۰٪ کالای خاصی را خریدار هستند، تخمین فاصله‌ای (فاصله اطمینان) ۹۵٪ برای P یعنی نسبت کل بازدیدکنندگان مشخص نمائید. ارقام زیر را به خاطر بسپارید.

$\alpha = ٪۵$ $\quad p = ٪۹۵$ $\quad Z_{\alpha/2} = Z_{٪۲۵} = 1,96$




$\alpha = ٪۱۰$ $\quad p = ٪۹۰$ $\quad Z_{\alpha/2} = Z_{٪۵} = 1,65$




$n = 900$

$p = ٪۱۰$ $\quad p \pm Z\sqrt{\frac{p(1-p)}{n}}$

احتمال $= ٪۹۵$

حد شکنه (x) اشتغال

$$\Rightarrow ٪۱ \pm 1,96\sqrt{\frac{٪۱(1-٪۱)}{900}} = [٪۸۴ \text{ و } ٪۸۹۶]$$

رگرسیون و همبستگی:

جامعه یک بعدی از هر فرد جامعه فقط و فقط یک صفت یادداشت می‌شود که آن x می‌گوییم اما در جامعه دو بعدی از هر فرد جامعه دو صفت x و y یادداشت می‌شود. هدف بررسی این است که آیا بین ۲ صفت x و y همبستگی و رابطه‌ای وجود دارد یا خیر و در روش آماری آن را مورد بررسی قرار می‌دهیم



ب) فاصله اعتماد برای نسبت جامعه

در این مورد معمولاً حجم جامعه را بزرگ می‌گیرند اگر P نسبت نمونه باشد فاصله اعتماد برای نسبت جامعه طبق فرمول پائین بدست می‌آید.

$$p \pm Z\sqrt{\frac{p(1-p)}{n}}$$

اگر نمونه‌ی تصادفی از یک جامعه نرمال انتخاب شده باشد:

$\sum(x_i - \bar{x})^2 = 172,8$ و $\sum x = 1008$ و $n = 120$

فاصله‌ی اطمینان ۹۹٪ برای میانگین جامعه چقدر است؟

$1-\alpha/2$

راهنمایی: $\left[\sum Z_{\alpha/2} = ٪۹۵ = 2,575\right]$

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{172,8}{120-1} = 1,45 \rightarrow S = \boxed{1,2}$$

$$\bar{X} \pm 2\frac{S}{\sqrt{n}} \Rightarrow \bar{X} = \frac{\sum x}{n} = \frac{1008}{120} = \boxed{8,4}$$

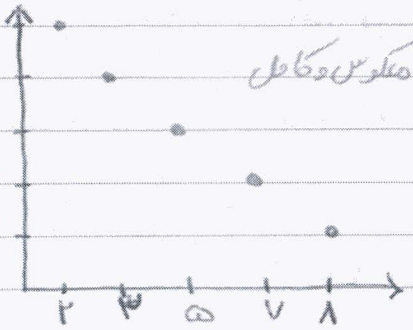
$$8,4 \pm \underbrace{2,575\frac{1,2}{\sqrt{120}}}_{d} \Rightarrow (8,118 \quad \bar{x} \quad 8,682)$$

$$\bar{X} = \frac{\sum x_i}{n} \quad \frac{1008}{120} = 8,4$$

Subject ..... Date .....

مثال: برای اطلاعات زیر دیاگرام پراکندگی را ترسیم کنید.

| x | y |
|---|---|
| ۲ | ۱۰ |
| ۳ | ۸ |
| ۵ | ۶ |
| ۷ | ۴ |
| ۸ | ۲ |




← همبستگی معکوس و کامل

ضریب همبستگی بهترین معیار

بهترین معیار تشخیص همبستگی و میزان آن و حتی جهت همبستگی ضریب همبستگی می‌باشد که فرمول‌های آن بدیع زیر است.

$$r_{\text{ضریب همبستگی}} = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum (x_i-\bar{x})^2 \sum (y_i-\bar{y})^2}}$$

$$r = \frac{\frac{1}{n}\sum (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum (x_i-\bar{x})^2 \sum (y_i-\bar{y})^2}}$$

بهترین فرمول

$$r = \frac{\frac{1}{n}\sum (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{\sum (x_i-\bar{x})^2}{n} \cdot \frac{\sum (y_i-\bar{y})^2}{n}}} = \frac{E(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x \sigma_y} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$r = \frac{\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)}{\sqrt{\left[\sum x_i^2 - \frac{1}{n}(\sum x_i)^2\right]\left[\sum y_i^2 - \frac{1}{n}(\sum y_i)^2\right]}}$$



Subject ..... Date .....

