فصل ششم

جامعه و توزیع نمونه گیری و برآورد (تئوری تخمین)

نمونه تصادفی: یک نمونه تصادفی با حجم n، نمونه ای است که هر زیر مجموعه n عضوی از جامعه دارای شانس انتخاب یکسان باشند.

تذکر: اگر $x_1 , x_2 , \dots , x_n$ متغیر های تصادفی را یک نمونه تصادفی با حجم n از جامعه f(x) می نامیم و توزیع احتمال آن عبارتست از: $f(x_1, \dots, x_n) = f(x_1)\, f(x_2) \dots f(x_n)$

آماره: هر تابعی از عضوهای نمونه تصادفی که شامل پارامترهای مجهول نباشد را یک آماره گویند.

مثلاً اگر $x_1 , x_2 , \dots , x_n$ یک نمونه تصادفی از متغیر تصادفی x باشند، توابع زیر آماره اند:

$$X_1 + 3X_2 - 1,\ \frac{X_1}{X_4},\ \frac{1}{n}\sum X_i,\ \frac{1}{n}\sum (X_1 - \overline{X})^2$$

و اگر μ و σ^2 مجهول باشند آنگاه توابع زیر آماره نیستند:

$$X_3 - \mu, \frac{\mu + \sigma^2}{X_2}, \frac{1}{n}\sum (X_i - \mu)^2$$

آماره ها را معمولا با حرف لاتین نشان داده و روی آن علامت (-) یا (^) یا (~) می گذارند مانند $\tilde{X}, \hat{X}, \overline{X}$,... وقتی نمونه گیری انجام می شود مقادیر $X_1, \dots, X_n$ را مشاهده کرد آنگاه مقدار مشاهده شده را با $\hat{P}$ نشان می دهیم.

بررسی چند آماره مفید:

گشتاور نمونه حول نقطه صفر: اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از متغیر تصادفی X باشند آنگاه rامین گشتاور نمونه حول مبدا به صورت زیر تعریف می شود:

$$M'r = \frac{1}{n}\sum_{i=1}^{n} x_i^r \quad , r = 1,2,3,...$$

اگر r=1 باشد میانگن نمونه به دست می اید که ان را با $\bar{X}n$ یا $\bar{X}$ نشان می دهند. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ و همچنین rامین گشتاور نمونه حول $\bar{X}n$ را به صورت زیر تعریف می کنند:

$$M_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^r \quad , r = 1,2,3,...$$

تذکر: اگر $X_1, ..., X_n$ یک نمونه تصادفی از جامعه ای با تابع احتمال f(x) باشد، آنگاه اگر μ_r' وجود داشته باشد داریم:

$$E(M_r') = \mu_r'$$

تذکر : اگر μ میانگین جامعه و σ^2 واریانس جامعه باشد آنگاه برای یک نمونه تصادفی $X_1, ..., X_n$ از جامعه ای با تابع احتمال f(x) داریم:

$$E(\bar{X}n) = \mu \quad , \operatorname{var}(\bar{X}n) = \frac{1}{n}\sigma^2$$

واریانس نمونه: برای نمونه تصادفی $X_1, X_2, ..., X_n$ از تابع احتمال f(x) واریانس نمونه به صورت زیر تعریف می شود:

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

(البته آماره $S'^2 = \frac{1}{n-1}\sum_{i=1}^{\wedge} (X_i - \bar{X})^2$ نیز برای واریانس نمونه تعریف می شود.)

توزیع نمونه : توزیع احتمال هر آماره را توزیع نمونه گویند. مثلا توزیع اماره $\bar{X}$ را توزیع نمونه میانگین و توزیع آماره S^2 را توزیع نمونه واریانس می گویند.

توزیع نمونه میانگین : اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از یک توزیع نرمال با میانگین μ و واریانس σ^2 باشد، آنگاه $\bar{X}$ دارای توزیع نرمال با میانگین μ و واریانس زیر است:

$$\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$$

تذکر: اگر تمام نمونه های تصادفی با حجم n از یک جامعه متناهی با حجم N و میانگین μ و واریانس σ^2 بدون حایگذاری انتخاب شوند، توزیع $\bar{X}$ تقریبا توزیع نرمال با میانگین μ و واریانس زیر است:

$$\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

قضیه حد مرکزی: اگر نمونه تصادفی n تایی از جامعه ای دلخواه با میانگین μ و واریانس σ^2 انتخاب شود آنگاه وقتی که n به حد کافی بزرگ باشد ($n \to \infty$) میانگین نمونه ای $\bar{X}$ به طور تقریبی دارای توزیع نرمال یا میانگین μ و واریانس $\frac{\sigma^2}{n}$ است و شکل حدی توزیع $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ نرمال استاندارد است با $\sigma^2 = 1,\ \mu = 0$

تذکر: اگر $n \geq 30$ باشد توزیع $\bar{X}$ بدون توجه به شکل توزیع جامعه، توزیع نرمال است. و اگر $n < 30$ باشد آنگاه در صورت نرمال بودن تقریبی جامعه توزیع $\bar{X}$ تقریبا نرمال است.

مثال: میانگین طول عمر یک لامپ ۷۸۰ ساعت و انحراف معیار آن ۳۶ ساعت است. یک نمونه تصادفی ۳۶ ایی از این لامپ با چه احتمالی میانگین طول عمری بیشتر از ۷۷۴ ساعت دارد؟

حل:

$$n=36 \qquad \mu_{\bar{x}}=\mu=780, \quad , \quad \sigma_{x}=\frac{\sigma}{\sqrt{n}}=\frac{36}{\sqrt{36}}=6$$

$$\Rightarrow P(\bar{X}>774)=1-P(\bar{X}<774)=1-P(Z<\frac{774-780}{6})$$

$$=1-P(Z<-1)=1-0/1587=0/8413$$

توزیع آماره $\bar{X}_1-\bar{X}_2$: (**تفاضل میانگین**) اگر نمونه های تصادفی مستقل با حجم های n_1و n_2از دو جامعه با مباتگین های μ_1 و μ_2 و واریانس های σ_1^2 و σ_2^2 انتخاب می شوند، توزیع نمونه تفاضل میانگین ها $(\bar{X}_1-\bar{X}_2)$ تقریبا توزیع نرمال با میانگین $\mu_{\bar{X}_1-\bar{X}_2}=\mu_1-\mu_2$ و واریانس $\sigma^2_{\bar{X}_1-\bar{X}_2}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$ می باشد و همچنین داریم:

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

مثال: یک نمونه تصادفی ۲۵ تایی از جامعه ای نرمال با میانگین ۸۰ و انحراف معیار ۵ و از جامعه نرمال دیگری با میانگین ۷۵ و انحراف معیار ۳ نمونه ای ۳۶ تایی انتخاب می کنیم. اگر $\bar{X}_1$ و به $\bar{X}_2$ به ترتیب میانگین های نمونه جامعه اول و دوم باشند، مطلوبست $P(3/4<\bar{X}_1-\bar{X}_2<8/9)$ در صورتیکه :

الف) میانگین ها تا هر میزان دقتی اندازه گیری شده باشند.

ب) میانگین ها تا یکدهم تقریب اندازه گیری شده باشند.

حل: $\mu_2 = 75$ و $\sigma_2 = 3, n_2 = 36$ $\mu_1 = 80, \sigma_1 = 5, n_1 = 25$

الف) $\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{25}{25} + \frac{9}{36} = 1 + \frac{1}{4} = \frac{5}{4}$ $\mu_{\bar{X}_1 - \bar{X}_2} = 80 - 75 = 5$

$$\Rightarrow P(3/4 < \bar{X}_1 - \bar{X}_2 < 8/9) P(\frac{3/4 - 5}{\sqrt{\frac{5}{4}}} < Z < \frac{8/9 - 5}{\sqrt{\frac{5}{4}}}) = 0/9236$$

ب)

$$\Rightarrow P(3/45 < \bar{X}_1 - \bar{X}_2 < 8/85) P(\frac{3/45 - 5}{\sqrt{\frac{5}{4}}} < Z < \frac{8/85 - 5}{\sqrt{\frac{5}{4}}}) = 0/9162$$

توزیع آماره S^2 : **(واریانس)** اگر S^2 واریانس نمونه تصادفی با حجم n از یک جامعه نرمال با واریانس σ^2 باشد آنگاه توزیع آماره $X^2 = \frac{(n-1)S^2}{\sigma^2}$ توزیع می توان دوم کای χ^2 با درجه آزادی $V = n - 1$ است.

مثال: احتمال اینکه یک نمونه تصادفی ۲۵ تایی از یک جامعه نرمال با واریانس σ^2 دارای واریانس s^2

الف) بزرگتر از ۹/۱ باشد چقدر است.

ب) بین ۳/۴۶۲ تا ۱۰/۷۴۵ باشد چقدر است.

حل: $n = 25 \rightarrow V = 25 - 1 = 24$

الف) $P(S^2 > 9/1) = 1 - P(S^2 < 9/1) = 1 - P(\frac{25-1}{6} s^2 < (\frac{25-1}{6}) 9/1)$

$$= 1 - P(X^2 < 36/4) = 1 - 0/95 = 0/05$$

ب) $P(3/462 < S^2 < 10/745) = P(S^2 < 10/745) - P(S^2 < 3/462)$

$$= P(X^2 < \frac{25-1}{6} \times 10/745) - P(X^2 < \frac{25-1}{6} \times 3/462) = P(X^2 < 42/98) - P(X^2 < 13/848)$$

=0/99-0/05=0/94

رابطه توزیع $\bar{X}$ با توزیع t: اگر حجم جامعه کمتر از ۳۰ باشد و واریانس جامعه نامعلوم باشد دیگر نمی توان از S به جای σ استفاده نمود زیرا در این صورت مقادیر S در نمونه های مختلف دارای نوسان قابل ملاحظه ای است و متغیر تصادفی $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ دارای توزیع t با درجه ازادی $V = n-1$ می باشد چون:

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{V}{n-1}}}$$

که در آن $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ می باشد و دارای توزیع نرمال استاندارد و واریانس $\frac{(n-1)}{\sigma}$ دارای توزیع χ^2 با درجه آزادی $V = n-1$ است و متغیر تصادفی T دارای توزیع t با درجه آزادی $V = n-1$ است.

مثال: از یک جامعه نرمال با میانگین $\mu = 20$ و واریانس مجهول، یک نمونه تصادفی ۹ تایی انتخاب می کنیم. آیا به نظر می رسد که میانگین و انحراف معیار این نمونه ۹ تایی به ترتیب ۲۴ و ۴/۱ باشد.

حل: ابتدا از جدول توزیع t مقدار $t_{0/975}$ یعنی $P(-2/31 < T < 2/31) = 0/95$ و نیز مقدار $t = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ را می یابیم.

$$t = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{24-20}{\frac{4/1}{3}} = 2/92$$

چون مقدار t محاسبه شده در بازه نیست یعنی $2/92 \notin (-2/31, 2/31)$ پس جواب منفی است.

