آمار و کاربرد آن در مدیریت

میانگین یا متوسط حسابی: مجموع اعداد تقسیم بر تعداد آن‌ها را میانگین گویند و با $\bar{X}$ و یا μ نشان می‌دهیم و از رابطه زیر بدست می‌آید:

$$\bar{X} = \frac{\sum x_i}{n}$$ «میانگین نمونه»

$n \rightarrow$ تعداد نمونه

$$\mu = \frac{\sum x_i}{N}$$ «میانگین جامعه»

$N \rightarrow$ تعداد جامعه

واریانس:

مجموع مجذورات تفاضل از میانگین را تقسیم بر تعداد واریانس گویند.

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$ «واریانس نمونه»

$$var(X) = \delta_x^2 = \frac{\sum (x_i - \mu)^2}{N}$$ «واریانس جامعه»

انحراف معیار: جذر واریانس را انحراف معیار گویند. انحراف

$$S = \sqrt{S^2}$$ «انحراف معیار نمونه»

$$\delta = \sqrt{\delta^2}$$ «انحراف معیار جامعه»

سوال: برای داده‌های زیر، میانگین، واریانس و انحراف معیار را بدست آورید.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
2	-4	16
4	-2	4
6	0	0
8	2	4
10	4	16
30		

$$\bar{X} = \frac{\sum x_i}{n} = \frac{30}{5} = 6$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{40}{4} = 10$$

$$S = \sqrt{S^2} = \sqrt{10} = 3.16$$

SUBJECT:
Year( )Month( )Date( )

نکته: در صورتیکه میانگین (اعشاری) باشد بهتر است با استفاده از یکی از روابط زیر واریانس را بدست آوریم.

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

روش اول راحتتر است

سوال) میانگین، واریانس و انحراف معیار اعداد زیر را بدست آورید.

x_i	x_i^2
1	1
3	9
4	16
7	49
9	81
24	156

$$\bar{X} = \frac{\sum x_i}{n} = \frac{24}{5} = 4.8$$

$$S^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{156 - 5\times 4.8^2}{4} = 10.2$$

$$S = \sqrt{S^2} = \sqrt{(10.2)^2} = 3.17$$

طبقه‌بندی اعداد:

1. دامنه تغییرات را بدست می‌آوریم.
2. تعداد طبقات را مشخص می‌کنیم.
3. طول رده را بدست می‌آوریم.

$$R = X_{max} - X_{min}$$

$$K = 1 + 3.32 \log n$$

$$I = \frac{R}{K}$$

نکته: واریانس تقریبی بین $\frac{R}{4}$ تا $\frac{R}{6}$ است.

* محاسبه میانگین، واریانس و انحراف معیار از روی جدول فراوانی:

$$\bar{X} = \frac{\sum f_i x_i}{n} \qquad S^2 = \frac{\sum f_i x_i^2 - n\bar{x}^2}{n-1} \qquad S = \sqrt{S^2}$$

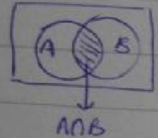
$$n = \sum f_i$$

STAEDTLER

SUBJECT: جلسه ۲
Year( )Month( )Date( )

* پیشامد قرینه - متمم :
هرگاه حل یک پیشامد قرینه شود از متمم استفاده می کنیم. به عنوان مثال اگر شخصی تیری را با احتمال ۷۰٪ به هدف بزند با احتمال ۳۰٪ به هدف نمی زند پس

$P(A') = 1 - P(A)$



محاسبه اجتماع دو پیشامد — اجتماع = یا

هرگاه احتمال وقوع A یا B را بخواهیم از احتمال اجتماع دو پیشامد استفاده می کنیم.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$\begin{cases} P(A \cap B) = 0 & \text{ناسازگار} \\ P(A \cap B) = P(A) \cdot P(B) & \text{مستقل} \end{cases}$$

* در حالت پیوسته به جای شمارش از مساحت استفاده می شود.

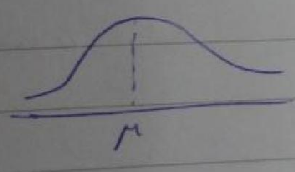
ص ۲۴

توزیع نرمال :
این توزیع یک توزیع متقارن است که در آن میانگین، میانه و مد برهم منطبق هستند. تابع چگالی آن بصورت زیر می باشد.

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

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

توزیع — متغیر تصادفی — نرمال — میانگین — واریانس



که محاسبات مربوط به احتمال آن مشکل است. به همین دلیل با تغییر متغیر z، متغیر تصادفی را به نرمال استاندارد با میانگین صفر و واریانس ۱ تبدیل می کنیم. در این صورت تابع چگالی نرمال استاندارد بصورت زیر می باشد:

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

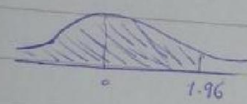
$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$$

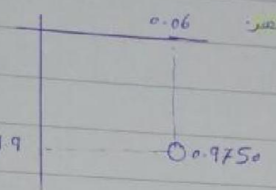
$Z \sim N(0, 1)$

در این صورت می توان مساحت سمت چپ یا احتمال کوچکتر Z را از جدول صفحه (142) بدست آورد.

با اعداد یکان و دهم را از ستون اول و عدد صدم را از سطر اول بدست می‌آوریم. محل برخورد دو عدد احتمال

$P(Z<1.96)$





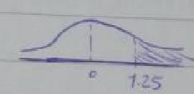
عدد سطر را نشان می‌دهد:

$P(Z<1.96)=\underset{\downarrow\ G}{\Phi}(1.96)=0.9750$

سوال) احتمال‌های زیر را بدست آورید:

$P(Z<2.54)=\Phi(2.54)=0.9945$

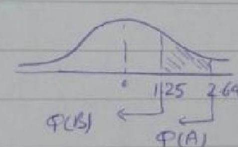
$P(Z>1.25)=1-\Phi(1.25)=1-0.8944=0.1056$



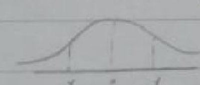
$P(Z)<-1.64)=P(Z>1.64)=1-\Phi(1.64)=0.0505$

، چون توزیع متقارن است،

$P(1.25<Z<2.64)=\overset{A}{\Phi}(2.64)-\overset{B}{\Phi}(1.25)=0.1015$



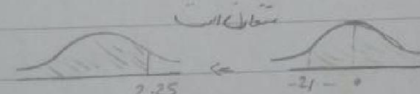
سوال)



$P(-1<Z<1)=\Phi(1)-\Phi(-1)$

$=0.8413-(1-\Phi(1))=0.8413-1+0.8413=0.6826$

متقارن است



$P(Z>-2.25)=\Phi(2.25)=0.9878$

« احتمال نه منفی است و نه بزرگتر از 1 »

چند نکته:

1) $P(Z<Z_0)=\Phi(Z_0)$

2) $P(Z>Z_0)=1-\Phi(Z_0)$

3) $P(Z<-Z_0)=1-\Phi(Z_0)$

4) $P(Z_1<Z<Z_2)=\Phi(Z_2)-\Phi(Z_1)$

5) $P(Z>-Z_0)=\Phi(Z_0)$

6) $P(-Z_0<Z<Z_0)=2\Phi(Z_0)-1$

SUBJECT

Year( )Month( )Date( )

مثال) قد دانش آموزان دوره ابتدایی دارای توزیع نرمال با میانگین ۱۴۰ و انحراف معیار ۱۰ است. اگر دانش آموزی انتخاب شود چقدر احتمال دارد که قد آن بیش از 152 cm باشد.

$\mu = 140 \qquad \delta = 10$

$$P(x > 152) = P\left(Z > \frac{152-140}{10}\right) = P(Z > 1.2) = 1 - \Phi(1.2) = 0.1151$$

$$Z = \boxed{\frac{x-\mu}{\delta}}$$ ← «استاندارد سازی»

استاندارد کردن:

$$Z = \frac{\text{متغیر} - \text{میانگین}}{\text{انحراف معیار}}$$ (نحوه استاندارد)

$$Z_x = \frac{x-\mu}{\delta}$$ (استاندارد X)

$$Z_{\bar{x}} = \frac{\bar{x}-\mu}{\frac{\delta_x}{\sqrt{n}}}$$ (استاندارد میانگین)

P: احتمال پیروزی $\qquad q = 1-p$ (احتمال شکست)

$$Z_{\text{دوجمله‌ای}} = \frac{x-np}{\sqrt{npq}}$$ (شکست، پیروزی)

در توزیع نمائی میانگین و انحراف معیار برابرند. $\left(\mu = \delta = \frac{1}{\lambda}\right)$

$$Z_{\text{نمائی}} = \frac{x-\frac{1}{\lambda}}{\frac{1}{\lambda}}$$

در توزیع پواسون واریانس و میانگین برابر است (λ) $\quad (\mu = \delta^2) \quad \delta = \sqrt{\mu}$

$$Z_{\text{پواسن}} = \frac{x-\mu}{\sqrt{\mu}}$$

وزن کنسروها دارای توزیع نرمال با میانگین ۴۰۰ و انحراف معیار ۱۰ می‌باشد. اگر یک نمونه 36 تایی انتخاب کنیم چقدر احتمال دارد میانگین وزن کنسروها کمتر از 397 گرم باشد.

$\mu = 400 \qquad \delta = 10$ (استاندارد میانگین)

$$P(\bar{x} < 397) = P\left(Z < \frac{397-400}{10/6}\right) = P(Z < 1.8) = 1 - \Phi(1.8) = 0.0359$$

$$Z_{\bar{x}} = \frac{\bar{x}-\mu}{\frac{\delta_x}{\sqrt{n}}}$$

STAEDTLER

محاسبه Z_α :

ابتدا $(1-\alpha)$ را بدست می‌آوریم. سپس در متن جدول توزیع نرمال عدد $(1-\alpha)$ را پیدا می‌کنیم و از روی آن Z را تعیین می‌کنیم. در صورتیکه عدد $(1-\alpha)$ در متن جدول موجود نباشد متوسط دو عدد نزدیک را بدست می‌آوریم.

مثال)

$Z_{0.025} = ?$ $\quad 1-\alpha = 0.975$ $\quad Z_{0.025} = 1.96$

	0.06
1.9	0.975

$Z_{0.05} = ?$ $\quad 1-\alpha = 1-0.05 = 0.95$

$$Z_{0.05} = \frac{1.64 + 1.65}{2} = 1.645$$

	0.04		0.05
1.6	0.9495	0.95	0.9505

مثال)

$Z_{0.1} = ?$ $\quad Z_{0.08} = ?$

$1-\alpha = 1-0.1 = 0.9$ $\quad \frac{1.29 + 1.28}{2} = 1.285$

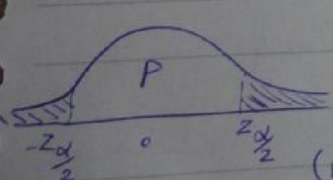
$1-\alpha = 1-0.08 = 0.92$ $\quad (1.41 + 1.4)/2 = 1.405$

• برآورد فاصله میانگین جامعه وقتی واریانس جامعه معلوم باشد :

برای صورت از توزیع Z یا نرمال استفاده می‌کنیم.

α : سطح خطا $(1-P)$

P : سطح اطمینان



(چون خطا که α است را نصف می‌کنیم نصف + و نصف منفی ⇐ $\frac{\alpha}{2}$ را حساب می‌کنیم)

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

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

STAEDTLER

مثال) در یک جامعه با توزیع نرمال و واریانس 64 یک نمونه 36 تایی انتخاب می‌کنیم که میانگین نمونه 45 است.
یک برآورد فاصله‌ای 95% برای میانگین واقعی جامعه بنویسید.

$\delta^2 = 64 \rightarrow \delta = 8$ $\quad n = 36 \quad \bar{x} = 45$

$\mu : \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}$

$P = 95\%$

$\alpha = 1 - P = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \rightarrow Z_{0.025} = 1.96$

$\mu : 45 \pm 1.96 \frac{8}{6}$ $\quad UCL : 47.61 \quad LCL : 42.39$

مهم

$Z_{0.025} = 1.96$

$Z_{0.05} = 1.645$

* برآورد فاصله میانگین وقتی واریانس جامعه معلوم نباشد

در این صورت با استفاده از اطلاعات نمونه، مسئله را حل می‌کنیم که دو حالت امکان دارد رخ دهد:

۱- حجم نمونه بیشتر از 30 باشد: در این صورت طبق قضیه حد مرکزی از توزیع Z یا نرمال استفاده می‌کنیم

به جای واریانس جامعه از واریانس نمونه استفاده می‌شود.

جامعه	نمونه
μ	$\bar{x}$
δ	S

$\mu = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$ → انحراف معیار نمونه

یک نمونه 49 تایی میانگین نمونه 95 و واریانس نمونه 81 است. یک برآورد فاصله‌ای 90% برای میانگین واقعی جامعه بنویسید.