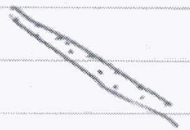
تعبیر: ضریب همبستگی یعنی r همواره بین $-1 \leq r \leq +1$ قرار دارد.

تفسیر ضریب همبستگی:

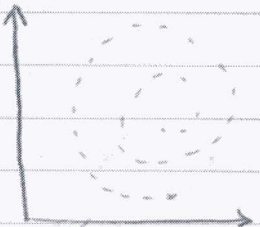
$r = 1$ همبستگی مستقیم و کامل است.

$0 < r < 1$ همبستگی مستقیم و ناقص است.

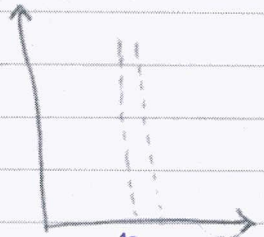
$r = -1$ همبستگی معکوس و کامل است.



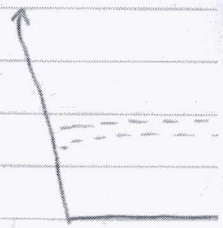
$-1 < r < 0$ همبستگی معکوس و ناقص است



عدم وجود همبستگی




عدم وجود همبستگی



عدم وجود همبستگی

Subject ..... Date .....

ضریب همبستگی برای مثال زیر را بدست آورید:

| $x$ | $y$ | $(x_i-\bar{x})$ | $(y_i-\bar{y})^2$ | $(x_i-\bar{x})^2$ | $(y_i-\bar{y})^2$ | $(x_i-\bar{x})(y_i-\bar{y})$ |
|---|---|---|---|---|---|---|
| ۲ | ۱۰ | -۳ | ۴ | ۹ | ۱۶ | -۱۲ |
| ۳ | ۸ | -۲ | ۲ | ۴ | ۴ | -۴ |
| ۵ | ۶ | ۰ | ۰ | ۰ | ۰ | ۰ |
| ۷ | ۴ | ۲ | -۲ | ۴ | ۴ | -۴ |
| ۸ | ۲ | ۳ | -۴ | ۹ | ۱۶ | -۱۲ |
| ۵ | ۶ | ۰ | ۰ | ۲۶ | ۴۰ | -۳۲ |
| $\bar{x}$ | $\bar{y}$ | | | | | |

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} = \frac{-۳۲}{\sqrt{۲۶\times ۴۰}} = \boxed{-/۹۹}$$

همبستگی معکوس و ناقص

$$\bar{x} = \frac{\sum x_i}{n} = \frac{۲۵}{۵} = ۵$$

اگر ۱+ و ۱- همبستگی کامل

$$-1 < r < 0$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{۳۰}{۵} = ۶$$

معادله خط رگرسیون:

به خطی می‌گویند که از لابه‌لای نقاط می‌گذرد و مجموع مربعات فواصل نقاط از این خط حداقل می‌باشد، خط رگرسیون همواره از نقطه میانگین می‌گذرد و شیب آن را $b$ (ضریب زاویه) می‌باشد ضرایب خط رگرسیون از فرمولهای زیر بدست می‌آید. ✓ حفظ الزامیست

*

$$y = a + bx \longrightarrow a = \bar{y} - bx \checkmark$$

$$b = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \qquad b = \frac{\sum(x_i y_i) - \frac{1}{n}(\sum x_i)(\sum y_i)}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}$$



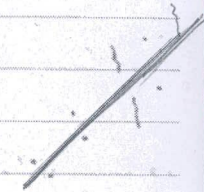
Subject ..... Date .....

$$b = \frac{Cov(x,y)}{\sigma_x^2} \qquad b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

در مثال بالا خط رگرسیون را پیدا کرده و به ازای $x = ۶$ مقدار $y$ را تخمین

$$y = a + bx \longrightarrow a = \bar{y} - b\bar{x} \Rightarrow y = ۱۲,۱۵ - ۱,۲۳(۶) = ۴,۷۷$$