توزیع آماره $\frac{S_1^2}{S_2^2}$ (مقایسه واریانس در جامعه): برای مقایسه واریانس دو جامعه از توزیع F با درجات آزادی $V_1 = n_1 - 1$ و $V_2 = n_2 - 1$ استفاده می کنیم که در آن n_2, n_1 حجم های دو نمونه مستقل اند که از دو جامعه نرمال انتخاب شده اند و متغیرهای مستقل U,V که:

$$U = \frac{(n_1 - 1)S_1^2}{\sigma_1^2}, \quad V = \frac{(n_2 - 1)S_2^2}{\sigma_2^2}$$

دو متغیر تصادفی مستقل توان دوم کای اند پس متغیر تصادفی

$$F = \frac{(n-1)S_1^2}{\sigma_1^2(n_1 - 1)} / \frac{(n_2 - 1)}{\sigma_2^2(n_2 - 1)} = \frac{S_1^2}{S_2^2} . \frac{\sigma_2^2}{\sigma_1^2}$$

دارای توزیع F با درجات آزادی V_2, V_1 است.

مثال: اگر S_2^2, S_1^2 واریانس های دو نمونه تصادفی مستقل با حجم های $n_1 = 25 \, , n_2 = 31$ از دو جامعه نرمال با واریانس های $\sigma_1^2 = 10 \quad , \quad \sigma_2^2 = 15$ باشند، مطلوبست محاسبه $P(\frac{S_1^2}{S_2^2} > 1/26)$

حل: $n_1 = 25 \rightarrow V_1 = 24, \; n_2 = 31 \rightarrow V_2 = 30$

$$P(\frac{S_1^2}{S_2^2} > 1/26) = P(\frac{S_1^2}{S_2^2} , \frac{\sigma_2^2}{\sigma_1^2} \geq \frac{15}{10} \times 1/26)$$

$$= P(F(24,30) > 1/89) = 1 - P(F(24,30) < 1/89) = 1 - 0/95 = 0/05$$

نظریه برآورد کردن (تئوری تخمین): برآورد در حالت کلی بر دو دسته زیر تقسیم می شود:

الف) برآورد نقطه ای   ب) برآورد فاصله ای

الف) برآورد نقطه ای: هدف از برآورد نقطه ای آن است که از روی نمونه تصادفی ، عددی به دست آوریم که انتظار داریم به مقدار نامعلوم پارامتر جامعه نزدیک باشد. θ را برای نشان دادن پارامتر به کار می بریم.

برآوردگر (تابع تصمیم): آماره ای که برای تخمین نقطه ای استفاده می شود را برآوردگر گویند و با $\hat{\theta}$ نشان می دهند.

برآورد نااریب: برآوردگر $\hat{\theta}$ را نااریب گوئیم هر گاه $\mu = E(\hat{\theta}) = \theta$

مثال: نشان دهید که برآوردگر $Y = \frac{1}{6}(X_1 + 2X_2 + 3X_3)$ که X_1, X_2, X_3 یک نمونه تصادفی با حجم n=3 از جامعه ای نرمال با میانگین μ و واریانس σ^2 می باشد، نااریب است.

حل: باید نشان دهیم:

$$\mu_Y = E(Y) = \mu$$

$$E(Y) = E(\frac{1}{6}(X_! + 2X_2 + 3X_3)) = \frac{1}{6}(\mu + 2\mu + 3\mu) = \mu$$

تذکر: برآورد نقطه ای میانگین جامعه μو میانگین نمونه $\overline{X}$ به صورت زیر است:

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{\wedge} Xi$$

۲- برآورد نقطه ای نسبت موفقیت در توزیع دو جمله ای (P) نسبت به نمونه مشاهده شده به صورت زیر است:

$$\hat{P} = \overline{P} = \frac{X}{n}$$

که x تعداد موفقیت در n بار آزمایش برنولی است.

۳- برآورد نقطه ای واریانس جامعه (σ^2) و واریانس نمونه به صورت زیر است:

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{\wedge}(X_i - \overline{X})^2$$

مثال : از جامعه کارگران بیمه شده نزد سازمان تامین اجتماعی ۱۶ کارگر را به عنوان نمونه انتخاب کردیم و تعداد فرزندان آنها به شکل زیر است:

۱و۲و۳و۶و۲و۴و۰و۶و۱و۲و۵و۷و۴و۴و۴و۳

بر آورد نقطه ای را برای متوسط واقعی تعداد فرزندان و انحراف معیار را بیابید.

حل:

$$\hat{\mu} = \bar{x} = 3/375$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = S^2 = 3.98 \quad S = \sqrt{3/98}$$

ب) برآورد فاصله ای: یک برآورد فاصله ای پارامتر θ فاصله ای به شکل$L<\theta<u$است که u و L توابعی از نمونه تصادفی x_n و ... و x_2 و x_1 می باشند به طوریکه

$$p(L<\theta<u)=1-\alpha$$

برای مقدار مشخص $1-\alpha$ این فاصله را یک فاصله اطمینان$(1-\alpha)$۱۰۰٪ برای θ می نامیم همچنین $1-\alpha$ درجه اطمینان (ضریب اطمینان) نیز نامیده می شود. معمولا برای $1-\alpha$ مقادیر ۰/۹۰ یا ۰/۹۵ یا ۰/۹۹ در نظر گرفته می شود. بر آورد فاصله ای به قرار زیر است.:

۱- برآورد فاصله ای میانگین

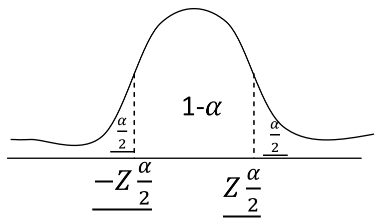
برای برآورد فاصله ای میانگین جامعه μ حالتهای زیر وجود دارد:

الف) اگر نمونه تصادفی به اندازه n از جامعه نرمال با واریانس معلوم σ^2 انتخاب شود، آنگاه $z=\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$

دارای توزیع نرمال استاندارد خواهد بود و از طرفی داریم:

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

که $Z_{\frac{\alpha}{2}}$ مقداری است که ازجدول نرمال استاندارد به صورت زیر بدست می آید $P(Z>Z_{\frac{\alpha}{2}})=\frac{\alpha}{2}$



با جایگذاری آماره $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ به جای Z در رابطه بالا خواهیم داشت:

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

و در نتیجه یک فاصله اطمینان $(1-\alpha)$ برای μ با فرض σ^2 معلوم به صورت زیر است :

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

مثال : توزیع وزن محصولات تولید شده یک کارخانه نرمال با انحراف معیار ۲۱ تن می باشد یک نمونه ۵۰ روزه از تولیدات انتخاب شده که میانگین وزن آن ۸۷۱ تن است. در سطح اطمینان ۹۰ درصد میانگین واقعی وزن محصولات تولید شده طی یک روز را بر آورد کنید.

حل:

$$n=۵۰ \quad \text{و} \quad \bar{x}=۸۷۱ \quad \text{و} \quad \sigma=۲۱$$

$$1-\alpha = ۰/۹۰ \rightarrow \alpha = ۰/۱ \rightarrow Z_{\frac{\alpha}{2}} = Z_{0/05}= 1/645$$

$$\rightarrow 871 - 1/645 \times\frac{21}{\sqrt{50}}<\mu< 871 + 1/645 \times\frac{21}{\sqrt{50}}\rightarrow (866/1 , 875/9)$$

ب) اگر نمونه تصادفی به اندازه n از جامعه نرمال با واریانس نامعلوم σ^2 انتخاب شود و $n < 30$ باشد

آنگاه آماره $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ دارای توزیع t با درجه آزادی $V = n - 1$ می باشد.

و نیز داریم:

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

و با جایگذاری $\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ به جای t داریم:

$$P\left(\bar{x} - t_{\frac{\alpha}{2}}.\frac{s}{\sqrt{n}}<\mu<\bar{x} + t_{\alpha/2}.\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

مثال : یک سازنده رنگ می خواهد متوسط زمان خشک شدن رنگ جدید دیوارهای داخلی ساختمان را معین کند اگر برای ۱۲ سطح آزمایشی با مساحت های برابر، میانگین و انحراف معیار زمان خشک شدن را به ترتیب ۶۶/۳ و ۸/۴ دقیقه بدست آورده باشد یک فاصله اطمینان ۹۵ درصد برای میانگین واقعی بیابید.

حل: $1-\alpha = ۰/۹۵ \rightarrow \alpha = ۰/۰۵ \rightarrow \alpha/2 = ۰/۰۲۵$ $\quad v= 12-1=11$

$\bar{x} = ۶۶/۳$ و $s = ۸/۴$ و $t_{0/025} = ۲/۲۰۱$

$$66/3 - 2/201 \times\frac{8/4}{\sqrt{12}}<\mu< 66/3 + 2/201 \times\frac{8/4}{\sqrt{12}}\rightarrow (61 , 71/6)$$

ج) اگر نمونه تصادفی به اندازه n از جامعه ای باواریانس نا معلوم σ^2 انتخاب شود و $n \geq ۳۰$ آنگاه آماری $z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است. یک فاصله اطمینان $(1-\alpha)$ ۱۰۰٪ برای μ با فرض σ^2 نامعلوم و $n < 30$ به صورت زیر بدست می آید.

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

مثال : برای برآورد متوسط درآمد هفتگی کارکنان رستوران در شهری داده های درآمد هفتگی را از نمونه ای مرکب از ۷۵ نفر از کارکنان رستوران گرد آوری شده است که میانگین و انحراف معیار به ترتیب ۱۲۷۰ و ۱۵۰ تومان بدست آمده یک فاصله اطمینان ۹۰ درصدی برای میانگین درآمد هفتگی بیابید.

حل:

$1-\alpha = ۰/۹ \rightarrow \alpha = ۰/۱ \rightarrow \alpha/2 = ۰/۰۵$

$n = ۷۵$ و $\bar{x} = ۱۲۷۰$ و $s = ۱۵۰$ و $z_{0/05} = ۱/۶۴۵$

$1270 - 1/645 \times\frac{150}{\sqrt{75}} < \mu < 1270\ 1/645 \times\frac{150}{\sqrt{75}} \rightarrow (1242 , 1298)$

۲- برآورد فاصله ای تفاضل بین میانگین ها:

برای دو نمونه تصادفی $n_۱$ و $n_۲$ از جامعه های آماری نرمال آماره

$$Z = \frac{(\bar{x}_۱ - \bar{x}_۲)-(\bar{\mu}_۱ - \bar{\mu}_۲)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_1}}}$$

دارای توزیع نرمال استاندارد خواهد بود که $P(-z\, \alpha/2 < Z < Z_{\frac{\alpha}{2}}) = 1-\alpha$

و فاصله اطمینان($1-\alpha$) ۱۰۰٪ به صورت زیر است :

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

مثال : یک فاصله اطمینان ۹۵ درصدی برای تفاضل واقعی بین طول عمرهای متوسط دو نوع لامپ بسازید با این فرض که نمونه ای تصادفی از ۴۰ لامپ روشنایی از یک نوع به طور متوسط ۴/۸ ساعت و ۷۵۰ لامپ از نوع دوم به طور متوسط ۴۰۲ ساعت در استفاده مستمر دوام آورده اند. و همچنین

σ_1 = ۲۶ و ۲۲ = σ_2 باشد.