$\bar{x} = 95 \quad n = 49 > 30 \Rightarrow S^2 = 81 \rightarrow S = 9$

$P = 90\% \rightarrow \alpha = 1 - 0.90 = 0.1 \quad \frac{\alpha}{2} = 0.05 \quad Z_{0.05} = 1.645$

$\mu = 95 \pm 1.645 \frac{9}{7}$ $\quad 97.11 \quad 92.89$

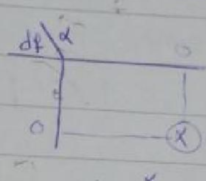
2. برآورد فاصله میانگین وقتی واریانس جامعه مجهول باشد و حجم نمونه کمتر از 30 باشد:

برای صورت از توزیع t جدول صفحه 143 با درجه آزادی n-1 برآورد فاصله‌ای را تشکیل می‌دهیم. منحنی t نسبت به منحنی Z پهن‌تر است.

$df = \nu = r = n-1$ = درجه آزادی

(ν = نو)

$$\mu = \bar{x} \pm t_{\frac{\alpha}{2}, df} \frac{S}{\sqrt{n}}$$



مثال) در یک نمونه 16 تایی میانگین نمونه 47 و واریانس نمونه 25 است. یک برآورد فاصله‌ای ٪95 برای میانگین واقعی جامعه بنویسید.

$n = 16 < 30$ درجه آزادی $= 15$ $\bar{x} = 47$ $S^2 = 25 \Rightarrow S = 5$

$P = 95\% \Rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$

$$\mu = 47 \pm 2.131\left(\frac{5}{4}\right) = 47 \pm 2.66375$$

* برآورد فاصله‌ای نسبت‌ها:

اغلب از توزیع Z استفاده می‌شود.

$\bar{P} = \frac{x}{n}$ $\frac{\rightarrow \text{تعداد حالات مساعد}}{\rightarrow \text{کل}}$ پیروزی $\bar{q} = 1 - \bar{p}$ (شکست)

$$P = \bar{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}}$$

از بین 100 کالا 20 کالا معیوب است. یک برآورد فاصله‌ای ٪90 برای نسبت کالاهای معیوب بنویسید.

$n = 100$ $\bar{p} = \frac{20}{100} = 0.2$ $\bar{q} = 0.8$

$P = 0.90$ $\alpha = 1 - 0.9 = 0.1$ $\frac{\alpha}{2} = 0.05$ $Z_{0.05} = 1.645$

$$P : 0.2 \pm (1.645)\sqrt{\frac{0.2(0.8)}{100}} = \begin{cases} 0.2658 \\ 0.1342 \end{cases}$$

SUBJECT: جلسه سوم
Year(89)Month(۱۱)Date(15)

برآورد فاصله واریانس و انحراف معیار :

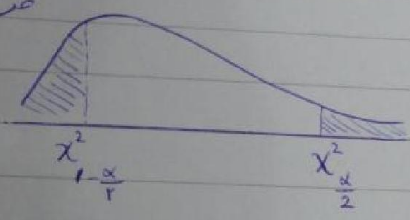
همانطور که می‌دانیم $\chi^2 \sim \frac{(n-1)S^2}{\delta^2}$ دارای توزیع کای مربع ، خی دو ، کای دو ، کای اسکوئر ، مجذور کای

کای مربع

با درجه آزادی $(n-1)$ است که این توزیع از مجذور نرمال‌ها به توان ۲ بدست می‌آید. منحنی آن چوله به راست است. که میانگین آن برابر درجه آزادی و واریانس آن دو برابر درجه آزادی است.

$$\begin{cases} \mu = df \\ \delta^2 = 2df \end{cases}$$

ص ۱۴۴



کل منحنی مساحت 1 دارد و چون متقارن نیست پس

$$\star\ \delta_L^2 = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},df}}$$ «حد پایین واریانس»

$$\star\ \delta_u^2 = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},df}}$$ «حد بالای واریانس»

برآورد فاصله‌ای واریانس: $\delta_L^2 < \delta^2 < \delta_u^2$

برآورد فاصله‌ای انحراف معیار: $\sqrt{\delta_L^2} < \delta < \sqrt{\delta_u^2}$

مثال) در یک نمونه ۱۰ تایی ، واریانس نمونه 40 و میانگین نمونه 50 می‌باشد. یک برآورد فاصله‌ای ۹۰٪ برای واریانس واقعی جامعه بنویسید.

$n=10 \qquad S^2=40 \qquad \bar{X}=50 \qquad \alpha = 1-P = 1-0.9 = 0.1$

$\frac{\alpha}{2} = 0.05 \qquad 1-\frac{\alpha}{2} = 0.95$

$\chi^2_{0.05,9} = 16.919 \qquad \chi^2_{0.95,9} = 3.325$

$$\delta_u^2 = \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},df}} = \frac{9\times 40}{3.325} = 108.27$$

$$\delta_L^2 = \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},df}} = \frac{360}{16.919} = 21.27$$

$\delta_L^2 < \delta^2 < \delta_u^2$

$$\boxed{4.61 < \delta < 10.40}$$

★ برآورد فاصله‌ای اختلاف میانگین ۲ جامعهٔ نرمال مستقل وقتی واریانس معلوم باشد:

- زمانی صورت بدون توجه به حجم نمونه از توزیع Z یا نرمال استفاده می‌کنیم.

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$Z_{0.025} = 1.96$$

$$Z_{0.05} = 1.645$$

$$\mu_1 - \mu_2 : \bar{X}_1 - \bar{X}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

مثال) از دو جامعهٔ A و B نمونه‌های زیر بدست آمده است. یک برآورد فاصله‌ای ٪۹۰ برای اختلاف میانگین دو جامعه بنویسید و توضیح دهید اختلاف برابر عامل تصادفی است.

A	B
$\bar{x}_1 = 100$	$\bar{x}_2 = 95$
$n_1 = 10$	$n_2 = 15$
$\sigma_1^2 = 30$	$\sigma_2^2 = 80$

$$\alpha = 1 - P = 0.1 \qquad \frac{\alpha}{2} = 0.05 \qquad 1 - \frac{\alpha}{2} = 0.95$$

$$Z_{\frac{\alpha}{2}} = 1.645$$

$$\mu_1 - \mu_2 : (100 - 95) \pm 1.645 \sqrt{3 + \frac{80}{15}} \left\langle \begin{array}{l} +9.93 \\ +0.07 \end{array} \right.$$

نکته) در صورتیکه در بازهٔ بدست آمده صفر وجود داشته باشد یعنی یکی مثبت و دیگری منفی باشد اختلاف برابر عامل تصادفی است. ولی اگر بازهٔ بدست آمده هر دو هم علامت باشد اختلاف واقعی است.

- + و − ⟸ عامل تصادفی
- هم علامت ⟸ اختلاف واقعی

• برآورد فاصله‌ای اختلاف میانگین دو جامعه مستقل (۲ وقتی واریانس جامعه مجهول باشد:

همانگونه که می‌دانیم از توزیع Z یا t بسته به حجم نمونه استفاده می‌کنیم.

الف) حجم نمونه‌ها بیشتر از 30 باشد:
در این صورت طبق قضیه حد مرکزی از توزیع Z استفاده می‌شود. (به جای σ، S می‌نویسیم)

$$\mu_1 - \mu_2 : \bar{X}_1 - \bar{X}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

مثال) برای اختلاف میانگین دو جامعه زیر یک برآورد فاصله‌ای 95% بنویسید و توضیح دهید آیا اختلاف بر اثر عامل تصادفی است؟

A	B
$\bar{X}_1 = 50$	$\bar{X}_2 = 55$
$n_1 = 40$	$n_2 = 50$
$S_1^2 = 200$	$S_2^2 = 200$

$\alpha = 1 - P = 0.05 \qquad \frac{\alpha}{2} = 0.025$

$$\mu_1 - \mu_2 = 5 \pm 1.96 \sqrt{\underbrace{\frac{200}{40}}_{5} + \underbrace{\frac{200}{50}}_{4}}_{=9} = \begin{cases} 10.88 \\ -0.88 \end{cases}$$

* صفر درون بازه است پس اختلاف براساس عامل تصادفی است با احتمال 95%

مثال) معتبر است سعی کنید به گونه‌ای این اختلاف را بدست آورید که نتیجه ($\bar{X}_1 , \bar{X}_2$) مثبت باشد. (تفاوت مطلق)

ب) حجم نمونه‌ها کمتر یا برابر 30 باشد:

در این صورت از توزیع t با درجه آزادی $df = n_1 + n_2 - 2$ استفاده می‌شود که با فرض برابری واریانس‌ها به ترتیب زیر عمل می‌کنیم.

واریانس ادغام شده
یا
واریانس مشترک

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

$$\mu_1 - \mu_2 = \begin{cases} \bar{X}_1 - \bar{X}_2 \pm t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} \\ \bar{X}_1 - \bar{X}_2 \pm t_{\frac{\alpha}{2}, df} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \end{cases}$$

مثال) برای اختلاف تجارب A و B یک برآورد فاصله‌ای ۹۰٪ بنویسید و توضیح دهید آیا اختلاف بر اساس عامل تصادفی است؟

A	B
$\bar{x}_1 = ۲۵$	$\bar{x}_2 = 28$
$n_1 = 10$	$n_2 = 5$
$S_1^2 = 20$	$S_2^2 = 30$

$n \leq 30 \Rightarrow$ توزیع t

$\alpha = 1 - P = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$

$d_f = n_1 + n_2 - 2 = 13$

$\boxed{t_{0.05} = 1.771}$

$$\mu_1 - \mu_2 = 3 \pm 1.771 \times 4.8\sqrt{\frac{1}{10}+\frac{1}{5}} = 3 \pm 4.65 \begin{cases} +7.65 \\ -1.65 \end{cases}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{13} = \frac{9\times 20 + 4\times 30}{13} = \frac{180+120}{13} = \frac{300}{13} = 23.07$$

$\boxed{S_p = 4.8}$

* چون عدد صفر داخل بازه قرار دارد اختلاف بر اثر عامل تصادفی است.

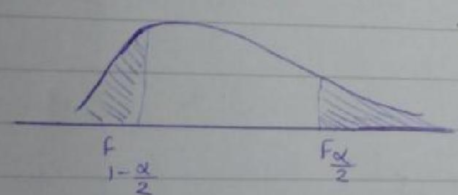
برآورد فاصله‌ای نسبت دو واریانس:

می‌دانیم نسبت دو واریانس دارای توزیع F یا فیشر با درجات آزادی v_1 و v_2 (نو) می‌باشد

$v_1 = n_1 - 1$

$v_2 = n_2 - 1$

که منحنی آن یک منحنی کشیده به راست می‌باشد.



$$\star\ \frac{1}{F_{\frac{\alpha}{2}, v_1, v_2}} \cdot \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2}, v_1, v_2}}$$

نکته

$$F_{1-\alpha, v_1, v_2} = \frac{1}{F_{\alpha, v_2, v_1}}$$

صد در صد سوال امتحانی است

STAEDTLER

$$\star \rightarrow \frac{1}{F_{\frac{\alpha}{2}, v_1, v_2}} \cdot \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \cdot F_{\frac{\alpha}{2}, v_2, v_1}$$

SUBJECT:
Year( )Month( )Date( )

عدد دوم v_2 \ v_1 عدد اول	7	10
7		3.6362
10	3.1355	

جدول فیشر

$F_{0.05, 10, 7} = 145$ جدول صفر

$$F_{0.95, 10, 7} = \frac{1}{F_{0.05, 7, 10}} = \frac{1}{3.1315}$$

مثال) از دو جامعهٔ A و B نمونه‌های زیر بدست آمده است. یک برآورد فاصله‌ای ٪90 برای نسبت دو واریانس بنویسید

A	B
$\bar{x}_1 = 50$	$\bar{x}_2 = 60$
$n_1 = 7$	$n_2 = 9$
$S_1^2 = 100$	$S_2^2 = 80$

$\alpha = 1 - P = 0.1 \quad \frac{\alpha}{2} = 0.05$ و توضیح بدهید آیا دو واریانس با یکدیگر برابر است.

$v_1 = n_1 - 1 = 6$

$v_2 = 8$

$$\frac{1}{F_{\frac{\alpha}{2}, v_1, v_2}} \cdot \frac{S_1^2}{S_2^2} < \frac{\delta_1^2}{\delta_2^2} < \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2}, v_1, v_2}}$$

$$F_{0.05, 6, 8} = 3.5806$$

$$F_{0.95, 6, 8} = \frac{1}{F_{0.05, 8, 6}} = \frac{1}{4.1488}$$

$$\rightarrow \frac{1}{3.5806} \cdot \frac{100}{80} < \frac{\delta_1^2}{\delta_2^2} < \frac{100}{80} \cdot 4.1468$$

$$\boxed{0.349 < \frac{\delta_1^2}{\delta_2^2} < 5.187}$$

نکته) در صورتیکه در بازهٔ بدست آمده عدد 1 وجود داشته باشد نتیجه می‌گیریم دو واریانس با احتمال P با یکدیگر برابرند

بنابراین در این مسأله با احتمال ٪90 دو واریانس با یکدیگر برابر هستند.