① $$b = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{-۳۲}{(۲۶)^2} = -۱,۲۳$$



$$\boxed{y = ۱۲,۱۵ - ۱,۲۳x}$$

معادله رگرسیون

* $y = ۴,۷۷ + ۱,۲۳x$ اشتباه

ضریب تشخیص یا ضریب تعیین

چند درصد از تغییرات $y$ ناشی از تغییرات $x$ می‌باشد و چند درصد از تغییرات $y$ ارتباطی به تغییرات $x$ ندارد می‌توان از ضریب تشخیص و یا تعیین به شرح زیر استفاده نمود.

$$R = (r)^2$$

در مثال بالا ضریب تشخیص را بدست آورده و آن را تفسیر نمائید:

$$R = (-/۹۹)^2 = /۹۸$$

/۹۸ از تغییرات $y$ ناشی از تغییرات $x$ می‌باشد

تفسیر آن این است که ۹۸٪ از تغییرات $y$ ناشی از تغییرات $x$ می‌باشد

Subject ..... Date .....

و ۲٪ از تغییرات $y$ مربوط به تغییرات $x$ نیست.

آزمون فرضیه، توزیع‌ها { پیوسته / گسسته

پیوسته: نرمال، لاپلاس، استیودنت، $\chi^2$، F، اسنوکر

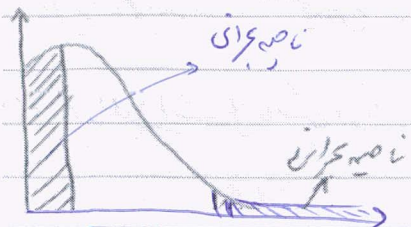
گسسته: دوجمله‌ای، پواسون، دوجمله‌ای هندسی، فوق هندسی، هندسی

$\chi^2$ توزیع چی دو: یک توزیع غیرقرینه، پیوسته و همیشه مثبت است

تعریف: اگر یک سری متغیر تصادفی داشته باشیم که هر کدام دارای توزیع نرمال باشند اگر اینها را استاندارد کنیم و بعد مجذور کنیم مجموع این متغیرهای مجذور شده دیگر دارای توزیع نرمال نیست و دارای توزیع جدیدی به نام $\chi^2$ چی دو یا خی دو ...

مثال: در نمونه به حجم ۵ و در سطح خطای ۵٪ مطلوب است تعیین نواحی بحرانی در توزیع «کای دو»

نکته: درجه آزادی در توزیع کای دو همیشه و همیشه $n-1$ می‌باشد.






Subject ..... Date .....

درجه آزادی:

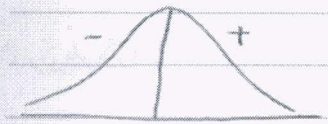
تعداد متغیرهایی که بتوانند آزادانه مقدار اختیار کنند را درجه آزادی می‌گویند.

$\chi^2$ از جدول کوانتیل‌ها

$n = 5$

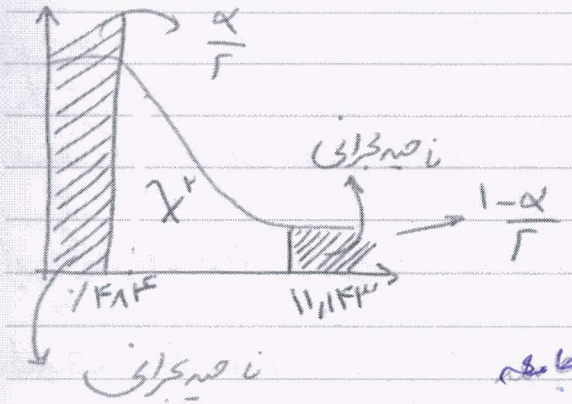
$\alpha = 0.05$

$\alpha/2 = 0.025$




$$\begin{cases} \chi^2_{\alpha/2,\, n-1} = \chi_{0.025,\,4} = 0.484 \\ \chi_{1-\alpha/2,\, n-1} = \chi_{0.975,\,4} = 11.143 \end{cases}$$

چی دو - خی دو - کای دو $\chi^2$




تعیین بازه فاصله‌ای برای میانگین جامعه

در یک کارخانه اتومبیل تحقیقی در مورد جمع کردن جعبه دنده‌ها انجام شده است نمونه تصادفی انتخاب شده و بصورت زیر است: به فرض اینکه مدت زمان جمع کردن جعبه دنده‌ها دارای توزیع نرمال با انحراف معیار ۳ ساعت باشد فاصله‌ی اطمینان با احتمال اعتماد ۹۸٪ را برای متوسط زمان جمع کردن جعبه دنده‌ها بدست آورید.