حل:

$$1-\alpha = ۰/۹۵ \rightarrow \alpha = ۰/۰۵ \rightarrow z_{0/25} Z_{\frac{\alpha}{2}} = \;= ۱/۹۶$$

$$\bar{x}_1 = ۴۱۸ \quad و \quad \bar{x}_2 = ۴۰۲ \quad n_1=40 , n_2= 50 \quad و \quad \sigma_1 = ۲۶ \sigma_2 = ۲۲$$

$$\xrightarrow{\text{فاصله اطمینان}} (418-402) - 1/96 \times \sqrt{\frac{26^2}{40}+\frac{22^2}{50}} < \mu_1 - \mu_2 < (418-402) + 1/96 \times \sqrt{\frac{26^2}{40}+\frac{22^2}{50}}$$

$$\rightarrow \ (۶/۱ \text{ و } ۲۵/۷)$$

۳- برآورد فاصله ای نسبت موفقیت:

برای برآورد پارامتر دو جمله ای و n بزرگ آماره $Z=\frac{\bar{P}-P}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}}$ برای توزیع تقریبی نرمال استاندارد خواهد بود که: $P(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}) = 1-\alpha$

و یک فاصله اطمینان $(1-\alpha)$ ۱۰۰٪ برای نسبت موفقیت به صورت زیر است:

$$\bar{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} < p < \bar{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

که در آن $\bar{P} = \frac{x}{n}$

مثال یک نمونه ۱۰۰ نفری که به طور تصادفی از تمام رأی دهندگان ناحیه ای انتخاب شده اند، نشان داد که ۵۵٪ رأی دهندگان به یک نامزد خاص رای دادند. در سطح اطمینان ۹۵ درصد نسبت تمام رای دهندگان که به این نامزد رای داده اند را بر آورد کنید.

حل: $\bar{P}$ = ۰/۵۵ و $Z_{0/25}$ = 1/96

$$\rightarrow \; ۰/۵۵ - ۱/۹۶ \times \sqrt{\frac{0/55\times0/45}{100}} < p < ۰/۵۵ + ۱/۹۶ \times \sqrt{\frac{0/55\times0/45}{100}}$$

$$\xrightarrow{\text{فاصله اطمینان}} (\,0/45\,,0/65)$$

۴- بر آورد فاصله ای تفاضل نسبت و موفقیت در دو جامعه.

برای دو نمونه تصادفی $n_۲$ و $n_۱$ و از دو جامعه آماری، آماره

$$Z = \frac{\left(\bar{P}_۱-\bar{P}_۲\right)-\left(P_۱-P_۲\right)}{\sqrt{\frac{\bar{P}_۱\left(1-\bar{P}_۱\right)}{n_1}+\frac{\bar{P}_2(1-\bar{P}_2)}{n_2}}}$$

دارای توزیع نرمال استاندارد تقریبی است که $P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = 1-\alpha$

و یک فاصله اطمینان $(1-\alpha)$ ۱۰۰٪ برای تفاضل بین نسبتها به صورت زیر است:

$$(\bar{p}_1-\bar{p}_2)-Z_{\frac{\alpha}{2}}.\sqrt{\frac{\bar{P}_۱\left(1-\bar{P}_۱\right)}{n_1}+\frac{\bar{P}_2(1-\bar{P}_2)}{n_2}} < P_1-P_2 < (\bar{p}_1-\bar{p}_2)+ Z_{\alpha/2}.\sqrt{\frac{\bar{P}_۱\left(1-\bar{P}_۱\right)}{n_1}+\frac{\bar{P}_2(1-\bar{P}_2)}{n_2}}$$

مثال : اگر ۱۳۲ نفر از ۲۰۰ رای دهندگان مذکر و ۹۰ نفر از ۱۵۹ رای دهنده مونث موافق کاندیدای خاصی برای انتخابات باشند، یک فاصله اطمینان ۹۹٪ برای تفاضل بین نسبتهای واقعی رای دهندگان مرد و زن که موافق این کاندیدا هستند بیابید.

$$\bar{P}_1 = \frac{132}{200} = ۰/۶۶ \quad \text{و} \quad \bar{P}_2 = \frac{90}{159} = ۰/۶$$

$$1-\alpha = ۰/۹۹ \rightarrow \alpha = ۰/۰۱ \rightarrow Z_{\frac{0}{005}} = Z_{\frac{\alpha}{2}} = ۲/۵۷۵$$

فاصله اطمینان (۰/۱۹۴ و ۰/۰۷۴ -)

۵- برآورد فاصله ای واریانس:

با مفروض بودن نمونه n تایی از جامعه نرمال آماره $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ دارای توزیع کای دو با درجه آزادی V= n-1 است و

$$P\left(x_{\frac{\alpha}{2}}^{\ 2} < x^2 < x_{1-\frac{\alpha}{2}}^{\ 2} \right) = ۱-\alpha$$

و یک فاصله اطمینان $(1-\alpha)$ ۱۰۰٪ برای واریانس جامعه به صورت زیر است:

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

مثال: در ۱۶ بار کار آزمایشی یک موتور تحت آزمایش، مصرف بنزین آن دارای انحراف معیار ۲/۲ بوده است. یک فاصله اطمینان ۹۹٪ برای واریانس بازید که میزان تغییر پذیری مصرف بنزین را بسنجد.

حل:

$$V= n-1= 16-1=15$$

$$1-\alpha = ۰/۹۹ \rightarrow \alpha = ۰/۰۱ \rightarrow x_{\frac{\alpha}{2}}^{\ 2} = x_{0/005}^2 = ۳۲/۸۰۱ \text{ و } x_{1-\frac{\alpha}{2}}^{\ 2} = x_{0/195}^2 = ۴/۶۰۱$$

$$\rightarrow \frac{(16-1)\,(2/2)}{4/601} < \sigma^2 < \frac{(16-1)\,(2/2)^2}{32/801} \xrightarrow{\text{فاصله اطمینان}} ۲/۲۱ < \sigma^2 < ۱۵/۷۸$$

فاصله اطمینان برای انحراف معیار نیز برابر است با: $\rightarrow$ ۱/۴۹ $<\sigma<$ ۳/۹۷

۶- برآورد فاصله ای نسبت واریانس :

اگر s_1^2 و $s_2{}^2$ واریانس نمونه های تصادفی مستقل n_2 و n_1 از جامعه های نرمال باشند، آنگاه

متغیر $\frac{s_1{}^2\sigma_2^2}{s_2{}^2\sigma_1^2}$ دارای توزیع F با درجات آزادی $V_1= n_1-1$ و $V_2= n_2-1$ است و

$$P\,(\, f_{1-\frac{\alpha}{2}} < F < f_{\frac{\alpha}{2}} \,) = ۱-\alpha$$

و یک فاصله اطمینان $(1-\alpha)$ ۱۰۰٪ برای نسبت واریانس ها به صورت زیر است:

$$\frac{1}{f_{\frac{\alpha}{2}}\ (V_1 \text{ و } V_2)}\frac{s_1{}^2}{s_2{}^2} < \frac{s_1{}^2}{s_2{}^2} < \frac{\sigma_1{}^2}{\sigma_2{}^2}\,.\ \ f_{\frac{\alpha}{2}}\ (V_2 \text{ و } V_1)$$

مثال : برای مقایسه پراکندگی نمرات دانشجویان دو دانشکده الف و ب یک نمونه ۱۶ تایی از دانشگاه الف و یک نمونه ۱۰ تایی از دانشکده به انتخاب نمودیم که $s_1{}^2=$ ۱۶ و $s_2{}^2=$ ۱۲ در سطح اطمینان ۹۰ درصد پراکندگی نمرات این دو دانشکده را مقایسه کنید.

$$1-\alpha = ۰/۹۰ \rightarrow \alpha = ۰/۱ \qquad s_1^2 = ۱۶ \ \text{ و } \ s_2{}^2 = ۱۲$$

$$V_1= n_1-1= 16-1=15 \qquad V_2 = n_2-1= 10-1= 9$$

$$\rightarrow f_{\frac{\alpha}{2}}\ (\,15\,,\,9\,) = 3/01\ ,\ f_{\frac{\alpha}{2}}\ (\,9\,,\,15\,) = 2/59$$

$$\longrightarrow \qquad 0/44 < \frac{\sigma_1^2}{\sigma_2^2} < 3/45$$

تحلیل واریانس (آنالیز واریانس):

به طور کلی تحلیل یک فاصله اطمینان نسبت واریانس ها در یکی از موارد زیر امکان پذیر است : الف) اگر کران های بالا و پایین اطمینان بزرگتر از یک باشد در سطح اطمینان مورد نظر می توان گفت σ_1^2 بیشتر از σ_2^2 است.

ب) اگر کران های بالا و پایین اطمینان کوچکتر از یک باشد، دز سطح اطمینان مورد نظر می توان گفت که σ_1^2 کوچکتر از σ_2^2 است.

ج) در غیر این حالت ها یعنی اگر فاصله بدست آورده شامل عدد یک باشد، نمی توان ادعا کرد که تفاوت معنی داری بین واریانس دو جامعه وجود دارد.

مثال: در مثال قبل آیا تفاوت معنی داری بین دو دانشکده وجود دارد؟

پس تفاوت معنی داری ندارد. (۳/۴۵ و ۰/۴۴)∋۱

تذکر : افزایش حجم نمونه باعث کاهش خطای معیار شده و فاصله اطمینان کوتاهتر می شود.

مثال های فصل ششم

۱- توزیع نمره های ارزشیابی کارمندان یک سازمان، نرمال با میانگین ۱۵ و انحراف معیار ۳ می باشد یک نمونه تصادفی ۹ تایی با چه احتمالی میانگین نمره ارزشیابی حداقل ۱۸ دارد؟

حل:

$$p(\bar{x} \geq 18) = p(Z \geq 3) = 0.0013$$

۲- توزیع جامع نیروی بازوی کارگران دارای میانگین ۱۱۰ و انحراف معیار ۵ می باشد. برای نمونه تصادفی ۷۵ تایی، با چه احتمالی میانگین نمونه بین ۱۰۹ و ۱۱۱ می باشد.