آزمون فرض آماری :

$$\begin{cases} H_0 : \text{«آنچه از قبل وجود دارد»} : \text{«بی‌گناه»} \rightarrow \text{آزاد} \\ H_1 : \text{«ادعا»} \quad \text{«قاتل»} \rightarrow \text{اعدام} \end{cases}$$

خطای نوع اول : $\alpha = P\left(\frac{\text{رد } H_0}{\text{اگر } H_0 \text{ درست باشد}}\right)$ — مانند : $\alpha = P\left(\frac{\text{قاتل}}{\text{بی‌گناه}}\right)$

خطای نوع دوم : $\beta = P\left(\frac{\text{پذیرش } H_0}{\text{اگر } H_0 \text{ نادرست باشد}}\right)$ — مانند : $\beta = P\left(\frac{\text{بی‌گناه}}{\text{قاتل}}\right)$

* هرگاه بخواهیم براساس اطلاعات قبلی تصمیم‌گیری کنیم از «آزمون فرض آماری» استفاده می‌کنیم. ادعا را با H_1 و نقیض آن را با حرف H_0 نشان می‌دهیم. در تصمیم‌گیری‌های آماری دو نوع خطا امکان وقوع دارد.

۱- خطای نوع اول (α) ← مثل اینکه فرد بی‌گناه را گناهکار تشخیص داده و اعدام کنیم در جای بازرسی [illegible]

بنابراین خطای نوع اول خطرناک است.

۲- خطای نوع دوم (β) ← مثل اینکه فرد گناهکار را بی‌گناه تشخیص دهیم و آزاد کنیم.

- توان آزمون :

احتمال رد H_0 درصورتیکه H_0 نادرست باشد. در واقع توان آزمون متمم خطای نوع دوم است

یعنی :

$$\pi = P\left(\frac{\text{رد } H_0}{\text{اگر } H_0 \text{ نادرست}}\right) \rightarrow \boxed{\pi = 1-\beta}$$

مثال) درصورتیکه خطای نوع دوم 0.2 باشد توان آزمون را بدست آورید.

$\pi = 1-\beta = 0.8$

21

SUBJECT:
Year( )Month( )Date( )

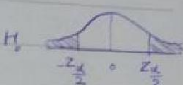
★ آزمون فرض میانگین با مقدار ثابت وقتی واریانس جامعه معلوم باشد:

برای صورت آزمون از Z استفاده می‌شود.

1) آزمون دوطرفه $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

$|Z| > Z_{\frac{\alpha}{2}} \Rightarrow$ H_0 رد می‌شود

یا $Z > Z_{\frac{\alpha}{2}}$ یا $Z < Z_{\frac{\alpha}{2}}$ H_0 رد می‌شود



2) آزمون یک طرفه چپ

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

«کمتر یا حداکثر»

$Z < -Z_{\alpha}$
H_0 رد می‌شود

3) آزمون یک طرفه راست

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

«بیشتر - حداقل»

$Z > Z_{\alpha}$
H_0 رد می‌شود

آماره آزمون (محاسبه شده) $\Leftarrow$ $Z = \dfrac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

از یک جامعه نرمال با واریانس 64 یک نمونه 25 تایی انتخاب می‌کنیم که میانگین نمونه 54 است. آیا در سطح خطای 0.05 (α) می‌توان پذیرفت که میانگین کمتر از 55 است. ابتدا با مشخص نوع این آزمون از بین 3 تای بالایی است.

حل: $\begin{cases} H_0 : \mu \geq 55 \\ H_1 : \mu < 55 \end{cases}$

$\begin{cases} \delta^2 = 64 \\ n = 25 \\ \bar{X} = 54 \end{cases}$

$Z = \dfrac{54 - 55}{\frac{8}{5}} = -0.625$

$Z < -Z_{\alpha}$ H_0 رد می‌شود

$Z_{0.05} = 1.645 \Rightarrow -0.625 \not< -1.645$

H_0 پذیرفته می‌شود یعنی با احتمال 95% میانگین بزرگتر مساوی 55 است.

$Z_{0.05}$

STAEDTLER

علامت ناحیه رد با H_1 یکی است.

* آزمون فرض میانگین با معیار ثابت وقتی واریانس جامعه مجهول باشد

در این صورت با توجه به حجم نمونه از توزیع Z یا t استفاده می‌شود

* الف) حجم نمونه کمتر از 30 باشد:

در این صورت از توزیع t با درجه آزادی $n-1$ استفاده می‌شود

$df = n-1$

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad |t| > t_{\frac{\alpha}{2}, df} \rightarrow H_0 \text{ رد می‌شود}$$

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$ علامت کوچکتر $\qquad t < -t_{\alpha, df} \Rightarrow H_0$ رد می‌شود

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$ علامت بیشتر $\qquad t > t_{\alpha, df} \Rightarrow H_0$ رد می‌شود

مثال) در یک نمونه 16 تایی میانگین نمونه 25 و واریانس نمونه 64 می‌باشد. در سطح خطای 5٪ آزمون کنید آیا می‌توان این ادعا را پذیرفت که میانگین 20 است.

$n = 16 \qquad \mu_0 =$

$$\begin{cases} H_0 \leq 20 \\ H_1 > 20 \end{cases}$$ ادعا

$$t = \frac{25-20}{\frac{8}{\sqrt{16}}} = 2.5$$

$\alpha = 0.05 \qquad df = n-1 = 15$

$t > t_{\alpha, df} \qquad 2.5 \ngtr 2.947$

H_0 رد می‌شود

H_0 پذیرفته است $\rightarrow$ یعنی با احتمال 95٪ ادعای میانگین جامعه 20 است

1.753

* ب) حجم نمونه بیشتر از 30 باشد: در این صورت از توزیع Z استفاده می‌کنیم.

$$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad |Z| > Z_{\frac{\alpha}{2}} \rightarrow H_0 \text{ رد می‌شود}$$

SUBJECT:
Year( )Month( )Date( )

$$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad Z > Z_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

$$\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad Z < -Z_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

مثال) در یک نمونه 36 تایی میانگین نمونه 50 و واریانس 81 است. در سطح خطای 5% آزمون کنید آیا میانگین جامعه 52 است؟

$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{50 - 52}{\frac{\sqrt{81}}{\sqrt{36}}} = \frac{-12}{9} = -1.33$$

«آماره آزمون»

$$\begin{cases} H_0: \mu = 52 \\ H_1: \mu \neq 52 \end{cases}$$

$Z_{0.025} = 1.96$

$Z_{0.05} = 1.645$

اگر $|Z| > Z_{\frac{\alpha}{2}} \Rightarrow H_0$ رد می‌شود

$|-1.33| \ngtr 1.96$

H_0 پذیرفته می‌شود یعنی با احتمال 95% میانگین جامعه 52 است.

* آزمون فرض نسبت‌ها :

$$\begin{cases} H_0: P = P_0 \\ H_1: P \neq P_0 \end{cases} \quad |Z| > Z_{\frac{\alpha}{2}} \Rightarrow H_0 \text{ رد می‌شود}$$

$$\begin{cases} H_0: P \geq P_0 \\ H_1: P < P_0 \end{cases} \quad Z < -Z_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

حالت کمتر

$$\begin{cases} H_0: P \leq P_0 \\ H_1: P > P_0 \end{cases} \quad Z > Z_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

حالت بیشتر

$\bar{P} = \frac{x}{n}$ ← x: تعداد حالات مساعد، n: کل حالات

$q_0 = 1 - P_0$

$$Z = \frac{\bar{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$$

ریشه = انحراف معیار

مثال) از بین ۱۰۰ کالا ۲۰ کالا معیوب است آیا در سطح خطای 5% می‌توان پذیرفت حداکثر کالاهای معیوب 25% است؟

$\bar{P} = \frac{X}{n} = \frac{20}{100} = 0.2 \qquad q_0 = 1 - P_0 = 1 - 0.25 = 0.75$

آماره آزمون: $Z = \frac{\bar{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.2 - 0.25}{\sqrt{\frac{0.25 \times 0.75}{100}}} = \frac{-0.05}{0.0433} = -1.15$

$$\begin{cases} H_0: P \geq 0.25 \\ H_1: P < 0.25 \end{cases}$$

$Z < -Z_\alpha \Rightarrow$ H_0 رد می‌شود

$-1.15 \not< -Z_{0.05} = -1.645$

H_0 پذیرفته می‌شود با احتمال 95% نسبت کالاهای معیوب برابر 25% است.

* آزمون فرض واریانس و انحراف معیار:

همانطور که می‌دانیم $\frac{(n-1)S^2}{\delta_0^2}$ دارای توزیع کای دو با درجه آزادی $n-1$ است. بنابراین از توزیع کای دو χ^2 استفاده می‌شود.

آماره آزمون: $\chi^2 = \frac{(n-1)S^2}{\delta_0^2}$

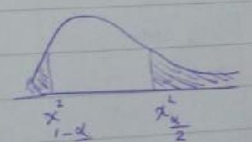
(اگر انحراف معیار بود باید به واریانس تبدیل شود)

$$\begin{cases} H_0: \delta^2 = \delta_0^2 \\ H_1: \delta^2 \neq \delta_0^2 \end{cases}$$

اگر $\chi^2 > \chi^2_{\frac{\alpha}{2}, df}$ یا $\chi^2 < \chi^2_{1-\frac{\alpha}{2}, df} \Rightarrow$ H_0 رد می‌شود

$$\begin{cases} H_0: \delta^2 \geq \delta_0^2 \\ H_1: \delta^2 < \delta_0^2 \end{cases}$$

$\chi^2 < \chi^2_{1-\alpha, df} \Rightarrow$ H_0 رد می‌شود



$$\begin{cases} H_0: \delta^2 \leq \delta_0^2 \\ H_1: \delta^2 > \delta_0^2 \end{cases}$$

$\chi^2 > \chi^2_{\alpha, df} \Rightarrow$ " "

مثال) در یک نمونه ۱۰ تایی واریانس نمونه 81 و میانگین نمونه 25 است در سطح خطای 5% آزمون کنید آیا حداکثر ریشه 8 است.

انحراف معیار $= 8 \Rightarrow$ واریانس $= 64$

$$\begin{cases} H_0: \delta^2 \leq 64 \\ H_1: \delta^2 > 64 \end{cases}$$

$\chi^2 > \chi^2_\alpha$ رد H_0

$\chi^2 = \frac{9(81)}{64} = 11.39$

$d_f = n - 1 = 9$

$\chi^2_{0.05, 9} = 16.919$

$\rightarrow 11.39 \not> 16.919 \Rightarrow$ با احتمال 95% واریانس جامعه کمتر مساوی 64 است یا ریشه کمتر از 8 است.

SUBJECT:
Year: Month: Date:

* آزمون فرض اختلاف میانگین دو جامعه وقتی واریانس ها معلوم باشد و دو جامعه مستقل باشد.

(1) [illegible] بدون توجه به حجم نمونه از توزیع Z استفاده می شود.

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases} \qquad \begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \qquad |Z| > Z_{\frac{\alpha}{2}} \Rightarrow H_0 \text{ رد می شود}$$

$$\begin{cases} H_0: \mu_1 - \mu_2 \geq 0 \\ H_1: \mu_1 - \mu_2 < 0 \end{cases} \qquad \begin{cases} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases} \qquad Z < -Z_{\alpha} \Rightarrow H_0 \text{ رد می شود}$$

$$\begin{cases} H_0: \mu_1 - \mu_2 \leq 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{cases} \qquad \begin{cases} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases} \qquad Z > Z_{\alpha} \Rightarrow H_0 \text{ رد می شود}$$

مستقل اند ⇐ 0

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

مثال) از دو جامعه A و B نمونه های زیر بدست آمده است. در سطح خطای 5% آزمون کنید آیا میانگین دو جامعه برابر است.
چون σ^2 است یعنی واریانس معلوم است.

A	B
$\bar{x}_1 = 100$	$\bar{x}_2 = 105$
$n_1 = 10$	$n_2 = 20$
$\sigma_1^2 = 50$	$\sigma_2^2 = 80$

$$Z = \frac{105 - 100}{\sqrt{\frac{50}{10} + \frac{80}{20}}} = \frac{5}{\sqrt{5+4}} = \frac{5}{3} = 1.66 \qquad Z_{0.025} = 1.96$$

$$|Z| > Z_{\frac{\alpha}{2}} \qquad 1.66 \ngtr 1.96 \Rightarrow$$ با احتمال [illegible] میانگین دو جامعه برابر است.

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

آزمون فرض اختلاف میانگین دو جامعه مستقل نرمال وقتی واریانس مجهول باشد.

حالتی که می‌دانیم ۲ حالت امکان وقوع دارد:

با استفاده از واریانس نمونه

A) حجم نمونه‌ها بیشتر از 30 باشد: در این صورت از توزیع Z استفاده می‌شود.

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

مثال) در سطح خطای 10% آزمون کنید میانگین دو جامعه A و B با یکدیگر برابر است؟

A	B
$\bar{x}_1 = 52$	$\bar{x}_2 = 56$
$n_1 = 40$	$n_2 = 50$
$S_1^2 = 200$	$S_2^2 = 200$

چون S داریم پس واریانس جامعه مجهول است.