Subject ..... Date .....

$x$: ۳٫۱ و ۳٫۸ و ۳٫۵ و ۴ و ۴٫۲ و ۳٫۶

$\pm Z \frac{\sigma}{\sqrt{n}}$ => اشتباه میانگین بدست می‌آید

| | |
|---|---|
| | ۰/۳۳ |
| ۲ | ۰/۹۹ |

$\bar{x} = \frac{۳٫۱ + ۳٫۸ + ۳٫۵ + ۴ + ۴٫۲ + ۳٫۶}{۶} = ۳٫۷$

از جدول: $1-\alpha/2 = ۰/۹۹ \leftarrow \alpha = ۰/۰۲ \leftarrow \gamma = ۰/۹۸$

$\bar{x} \pm Z \frac{\sigma/s}{\sqrt{n}} = ۳٫۷ \pm ۲٫۳۳ \frac{۰/۳}{\sqrt{۶}} \rightarrow ۳٫۴۱ \leq \mu \leq ۳٫۹۸$

نتیجه: اگر در این بازه قرار گیرد استاندارد است

مثال: یکی از کارخانه‌های خودروسازی می‌خواهد برای جدیدی را به بازار عرضه نماید

مطلوبست: فاصله اعتماد با احتمال ۹۰٪ برای متوسط مصرف بنزین کلیه خودروهای جدید لذا به این منظور کارخانه مصرف ۱۶ خودرو تولیدی خود را مورد بررسی قرار داد و میانگین مصرف ۱۳٫۹ لیتر بدست آمد (با انحراف معیار ۲٫۳)

$\gamma = ۰/۹۰$ میزان خطا

چون نمونه کمتر از ۳۰ تاست از فرمول ts استفاده می‌کنیم

$\alpha = ۰/۱۰ \quad t_{1-\alpha/2, n-1} = t_{۰/۹۵, ۱۵} = ۱٫۷۵۳$

$\alpha/2 = ۰/۰۵$

$\bar{x} \pm t_s \frac{\sigma}{\sqrt{n}} \quad ۱۳٫۹ \pm ۱٫۷۵۳ \frac{۲٫۳}{\sqrt{۱۶}} = ۱۲٫۸ \leq \mu \leq ۱۴٫۹$



Subject ..... Date .....

تخمین بازه فاصله‌ای برای واریانس جامعه: بسیار مهم

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

مثال: نمونه زیر از یک جامعه نرمال انتخاب شده است. مطلوبست تخمین بازه فاصله‌ای واریانس و انحراف معیار جامعه با احتمال ۹۹٪

$x_i = ۲, ۳, ۴, ۵, ۶$

| $x_i$ | $x_i^2$ |
|---|---|
| ۲ | ۴ |
| ۳ | ۹ |
| ۴ | ۱۶ |
| ۵ | ۲۵ |
| ۶ | ۳۶ |
| $\sum x = ۲۰$ | $\sum x^2 = ۹۰$ |

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

$s^2 = \frac{۹۰ - \frac{(۲۰)^2}{۵}}{۵-1} = ۲٫۵ \Rightarrow s = \sqrt{۲٫۵} = ۱٫۵۸$

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

$\chi_{\alpha/2, n-1} = \chi^2_{۰٫۰۵, ۴} = ۰/۰۷$

$\chi^2_{1-\alpha/2, n-1} = \chi^2_{۰٫۹۵, ۴} = ۱۴٫۹$

$I = \left(\frac{(۵-1)۲٫۵}{۱۴٫۹} \leq \sigma^2 \leq \frac{(۵-1)۲٫۵}{۰/۰۷}\right) \Rightarrow I(۰/۶۷ \leq \sigma^2 \leq ۱۴۸٫۳)$

$\left(I\left(\sqrt{۰/۶۷} \leq \sigma \leq \sqrt{۱۴۸٫۳}\right)\right)$

$(n-1)s^2$

Subject ..... Date .....