حل:

$$p(109 \leq \bar{x} \leq 111) = p(-1.73 \leq Z \leq 1.73) = 0.9164$$

۳- اگر $\bar{x}$ میانگین نمونه ۲۵ تایی از جامعه ای با تابع چگالی احتمال زیر باشد مطلوبست $p(1.5 < \bar{x} < 1.65)$

$$\mathrm{f(x)} = \begin{cases} \dfrac{x^3}{4} & \cdot < x < 2 \\ \cdot & \text{نقاطسایر} \end{cases}$$

حل:

$$E(X) = 1.6 \qquad \sigma^2 = \frac{8}{75}$$

$$p(1.5 < \bar{x} < 1.65) = p(\frac{1.5-1.6}{\sqrt{\frac{\frac{8}{75}}{25}}} < Z < \frac{1.65-1.6}{\sqrt{\frac{\frac{8}{75}}{25}}}) = 0.7176$$

۴- میانگین نمرات دانشجویان یک دانشکده ۵۴۰ و انحراف معیار آن ۵۰ می باشد. دو نمونه تصادفی با حجم $n_1=32$ و $n_2=50$ انتخاب می کنیم احتمال اینکه تفاضل میانگین نمرات این دو نمونه بیش از ۲۰ باشد چقدر است

حل:

$$P\ (\ |\bar{x}_1 - \bar{x}_2|)=0.0768> 20)$$

۵- تولیدکننده ای ادعا می کند که محصولات کارخانه اش دارای حد متوسط ۱۸/۳ کیلوگرم است. یک نمونه ۸ تایی از محصولات را وزن کردیم. نتایج زیر بدست آمد:

۲۰ و۱۷و۲۱و۱۹و۲۲و۲۱و۲۰و۱۶

آیا با ادعای کارخانه دار موافقید؟

حل: $t =\frac{19/5 - 18/3}{\sqrt{\frac{4/28}{8}}}=1.63$

$n = 9 \rightarrow \qquad V= n-1= 9-1= 8 \rightarrow t_{0/975} = 2/36$

۲/۳۶ > ۱/۶۳ پس ادعا درست است.

۶- شهری دارای ۲۰۰/۰۰۰ نفر جمعیت است.می خواهیم در صد افرادی که به کاندیدای خاصی رای می دهند را بر آورد کنیم. بدین منظور نمونه ای ۲۰/۰۰۰ نفری انتخاب می کنیم گه ۸۰۰۰ نفر آنها نسبت به کاندیدا نظر مثبت دارند. درصد آرای این کاندیدا در این شهر را برآورد کنید؟

حل: $P=\frac{x}{n} = \frac{8000}{20000} = 0/4$ یا ۴۰ درصد

۷- در یک نمونه تصادفی ۳ تایی از کارمندان حقوق ماهیانه ۹۱و۸۹ و ۹۰ هزار تومان بوده است فاصله اطمینان ۹۰ درصدی برای میانگین حقوق ماهیانه کارمندان این شرکت را بیابید؟

حل: $1-\alpha = ۰/۹۰ \rightarrow \alpha = ۰/۱$

$\bar{x}= \frac{91+89+90}{3} = 90 \qquad S^2 = \frac{1}{3-1}(91-90)^2 + (89-90)^2 + (90-90)^2 = 1$

$\frac{1}{\sqrt{3}}\times 90 + 2/9 < \mu<\frac{1}{\sqrt{3}}\times 90-2/9 \rightarrow \frac{1}{\sqrt{3}}\times < \mu< 90+ t_{0/05}\frac{1}{\sqrt{3}}90-t_{0/05} \times \xrightarrow{\text{توزیع } t}$

$\rightarrow \left(90 -\frac{2/9}{\sqrt{3}} \text{ و } 90 +\frac{2/9}{\sqrt{3}} \right)$

۸- نمونه های تصادفی از جامعه ای با $n_1= 100$ و $n_2= 150$ به قرار زیر است:

$\bar{x}_1 = 40 \qquad \sigma_1 =13 \bar{x}_2= 32 \qquad \sigma_2 =16$

یک فاصله اطمینان ۹۹ درصدی برای تفاضل میانگین بیابید.

حل:

$$\left((40-32) - 2/575 \times\sqrt{\frac{13^2}{100}+\frac{16^2}{150}} \; , (40-32) + 2/575 \times\sqrt{\frac{13^2}{100}+\frac{16^2}{150}} \right)$$

$= (3/26 . 12/74)$

چون هر دو کران مثبت است پس میانگین جامعه اول بیشتر است.

۹- در ۴۰ مرتبه پرتاب یک سکه ۲۴ مرتبه شیر ظاهر شده است. فاصله اطمینان ۹۸ درصدی را برای نسبت شیرهای که در تعداد نا محدود پرتاب سکه ظاهر می شوند بدست آورید.

حل: $p=\frac{x}{n} = \frac{24}{40} = 0/60$

$$\left(0/6 - 2/325 \times\sqrt{\frac{0/6\times 0/4}{40}} \; , \; 0/6 + 2/325 \times\sqrt{\frac{0/6\times 0/4}{40}} \right) =(0.42,0.78)$$

فصل هفتم

آزمون فرض ها

فرض آماری: فرض یا بیان حدسی درباره توزیع جامعه یا پارامترهای جامعه را فرض آماری گویند که بر دو نوع زیر است:

فرض ساده: اگر در فرض مقدار دقیقی پارامترهای جامعه باشد، آن فرض را ساده و در غیر این صورت آنرا مرکب گویند. مثلاً

فرض ساده $\sigma = 4$ و $\mu = 6$ و $P=\frac{1}{2}$

فرض مرکب $\sigma<3$ و $\mu > 16$ و $P<\frac{1}{4}$

تذکر: شروع یک آزمون فرض با دو فرض آماری می باشد که در مقابل هم قرار می گیرند که به صورت زیرند:

الف) فرض صفر: فرض آماری که برای رد شدن تنظیم می شود را فرض صفر گویند و با H_0 نشان می دهند.

ب) فرض مقابل: فرض آماری است که در مقابل فرض صفر قرار می گیرد و آنرا با H_1 نشان می دهند.

تذکر: همواره در صدد رد H_0 به نفع H_1 هستیم، چون رد یا قبول فرض آماری به کمک نمونه انجام می شود پس وقوع خطا اجتناب ناپذیر است. این خطاها به صورت زیرند:

الف) خطای نوع اول: رد H_0 در حالی که درست باشد را خطای نوع اول گویند و احتمال آنرا با $\propto$ نشان می دهند که در آن

$$\propto = P\left(\text{رد } H_0 \middle| \text{درست } H_0\right)$$

ب) خطای نوع دوم: قبول $H_.$ در حالی که نادرست باشد را خطای نوع دوم گویند و احتمال آنرا با β نشان می دهند که در آن

$$\beta = P\left(H_.\text{قبول} \,\middle|\, H_.\text{نادرست}\right)$$

تذکر: ∝ را میزان معنی داری یک آزمون یا سطح تشخیص و β -۱ را توان آزمون می گویند.

ناحیه بحرانی و قبولی: ناحیه ای باعث رد $H_.$ شود را ناحیه رد و ناحیه ای که باعث قبولی $H_.$ شود را ناحیه قبولی و مرز بین این دو ناحیه را ناحیه بحرانی گویند.

مثال: متغیر تصادفی x خطای اندازه گیری با یک دستگاه فیزیکی است که دارای توزیع نرمال با واریانس ۴ باشد. برای آزمون فرض $\begin{cases} H_.: \mu = ۰ \\ H_1: \mu = 1 \end{cases}$ یک نمونه تصادفی ۲۵ تایی را در نظر می گیریم. اگر ناحیه بحرانی به صورت $\bar{x} \geq ۰/۴$ باشد، خطاهای نوع اول و دوم را بیابید.

حل:

$$\propto \bar{x} = P(\geq ۰/۴) = P\left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{0/4-0}{\frac{2}{\sqrt{25}}}\right)$$

$$= P(z > 1) = 1-P(z<1) = 1- 0/8413 = 0/1587$$

$$\beta = P\left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{0/4-1}{\frac{2}{\sqrt{25}}}\right) = 0.0668$$

تذکر: آزمونی معنی دار است که $\propto \leq ۰/۰۵$ باشد.

آزمون های یک طرفه و دو طرفه:

یک آزمون فرض را یک طرفه گویند هرگاه فرض مقابل آن یک طرفه باشد یعنی به یکی از دو صورت زیر باشد:

$$\begin{cases} H_.:\ \theta = \theta_. \\ H_1:\ \theta > \theta_. \end{cases} \quad \text{یا} \quad \begin{cases} H_.:\ \theta = \theta_. \\ H_1:\ \theta < \theta_. \end{cases}$$

و یک آزمون فرض را دو طرفه گویند هرگاه فرض مقابل دو طرفه باشد

$$\begin{cases} H_.:\ \theta = \theta_. \\ H_1:\ \theta \neq \theta_. \end{cases}$$

آزمون میانگین و واریانس:

برای آزمون میانگین جامعه در سطح تشخیص $\propto$ با معلوم بودن واریانس جامعه، آزمونی به صورت $\begin{cases} H_.: \mu = \mu_. \\ H_1:\ \mu \neq \mu_. \end{cases}$ با آماره مناسب مثلا $Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ را می بایم و سپس مقادیر $Z_{\frac{\propto}{2}}$ و $Z_{1-\frac{\propto}{2}}$ را می یابیم. ناحیه قبول $-Z_{1-\frac{\propto}{2}} < Z < Z_{1-\frac{\propto}{2}}$ می باشد و ناحیه بحرانی $Z < Z_{\frac{\propto}{2}}$ و $Z > Z_{1-\frac{\propto}{2}}$ می باشد و اگر $n \leq 30$ از توزیع t یا χ^2 استفاده می شود.

مثال: یک نمونه ۵ تایی از جامعه ای نرمال انتخاب کردیم. نتایج به صورت زیر است:

۱.۹و۲.۴و۳.۵و۴.۲و۳

آیا می توان گفت میانگین جامعه از ۳/۵ کمتر است؟

حل: $s = ۰/۹$ و $\begin{cases} H_.: \mu = 3/5 \\ H_1:\ \mu < 3/5 \end{cases}$ $\bar{x} = \frac{3+4.2+3.5+2.4+1.9}{5} = 3$

$$T= \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{3-\frac{3}{5}}{\frac{\frac{0}{9}}{\sqrt{5}}} = -\frac{1}{238}$$

$n=5 \longrightarrow v=n-1=5-1=4$

$\propto= ۰/۰۵ \ \rightarrow\ = t_{1-\propto} = t_{0.95=-2.132}$

-2/132 >-1/238 پس فرض $H_.$ رد نمی شود.

مثال : اگر انحراف معیار نمونه ۲۵ تایی از یک نرمال ۲۸/. باشد ، آیا می توان ادعا کرد که انحراف معیار این جامعه کمتر از ۰/۳ است؟

حل: $n = 25$ و $\propto= ۰/۰۵$ $\begin{cases} H_.: \sigma = 0/3 \\ H_1:\ \sigma < 0/3 \end{cases}$

$$\mathcal{X}^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)(0/28)^2}{(0/3)^2}=20.907$$

$n=25 \longrightarrow v=n-1=25-1=24$

$\propto= ۰/۰۵ \ \mathcal{X}_{\propto}^{2} \longrightarrow \mathcal{X}_{0/05}^{2}= 13/8$

۱۳/۸ < ۲۰/۹۰۷ پس $H_.$ رد نمی شود.

آزمون نسبت:

اگر در یک جامعه دو جمله ای احتمال پیروزی یعنی P را مورد آزمون قرار دهیم به آن آزمون نسبت گویند که درآن $Z = \frac{x-np}{\sqrt{np_0 q_0}}$ آماره مناسب است و

اگر $H_1 : P \neq P.$ آنگاه ناحیه بحرانی به صورت $Z>Z_{1-\frac{\propto}{2}}$ و $Z<Z_{\frac{\propto}{2}}$ است

اگر $H_1: P < P.$ آنگاه ناحیه بحرانی به صورت $Z < -Z_{1-\propto}$ است

اگر $H_1: P > P.$ آنگاه ناحیه بحرانی به صورت $Z > Z_{1-\propto}$ است.