نمونه‌ها $> 30 \Rightarrow Z$

آماره آزمون

$$Z = \frac{4}{\sqrt{\frac{200}{40} + \frac{200}{50}}} = \frac{4}{3} = 1/33$$

$|Z| > Z_{\frac{\alpha}{2}}$ H_0 پذیرفته

$1.33 \not> 1.645 \rightarrow$ H_0 پذیرفته می‌شود و با احتمال 90% میانگین دو جامعه برابر است.

B) حجم نمونه کمتر از 30 باشد: در این صورت از توزیع t با درجه آزادی $(n_1 + n_2 - 2)$ استفاده می‌کنیم.

واریانس ادغام شده: $S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$ $\qquad S_p = \sqrt{S_p^2}$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)}$$

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \qquad |t| > t_{\frac{\alpha}{2}, df} \Rightarrow H_0 \text{ رد می‌شود}$$

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \qquad t < -t_{\alpha, df} \Rightarrow \text{"}$$

$$\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \qquad t > t_{\alpha, df} \Rightarrow \text{"}$$

SUBJECT:
Year( )Month( )Date( )

$416 \,|\, 16$, 32, 96, 26

$H_1 = $ ادعا

مثال) از دو جامعه A و B نمونه‌های زیر بدست آمده است. در سطح خطای 5% آزمون کنید آیا میانگین جامعه A بزرگتر از جامعه B است؟ نمونه کمتر از 30 ← توزیع t

A	B
$\bar{x}_1 = 85$	$\bar{x}_2 = 79$
$n_1 = 10$	$n_2 = 8$
$S_1^2 = 40$	$S_2^2 = 50$

$d_f = 16$

$H_1 : \mu_1 > \mu_2$

$$S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sqrt{(n_1+n_2-2)^2}} = \frac{9 \times 40 + 7 \times 50}{16} = \frac{360 + 350}{16} = 44/37$$

$$S_p = 6.661$$

$$t = \frac{6}{6.661 \times \sqrt{\frac{1}{10} + \frac{1}{8}}} \simeq 4.05 \qquad \alpha = 0.05$$

$t_{0.05,16} = 1.746$ ناحیه بحرانی

$t > t_{\alpha, df} \Rightarrow 4.05 > 1.746 \Rightarrow$ H_0 رد می‌شود. با احتمال 95% میانگین جامعه A از جامعه B بزرگتر است.

1 مقایسه‌های زوجی:

هرگاه افراد مورد آزمایش ثابت باشند و حالت قبل و بعد یا آزمایشگاه 1 و 2 در صورت سوال مطرح شود در این صورت از مقایسه‌های زوجی استفاده می‌کنیم با درجه آزادی $(n-1)$ که برای انجام این آزمون اختلاف دو حالت را بدست می‌آوریم سپس میانگین، واریانس و انحراف معیار را تعیین می‌کنیم و بر اساس آن آزمون فرض را انجام می‌دهیم.

$d = x - y$

$$\bar{d} = \frac{\sum d_i}{n}$$

$$S_d^2 = \frac{\sum d_i^2 - n\bar{d}^2}{n-1}$$

$$t = \frac{\bar{d} - 0}{\frac{S_d}{\sqrt{n}}}$$

$\begin{cases} H_0 : \bar{d} = 0 \\ H_1 : \bar{d} \neq 0 \end{cases}$ $|t| > t_{\frac{\alpha}{2}, df} \rightarrow$ H_0 رد می‌شود

$\begin{cases} H_0 : \bar{d} \geq 0 \\ H_1 : \bar{d} < 0 \end{cases}$ $t < -t_{\alpha, df} \rightarrow$ 〃

$\begin{cases} H_0 : \bar{d} \leq 0 \\ H_1 : \bar{d} > 0 \end{cases}$ $t > t_{\alpha, df} \rightarrow$ 〃

مثال) نمرات ۷ نفر در درس آمار قبل از شرکت در کلاس و بعد از شرکت در کلاس ثبت شده است. در سطح خطای 5% آزمون کنید آیا کلاس مؤثر است؟

x قبل	8	12	10	16	15	19	20
y بعد	10	13	9	15	18	17	19
$d = y - x$	2	1	-1	-1	3	-2	-1

$$\bar{d} = \frac{2+1-1-1+3-2-1}{7} = \frac{1}{7}$$

$$S_d^2 = \frac{\sum d_i^2 - n\bar{d}^2}{n-1} = \frac{(21 - 7(\frac{1}{7})^2)}{6} = \frac{20.86}{6} = 3.47 \qquad S_D = \sqrt{3.47} = 1.86$$

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{.14}{(1.86/\sqrt{7})} = 0.199 \qquad t_{0.05,6} = 1.943$$

ناحیه رد: $t > t_{\alpha, df} \Rightarrow 0.199 \not> 1.943 \Rightarrow$ با احتمال 95% کلاس مؤثر نیست.

۲۹ آذر ۸۹

- آزمون فرض واریانس با مقدار ثابت:

$$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases} \qquad \chi^2$$

$$\begin{cases} H_0: \sigma^2 \geq \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases}$$

$$\begin{cases} H_0: \sigma^2 \leq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases}$$

همانطور که می‌دانیم برای آزمون فرض واریانس و انحراف معیار از توزیع کای دو استفاده می‌شود.

SUBJECT:
Year( )Month( )Date( )

آماره آزمون: $F = \dfrac{S_1^2}{S_2^2}$

آزمون فرض نسبت دو واریانس:

همانطور که می دانیم نسبت واریانس ها از توزیع F یا فیشر تبعیت می کند با درجه آزادی ν_1 و ν_2

$$\begin{cases} H_0: \dfrac{\delta_1^2}{\delta_2^2} = 1 \\ H_1: \delta_1^2/\delta_2^2 \neq 1 \end{cases}$$

$F > F_{\frac{\alpha}{2}, \nu_1, \nu_2}$ یا $F < F_{1-\frac{\alpha}{2}, \nu_1, \nu_2} = \dfrac{1}{F_{\frac{\alpha}{2}, \nu_2, \nu_1}}$ $\Rightarrow$ H_0 رد می شود

$$\begin{cases} H_0: \dfrac{\delta_1^2}{\delta_2^2} \geqslant 1 \\ H_1: \dfrac{\delta_1^2}{\delta_2^2} < 1 \end{cases}$$

$F < F_{1-\alpha, \nu_1, \nu_2} = \dfrac{1}{F_{\alpha, \nu_2, \nu_1}}$ $\Rightarrow$ H_0 رد می شود

$$\begin{cases} H_0: \dfrac{\delta_1^2}{\delta_2^2} \leqslant 1 \\ H_1: \dfrac{\delta_1^2}{\delta_2^2} > 1 \end{cases}$$

$F > F_{\alpha, \nu_1, \nu_2}$ $\Rightarrow$ H_0 رد می شود

$(\nu_2 = n_2 - 1)$ و $(\nu_1 = n_1 - 1)$

مثال) از دو جامعه A و B نمونه های زیر بدست آمده است. در سطح خطای 10% آزمون کنید آیا واریانس دو جامعه برابر است؟

A	B
$\bar{x}_1 = 100$	$\bar{x}_2 = 105$
$n_1 = 8$	$n_2 = 5$
$S_1^2 = 40$	$S_2^2 = 25$

$H_1 = \dfrac{\delta_1^2}{\delta_2^2} \neq 1$

$F = \dfrac{40}{25} = 1.6$ $\quad \nu_1 = 7 \quad \nu_2 = 4$

$\dfrac{\alpha}{2} = 0.05$ $\quad F_{0.05, 7, 4} = 6.0942$

$F_{1-\frac{\alpha}{2}, \nu_1, \nu_2} = \dfrac{1}{F_{\frac{\alpha}{2}, 4, 7}} = \dfrac{1}{4.1203} = 0.24$

$\left\{ 1.6 \not< 6.0942 \right.$, $\left. 1.6 \not< 0.24 \right\}$ H_0 تایید می شود یعنی با احتمال 90% واریانس دو جامعه برابر است.

تعریف ضریب همبستگی: ضریب همبستگی معیاری است برای تعیین ارتباط بین دو متغیر مستقل (X) و وابسته (Y). ضریب همبستگی عددی بین (+1 , -1) است و بدون واحد است. در صورتیکه ضریب همبستگی منفی باشد یعنی رابطه دو متغیر معکوس است مانند رابطه فشار هوا با ارتفاع از زمین و در صورتیکه ضریب همبستگی مثبت باشد رابطه دو متغیر مستقیم است.

$|r| \leq 1 \Rightarrow -1 \leq r \leq 1$

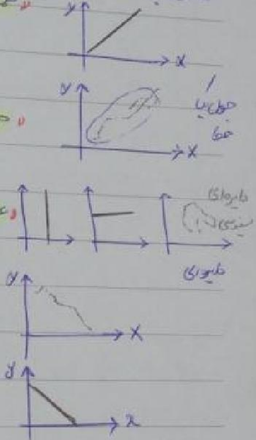
1) $r = +1$: همبستگی مستقیم کامل — مانند رابطه قد و وزن

2) $0 < r < 1$: همبستگی مستقیم

3) $r = 0$: عدم همبستگی خطی

4) $-1 < r < 0$: همبستگی معکوس

5) $r = -1$: همبستگی معکوس کامل



نکته: به مجذور ضریب همبستگی، ضریب تعیین گویند که نشان دهنده آن است که چند درصد تغییرات y توسط x بیان می‌شود.

ضریب تعیین $= R^2$

* چگونگی محاسبه ضریب همبستگی پیرسون:

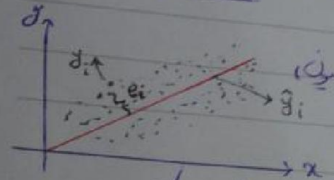
$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum xy - n\bar{x}\bar{y} = \sum xy - \frac{\sum x \sum y}{n}$$

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

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y^2 - n\bar{y}^2 = \sum y^2 - \frac{(\sum y)^2}{n}$$

معادله خط رگرسیون: در رگرسیون هدف تعیین معادله خطی است که کمترین فاصله را با نقاط دیاگرام پراکنش داشته باشد.
منظور از رگرسیون بازگشتی باشد.



دیاگرام پراکنش

e

خطا $e_i = y_i - \hat{y}_i$

$e_i \sim N(0, \delta_e^2)$

y_i ← مقدار واقعی
$\hat{y}_i$ ← مقدار برآورد شده

$$\delta_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

شیب خط رگرسیون ۱) $\hat{b} = \frac{S_{xy}}{S_{xx}}$

عرض از مبدأ ۲) $\hat{a} = \bar{y} - \hat{b}\bar{x}$

معادله خط رگرسیون ۳) $\hat{y} = \hat{a} + \hat{b}x$

مقدار پیش بینی ۴) $y^* = \hat{a} + \hat{b}x^*$

مجموع مربعات خطا ۵) $SS_E = S_{yy} - b^2 S_{xx} = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$

واریانس خطا ۶) $\delta_e^2 = \frac{SS_E}{n-2}$

مثال) برای دو متغیر مستقل x و وابسته y تعیین کنید: ۱) ضریب همبستگی ۲) ضریب تعیین
۳) معادله خط رگرسیون ۴) درصورتیکه مقدار x برابر 6 باشد مقدار y را برآورد کنید. ۵) مجموع مربعات
خطا و واریانس خطا

	x	y	
۴	۲	۱	۱
۱۶	۴	۳	۹
۲۵	۵	۴	۱۶
۴۹	۷	۸	۴۲
۸۱	۹	۶	۳۶
۱۷۵	۲۷	۲۲	

$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = ۰٫۸۶$

$S_{xy} = \sum xy - n\bar{x}\bar{y} = ۱۲۴ - ۵(۵.۴)(۴.۴) = 25.2$

$\bar{x} = \frac{\sum x_i}{n} = \frac{۲۷}{۵} = ۵٫۴$

$\bar{y} = \frac{۲۲}{۵} = ۴٫۴$

STAEDTLER

SUBJECT:
Year/ Month/ Date/

$$S_{xx} = \sum x^2 - n\bar{x}^2 = ۱۷۵ - ۵(۵.۴)^۲ = \boxed{۲۹.۲}$$
$$S_{yy} = \sum y^2 - n\bar{y}^2 = ۱۲۶ - ۵(۴.۴)^۲ = \boxed{۲۹.۲}$$

۲) ضریب تعیین: $R^2 = (۰٫۸۶)^۲ = \boxed{۰٫۷۴۴}$

$$۳)\ \hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{۲۵٫۲}{۲۹٫۲} = \boxed{۰٫۸۶۳}$$

$$۴)\ \hat{a} = \bar{y} - \hat{b}\bar{x} = ۴٫۴ - (۰٫۸۶۳)(۵٫۴) = \boxed{-۰٫۲۶}$$

$$۵)\ \hat{y} = \hat{a} + \hat{b}x \Rightarrow \hat{y} = (-۰٫۲۶) + (۰٫۸۶۳)x$$

$$۶)\ x^* = ۶ \Rightarrow y^* = (-۰٫۲۶) + (۰٫۸۶۳)(۶) = ۴٫۹۱$$

$$۷)\ SS_E = S_{yy} - b^2 S_{xx} = ۲۹٫۲ - (۰٫۸۶۳)^۲(۲۹٫۲) = \boxed{۷٫۴۵}$$

$$۸)\ s_e^2 = \frac{SS_E}{n-2} = \frac{۷٫۴۵}{۵-۲} = \boxed{۲٫۴۸}$$

توزیع شیب خط:

شیب خط دارای توزیع نرمال $b \sim N(0, \frac{s_e^2}{S_{xx}})$ است.

$$\begin{cases} \mu = \bar{x} \pm t_{\frac{\alpha}{2}, df} \frac{s}{\sqrt{n}} \\ df = n-1 \end{cases}$$

برآورد فاصله‌ای شیب خط:

$$B: \hat{b} \pm t_{\frac{\alpha}{2}, df} \frac{s_e}{\sqrt{S_{xx}}} \qquad df: n-2$$

مثال) در مسأله قبل یک برآورد فاصله‌ای ۹۵٪ برای شیب خط رگرسیون بنویسید و توضیح دهید آیا شیب خط صفر است؟

$$n = 5 \Rightarrow df = 3 \qquad \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow t_{0.025, 3} = ۳٫۱۸۹$$

$$s_e^2 = ۲٫۴۸ \Rightarrow s_e = ۱٫۵۷ \qquad \Rightarrow B = ۰٫۸۶۳ \pm (۳٫۱۸۹)\frac{۱٫۵۷}{\sqrt{۲۹٫۲}} = \begin{cases} ۱٫۸۵ \\ -۰٫۱۱ \end{cases}$$

۰٫۹۲۹   ۱٫۰۷۸

چون صفر در این بازه وجود دارد پس شیب خط با احتمال ۹۵٪ برابر صفر است.