توزیع فیشر fisher

یک توزیع غیرمرضیه است و مثل توزیع $\chi^2$ دارای ۲ ناحیه بحرانی است ولی

درجه آزادی متفاوت می‌باشد

$$b = \frac{\frac{\chi^2_{v_1}}{v_1}}{\frac{\chi^2_{v_2}}{v_2}}$$

مثال: اولین عدد همیشه $v_1$ و دومین عدد همیشه $v_2$ است.

$v_1$ افقی $F_{.95}$ و ۳ و ۲ $= 9,552$

$v_2$ عمودی

$F_{.95}$ و ۳ و ۲ $= 19,164$

$v_1$ افقی

$v_2$ عمودی

مثال:

$$I\left(\frac{S_1^2}{S_2^2} \times \frac{1}{F_{1-\alpha/2,\, v_1, v_2}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} F_{1-\alpha/2,\, v_1, v_2}\right)$$

صورت

مخرج

مثال: دو نمونه تصادفی $n_1 = 16$ و $n_2 = 21$ انتخاب کرده‌ایم

و واریانس نمونه به ترتیب $S_1^2 = 100$ و $S_2^2 = 25$ محاسبه شده است

مطلوبست فاصله اعتماد برای نسبت واریانس‌های جامعه اصلی با احتمال اعتماد ۰/۹



Subject ..... Date .....

$S_1^2 = 100 \quad \gamma = 0/9 \quad \begin{cases} v_1 = n_1 - 1 \\ v_2 = n_2 - 1 \end{cases} \quad \begin{cases} v_1 = 16 - 1 = 15 \\ v_2 = 21 - 1 = 20 \end{cases}$

$S_2^2 = 25 \quad \alpha = 0/1$

I) از جدول توزیع

$F_{.95,\, v_1 = 15,\, v_2 = 20} = \boxed{2,233}$

سوال X

$F_{.95,\, v_2 = 20,\, v_1 = 15} = \boxed{2,203}$

$\alpha = 0/1$

$\gamma = 0/9$

$$I\left(\frac{100}{25} \times \frac{1}{2,2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{100}{25} \times 2,33\right)$$

$v_1, v_2$ $\quad$ $v_2, v_1$

$$I\left(1,716 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 8,80\right)$$

آزمون فرضیه مقایسه یکسان بودن روش‌ها «فریدمن» اعتبارسنجی پرسشنامه

فرضیه‌سازی

1) $H_0: \beta_1 = \beta_2 = \dots = \beta_k$

$H_1: \beta_i \neq \beta'_i \quad i, j = 1, 2, \dots, k$

زمانی که چن تا فرضیه داشته باشی هر چی بود فقط ۲ برجدید

2) $k = \frac{12}{nk(k+1)} \sum R^2 - 3n(k+1) = \chi^2_{1-\alpha,\, k-1}$

ملاک آزمون

تعیین میزان خطا

3) $\alpha = \dots$

$1 - \alpha = \dots \quad \chi^2_{1-\alpha,\, k-1}$

4) $W(\chi^2, \chi^2_{1-\alpha,\, k-1}) \to Q_c = K$ تعیین نواحی بحرانی

5) محاسبه عددی

6) تصمیم‌گیری Dec

Subject ..... $k=3$ Date .....

مثال: ۱۸ کارمند به سه گروه تقسیم می‌شوند سپس اعضای داخل گروه‌ها بطور تصادفی

$k=3$

تحت سه روش رفتاری قرار داده می‌شوند معلومست:

روش

۱) آزمون یکسان بودن روش‌ها با توجه به آنکه نظر بدست آمده در فیمه‌ی تاثیر این

روش

در این مسئله گروه بصورت زیر می‌باشد: اگر مسئله α روش تعیین نکرده بود α: بطور پیش فرض ۰/۵ هستش