مثال : تیراندازی مدعی است که ۶۰ درصد تیرهایش به هدف می خورد. اگر ۵۵ تیر از صد تیر که شلیک کرده به هدف خورده باشد، در مورد این ادعا چه اظهار نظری می توان نمود؟

حل:

$$\begin{cases} H.: P = 0/6 \\ H_1: P \neq 0/6 \end{cases}$$

$\propto =$ ۰/۰۵ $\rightarrow \frac{\propto}{2} =$ ۰/۰۲۵ $Z_{1-\frac{\propto}{2}} = 1.96$

$x= 55$ و $P. = \frac{6}{10}$ و $n= 100$ و $Z_{\frac{\propto}{2}} = -1.96$

$$Z = \frac{X - nP_0}{\sqrt{nP_0q_0}} = \frac{55-100\times0/6}{\sqrt{100\times0/6\times0/4}} = -1/02 \qquad \in (-1/96, 1/96)$$

پس $H.$ رد نمی شود. یعنی دلیلی برای رد ادعای تیر انداز نداریم.

آزمون تفاضل دو نسبت :

اگر در دو جامعه آماری احتمال پیروزی به ترتیب P_1, P_2 باشد برای آزمون از آماره $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$ استفاده می کنیم که در آن $\hat{P}_1 = \dfrac{X_1}{n_1}, \hat{P}_2 = \dfrac{X_2}{n_2}$ و $\hat{P} = \dfrac{X_1 + X_2}{P_1 + P_2}$ و

اگر $H_1 : P_1 < P_2$ آنگاه ناحیه بحرانی $Z < Z_\alpha$ است

اگر $H_1 : P_1 > P_2$ آنگاه ناحیه بحرانی $Z > Z_{1-\alpha}$ است

اگر $H_1 : P_1 \# P_2$ آنگاه ناحیه بحرانی $Z < Z_{\frac{\alpha}{2}}$ و $Z < Z_{1-\frac{\alpha}{2}}$ است.

مثال : کارخانه ای دو نوع نوشابه تولید می کند معلوم شده که ۵۶نفر از ۲۰۰ نفر مصرف کننده نوشابه نوع A و ۲۹ نفر از ۱۵۰ نفر دیگر که مصرف کننده نوشابه هستند نوع را ترجیح می دهند.آیا می توان گفت که نوشابه نوع A از نوع B بهتر است؟ $\alpha = 0/06$ بگیرید.

حل:

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

$$\alpha = 0/06 \rightarrow Z_{1-\alpha} = Z_{0/94} = 1/555$$

$$\hat{P}_1 = \frac{X_1}{n_1} = \frac{56}{200} = 0/28$$

$$\hat{P}_2 = \frac{X_2}{n_2} = \frac{29}{150} = 0/19 \qquad \hat{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{56+29}{200+150} = 0/24$$

$$Z = \frac{0/28 - 0/19}{\sqrt{(0/24)(0/76)(\frac{1}{200} + \frac{1}{150})}} = 1/95 \rightarrow Z > Z_{1-\alpha}$$

پس فرض صفر رد می شود یعنی جواب مثبت است.

P- مقدار :

در بسیاری از کتابهای آمار اصطلاحی به نامp-value استفاده می شودکه آنرا **p- مقدار** یا مقدار احتمال گویند.متناظر با مقدار مشاهده شده آماره آزمون، P- مقدار عبارتست ازکمترین سطرح معنی داری که می توان فرض صفر را در آن رد کرد.اگر θآماره آزمون و u مقدار مشاهده باشد آنگاه :

الف) برای فرض مقابل $H_1 : \theta > \theta_0$ داریم: $P = P_{\theta_0}(U \geq u)$ - مقدار

ب)برای فرض مقابل $H_1 : \theta < \theta_0$ داریم: $P = P_{\theta_0}(U \leq u)$ - مقدار

ج) برای فرض مقابل $H_1 : \theta \neq \theta_0$ داریم: $P = 2\min\{P_{\theta_0}(U \leq u), P_{\theta_0}(U \geq u)\}$ - مقدار

مثال : تعداد اعضای هر خانواده در شهری بزرگ،متغیر تصادفی با توزیع $N(\mu,4)$ است. در یک نمونه تصادفی ۱۰۰ تایی از خانواده ها در این شهر،متوسط ۴/۲ بدست آمده است.فرض

$\begin{cases} H_0 : \mu = 3/9 \\ H_1 : \mu > 3/9 \end{cases}$را در سطح معنی داری ۰/۰۵ و ۰/۱ آزمون کنید.

حل: $n = 100, \bar{X} = 4/2 \quad , \ \sigma^2 = 4$

$$\text{مقدار} - P = P(\bar{X} > 4/2) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{4/2 - 3/9}{\frac{2}{\sqrt{100}}}\right) = P(Z > 1/5)$$

$$= 1 - P(Z < 1/5) = 1 - 0.9332 = 0/0668$$

با توجه به P– مقدار، فرض صفر در سطح تشخیص $\alpha = 0/05$ رد نمی شود ولی در سطح تشخیص $\alpha = 0/1$ رد می شود.

مثال : مثال قبل را به صورت یک آزمون دو طرفه انجام دهید.

حل:

$$\begin{cases} H_0 : \mu = 3/9 \\ H_1 : \mu \neq 3/9 \end{cases} \quad \xrightarrow{\text{مقدار}} -P = 2\min\{P(\bar{X} > 4/2), P(\bar{X} < 4/2)\}$$

$$= 2\min\{P(Z > 1/5), P(Z < 1/5)\} = 2\min\{0.0668, 0.9332\} = 2 \times 0/0668 = 0/1336$$

برای هر دو مقدار $\alpha = 0/05$ و $0/1$ فرض صفر رد نمی شود.

آزمون فرض واریانس:

هر گاه فرضیه ای درباره پراکندگی جامعه باشد برای صحت آن آزمونهای زیررا انجام می دهیم.

اگر $H_1 : \sigma^2 \neq \sigma_0^2$ آنگاه ناحیه بحرانی $\chi^2 > \chi^2_{\frac{\alpha}{2}}$, $\chi^2 < \chi^2_{1-\frac{\alpha}{2}}$

اگر $\sigma^2 > \sigma_0^2$ آنگاه بحرانی $\chi^2 < \chi^2_{1-\alpha}$ می باشد و $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

مثال : فرایند تولید قطعه ای تحت کنترل تلقی می شود اگر تغییر پذیری ضخامت ها،واریانس نا بیشتر از ۰/۳۶ داشته باشد.ضخامت یک نمونه تصادفی ۱۸ تایی دارای واریانس ۰/۶۸ بوده است.در سطح معنی داری ۰/۰۵ آزمون کنید آیا فرایند تحت کنترل است؟

حل: $n = 18, \ S^2 = 0/68, \ \sigma_0^2 = 0/36$ $\quad \begin{cases} H_0 : \sigma^2 \leq 0/36 \\ H_1 : \sigma^2 > 0/36 \end{cases}$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(18-1)(0/68)}{0/36} = 32/11$$

$$\begin{cases} \alpha = 0/05 \rightarrow \chi^2_\alpha = \chi^2_{0/95} = 27/6 \\ v = n-1 = 18-1 = 17 \end{cases}$$

چون $\chi^2 > \chi^2_\alpha$ پس H_0 رد می شود یعنی می توان گفت فرایند تولید تحت کنترل نیست.

آزمون فرض مقایسه واریانس دو جامعه آماری :

برای صحت فرضیه ای درباره پراکندگی دو جامعه از آزمون های زیر استفاده می کنیم.

اگر $H_1:\sigma_1^2>\sigma_2^2$ آنگاه ناحیه بحرانی $\frac{S_1^2}{S_2^2}>f_{\alpha,V_1,V_2}$ می باشد.

اگر $H_1:\sigma_1^2<\sigma_2^2$ آنگاه ناحیه بحرانی $\frac{S_1^2}{S_2^2}>f_{\alpha,V_2,V_1}$ می باشد.

اگر $H_1:\sigma_1^2\neq\sigma_2^2$ آنگاه ناحیه بحرانی برای حالت $S_1^2>S_2^2$ به صورت $\frac{S_1^2}{S_2^2}>f_{\frac{\alpha}{2},V_1,V_2}$ و اگر $S_1^2<S_2^2$ به صورت $\frac{S_1^2}{S_2^2}>f_{\frac{\alpha}{2},V_2,V_1}$ می باشد.

مثال: در گزارش مامور کنترل آمده که پراکندگی وزن محصولات تولید شده توسط ماشین A از ماشین B بیشتر است.مدیر کارخانه به منظور بررسی گزارش مامور کنترل از ماشین های AوB نمونه هایی انتخاب کرده که اطلاعات آن در جدول زیر آمده است:

ماشین A	ماشین B
$n_A=16$	$n_B=8$
$S_A=4/5$	$S_B=2/25$

با فرض نرمال بودن توزیع وزن محصولات تولید شده به وسیله ماشین های AوB،صحت گزارش را در سطح تشخیص ۵ در صد آزمون کنید.

حل:

$$\begin{cases} H_0: \sigma_A^2 = \sigma_B^2 \\ H_1: \sigma_A^2 > \sigma_B^2 \end{cases} \qquad F = \frac{S_A^2}{S_B^2} = \frac{(4/5)^2}{(2/25)^2} = 4$$

$$\begin{cases} \alpha = 0/05 \to f_{\alpha,15,7} = 3/511 \\ v_1 = 15, v_2 = 7 \end{cases}$$

$3/511 < 4$ پس فرض برابری واریانس ها رد می شود.

آزمون نیکویی بر ارزش:

اگر بخواهیم یک توزیع فراوانی مشاهده شد را با مقادیر متناظر توزیع مورد انتظار مقایسه کنیم از آزمون نیکویی بر ارزش استفاده می کنیم.فرض صفر در این آزمون به شکل زیر است:

$$H_0: P_1 = P_{1_0}, \quad P_2 = P_{2_0}, \quad ,..., P_K = P_{K_0}$$

که در آن مقادیر $P_{1_0}, P_{2_0,...,}, P_{K_0}$ رابطه $P_{1_0} + ... + P_{K_0} = 1$ صدق می کنند هدف این آزمون برازندگی مدل داده شده به وسیله فرض H_0 است.

فراوانی های مورد انتظار به صورت $e_i = nP_i$ محاسبه می شود.آماره ای که برای تعیین این اختلاف می توان استفاده کرد عبارتست از :

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nP_{1_0})^2}{nP_{1_0}} = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}$$

که در آن e_i, n_i به ترتیب نمادهایی برای فراوانی مشاهده شده و فراوانی مورد انتظار هستند. ناحیه بحرانی به صورت $\chi^2 > \chi^2_{\alpha k, -1}$ می باشد.