STAEDTLER

$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \quad (n<30) \quad df = n-1$

SUBJECT:
Year(Month(Date( )

روش دیگر: آزمون فرض شیب خط رگرسیون:

آماره آزمون: $t = \frac{b}{\frac{Se}{\sqrt{S_{xx}}}} \sim t \quad df = n-2$

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

$|t| > t_{\frac{\alpha}{2}, df} \Rightarrow$ H_0 رد می‌شود یعنی شیب خط معنی‌دار است.
(مخالف صفر و معنی‌دار)

مثال) در مسأله قبل آماره آزمون شیب خط را بدست آورده و توضیح دهید آیا شیب خط برابر صفر است؟

$$t = \frac{b}{\frac{Se}{\sqrt{S_{xx}}}} = \frac{۰/۸۴}{\frac{\sqrt{۲٫۴۸}}{\sqrt{۲۹٫۲}}} = ۲/۹۶$$

$|t| > t_{\frac{\alpha}{2}, df} \quad ۲٫۹۶ \ngtr ۳٫۱۸۲$

H_0 تایید می‌شود. با احتمال ۹۵٪ شیب برابر صفر است.

آزمون فرض ضریب همبستگی:

آماره آزمون: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$ رد $\Rightarrow$ (جامعه)

$|t| > t_{\frac{\alpha}{2}, df} \Rightarrow H_0$ رد می‌شود

مثال) در مسأله قبل:

$$t = \frac{۰٫۸۴(\sqrt{۵-۲})}{\sqrt{۱-(۰٫۸۴)^۲}} = ۲٫۹۷$$

$|۲٫۹۷| \ngtr ۳٫۱۸۲$

بنابراین H_0 تایید می‌شود یعنی با احتمال ۹۵٪ ضریب همبستگی صفر است.

نکته) نتیجه آزمون شیب خط و ضریب همبستگی همیشه مشابه است.

برآورد فاصله‌ای عرض از مبدأ:

$$A : a \pm t_{\frac{\alpha}{2}, df} \cdot S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

$$A \sim N\left(0, S_e^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

Year(Month(Date( )

$t_{0.025, 3} = 3.189$

مثال) در مسأله قبل یک برآورد فاصله‌ای ٪95 برای عرض از مبدأ بنویسید.

$$A : (-0.26) \pm 3.189 \times \sqrt{1.48}\left(\sqrt{\frac{1}{5} + \frac{(5.4)^2}{29.2}}\right) = -0.26 \pm 5.49$$

$-5.75 \quad 5.23$

شامل صفر می‌شود یعنی با احتمال ٪95 عرض از مبدأ برابر صفر است.

* آزمون فرض عرض از مبدأ :

آماره آزمون : $t = \dfrac{a}{S_e\left(\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right)}$

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

$|t| > t_{\frac{\alpha}{2}, df} \Rightarrow$ H_0 رد می‌شود.

مثال) در مسأله قبل آزمون کنید آیا عرض از مبدأ در سطح خطای ٪5 برابر صفر است؟

$$t = \frac{-0.26}{1.57\sqrt{\frac{1}{5} + \frac{(5.4)^2}{29.2}}} = -0.26 \qquad |-0.26| \not> 3.182$$

H_0 تأیید می‌شود یعنی با احتمال ٪95 عرض از مبدأ برابر صفر است.

نکته) معادله خط رگرسیون از نقطه $\bar{x}$ و $\bar{y}$ عبور می‌کند.

* برآورد فاصله‌ای میانگین پاسخ :

داشتیم : $y^* = a + bx^*$

$$\mu_y = \underset{\text{امید}}{E(y \mid x = x^*)} = y^* \pm t_{\frac{\alpha}{2}} S_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

فرمول مهم

مثال) یک برآورد فاصله‌ای ٪95 برای میانگین پاسخ به ازای $x = 10$ بنویسید.

$$y^* = -7.26 + (0.86)(10) = 8.34$$

$$\mu_y = 8.34 \pm \left(3.182 * 1.57\sqrt{\frac{1}{5} + \frac{(10 - 5.4)^2}{29.2}}\right)$$

STAEDTLER

* برآورد فاصله‌ای تک پاسخی :

$$\Rightarrow y^* \pm t_{\frac{\alpha}{2}, df}\, \delta_e \left(\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} \right)$$

مثال) در مسأله قبل یک برآورد فاصله‌ای بازای $x = 10$ بنویسید. که میانگین نداریم

$$\text{برآورد فاصله‌ای تک مقداری} = 8.34 \pm 3.182 \left(1.57 \sqrt{1 + \frac{1}{5} + \frac{(10 - 5.4)^2}{29.2}} \right)$$

* رابطهٔ ضریب همبستگی و شیب خط :

$$b = r \frac{\delta_y}{\delta_x} \quad \text{or} \quad b = r \sqrt{\frac{S_{yy}}{S_{xx}}}$$

* معادلهٔ خط رگرسیون :

$$\boxed{y - \bar{y} = r \frac{\delta_y}{\delta_x} (x - \bar{x})}$$

مثال) از متغیرهای تصادفی x و y دارای توزیع نرمال دو متغیره باشند همچنین $\mu_x = 40$ و $\mu_y = 80$ $\delta_y^2 = 4\delta_x^2$

چنانچه ضریب همبستگی بین x و y برابر ۱ باشد در این صورت معادلهٔ خط رگرسیون y برحسب x را بنویسید.

$$y - 80 = 2(x - 40) \qquad y - 80 = 2x - 80 \qquad \boxed{y = 2x}$$

نکته) در معادلهٔ خط رگرسیون به اندازهٔ ضریب R^2 (ضریب تعیین) مقادیر y توسط خط رگرسیون تبیین (برآورد) می‌شود

اگر ضریب همبستگی بین دو متغیر x و y برابر ۰/۷ باشد چند درصد از تغییرات y تحت تأثیر x نیست؟

$$R^2 = (0.7)^2 = 0.49 \Rightarrow \text{تحت تأثیر هست}$$

تحت تأثیر نیست: $\Rightarrow 1 - 0.49 = 0.51$

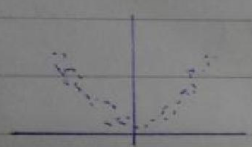
STAEDTLER

نکته) ضریب همبستگی بین دو متغیر x و y همیشه ثابت است. یعنی در هر ضریبی ضرب شود یا مقداری اضافه یا کم شود، ضریب همبستگی تغییر نمی‌کند مگر اینکه علامت آن تغییر کند یعنی یکی از ضرایب مثبت و دیگری منفی باشد.

$$r_{x,y} = \rho$$

$$r(ax+b,\ cy+d) = \rho \left(\frac{|ac|}{ac}\right)$$ ← صرفاً جهت تعیین علامت

نکته) برای تعیین معادلهٔ خط رگرسیون ابتدا نمودار پراکنش را رسم می‌کنیم و براساس آن یک مدل فرضی در نظر می‌گیریم. سپس برای بدست آوردن مدل خطی تغییر متغیر مورد نیاز را انجام می‌دهیم



1) $y = bx^2 + a$ $\qquad X = x^2$

$y = bX + a$ $\Rightarrow$ در حقیقت به جای X ، $\sqrt{x}$ قرار می‌دهیم

2) $y = ce^{Bx}$ $\qquad \ln y = \ln ce^{Bx} \Rightarrow \ln y = \ln c + \ln e^{Bx}$

$$\Rightarrow y^* = c^* + Bx$$

3) $y = A\sin x + B$

$\frac{y}{A} = \sin x + \frac{B}{A} \Rightarrow \sin^{-1}\frac{y}{A} = x + \sin^{-1}\frac{B}{A}$

* رگرسیون چندمتغیره:

هرگاه بخواهیم تأثیر چند متغیر مستقل را بر روی متغیر وابسته نشان دهیم از رگرسیون چندمتغیره استفاده می‌شود.

قیمت نوشابه $y = \underbrace{b_0}_{\text{عرض از مبدا}} + \underbrace{b_1}_{\text{شیب دستمزد}}\underbrace{x_1}_{\text{دستمزد}} + \underbrace{b_2}_{\text{شیب فاصله}}\underbrace{x_2}_{\text{فاصله}} + \dots$

K: تعداد متغیر مستقل

$K+1 = P$

تعداد کل متغیرها

ANOVA

منبع تغییر	SS (مجموع مربعات)	df	MS واریانس	F
رگرسیون	$SS_R = b^2 S_{xx}$	K	$MS_R = \frac{SS_R}{K}$	$\frac{MS_R}{MS_E}$
خطا	SS_E	$N-K-1$ $N-(K+1)$ $N-P$	$MS_E = \frac{SS_E}{N-P}$	
کل	S_{yy}	$N-1$		

اگر $F > F_{\alpha, k, N-P}$ ⇐ H_0 رد می‌شود ⇐ شیب خط خلاف صفر است.

$$SS_y = SS_R + SS_E \qquad SS_E = S_{yy} - b^2 S_{xx} \quad (b^2 S_{xx} = SS_R)$$

ضریب تعیین
$$\boxed{R^2 = \frac{SS_R}{S_{yy}}} = \frac{S_{yy} - SS_E}{S_{yy}}$$

$$\boxed{R^2 = 1 - \frac{SS_E}{S_{yy}}}$$

مثال) ضریب رگرسیون خطی دوگانه معادله برآوردی عبارت است از:

$$\underset{\text{پس‌انداز}}{\hat{y}} = \underbrace{3{,}9365}_{\text{ثابت}} + {,}42\,\underset{\text{درآمد}}{x_1} + \underbrace{41005\,x_2}_{\text{ثابت}}$$

دارای صورت مفهوم $b_1 = {,}42$ آن است که:

الف) همواره ۴۲ر۰ سکه پس‌انداز می‌شود.

ب) به ازای کاهش یک فرزند ۴۲ر۰ درآمد پس‌انداز می‌شود.

ج) به فرض ثابت بودن تعداد فرزندان ۴۲ر۰ سکه پس‌انداز می‌شود.

د) به ازای افزایش یک واحد درآمد ۴۲ر۰ سکه پس‌انداز می‌شود.

y = پس‌انداز خانواده x_1 = درآمد خانواده x_2 = تعداد فرزندان

هر زمان که اثر یک عامل را می‌خواهیم بدانیم باید بقیه عوامل را ثابت در نظر بگیریم

مثال) در جدول تحلیل رگرسیون زیر، ضریب تعیین را بدست آورید و مقدار آماره آزمون را بدست آورید.

منبع تغییر	SS	df	MS واریانس	F
رگرسیون	۴۰	۲	$MS = \frac{SS}{df}$ $\frac{40}{2} = 20$	
خطا	۶۰	۵	$\frac{60}{5} = 12$	$F = \frac{20}{12}$
کل	۱۰۰	۷		

$$R^2 = \frac{SSR}{S_{yy}} = \frac{40}{100} = 0.4$$

$$F = \frac{MS_R}{MS_E} = \frac{20}{12} = 1.67$$

ناحیه بحرانی: اگر $F > F_{\alpha, k, n-p} \Leftarrow F_{0.05, 2, 5}$

تحلیل واریانس

هرگاه بخواهیم میانگین بیش از دو جامعه را با یکدیگر مقایسه کنیم از تحلیل واریانس استفاده می‌کنیم. زیرا اگر بخواهیم K جامعه را از لحاظ میانگین با یکدیگر مقایسه کنیم به اندازه $K\left(\frac{K-1}{2}\right)$ آزمون t باید انجام شود. عواملی که با یکدیگر مقایسه می‌شوند را تیمار گویند.