از مرحله ۱ شروع می‌کنیم تا مرحله آخر پیش می‌رویم

درجه اهمیت بزرگترین رقم به ترتیب در هر ردیف

| گروه | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| ۱ | ۱۵ (3) | ۱۲ (2) | ۱۱ (1) |
| ۲ | ۱۹ (3) | ۱۵ (2) | ۱۴ (1) |
| ۳ | ۲۳ (3) | ۱۸ (2) | ۱۵ (1) |
| ۴ | ۲۰ (1) | ۲۲ (3) | ۲۱ (2) |
| ۵ | ۲۵ (2) | ۱۸ (2) | ۱۷ (1) |
| ۶ | ۲۷ (3) | ۲۲ (1) | ۲۴ (2) |
| $R_j$ | ۲۵۹ (16) | ۱۴۴ (12) | ۶۴ (8) |

$K = \frac{12}{n k(k+1)} \sum R_j^2 - 3n(k+1) = \chi^2$

n ← تعداد نمونه

k ← تعداد روش‌ها

$k=3 \quad \alpha = 0.05 \quad 1-\alpha = 0.95$

$\chi^2 = 5.99$ → $0.95$ و $2 \to k-1$ مزد

$\sum R_j^2 = 464$

n چرا اینجا (۶)؟

$3(6)$

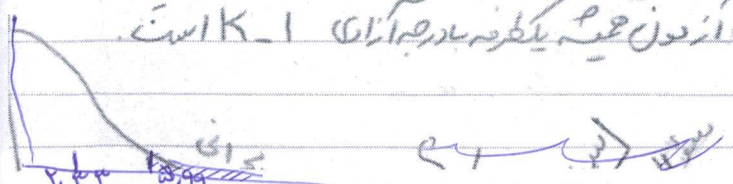
فرضیه $H_0$ رد می‌شود.

$K = \frac{12}{18 \times 3(3+1)} 464 - 3(18)(4) = 2.33$ *

$H_0$ حفظ می‌شود

یعنی سه روش رفتاری یکسان است.

نکته: این آزمون همیشه یکطرفه با درجه آزادی $k-1$ است.





Subject ..... Date .....

آنالیز واریانس ANOVA

جدول آنالیز واریانس:

| منبع تغییرات | مجموع توان دوم‌ها | درجه آزادی $d_f$ | واریانس $M_S$ | F |
|---|---|---|---|---|
| بین گروه‌ها | SSR | $v_1 = k-1$ | MSR | $\frac{MSR}{MSE}$ |
| درون گروه‌ها | SSE | $v_2 = N-k$ | MSE | |
| کل گروه‌ها | SST | $N-1$ | MST | |

$SSR = \sum_{i=1}^{k} n(x_i - \bar{x})^2$ $\qquad F = \frac{MSR}{MSE} = n s_{\bar{x}}^2$

$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} (x_{ij} - \bar{x})^2$ $\qquad v_2 = k-1$

$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n} (x_{ij} - X_i)^2$ $\qquad v_1 = N-k$

$MSR = \frac{SSR}{k-1} = n s_{\bar{x}}^2$

$MSE = \frac{SSE}{N-k} = s^2 p$ $\qquad SSE = SST - SSR$

$MST = \frac{SST}{N-1} = s^2$ $\qquad$ فرض $H_0$ را رد می‌کنیم $\qquad$ اگر $F > F_{(v_1, v_2, \alpha)}$

$SSR = \sum_{i=1}^{k} \frac{T_i^2}{n} - \frac{T^2}{N}$ $\qquad$ فرض $H_0$ را می‌پذیریم $\qquad$ اگر $F \leq F_{(v_1, v_2, \alpha)}$

$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \frac{T^2}{N}$

Subject ..... Date .....

مروره جلسه بعد بیاورید

## آزمون چی دو کارل پیرسون

برای پی بردن به validity (روایی) یا اعتبار پرسشنامه همیشه از آزمون استقلال

ی دو کارل پیرسون

چگالی نمی

$$1)\begin{cases} H_0 = \varphi(x,y) = \varphi(x)\cdot\varphi(y) \\ H_1 = \varphi(x,y) \neq \varphi(x)\cdot\varphi(y) \end{cases}$$