مثال: یک تاس در صورتی سالم که هر یک از شماره ها در $\frac{1}{6}$ دفعات ظاهر شوند.تاسی را۶۰ بار پرتاب کردیم و نتایج زیر بدست آمد:

شماره	1	2	3	4	5	6
تعداد	7	11	8	14	9	11

در سطح معنی داری $\alpha = 0/05$ آیا می توان گفت تاس سالم است؟

حل: $H_0 : P_i = \frac{1}{6} \quad , \; i = 1,2,\ldots,6$

شماره	1	2	3	4	5	6	جمع
n_i	7	11	8	14	9	11	60
P_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1
$e_i = n_i P_i$	10	10	10	10	10	10	60

$$\chi^2 = \sum_{i=1}^{6} \frac{(n_i - e_i)^2}{e_i} = \frac{(7-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} = 3/2$$

$$\alpha = 0/05 \rightarrow \chi^2_{\alpha k, -1} = \chi^2_{0/05,5} = 11/07$$

چون $\chi^2 \ngtr \chi^2_\alpha$ پس H_0 رد نمی شود یعنی تاس سالم است.

مثال: جدول زیر وضعیت عدد تصادفی است که به کمک کامپیوتر تولید شده است:

اعداد	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
فراوانی	13	8	12	11	10	5	7	11	11	12

در سطح تشخیص $\alpha = 0/05$ آزمون کنید ایا اعداد تولید شده از یک توزیع یکنواخت حاصل شده است؟

حل: اعداد تصادفی فوق از یک توزیع یکنواخت تولید شده اند : H_0

$$H_0 : P_j = \frac{1}{10}, i = 1,2,...,10$$

اعداد تصادفی	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	جمع
n_i	13	8	12	11	10	5	7	11	11	12	100
P_i	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	1
$e_i = n_i P_i$	10	10	10	10	10	10	10	10	10	10	100

$$\chi^2 = \frac{(13-10)^2}{10} + ... + \frac{(12-10)^2}{10} = 5/8$$

$$\alpha = 0/05,\ k = 10 \rightarrow \chi^2_{0/05,q} = 16/919$$

چون $\chi^2 \not> \chi^2_\alpha$ پس H_0 رد نمی شود یعنی تولید شده از یک توزیع یکنواخت است.

مثال های فصل هفتم

۱- تولید کننده باتری ادعا می کند که طول عمر آنها دارای توزیع نرمال با میانگین ۵۵ روز و انحراف معیار ۱۰ روز است.متوسط طول عمر ۲۵ عدد از این باتری ها ۵۰ روز بوده است.ادعای این تولید کننده را در سطح معنی داری ۰/۰۵ آزمون کنید.

حل:
$$\begin{array}{l} H_{.}:\ \mu = 55 \\ H_1:\ \mu < 55 \end{array}$$

$$P(\bar{x} \leq ۵۰ \mid \mu = ۵۵) = P\left(Z \leq \frac{50-55}{\frac{10}{\sqrt{25}}}\right) = P(Z \leq -2/5) = 0/0062$$

که $0/0062 < 0/05$ پس فرض صفر رد می شود.

۲-اگر میانگین نمونه تصادفی ۶۴ تایی از یک جامعه ۷۸/۸ و انحراف معیار جامعه ۱۲/۸ باشد آیا در میزان معنی داری ۰/۰۲ می توان گفت میانگین جامعه از ۷۵ بیشتر است؟

حل: $\bar{x}$= ۷۸/۸ و α = ۰/۰۲

$$\begin{array}{l} H_{.}:\ \mu = 75 \\ H_1:\ \mu > 75 \end{array}$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{78/8 - 75}{\frac{12/8}{\sqrt{64}}} = 2/375$$

$$\alpha = 0/02 \rightarrow Z_{1-\alpha} = Z_{0/98} = 2/054 \quad 2/375 > 2/054$$

پس H_0 رد می شود یعنی $\mu > 75$ است.

۳-از دو کلاس ۴۰ و ۵۰ نفری امتحان مشابهی گرفتیم.نتایج میانگین و واریانس این دو گروه به صورت زیر است:

$$\bar{X}_1 = 74,\ S_1^2 = 64,\ \bar{X}_2 = 78,\ S_2^2 = 49$$

آیا می توان این دو کلاس در سطح تشخیص ۰/۰۵ اختلاف دارند؟ در سطح تشخیص ۰/۰۱ چطور؟

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

$$, Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{74 - 78}{\sqrt{\frac{64}{40} + \frac{49}{50}}} = -2/49$$

الف) $\alpha = 0/05 \rightarrow \frac{\alpha}{2} = 0/025 \rightarrow Z_{1-\frac{\alpha}{2}} = Z_{0/975} = 1/96, -Z_{0/975} = -1/96$

$\rightarrow -2/49 \notin (-1/96 و 1/96)$ پس H_0 رد می شود.

ب) $\alpha = 0/01 \rightarrow \frac{\alpha}{2} = 0/005 \rightarrow Z_{1-\frac{\alpha}{2}} = Z_{0/995} = 2/575$

$\rightarrow -2/49 \in (-2/575, 2/575)$ پس H_0 رد نمی شود.

۴- مدیر عامل بورس ادعا کرده که ریسک (انحراف معیار) بازده سهام شرکت های عرضه کننده سهام کمتر از ۵ تومان است.بدین منظور یکی از کارگزاران ،۲۵ شرکت را به طور تصادفی انتخاب کرده که انحراف معیار بازده آن ۴ تومان بوده است اگر بازده شرکت ها از توزیع نرمال برخوردار باشد ، ادعا را در سطح $\alpha = 0/05$ آزمون کنید

حل:
$$\begin{cases} H_0 : \sigma^2 = 25 \\ H_1 : \sigma^2 < 25 \end{cases} \qquad n = 25, s = 4$$

$$\rightarrow \chi^2 = \frac{(25-1) \times 4^2}{25} = 15/36$$

$$\alpha = 0/05 \rightarrow \chi^2_{1-\alpha} = \chi^2_{0/95} = 13/85$$

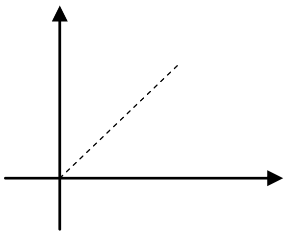
$\chi^2 \nless \chi^2_{1-\alpha}$ پس H_0 رد نمی شود یعنی ادعا مدیر عامل تائید نمی شود.

فصل هشتم

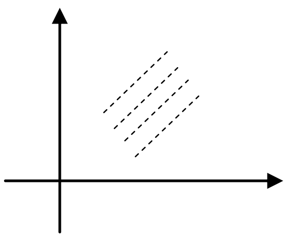
ضریب همبستگی و خط رگرسیون

مفهوم همبستگی:مطالعه رابطه بین متغیرها را در اصطلاح آماری مطالعه همبستگی می نامند که ممکن است این همبستگی مثبت و یا منفی باشد و یا اصلا وجود نداشته باشد.اندازه همبستگی بین متغیرها را ضریب همبستگی می گویند که معمولا در بازه [۱و۱-] قرار می گیرد و به صورت زیر تفسیر می شود:

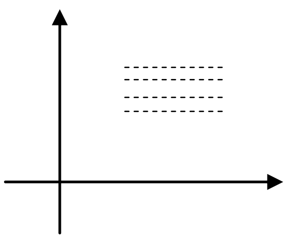
۱- اگر ضریب همبستگی برابر یک شود،آنگاه همبستگی را کامل و مستقیم نامند:



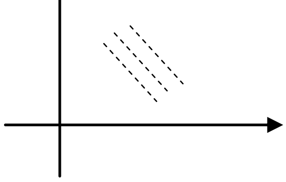
۲-اگر ضریب همبستگی در فاصله (۱و۰) باشد،همبستگی را مستقیم و ناقص گویند:



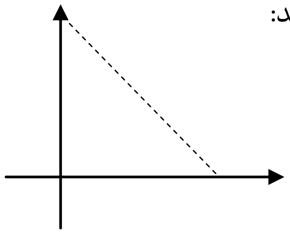
۳- اگر ضریب همبستگی صفر شود آنگاه دو متغیر همبستگی ندارند :



۴-اگر ضریب همبستگی درفاصله (۰و۱-) باشد،همبستگی را ناقص و معکوس گویند:



۵- اگر ضریب همبستگی ۱- شود آنگاه همبستگی را معکوس و کامل گویند:



انواع ضریب همبستگی : ضریب همبستگی بین متغیرها را با توجه به مقیاس آنها از روش های مختلف می توان محاسبه نمود در زیر نمونه هایی از این روش ها را بررسی می کنیم:

۱-ضریب همبستگی خطی پیرسن:

هنگامی که داده ها دارای مقیاس فاصله ای ویا نسبی هستند،ضریب همبستگی بین دو متغیر را می توان از طریق محاسبه ضریب همبستگی خطی پیرسن بدست آورد.به طوریکه ابتدا از دوصفت x,y نمونه های مستقل n تایی (x_1وy_1) و...و(X_nوy_n) استخراج نموده و سپس به شکل زیر ضریب همبستگی خطی r_{xy} یا r را می یابیم:

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^2-(\sum x)^2}\sqrt{n\sum y^2-(\sum y)^2}}$$

مثال: دانشجویی در ترم گذشته نمرات زیر را کسب نمود. ضریب همبستگی خطی را بیابید چه نتیجه ای می گیرید.

درس	میان ترم x	پایان ترم
زبان	۲۰	۲۰
آمار	۱۰	۱۱
معارف	۱۴	۱۵
ریاضی	۸	۱۲

حل: $n=4, \sum x=52, \sum y=58, \sum xy=816, \sum x^2=760, \sum y^2=890$

$$r=\frac{4\times 816-52\times 58}{\sqrt{4(760)-(52)^2}\sqrt{4(890)-(58)^2}}=0/966$$

پس همبستگی قوی و مستقیم است.

ضریب تعیین : پس از محاسبه rمی توان ضریب تعیین را از رابطه($cd=r^2(100)$)بدست آورد. ضریب تعیین نشان می دهد که چند در صد واریانس یا پراکنده متغیر وابسته y مربوط به متغیر مستقل x است و چند درصد تغییرات مربوط به خطا خواهد بود.