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_K \\ H_1 : \text{حداقل یکی از } \mu_i \text{ها متفاوت است.} \end{cases}$$

منبع تغییرات	SS	df	MS واریانس	آماره آزمون F
تیمار	SS_{Tr}	$df_{tr} = k-1$	$MS_{Tr} = \frac{SS_{Tr}}{k-1}$	$\frac{MS_{Tr}}{MS_E} = F$
خطا	SS_E	$df_e = N-K = k(n-1)$	$MS_E = \frac{SS_E}{k(n-1)}$	
کل	SS_t	$N-1$		

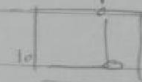
K: تعداد تیمار
n_i: تعداد هر تیمار
N: تعداد کل افراد

$F > F_{\alpha, k-1, k(n-1)} \Rightarrow H_0$ رد می‌شود $\Rightarrow$ با احتمال P میانگین‌ها متفاوت است و تیمار

مثال) تحلیل واریانس زیر را تکمیل کنید و در سطح خطای ۵٪ آزمون کنید آیا میانگین‌ها با یکدیگر برابر است؟

منبع تغییرات	SS	df	MS	F
تیمار	۴۰	۲	$MS_{tr} = \frac{SS_{tr}}{df_{tr}} = 20$	$F = \frac{MS_{tr}}{MS_E} = \frac{20}{16} = 1.25$
خطا	۱۶۰	۱۰	$MS_E = 16$	
کل	۲۰۰	۱۲		

$SS_t = SS_E + SS_{tr}$ $\qquad$ $df_e = df_t + df_{tr}$



اگر $F > F_{0.05,2,10} \Rightarrow$ H_0 رد می‌شود

$1.25 < 4.1028 \Rightarrow$ H_0 تایید می‌شود یعنی تیمارها با احتمال ۹۵٪ موثر نیست (محصول تغییر نکرده)

مثال) جدول تحلیل واریانس زیر را تکمیل کنید و در سطح خطای ۰/۰۵ آزمون کنید آیا میانگین‌ها برابر است؟

منبع تغییرات	SS	df	MS	F
تیمار	۱۵۰	۳	۵۰	
خطا				
کل	۹۰۰	۱۵		

$SS_{Tr} = MS \times df = 50 \times 3 = 150$

$SS_E = SS_t - SS_{tr} = 900 - 150 = 450$ & $df_e = 15 - 3 = 12 \Rightarrow MS_e = \frac{450}{12} = 37.5$

$F = \frac{MS_{tr}}{MS_e} = \frac{50}{37.5} = 1.333$ $\qquad$ $F > F_{\alpha,3,12} \Rightarrow$ H_0 رد می‌شود

$1.333 < 3.490 \Rightarrow$ با احتمال ۹۵٪، H_0 پذیرفته می‌شود

مثال) تحلیل واریانس زیر را تکمیل کنید و در سطح خطای ۵٪ آزمون کنید آیا میانگین ها با یکدیگر برابر است؟

منبع تغییرات	SS	df	MS	F
تیمار	۴۰	۲	$MS_{tr} = \frac{SS_{tr}}{df_{tr}} = ۲۰$	
خطا	۱۶۰	۱۰	$MS_E = ۱۶$	$F = \frac{MS_{tr}}{MS_E} = \frac{۲۰}{۱۶} = ۱/۲۵$
کل	۲۰۰	۱۲		

$SS_t = SS_E + SS_{tr}$ $df_e = df_t - df_{tr}$

اگر $F > F_{0/05,۲,۱۰} \Rightarrow$ H_0 رد می‌شود

$۱/۲۵ < ۴/۱۰۲۸ \Rightarrow$ H_0 تایید می‌شود یعنی با احتمال ۹۵٪ موثر نیست (محصول تغییر نکرده)

مثال) جدول تحلیل واریانس زیر را تکمیل کنید و در سطح خطای ۰/۰۵ آزمون کنید آیا میانگین ها برابر است؟

منبع تغییرات	SS	df	MS	F
تیمار	۱۵۰	۳	۵۰	
خطا				
کل	۶۰۰	۱۵		

$SS_{Tr} = MS \times df = ۵۰ \times ۳ = ۱۵۰$

$SS_E = SS_t - SS_{tr} = ۶۰۰ - ۱۵۰ = ۴۵۰$ & $df_e = ۱۵ - ۳ = ۱۲ \Rightarrow MS_e = \frac{۴۵۰}{۱۲} = ۳۷/۵$

$F = \frac{MS_{tr}}{MS_e} = \frac{۵۰}{۳۷/۵} = ۱/۳۳۳$

$F > F_{\alpha,۳,۱۲} \Rightarrow$ H_0 رد می‌شود

$۱/۳۳۳ < ۳/۴۹۰ \Rightarrow$ با احتمال ۹۵٪، H_0 پذیرفته می‌شود

محاسبه مجموع مربعات :

A	B	C
۱	۲	۳
۲	۱	۴
۰	۳	۵
۱	۰	۳

۱) $SS_{tr} = \left(\sum \frac{T_i^2}{n_i}\right) - \frac{T_{..}^2}{N}$

۲) $SS_t = \sum_i \sum_j x_{ij}^2 - \frac{T_{..}^2}{N}$

۳) $SS_E = SS_t - SS_{tr}$

T_i : جمع هر تیمار
$T_{..}$: جمع کل اعداد
N : تعداد " "
x_{ij} : مقدار اعداد

مثال) در سطح خطای ۰/۰۵ آزمون کنید آیا میانگین های جوامع با یکدیگر برابرند؟

$T_1 = ۱+۲+۰+۱ = ۴$
$T_2 = ۲+۱+۳+۰ = ۶$
$T_3 = ۳+۴+۵+۳ = ۱۵$

$T_{..} = \sum T_i = ۴+۶+۱۵ = ۲۵$ $N = ۴ \times ۳ = ۱۲$

$SS_{tr} = \sum \frac{T_i^2}{n_i} - \frac{T_{..}^2}{N} = \left(\frac{۴^2}{۴} + \frac{۶^2}{۴} + \frac{۱۵^2}{۴}\right) - \frac{۲۵^2}{۱۲} = ۱۰/۱۷$

$SS_t = \sum\sum x_{ij}^2 - \frac{T_{..}^2}{N} = (۱^2 + ۲^2 + \cdots + ۵^2 + ۳^2) - \frac{۲۵^2}{۱۲} = ۲۶/۹۲$

$SS_E = SS_t - SS_{tr} = ۲۶/۹۲ - ۱۰/۱۷ = ۱۶/۷۵$

منبع تغییرات	SS	df	MS	F
تیمار	۱۰/۱۷	تعداد تیمار $۳-۱=۲$	$\frac{۱۰/۱۷}{۲} = ۵/۱$	$\frac{۵/۱}{۱/۹}$
خطا	۱۶/۷۵	۹	$\frac{۱۶/۷۵}{۹} = ۱/۹$	
کل	۲۶/۹۲	$۱۲-۱=۱۱$		

$= ۲/۷ \ngtr F_{۰/۰۵,۲,۹} = ۴/۲۵$

H_0 پذیرفته می شود یعنی تیمار موثر نیست.

مثال) در سطح خطای ۰/۰۵ آزمون کنید آیا میانگین های جوامع A و B و C باهم متفاوت است؟

SUBJECT:
Year: Month: Date:

A	B	C
۲	۱	۴
۳	۲	۸
۴	۰	
	۱	
۹	۴	۱۲

$T_{..} = ۲۵$

$$SS_{tr} = \sum \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N}$$

$$= \left(\frac{۹^2}{۳} + \frac{۱۶}{۴} + \frac{۱۴۴}{۲}\right) - \frac{(۲۵)^2}{۹} = ۳۳{,}۶$$

$$SS_T = \sum\sum x_{ij}^2 - \frac{۲۵^2}{۹} = (۴+۹+۱۶+۰+۱+۱۶+۶۴) - ۸۷{,}۱۱$$

$$= ۴۵{,}۶$$

$$SS_E = ۴۵{,}۶ - ۳۳{,}۶ = ۱۲{,}$$

منبع تغییرات	SS	df	MS	F	
تیمار	۳۳,۶	۲	۱۶,۸	$\frac{۱۶,۸}{۲} = ۸,۴ > F_{۰,۰۵ \, ۲,۶}$ $۸,۴ > ۵,۱۴۳۳$	
خطا	۱۲	۶	۲	H_0 رد می‌شود یعنی میانگین‌ها با احتمال ۹۵٪ متفاوت است.	
کل	۴۵,۶	۸			

SS کل به دو قسمت تقسیم می‌شود (تیمار - خطا)

- طرح بلوک‌های تصادفی (طرح دوعاملی وقتی مشاهدات تکرار نداشته باشد)

از این طرح در مواقعی استفاده می‌شود که در سطول مقاطع یک عنصر محدود داشته باشد.

مدل طرح بلوک تصادفی بصورت زیر است:

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ یک‌عاملی

بلوک تصادفی

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

میانگین کل — اثر تیمار — اثر بلوک — اثر خطا

$$\mu = \frac{T_{..}}{N}$$

K: تعداد تیمار

n: تعداد بلوک

کارخانه (تیمار) →	A	B	C
۱			
۲			
۳			
۴			

(بلوک) (جاده)

STAEDTLER

SUBJECT:
Year( )Month( )Date( )

طرح بلوک‌های تصادفی ANOVA

منبع تغییرات	SS	df	MS	F
تیمار	SS_{tr}	$k-1$	$MS_{tr}=\frac{SS_{tr}}{k-1}$	$F_{tr}=\frac{MS_{tr}}{MS_E} > F_{\alpha,(k-1),(k-1)(n-1)}$
بلوک	SS_B	$n-1$	$MS_B=\frac{SS_B}{n-1}$	H_0 رد می‌شود یعنی تیمار موثر است
خطا	SS_E	$(k-1)(n-1)$	$MS_E=\frac{SS_E}{(k-1)(n-1)}$	$F_B=\frac{MS_B}{MS_E} > F_{\alpha,(n-1),(k-1)(n-1)}$
کل	SS_T	$N-1$		H_0 رد می‌شود یعنی بلوک موثر است

$$SS_{Tr}=\sum\frac{T_{i\cdot}^2}{n_i}-\frac{T_{\cdot\cdot}^2}{N} \qquad SS_T=\sum\sum x_{ij}^2-\frac{T_{\cdot\cdot}^2}{N}$$

$$SS_B=\sum\frac{T_{\cdot j}^2}{n_j}-\frac{T_{\cdot\cdot}^2}{N} \qquad SS_E=SS_t-SS_{tr}-SS_B$$

مثال) در سطح خطای ۰٫۰۵ آزمون کنید تیمار موثر است یا بلوک؟

بلوک	A	B	C	$T_{\cdot j}$
①	۱	۰	۲	۳
②	۱	۴	۶	۱۱
③	۳	۲	۵	۱۰
④	۱	۴	۷	۱۲
$T_{i\cdot}$	۶	۱۰	۲۰	$T_{\cdot\cdot}=۳۶$

$$SS_{Tr}=\frac{۶^2}{۴}+\frac{۱۰^2}{۴}+\frac{۲۰^2}{۴}-\frac{۳۶^2}{۱۲}=۲۶$$

$$SS_B=\frac{۳^2}{۳}+\frac{۱۱^2}{۳}+\frac{۱۰^2}{۳}+\frac{۱۲^2}{۳}-\frac{۳۶^2}{۱۲}=۱۶٫۶۷$$

$$SS_T=۵۴$$

$$SS_E=۵۴-۱۶٫۶۷-۲۶=۱۱٫۳۲$$

منبع تغییرات	SS	df	MS	F
تیمار	۲۶	۲	$\frac{۲۶}{۲}=۱۳$	$F_{tr}=\frac{۱۳}{۱٫۸۷}=۶٫۹۵ > F_{۰٫۰۵,۲,۶}$
بلوک	۱۶٫۶۷	۳	۵٫۵۳	$۶٫۹۵ > ۵٫۱۴$ ← H_0 رد می‌شود با اطمینان ۹۵٪ تیمار موثر است
خطا	۱۱٫۳۲	۶	۱٫۸۷	$F_B=\frac{۵٫۵۳}{۱٫۸۷}=۲٫۹۵ > F_{\alpha,۳,۶}$
کل	۵۴	۱۱		$۲٫۹۵ \ngtr ۴٫۷۵$ H_0 پذیرفته می‌شود یعنی بلوک موثر نیست

STAEDTLER

SUBJECT:
Year( )Month( )Date( )

$\begin{cases} H_0 \text{ رد می‌شود با احتمال ۹۵٪ تیمار موثر است} \\ H_0 \text{ تایید می‌شود یعنی با احتمال ۹۵٪ بلوک موثر نیست} \end{cases}$

* طرح دو عاملی
هرگاه در طرح بلوک‌ها در هر سلول m عدد داشته باشیم از طرح دو عاملی استفاده می‌شود. مدل این طرح بصورت زیر می‌باشد.