توزیع ناپیوسته گسسته

$$\begin{cases} H_0 \;\&\; P_{ij} = P_i P_j \\ H_1 \;\&\; P_{ij} \neq P_i P_j \end{cases} \quad \text{یا} \quad \begin{cases} H_0 \text{ مستقل هستند } x,y \\ H_1 \text{ وابسته هستند } x,y \end{cases}$$

$$2)\quad K = \sum\sum \frac{(M_{ij} - M'_{ij})^2}{M'_{ij}} - n = \sum\sum \frac{M_{ij}^2}{M'_{ij}} \Rightarrow$$

$$\chi^2_{1-\alpha,(S-1)(t-1)}$$

محاسبه چی دو در جدول پس از محاسبه جدولهای نظری ادغامهای احتمالی

فراوانی‌های

ادغام احتمالی مشخص می‌نمائیم. ادغام از جدول $M_{ij}$ آغاز می‌شود. پس همان مراحل

در جدول $M_{ij}$ نیز اعمال می‌شود. در صورتی که در جدول $M_{ij}$ عدد کمتر از ۵ باشد نیاز به ادغام می‌باشد



Subject ..... Date .....

مثال: جدول آنالیز واریانس ANOVA را برای داده‌های جدول زیر تشکیل داده

و فرض تساوی میانگین‌های ۴ جامعه را در سطح $\alpha = 0/1$ آزمون کنید.

| ۱ | ۲ | ۳ | ۴ | منبع تغییرات | مجموع توان دوم | درجه آزادی | واریانس | F |
|---|---|---|---|---|---|---|---|---|
| ۶۵ | ۷۵ | ۵۹ | ۹۴ | بین گروه‌ها | $SSR = 880,15$ | $v_1 = k-1$, $v_1 = 3$ | $MSR = 293,4$ | $\frac{MSR}{MSE}$ |
| ۸۷ | ۶۹ | ۷۸ | ۸۹ | درون گروه‌ها | $SSE = 951,6$ | | $MSE = 59,6$ | $F = 4,93$ |
| ۷۳ | ۸۳ | ۶۷ | ۸۰ | کل گروه‌ها | $SST = 1831,75$ | $v_2 = N-k$, $20-4=16$ | $\frac{951,6}{19}$ | |
| ۷۹ | ۸۱ | ۶۲ | ۸۸ | | | $N-1$, $20-1=19$ | MST نمی‌خواد | |
| ۸۱ | ۷۲ | ۸۳ | ۹۰ | | | | | |
| ۳۸۵ | ۳۸۰ | ۳۴۹ | ۴۹۱ | $T_i$ مجموع داده‌ها | | | | |
| ۷۷ | ۷۶ | ۶۹,۸ | ۸۸,۲ | $\bar{x}$ میانگین داده‌ها | | | | |
| ۲۹۲۵ | ۲۹۰۲۰ | ۲۴۷۸۷ | ۳۹۰۰۱ | $\sum x_i^2$ مجموع توان دوم داده‌ها | | | | |

$$T = T_1 + T_2 + T_3 + T_4 = 385 + 380 + 349 + 491 = 1555$$

$$\sum_{ij} x_{ij}^2 = 2925 + 29020 + 24787 + 39001 = 122722$$

$$SSR = \sum_{i=1}^{k} \frac{T_i^2}{n} - \frac{T^2}{N} = \frac{1}{5}\left((385)^2 + (380)^2 + (349)^2 + (491)^2\right) - \frac{1000}{20} = 880,15$$

بین گروه‌ها — کل گروه‌ها

$$SST = \sum x_{ij}^2 - \frac{T^2}{N} = 122722 - \frac{(1555)^2}{20} = 1831,75$$

مجموع توان دوم داده‌ها — میانگین

$$SSE = SST - SSR = 1831,75 - 880,15 = 951,6$$

$$F(v_1, v_2, \alpha) = F_{3,16,0/1} = 5,29$$

$$\begin{cases} H_0 = \mu_1, \mu_2, \mu_3, \mu_4, \mu_5 \\ H_1 = \mu_i \neq \mu_i \end{cases}$$

مرز $F = 4,93$

فرض میانگین جامعه رد می‌شود

$H_0$ رد می‌گردد

Subject ..... Date .....