مثال: در مثال قبل ضریب تعیین را بیابید:

حل: $r=0/966 \rightarrow cd=(0/966)^2\times 100=93/39$

۲- ضریب همبستگی رتبه ای اسپیرمن:

برای داده های کمی که نرمال بودن آنها ثابت نشده باشد از ضریب همبستگی اسپیرمن استفاده می شود.ابتدا به هر یک از داده ها در گروه خودش رتبه های nو...و۳و۲و۱ می دهیم.مثلا به x_iها رتبه u_i و به y_iها رتبه v_iمی دهیم.یعنی به جای نمونه تصادفی دو بعدی $(x_1,y_1),...,(x_n,y_n)$ از رتبه های $(u_1,v_1),...,(u_n,v_n)$ استفاده می کنیم با فرض $d_i=u_i-v_i$ ضریب همبستگی اسپیرمن را که با ρ یا r_sنشان می دهند از رابطه زیر بدست می آید:

$$\rho=1-\frac{y\sum_{i=1}^{n}di^2}{n(n^2-1)}$$

مثال : چهار نقاشی از یک نقاش توسط داوران امتیازات زیر را کسب کرده اند ضریب همبستگی رتبه ای اسپیرمن را بیابید:

حل :

نقاشی	امتیاز داور اول u_i	امتیازداور دوم v_i	$d_i = u_i - v_i$
اول	۱۰	۹	۱
دوم	۸	۸	۰
سوم	۹	۷	۲
چهارم	۸	۱۰	-۱

$$\rho = 1 - \frac{6(1^2 + 0^2 + 2^2 + (-1)^2)}{4(4^2 - 1)} = 0.1$$

نظر کارشناسان هم جهتند ولی شدت همبستگی ضعیف است.

آزمون معنی دار بودن ضریب همبستگی :

ضریب همبستگی پیرسن نیز همانند دیگر شاخص های آماری از نمونه ای به نمونه دیگر متفاوت است.برای معنی داری همبستگی با فرض نرمال بودن متغیر های مستقل y وx فرض های $\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$ که در آن ρ ضریب همبستگی جامعه باشد را با آماره $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ که دارای توزیع t با درجه آزادی ۲۸ است آزمون می کنیم.فرض صفر وقتی رد می شود که $|t| > t_{\frac{\alpha}{2}}$ باشد.

مثال:اگر(yوx) دارای توزیع نرمال توام باشند و برای نمونه ۲۵ تایی r=0/۶ بدست آمده باشد،در سطح α=۰/۰۵ استقلال yوx را بیازمائید.

حل :

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases} \qquad t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0/6\sqrt{25-2}}{\sqrt{1-(0/6)^2}} = 3/597$$

$$\alpha = 0/05 \rightarrow t_{\frac{\alpha}{2}, n-2} = t_{0/025,23} = 2/069$$

چون $3/597 > 2/069$ پس H_0 رد می شود یعنی بین دو متغیر yوx همبستگی معنی داری وجود دارد

تذکر:اگر در آزمون معنی داری ضریب همبستگی فرض به صورت $\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$باشد از توزیع فیشر

با آماره $Z = \dfrac{Z_r - E(Zr)}{\dfrac{1}{\sqrt{n-3}}}$ استفاده می کنیم که در آن:

$$Z_r = \frac{1}{2} Ln \frac{1+r}{1-r}$$

$$E(Z_r) = \frac{1}{2} Ln \frac{1+\rho}{1-\rho}$$

$$\mathrm{var}(Z_r) = \frac{1}{n-3}$$

مثال : اگر در نمونه ای تصادفی به اندازهn=103از (xوy) ضریب همبستگی r=0/5 بدست آید،آزمون

$\begin{cases} H_0 : \rho = 0/6 \\ H_1 : \rho \neq 0/6 \end{cases}$را با $\alpha = 0.05$آزمون کنید.

حل :

$$Z_r = \frac{1}{2} Ln \frac{1+r}{1-r} = \frac{1}{2} Ln \frac{1+0/5}{1-0/5} = 0/55$$

$$E(Z_r) = \frac{1}{2} Ln \frac{1+\rho}{1-\rho} = \frac{1}{2} Ln \frac{1+0/6}{1-0/6} = 0/69$$

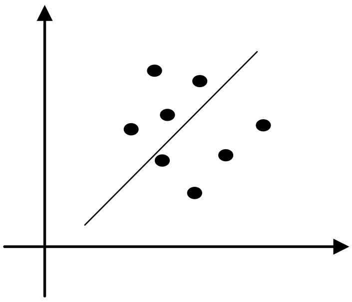
$$\mathrm{var}(Z_r) = \frac{1}{n-3} = \frac{1}{1.3-3} = 0/01$$

$$\rightarrow Z = \frac{Z_r - E(Z_r)}{\sqrt{\mathrm{var}(Z_r)}} = \frac{0/55 - 0/69}{\sqrt{0/01}} = -1/4 \ , \ Z_{\frac{\alpha}{2}} = Z_{0/025} = 1/96$$

$-1/4 \not> 1/96$ پس H_0 رد نمی شود.

رگرسیون خطی :

خط رگرسیون ابزاری است برای پیش بینی یک متغیر بر حسب متغیری که به آن وابسته است به عبارت دیگر اگر مجموعه ای از نقاط (xوy) در صفحه تعیین کنیم بهترین خطی که بتوان از این نقاط عبور داد که دارای کمترین خطا باشد را خط رگرسیون گویند.



معادله خط رگرسیون به صورت y=ax+b می باشد که در آن aوb از روابط زیر بدست می آیند:

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \quad , \quad b = \bar{y} - b\bar{x}$$

مثال : معادله خط رگرسیون مربوط به جدول زیر را بنویسید که در آن y متغیر وابسته است:

قدx	172	189	162	180	158
وزنy	84	80	75	93	67

حل:

$$n = 5, \sum xy = 69044, \sum x = 861, \sum y = 399, \sum x^2 = 148913, \sum y^2 = 32249$$

$$\rightarrow a = \frac{5 \times 69044 - 861 \times 399}{5 \times 148913 - (861)^2} = 0/51$$

$$b = \frac{399}{5} - 0/51 \times \frac{861}{5} = -8/022$$

$$\rightarrow y = 0/51x - 8/022$$

تذکر: اگر در معادله خط رگرسیون y=ax+b جای متغیر وابسته و مستقل را عوض کنیم آنگاه معادله خط رگرسیون به صورت $x=a'y+b'$ تغییر می کند که در آن

$$a'=\frac{n\sum xy-\sum x\sum y}{n\sum y^2-(\sum y)^2} \quad , \quad b'=\bar{x}-a\bar{y}$$

مثال : در مثال قبل اگر x متغیر وابسته باشد،آنگاه معادله خط رگرسیون را بنویسید:

حل :

$$a=\frac{5\times 69044-861\times 399}{5\times 32249-(399)^2}=0/82,\ b=\frac{861}{5}-0/82\times\frac{399}{5}=106/76$$

$$\rightarrow x=0/82y+106/76$$

تخمین بوسیله خط رگرسیون:

از خط رگرسیون می توان مقدار متغیر وابسته را بر حسب متغیر مستقل حساب نمود.مثلا اگر معادله خط رگرسیونی به صورت y=ax+b داده شده باشد آنگاه برآورد مقدار y بر حسب x نیز امکان پذیر است.

مثال : معادله خط رگرسیون به صورت $y=0/2x+3/18$ داده شده است.اگر **x=3/11** آنگاه مقدار yرا تخمین بزنید.

حل :

$$x=3/11\rightarrow y=0/2\times 3/11+3/18=3/802$$

رابطه بین ضریب همبستگی و خط رگرسیون :

اگر r ضریب همبستگی پیرسن و a ضریب xخط رگرسیون y=ax+b باشد آنگاه رابطه $a=\frac{rS_y}{S_x}$بین آنها برقرار است.

مثال:اگر $S_y = 3, S_x = 2/1, r = 0/6$ داده شده باشد و همچنین اگر $n = 10, \sum y = 4/5, \sum x = 321$ آنگاه معادله خط رگرسیون را بر حسب متغیر وابسته y بنویسید.

حل:

$$a = \frac{rS_y}{S_x} = \frac{0/6 \times 2/1}{3} = 0/42$$

$$b = \bar{y} - a\bar{x} = \frac{415}{10} - 0/42 \times \frac{321}{10} = 28/018$$

$$\rightarrow y = 0/42x + 28/018$$

خطای معیار برآورد:

اگر y مقدار مشاهده $\hat{y}_i$ مقدار پیش بینی شده باشد خطای پیش بینی را با $e_i = y_i - \hat{y}_i$ نشان می دهیم خطای استاندارد بر آورد از رابطه زیر بدست می آید:

$$S_e = S_y \sqrt{1 - r_{xy}^2}$$

که در آن r_{xy} ضریب همبستگی پیرسن است.

مثال : اگر $S_y = 2/5, \ r_{xy} = 0/8$ داده شده باشد، آنگاه خطای معیار یا خطای استاندارد پیش بینی را محاسبه کنید.

حل: $S_e = S_y \sqrt{1 - r_{xy}^2} = 2/5\sqrt{1 - (0/8)^2} = 1/5$

رگرسیون چند متغیره (چندگانه):

برای تغییر در یک متغیر وابسته بر حسب متغیرهای متعدد و پیش بینی آنها از رگرسیون چندگانه استفاده می کنیم.که مدل دو متغیره آن به صورت زیر است:

اگر x_1, x_2 متغیرهای مستقل و y متغیر وابسته باشد، مدل رگرسیون چندگانه پیش بینی کننده به شکل زیر است: $y = b_1x + b_2x + a$ که در آن:

$$b_1 = \frac{S_{x_1y}.SS_{x_2} - S_{x_1x_2}.S_{x_2y}}{SS_{x_1}.SS_{x_2} - (S_{x_1x_2})^2}$$

$$b_2 = \frac{S_{x_2y}.SS_{x_1} - S_{x_1x_2}.S_{x_1y}}{SS_{x_2}.SS_{x_1} - (S_{x_1x_2})^2}$$

و همچنین:

$$a = \frac{\sum y - b_1\sum x_1 - b_2\sum x_2}{n}$$

$$SS_{x_1} = \sum x_1^2 - \frac{(\sum x_1)^2}{n}, SS_{x_2} = \sum x_2^2 - \frac{(\sum x_2)^2}{n}$$

$$S_{x_1x_2} = \sum x_1x_2 - \frac{\sum x_1 \sum x_2}{n}, S_{xy} = \sum x_1 y - \frac{\sum x_1 \sum y}{n}$$

$$, S_{x_2y} = \sum x_2 y - \frac{\sum x_2 \sum y}{n}$$

مثال : اگر در جدول زیر y متغیر وابسته x_2, x_1 متغیر مستقل باشند آنگاه معادله خط رگرسیون چند گانه را بنویسید:

Y	2	3	2	5	5	7	5	6	7	8
X_1	1	2	2	3	4	5	6	8	9	10
X_2	0	0	0	1	0	1	0	1	1	1

حل:

$$\sum x_1 = 50, \sum x_2 = 5, \sum y = 45, \sum y^2 = 290, \sum x_1^2 = 340$$

$$\sum x_2^2 = 5, \sum x_1 y = 303, \sum x_1 x_2 = 35$$

$$\Rightarrow SSx_1 = 90, SSx_2 = 2/5, Sx_1x_2 = 10, S_{x_1y} = 78, S_{x_2y} = 10/5$$

$$, SSy = 87/5$$

$$\Rightarrow b_1 = \frac{78 \times 2/5 - 10 \times 10/5}{90 \times 2/5 - (10)^2} = -0/72$$

$$b_2 = \frac{10/5 \times 90 - 10 \times 78}{90 \times 2/5 - (10)^2} = 1/32$$

$$a = \frac{45 + 0/72 \times 50 - 1/32 \times 5}{10} = 0/24 \quad \rightarrow y = -0/72x_1 + 1/32x_2 + 0/24$$