$$y_{ijk} = \underbrace{\mu}_{\text{میانگین کل}} + \underbrace{\alpha_i}_{\text{اثر تیمار}} + \underbrace{\beta_j}_{\text{اثر بلوک}} + \underbrace{(\alpha\beta)_{ij}}_{\text{اثر متقابل}} + \underbrace{\varepsilon_{ijk}}_{\text{خطا}}$$

$$\mu = \frac{T_{\circ\circ}}{N}$$

$$N = n\,K\,\underbrace{m}_{\text{تکرار}}$$

	A	B	C
①	۱ ۲ ۴	۳ ۲ ۵	۶ ۱ ۴
②			
③			

منبع تغییرات	SS	df	MS	F	
تیمار	SS_{tr}	$k-1$	$MS_{tr} = \frac{SS_{tr}}{K-1}$	$F_{tr} = \frac{MS_{tr}}{MS_E}$	$F_{tr} > F_{\alpha, k-1, nk(m-1)}$ H_0 رد می‌شود، تیمار موثر است
بلوک	SS_B	$n-1$	$MS_B = \frac{SS_B}{n-1}$	$F_B = \frac{MS_B}{MS_E}$	$F_B > F_{\alpha, (n-1), nK(m-1)}$ H_0 رد می‌شود، بلوک موثر است
اثر متقابل	$SS_{tr/B}$	$(K-1)(n-1)$	$MS_{Tr/B} = \frac{SS_{tr/B}}{(K-1)(n-1)}$	$F_{Tr/B} = \frac{MS_{Tr/B}}{MS_E}$	$F_{Tr/B} > F_{\alpha, (k-1)(n-1), nK(m-1)}$ H_0 رد می‌شود و تیمار و بلوک اثر متقابل طرد
خطا	SS_E	$nK(m-1)$	$MS_E = \frac{SS_E}{nK(m-1)}$		
کل	SS_T	$N-1$			

STAEDTLER

SUBJECT:
Year( )Month( )Date( )

$F_{trF} > F_{\alpha, (K-1), nK(m-1)}$ H_0 رد می‌شود ⇐ تیمار موثر است.

$F_B > F_{\alpha, (n-1), nK(m-1)}$ H_0 〃 〃 ⇐ بلوک موثر است.

$F_{trB} > F_{\alpha, (n-1)(K-1), nK(m-1)}$ 〃 〃 〃 ⇐ تیمار و بلوک اثر متقابل دارند.

$$\begin{cases} H_0 : \alpha_i = 0 \\ H_1 : \alpha_i \neq 0 \end{cases} \qquad \begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 1 \end{cases} \qquad \begin{cases} H_0 : \alpha\beta = 0 \\ H_1 : \alpha\beta \neq 0 \end{cases}$$

• ناپارامتری: * در آزمون های ناپارامتری داده‌های مشاهده O_i (نمای) و مقادیر محاسبه شده نظری (e_i) است.

* از آزمون های ناپارامتری وقتی استفاده می‌شود به توزیع جامعه مشخص نباشد. آزمون هایی مانند **نیکویی برازش**

و تملی را با آزمون های ناپارامتری انجام می‌دهند.

* آزمون تملی یا برابری نسبت‌ها:

برای آزمون تملی ابتدا مقادیر تجربی را با استفاده از این رابطه بدست می‌آوریم: $e_{ij} = \dfrac{n_{i.} \times n_{.j}}{n_{..}}$

سپس مقدار آماره آزمون را از رابطه روبرو بدست می‌آوریم: $X^2 = \sum \dfrac{(O_{ij} - e_{ij})^2}{e_{ij}}$

سپس با مقدار کای مربع (X^2) و درجه آزادی $d_f = \underbrace{(r-1)}_{\text{سطر}}\underbrace{(C-1)}_{\text{تعداد ستون}}$ مقایسه می‌کنیم در صورتیکه مقدار

χ^2 بزرگتر از مقدار بحرانی باشد، در این صورت H_0 رد می‌شود و نسبت‌ها باهم برابر نیستند.

مثال: مدیر شرکتی در شمال ۱۰۰ کارمند است براساس جنسیت و میزان تحصیلات کارمندان را تقسیم بندی کرده‌ایم. در سطح خطای ۰/۰۵ آزمون کنید آیا جنسیت و میزان تحصیلات با یکدیگر نسبتشان برابر است؟

O_{ij} (نظری)

جنسیت \ مدرک	فوق دیپلم	لیسانس	فوق لیسانس	
خانم‌ها	۵	۱۰	۱۶	۳۱
آقایان	۱۵	۲۰	۳٤	۶۹
	۲۰	۳۰	۵۰	۱۰۰

$\begin{cases} H_0: \text{فراوانی نظری} = \text{فراوانی تجربی} \\ H_1: \quad " \neq " \end{cases}$

یا

$\begin{cases} H_0: O_{ij} = e_{ij} \\ H_1: \quad " \neq " \end{cases}$

یا

$\begin{cases} H_0: P_1 = P_2 \\ H_1: P_1 \neq P_2 \end{cases}$

e_{ij} (تجربی)

(۳۱ × ۲۰)/۱۰۰ = ۶٫۲	(۳۰×۳۱)/۱۰۰ = ۹٫۳	(۵۰×۳۱)/۱۰۰ = ۱۵٫۵
(۲۰×۶۹)/۱۰۰ = ۱۳٫۸	(۳۰×۶۹)/۱۰۰ = ۲۰٫۷	(۵۰×۶۹)/۱۰۰ = ۳٤٫۵

آماره آزمون : $\chi^2 = \sum \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \frac{(۵-۶٫۲)^2}{۶٫۲} + \frac{(۱۰-۹٫۳)^2}{۹٫۳} + \frac{(۱۶-۱۵٫۵)^2}{۱۵٫۵} + \frac{(۱۵-۱۳٫۸)^2}{۱۳٫۸}$

$+ \frac{(۲۰-۲۰٫۷)^2}{۲۰٫۷} + \frac{(۳٤-۳٤٫۵)^2}{۳٤٫۵} = ۰٫٤۳$

$df = (r-1)(c-1) = (۲-1)(۳-1) = ۲$

$\chi^2_{۰٫۰۵,۲} = ۵٫۹۹۱ \rightarrow ۰٫٤۳ \ngtr ۵٫۹۹۱ \Rightarrow$ H_0 تایید می‌شود یعنی میزان تحصیلات و جنسیت نسبتان برابر است.

O_{ij}

10	30	40
40	20	60
50	50	100

e_{ij}

۲۰	۲۰
۳۰	۳۰

تست ۶۴ ، تمرین ۱

SUBJECT
Year( )Month( )Date( )

۲۰/۱۲/۸۹

آزمون نیکویی برازش

هدف از انجام این آزمون آن است که مشخص کنیم آیا مقادیر ارائه شده از توزیع خاصی تبعیت می‌کند؟ یا خیر؟

برای بدست آوردن مقادیر نظری از روی داده‌ها استفاده می‌کنیم و برای بدست آوردن مقادیر تجربی از رابطه نسبت استفاده می‌کنیم.

$$P = \frac{x}{n} \rightarrow \boxed{x = pn = F_{ei}} \Leftarrow \text{مقادیر تجربی}$$

$$F_{oi} \Leftarrow \text{مقادیر نظری}$$

$$X^2 = \sum \frac{(F_{oi} - F_{ei})^2}{F_{ei}}$$

$$\begin{cases} H_0 : F_{oi} = F_{ei} \\ H_1 : F_{oi} \neq F_{ei} \end{cases}$$

$$X^2 > X^2_{\alpha, df} \rightarrow H_0 \text{ رد می‌شود}$$

$$df = k - m - 1$$

k ← تعداد ردها
m ← تعداد پارامترهایی که باید برآورد شوند

نرمال : $m = 2$
پواسن : $m = 1$
نمایی : $m = 1$
زمانی توزیع پارامتر نداشته باشد (مثال ۲) : $m = 0$

مثال) جدول توزیع فراوانی حقوق ماهانه ۱۰۰ کارمند یک وزارت خانه برحسب صدهزار تومان بصورت زیر می‌باشد:

الف) میانگین، واریانس و انحراف معیار را بدست آورید.

ب) تحقیق کنید در توزیع فراوانی حقوق ماهانه کارمندان این وزارتخانه می‌تواند از توزیع نرمال تبعیت نماید. سطح خطا ۵٪

Cl	۴۰-۶۰	۶۰-۸۰	۸۰-۱۰۰	۱۰۰-۱۲۰	۱۲۰-۱۴۰
F_i	۱۰	۲۰	۳۰	۲۵	۱۵
x_i	۵۰	۷۰	۹۰	۱۱۰	۱۳۰

۱- نشانه دسته را بدست می‌آوریم (مجموع دو حد تقسیم بر ۲)

۲- کل کارمندان شرکت ۱۰۰ جمع

$$\mu = \frac{\sum F_i x_i}{N}$$

$$\mu = \frac{\sum F_i x_i}{N} = 93$$

$$\delta^2 = \frac{\sum F_i x_i^2 - N\mu^2}{N} = 571 \qquad \delta = 23.9$$

STAEDTLER

پایین و بالای حدود طبقات l
غیر مبهم

	حدود طبقات	$Z = \frac{x-\mu}{\delta}$	$P(Z)$	$F_{ei} = NP$	$(F_{oi}-F_{ei})^2/F_{ei}$
۱	$x < ۶۰$	$Z < \frac{۶۰-۹۳}{۲۳٫۹} = -۱٫۳۸$	۰٫۰۸۳۸	۸٫۳۸	۰٫۳۱۳
۲	$۶۰ \leq x < ۸۰$	$-۱٫۳۸ \leq Z < -۰٫۵۴$	۰٫۲۱۰۸	۲۱٫۰۸	۰٫۰۵۵
۳	$۸۰ \leq x < ۱۰۰$	$-۰٫۵۴ \leq Z < ۰٫۲۹$	۰٫۳۱۹۵	۳۱٫۹۵	۰٫۱۱۹
۴	$۱۰۰ \leq x < ۱۲۰$	$۰٫۲۹ \leq Z < ۱٫۱۳$	۰٫۲۵۶۷	۲۵٫۶۷	۰٫۰۱۷
۵ = K	$۱۲۰ \leq x$	$۱٫۱۳ \leq Z$	۰٫۱۲۹۲	۱۲٫۹۲	۰٫۳۳۵

$\chi^2 = ۰٫۸۴$

$df = K - m - 1 = ۵ - ۲ - ۱ = ۲$

μ و σ^2 حساب نمودیم

$\boxed{\text{نقطه بحرانی} = \chi^2_{۰٫۰۵,۲}}$ $= ۵٫۹۹$ $\quad \chi^2 > \chi^2_{\alpha,df} \rightarrow$ رد H_0 $\rightarrow ۰٫۸۴ \ngtr ۵٫۹۹$

H_0 تایید می‌شود یعنی با احتمال ۹۵٪ از توزیع نرمال تبعیت می‌کند.

مثال) مندل در آزمایشی با مقداری نخود انجام داد مشاهدات زیر را بدست آورد:

F_{oi}

۱ چروکیده و سبز	۲ چروکیده و زرد	۳ گرد و سبز	۴ گرد و زرد
۳۲	۱۰۱	۱۰۸	۳۱۵

بنابر نظریه وراثت او نسبت اعداد باید بصورت زیر باشد:

۴	۳	۲	۱	
۹	۳	۳	۱	$\rightarrow \Sigma = ۱۶$

آیا دلیلی با اطمینان ۹۹٪ برای تردید نسبت به نظریه او وجود دارد یا خیر؟

مشاهدات	F_{oi}	$NP = F_{ei} = ۵۵۶ \times \frac{x}{۱۶}$	$\frac{(F_{oi}-F_{ei})^2}{F_{ei}}$
چروکیده سبز	32	$۵۵۶ \times \frac{۱}{۱۶} = ۳۴٫۷۵$	$\frac{(۳۲-۳۴٫۷۵)^2}{۳۴٫۷۵} = ۰٫۲$
" زرد	۱۰۱	$۵۵۶ \times \frac{۳}{۱۶} = ۱۰۴٫۲۵$	۰٫۱
گرد سبز	۱۰۸	$۵۵۶ \times \frac{۳}{۱۶} = ۱۰۴٫۲۵$	۰٫۱۳
" زرد	۳۱۵	$۵۵۶ \times \frac{۹}{۱۶} = ۳۱۲٫۷۵$	۰٫۰۲
	$\Sigma = ۵۵۶$		$\chi^2 = ۰٫۴۷$

Subject:
Year. Month. Date. / /

[illegible]

$$df = k - m - 1 = 4 - 0 - 1 = 3$$

$\alpha = 1 - P = 0.01$

$$\chi^2_{\alpha, df} \rightarrow \chi^2_{0.01, 3} = 11.34$$

$0.47 \ngtr 11.34 \rightarrow$ H_0 تایید می‌شود یعنی نظریه مندل

با احتمال ۰/۹۹ صحیح است.