۳) $\alpha = \dots\dots$ ۰/۵ و یا ۰/۱ همیشه یک طرفه $1-\alpha$

$$\chi^2_{1-\alpha,\,(s-1)(t-1)}$$

* ضرب

۴) $H_0$ $\chi^2 \geq \chi^2_{1-\alpha,\,(s-1)(t-1)}$

مثال: توزیع فراوانی توزیع یک موسسه برحسب جنس x و شغل y بیان شده است. مطلوبست آزمون فرضیه دو صفت x و y :

جدول اصلی M

| شغل y / جنس x | کارمند اداری | مهندس | تکنیسین | استادکار | کارگر | $M_{i.}$ |
|---|---|---|---|---|---|---|
| مرد | ۳۰ | ۱۱ | ۳۳ | ۷۶ | ۱۵۰ | ۳۰۰ |
| زن | ۱۰ | ۱ | ۹ | ۱۰ | ۷۰ | ۱۰۰ |
| $M_{.j}$ فراوانی تجربی صفت | ۴۰ | ۱۲ | ۴۲ | ۸۶ | ۲۲۰ | ۴۰۰ |

M'

| شغل y / جنس x | کارمند اداری | مهندس | تکنیسین | استادکار | کارگر | $M'_{i.}$ |
|---|---|---|---|---|---|---|
| مرد | ۳۰ | ۹ + | ۳۱,۵ | ۶۴,۵ | $\frac{۲۲۰ \times ۳۰۰}{۴۰۰} = ۱۶۵$ | ۳۰۰ |
| زن | ۱۰ | $\boxed{۳}$ + | ۱۰,۵ | ۲۱,۵ | ۵۵ | ۱۰۰ |
| | ۴۰ | ۱۲ | | | | ۴۰۰ |

این جدول ملاک ماست برای مرحله بعد

یک عدد کمتر از ۵ داریم با جلیقی ادغام می‌کنیم

| شغل y / جنس x | کارمند اداری | تکنیسین و مهندس | استادکار | کارگر | |
|---|---|---|---|---|---|
| مرد | ۳۰ | ۴۰,۵ | ۶۴,۵ | ۱۶۵ | ۳۰۰ |
| زن | ۱۰ | ۱۳,۵ (۱۰,۵+۳) | ۲۱,۵ | ۵۵ | ۱۰۰ |
| | ۴۰ | ۵۴ | ۸۶ | ۲۲۰ | ۴۰۰ |



Subject ..... Date .....

جدول اصلی باید با هم مکمل شوند کنیم سپس

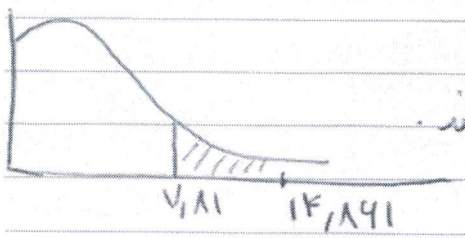
۳) $\chi^2_{1-\alpha,\,(s-1),\,(t-1)} = \chi^2_{0.95,\,3} = 7.81$ از جدول انتهای کتاب

$s-1$: تعداد سطر ، $t-1$: تعداد ستون ← از جدول ادغام شده بدست می‌آید

۴) $\chi^2 \geq 7.81$

| $M_{ij}$ جدول اصلی | $M'_{ij}$ جدول ادغام شده | $M_{ij} - M'_{ij}$ | $\frac{(M_{ij} - M'_{ij})^2}{M_{ij}}$ |
|---|---|---|---|
| ۱۵۰ | ۱۶۵ | -۱۵ | ۱,۳۶۴ |
| ۷۶ | ۶۴,۵ | ۱۱,۵ | ۲,۰۵ |
| ۴۴ * تو جدول اصلی مهندس و تکنیسین رو با هم جمع کردیم | ۴۰,۵ | ۳,۵ | ۰/۳ |
| ۳۰ | ۳۰ | ۰ | ۰ |
| ۷۰ | ۵۵ | ۱۵ | ۴,۹ |
| ۱۰ | ۲۱,۵ | -۱۱,۵ | ۶,۱۵ |
| ۱۰ * جمع مهندس و تکنیسین | ۱۳,۵ | -۳,۵ | ۰/۹۷ |
| ۱۰ | ۱۰ | ۰ | ۰ |
| | | | ۱۴,۸۶۱ |

$\chi^2$ محاسبه شده




فرضیه $H_0$ رد می‌شود و x و y وابسته به هم هستند.