تصحیح ییتس:

هرگاه در یک آزمون ناپارامتری « درجه آزادی » برابر ۱ شود از تصحیح ییتس استفاده می‌کنیم. در این صورت آماره آزمون از رابطه زیر بدست می‌آید:

$$\chi^2 = \sum \frac{(|F_{oi} - F_{ei}| - 0.5)^2}{F_{ei}}$$

مثال) مدیر کنترل کیفیت کارخانه هر روز یک اتومبیل را بطور تصادفی از بین تمام اتومبیل‌هایی که از خط مونتاژ خارج می‌شوند انتخاب و تحقیق می‌نماید که آیا اتومبیل انتخاب شده نیاز به تنظیم دارد یا خیر؟ نتیجه تحقیق در ۱ ماه نشان می‌دهد که اتومبیل‌های انتخاب شده در ۱۰ روز احتیاج به تنظیم دارد و در ۲۰ روز احتیاج به تنظیم ندارد. در سطح $\alpha = 0.05$ تحقیق کنید که آیا اتومبیل‌هایی که احتیاج به تنظیم روزانه دارند از توزیع برنولی با پارامتر $P = 0.2$ تبعیت می‌کنند یا خیر؟

وضعیت	F_{oi}	$P = 0.2$ $F_{ei} = NP$	$\frac{(\lvert F_{oi} - F_{ei}\rvert - 0.5)^2}{F_{ei}}$
نیاز به تنظیم دارد	10	$6 = 30 \times 0.2$	2.04
نیاز به تنظیم ندارد	20	$24 = 30 \times 0.8$	0.51
	+ 30		2.55

$$df = k - m - 1 = 2 - 0 - 1 = 1 \rightarrow$$ نیاز به تصحیح ییتس دارد

نیاز به محاسبه نبود

$$\chi^2_{0.05, 1} = 3.84 \rightarrow 2.55 \ngtr 3.84 \rightarrow$$ H_0 تایید می‌شود یعنی از توزیع برنولی تبعیت می‌کند

★ یادآوری توزیع برنولی:

$$\boxed{P(x) = p^x q^{1-x}}$$
↓
پیوسته

$$f(x) = 0/2^x \, 0/8^{1-x}$$
↓
گسسته

• آزمون رتبه علامت‌دار «ویلکاکسون»:

هرگاه نمونه‌ها بصورت زوجی باشد از آزمون رتبه علامت‌دار استفاده می‌کنیم که ترتیب انجام این آزمون به ترتیب زیر است: اگر بتوانیم از توزیع t و اگر نتوانیم از این آزمون استفاده می‌کنیم.

۱- اختلاف زوج‌ها را بدست می‌آوریم.

۲- اختلاف‌ها را در نظر گرفته و مرتب می‌کنیم.

۳- به قدرمطلق اختلاف‌ها رتبه می‌دهیم. (از کوچک به بزرگ) در صورتیکه اعداد مساوی داشته باشیم رتبه متوسط می‌دهیم.

۴- علامت رتبه‌ها را مشخص می‌کنیم.

۵- مجموع رتبه‌های مثبت را بدست می‌آوریم.

تا اینجا را سر امتحان می‌دهند ←

۶- میانگین، واریانس وانحراف معیار رتبه‌های مثبت را بدست می‌آوریم و با آماره z آزمون را انجام می‌دهیم.

$$T^+ = \sum R_i^+$$

$$E(T^+) = \frac{n(n+1)}{4}$$ امید، میانگین

$$Var(T^+) = \frac{n(n+1)(2n+1)}{24}$$

$$Z = \frac{T^+ - E(T^+)}{\sqrt{Var(T^+)}}$$

$|Z| > Z_{\frac{\alpha}{2}} \rightarrow$ H_0 رد می‌شود.

Subject:
Year.   Month.   Date.   ( )

سوال صفحه ۲۸

مثال) نمرات ۲۰ کارمند قبل از گذراندن یک دوره آموزشی و بعد از گذراندن دوره آموزشی در جدول زیر ثبت شده است.

در سطح خطای ۰/۰۱ آزمون کنید آیا کلاس ضمن خدمت موثر بوده است؟ (نوع توزیع معلوم نیست)

قبل ۱	۱۲	۱۰	۱۹	۱۱	۱۶	۱۴	۱۵	۱۲	۱۴	۱۲	۱۰	۱۳	۱۲
بعد ۲	۱۷	۱۹	۱۲	۱۶	۱۴	۱۵	۱۸	۲۰	۱۹	۱۴	۱۷	۱۹	۱۴
	۵	۶	-۷	۵	-۲	۱	۳	۸	۵	۲	۷	۶	۲

ادامه

قبل	۱۱	۱۵	۱۱	۱۲	۱۱	۱۲	۱۱
بعد	۱۵	۸	۱۸	۱۷	۱۶	۱۸	۱۸
	۴	-۷	۷	۵	۵	۶	۷

$d = (۲) - (۱)$ راه ← مرحله ۱

مرتب شده $\lvert d \rvert$	۱	۲	۲	۲	۳	۴	۵	۵	۵	۵	۵	۶	۶	۶	۷	۷	۷	۷	۷	۸
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
رتبه:	۱	۳	۳	۳	۵	۶	۹	۹	۹	۹	۹	۱۳	۱۳	۱۳	۱۷	۱۷	۱۷	۱۷	۱۷	۲۰
علامت:	+	−	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+

(دوتا +۲ و یکی −۲)

$$T^+ = \sum (R_i)^+ = ۱ + ۳ + ۳ + ۵ + ۶ + ۹(۵) + ۱۳(۳) + ۱۷(۳) + ۲۰ = ۱۷۳$$

جمع رتبه‌های مثبت

$$E(T^+) = \frac{n(n+1)}{4} = \frac{۲۰(۲۰+۱)}{۴} = ۱۰۵$$

$$Var(T^+) = \frac{n(n+1)(۲n+1)}{۲۴} = ۷۱۷/۵$$

$$Z = \frac{T^+ - E(T^+)}{Var(T^+)} = ۲/۵۳۸$$

$Z_{۰/۰۱} = ۲/۳۳۶ \rightarrow ۲/۵۳۸ > ۲/۳۳۶ \rightarrow$ H_0 رد می‌شود یعنی کلاس موثر است.

PAPCO

آزمون جمع رتبه‌ای U من ویتنی:

از این آزمون در مواقعی استفاده می‌شود که دو جامعه مستقل را بخواهیم از لحاظ میانگین با یکدیگر مقایسه کنیم.

برای انجام این آزمون ابتدا دو جامعه را با یکدیگر مخلوط می‌کنیم سپس رتبه هر یک از اعداد را مانند حالت قبل تعیین می‌کنیم. سپس با استفاده از آماره‌های زیر می‌توانیم آزمون را انجام دهیم:

$$\begin{cases} U_1 = n_1 n_2 + \dfrac{n_1(n_1+1)}{2} - R_1 \\ U_2 = n_1 n_2 + \dfrac{n_2(n_2+1)}{2} - R_2 \end{cases} \quad or \quad \begin{cases} U_1 = \underbrace{R_1}_{\text{جمع رتبه‌های جامعه ۱}} - \dfrac{n_1(n_1+1)}{2} \\ U_2 = \underbrace{R_2}_{\text{جمع رتبه‌های جامعه ۲}} - \dfrac{n_2(n_2+1)}{2} \end{cases}$$

$$E(U) = \frac{n_1 n_2}{2}$$

$$var(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

$$Z = \frac{U - E(U)}{\sqrt{var(U)}}$$

مثال) در سطح خطای ۰/۰۵ آزمون کنید آیا میانگین دو جامعه A و B با یکدیگر برابر است؟

A	B
13	12
16	10
15	14
14	17
0	19

اعداد را مرتب می‌کنیم

		رتبه	جامعه A یا B
1	10	1	B
2	12	2	B
3	13	3	A
4	14	4.5	A
5	14	4.5	B
6	15	6	A
7	16	7	A
8	17	8	B
9	19	9	B

$R_A \Rightarrow$ جمع رتبه‌های A

$$R_A = 3 + 4.5 + 6 + 7 = 20.5$$

$$U_A = 20.5 - \frac{4(4+1)}{2} = 10.5$$

$$E(U_A) = \frac{4 \times 5}{2} = 10$$

$$var(U_A) = \frac{4 \times 5 (4 + 5 + 1)}{12} = 16.66$$

$$Z = \frac{U - E(U)}{\sqrt{var(U)}} = \frac{10.5 - 10}{\sqrt{16.66}} = 0.123$$

$|Z| > Z_{\frac{\alpha}{2}}$ $\quad 0.123 \ngtr 1.96 \rightarrow$ H_0 تایید می‌شود

با احتمال ۹۵٪ میانگین دو جامعه برابر است.

Subject:

Year: Month: Date: ( )

جلسه نهم ۲۷، ۱۲، ۸۹

• ضریب همبستگی اسپیرمن «Spriman»

هرگاه توزیع جامعه نامشخص باشد از ضریب همبستگی اسپیرمن استفاده می‌کنیم. امید ریاضی این ضریب برابر صفر و واریانس آن $\frac{1}{n-1}$ می‌باشد.

$$E(r_S) = 0 \qquad Var(r_S) = \frac{1}{n-1}$$

برای انجام محاسبات آن به هر یک از متغیرهای مستقل و وابسته رتبه می‌دهیم (از کوچک به بزرگ) سپس اختلاف رتبه‌ها را بدست می‌آوریم و نام آن‌ها را D_i می‌گذاریم و از رابطه زیر ضریب همبستگی را بدست می‌آوریم.

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

مثال) ضریب همبستگی Spriman را بدست آورید.

x	y	r_x = رتبه x	r_y	$d = r_x - r_y$
۳	۲	۱	۱	۰
۴	۵	۲	۳	-۱
۷	۶	۵	۴	+۱
۶	۸	۴	۵	-۱
۵	۴	۳	۲	۱
۱۰	۹	۶	۶	۰

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6\times 4}{6(36-1)} = 0/885$$

$=$

* کولموگروف - اسمیرنوف : (مهم)

از این آزمون، برای آزمون فراوانی تجمعی نسبی نظری و تجربی استفاده می‌شود.

فراوانی مشاهده شده نسبی تجمعی

فراوانی نظری نسبی تجمعی

$$\begin{cases} H_0 : F_{oi} = F_{ei} \\ H_1 : F_{oi} \neq F_{ei} \end{cases}$$

برای انجام این آزمون اختلاف F_o و F_e را بدست می‌آوریم و قدر مطلق اختلاف آن‌ها را تعیین می‌کنیم. بزرگترین اختلاف به عنوان آماره آزمون در نظر گرفته می‌شود.

$$D = Max|F_o - F_e| \qquad if\ D > D_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

مثال) برای آزمون کولموگروف - اسمیرنوف آماره آزمون را بدست آورید.

PAPCO

F_{o_i}	F_{e_i}	$D = \lvert F_{o_i} - F_{e_i} \rvert$
۰/۱	۰/۱۵	۰/۰۵
۰/۳	۰/۲۵	۰/۰۵
۰/۴	۰/۳	۰/۱
۰/۶	۰/۷	۰/۱
۰/۷	۰/۷۸	۰/۰۸
۰/۹	۰/۹۳	۰/۰۷
۱	۱	۰

آماره آزمون $\boxed{D = ۰/۱۱}$

۰/۱ ← بزرگترین عدد، آماره آزمون است

تجمعی وقتی می‌نویسیم ! جدول همیشه ۱ است

• آزمون خطی بودن در رگرسیون ؟

$$F = \frac{\frac{X_1^2}{K-2}}{\frac{X_2^2}{n-K}} \qquad \text{if } F > F_{\alpha, K-2, n-K} \Rightarrow$$ H₀ رد می‌شود و خطی نیست

n : نمونه

K : تعداد رده

مثال) در آزمون خطی بودن برای مسئله خاص $n=14$ و $K=7$ ، $X_1^2 = 35/12$ و $X_2^2 = 8/05$

آماره آزمون را بدست آورید:

$$F = \frac{\frac{25/12}{5}}{\frac{8/05}{7}} = 4/36 > F_{0/05;5,7} = 3/97 \rightarrow$$ H₀ رد می‌شود

نکته) تعداد تغییرات همان تعداد ردیف است: برای مثال ۶ ردیف ←

$$\underbrace{TTT}_{1}\ \underbrace{FF}_{2}\ \underbrace{TT}_{3}\ \underbrace{FFF}_{4}\ \underbrace{T}_{5}\ \underbrace{FF}_{6}$$

$$E(R) = \frac{2n_1n_2}{n_1+n_2} + 1 \qquad var(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

* تبدیل متغیرهای مستقل کیفی:

تعداد متغیر مجازی برای یک متغیر کیفی = تعداد سطوح متغیر کیفی − ۱

$\left\{ \begin{array}{l} \text{مجرد} \\ \text{متاهل} \end{array} \right. \Rightarrow 2-1 = 1$

$\left\{ \begin{array}{ll} 1 & \text{خوب} \\ 2 & \text{متوسط} \\ 3 & \text{ضعیف} \end{array} \right. \Rightarrow 3-1 = 2$