رابطه ضریب همبستگی و رگرسیون چندگانه :

ضریب همبستگی بین سه متغیر را در دوبعد می توان از روابط زیر محاسبه نمود:

$$r_{yx_1} = \frac{Sx_1y}{\sqrt{SSx_1}\sqrt{SSy}}, r_{yx_2} = \frac{Sx_2y}{\sqrt{SSx_2}\sqrt{SSy}}, r_{x_1x_2} = \frac{S_{x_1x_2}}{\sqrt{SSx_1}\sqrt{SSx_2}}$$

هرگاه همبستگی بین متغیرهای مستقل صفر باشد ولی همبستگی هر یک از آنها با متغیر وابسته معنی دار باشد در این صورت مقدار ضریب تعیین از رابطه زیر بدست می آید:

$$R^2_{yx_1x_2} = r^2_{yx_1} + r^2_{yx_2}$$

و همچنین ضریب همبستگی جزیی یا نیمه تفکیکی $r_y(x_2, x_1)$ که همبستگی بین y با متغیر x_2 باشد در حالی که اثر متغیر x_1از متغیر x_2 حذف شده اما اثر آن روی y وحدف نشده است.همبستگی نیمه تفکیکی از رابطه زیر بدست می آید:

$$r_y(x_2,x_1)=\frac{r_{yx_2}-r_{yx_1}-r_{x_1x_2}}{\sqrt{1-r_{x_1x_2}^2}}$$

بنابراین R^2 از رابطه زیر قابل محاسبه است:

$$R_{yx_1x_2}^2=r_{yx_1}^2+r_{y(x_2x_1)}^2$$

مثال : در مثال قبل ضریب همبستگی بین سه متغیر را در دو بدو بیابید و ضریب همبستگی جزیی $r_y(x_2,x_1)$ و ضریب همبستگی چند متغیره $R_{y.x_1x_2}^2$ را بیابید.

حل :

$$r_{yx_1}=0/88, r_{yx_2}=0/71, r_{x_1x_2}=0/67, r_{y(x_2,x_1)}=0/16$$

$$R_{yx_1x_2}^2=r_{yx_1}^2+r_{y(x_2x_1)}^2=(0/88)^2+(0/16)^2=0/8$$

یعنی 80 در صد تغیرات y به وسیله x_2,x_1 تعیین می شود و بقیه به عوامل ناشناخته مربوط است.

مثال های فصل هشتم

۱- اگر ضریب همبستگی بین دو متغیر تصادفی ۰/۹ و ضریب همبستگی بین دو متغیر دیگر ۰/۳ باشد،همبستگی دو متغیر اول چند برابر قوی تر از متغیر دوم است؟

حل :

$$r_1 = 0/9, r_2 = 0/3 \rightarrow \frac{r_1^2}{r_2^2} = \frac{(0/9)^2}{(0/3)^2} = 9$$

۲- داده های زیر تعداد ساعات مطالعه ۱۰ دانشجو و نمره های انها در یکی از دروس آمار است.ضریب همبستگی رتبه ای اسپیرمن را بیابید:

تعدادساعاتX	۸	۵	۱۱	۱۳	۱۰	۵	۱۸	۱۵	۲	۸
نمره Y	56	44	79	72	70	54	94	85	33	65

حل : ابتدا uوv را در نظر می گیریم.داده هایی که تکراری باشند،میانگین رتبه ها را به عنوان رتبه مورد نظر در نظر می گیریم.به همین ترتیب r_i را در رتبه y قرار می دهیم:

X تعداد ساعات	Y نمره	u_i	V_i	$D_i=u_i-v_i$
8	56	4/5	4	0/5
5	44	2/5	2	0/5
11	79	7	8	-1
13	72	8	7	1
10	70	6	6	0
5	54	2/5	3	-0/5
18	94	10	10	0
15	85	9	9	0
2	33	1	1	0
8	65	4/5	5	-0/5

$$\rho = 1 - \frac{6((0/5)^2 + (0/5)^2 + (-1)^2 + 1^2 + (-0/5)^2 + (-0/5)^2)}{10(10^2 - 1)} = 0/98$$

یعنی میزان ساعات مطالعه با نمره رابطه مستقیم دارد و شدت این وابستگی زیاد است.

۳- معادله خط رگرسیون جدول زیر را بر حسب متغیر وابسته y نوشته و وزن شخصی را بر آورد کنید که سن او ۵۲ سال باشد.

X سن	45	42	56	48	42	35	58	40	39	50
Y وزن	65/3	63	95/2	75	69/9	59	94/9	62	65/5	87/2

حل :

$$n=10,\ \sum xy=34415/2, \sum x=455, \sum y=737, \sum x^2=21203$$

$$a=\frac{10\times 34415/2-455\times 737}{10\times 21203-(455)^2}=1/76$$

$$b=\frac{737}{10}-1/76\times\frac{455}{10}=-6/4 \rightarrow y=1/76x-6/4$$

$x=52 \rightarrow y=1/76\times 52-6/4=85/1$ کیلوگرم

فصل نهم

آزمون های ناپارمتریک

می دانیم که اگر توزیع جامعه نرمال باشد یا اینکه توزیع آماری متغیرها تقریبا نرمال باشند مثلا برای مقایسه میانگین دو گروه مستقل از آزمون t استفاده می کنیم.حال اگر این شرایط برقرار نباشد یا ندانیم که توزیع متغیر در جامعه آماری چیست دیگر نمی توان از آزمون استیودنت استفاده نمود.در این گونه مسائل باید از روش های دیگری که به توزیع جامعه آماری بستگی ندارند استفاده کنیم.این روش ها به روش های آمار ناپارامتری مشهورند این آزمون ها به قرار زیر است.

آزمون کلمو گروف – اسمیرنوف:

این آزمون یکی از انواع آزمونهای نیکویی برارزش است که در اکثر مواقع از این آزمون برای بررسی نرمال بودن توزیع جامعه استفاده می شود.به طور کلی دو نوع آزمون کلموگروف اسمیرنوف وجود دارد.اولین آزمون برای مقایسه توزیع یک جامعه با توزیع مفروض مانند توزیع نرمال می باشد و نوع دوم آن برای مقایسه توزیع دو جامعه با هم است.به نوع اول آزمکون کلموگروف - اسمیرنوف تک نمونه ای و به نوع دوم دو نمونه ای گویند.در این درس هدف استفاده از آزمون کلموگروف – اسمیرنوف تک نمونه ای برای بررسی نرمال بودن توزیع داده هاست. در این آزمون می خواهیم تابع توزیع تجربی داده های اخذ شده از جامعه را بدست آورده و آنرا با تابع توزیع نرمال مقایسه کنیم.

تابع توزیع تجربی را به صورت $\hat{F}(x) = \frac{\#(x_i \le x)}{n}$ که در آن # نشان دهنده تعداد n حجم نمونه است.

مثال : تابع توزیع تجربی داده های زیر را مشخص کنید:

۳/۹و ۳/۰و ۲/۸و ۲/۴و ۱/۲و ۰/۹

حل : کافی است با توجه به داده ها $\hat{F}(x)$ را بیابیم.

$$\hat{F}(x) = \begin{cases} 0 & x < 0/9 \\ \frac{1}{6} & 0/9 \le x < 1/2 \\ \frac{2}{6} & 1/2 \le x < 2/4 \\ \frac{3}{6} & 2/4 \le x < 2/8 \\ \frac{4}{6} & 2/8 \le x < 3/0 \\ \frac{5}{6} & 3/0 \le x < 3/9 \\ 1 & x \ge 3/9 \end{cases}$$

که تابعی است پله ای و پیوسته از راست

تذکر : آزمون کلموگروف – اسمیرنوف بر اساس مقایسه $\hat{F}(x), F(x)$ ساخته شده است. $F(x)$ یک تابع پیوسته و $\hat{F}(x)$ یک تابع پله ای است.پس آماره آزمون به صورت زیر تعریف می شود :

$$D = \underset{x}{SUP}\left|F(x) - \hat{F}(x)\right|$$

بکاربردن ***SuP*** به جایmax به این دلیل است که گاهی ماکسیمم مقادیر ممکن است موجود نباشد.

تذکر : توزیع دقیق D قابل حصول است و جداولی تحت عنوان جدول توزیع کلموگروف اسمیرنوف موجودند که مقادیر بحرانی D را معلوم می کنند . اگر d (مقدار عددی D) از عدد بحرانی جدول اول بزرگتر باشد فرضیه آماری (یعنی نرمال بودن توزیع داده ها) را رد می کنیم.

مثال : با آزمون نیکویی بر ارزش کلموگروف – اسمیرنوف بررسی کنید که آیا داده های جدول زیر از توزیع نرمال پیروی می کند :

0/6	0/8	1/1	1/2	1/4	1/7	1/8	1/9	2/2	2/4
2/5	2/9	3/1	3/4	3/4	3/9	4/4	4/9	5/9	5/2

x_i	$F(x_i)$	$\hat{F}(x_i)$	$F(x_i)-\hat{F}(x_i)$	x_i	$F(x_i)$	$\hat{F}(x_i)$	$F(x_i)-\hat{F}(x_i)$
۰/۶	۰/۱	۱/۲۰=۰/۰۵	۰/۰۵	۲/۵	۰/۴۲	۰/۵۵	-۰/۱۳
۰/۸	۰/۱۲	۰/۱	۰/۰۳	۲/۹	۰/۴۸	۰/۶	-۰/۱۲
۱/۱	۰/۱۸	۰/۱۵	۰/۰۳	۳/۱	۰/۵۲	۰/۶۵	-۰/۱۳
۱/۲	۰/۲	۰/۲	۰	۳/۴	۰/۵۷	۰/۷۵	-۰/۱۸
۱/۴	۰/۲۳	۰/۲۵	۰/۲	۳/۹	۰/۶۵	۰/۸	-۰/۱۵
۱/۷	۰/۲۸	۰/۳	-۰/۲	۴/۴	۰/۷۳	۰/۸۵	-۰/۱۲
۱/۸	۰/۳۰	۰/۳۵	-۰/۰۵	۴/۹	۰/۸۲	۰/۹	-۰/۰۸
۱/۹	۰/۳۲	۰/۴	-۰/۰۸	۵/۲	۰/۸۷	۰/۹۵	-۰/۰۸
۲/۲	۰/۳۷	۰/۴۵	-۰/۰۸	۵/۹	۰/۹۸	۱/۰	-۰/۰۲
۲/۴	۰/۴	۰/۵	۰/۱	–	–	–	–

حال آماره آزمون را به صورت $d = \left|-0/18\right| = 0/18$ محاسبه می کنیم.باید این مقدار را با مقدار بحرانی جدول کلموگروف – اسمیرنوف (قسمت دو طرفه و به ازای $n=20, \alpha=0/05$) مقایسه کنیم.

مقدار بحرانی از جدول مربوطه برابر ۰/۲۹۴ می باشد (جداول ضمیمه)

چون ۰/۲۹۴ > ۰/۱۸ پس فرضیه آماری در سطح خطای ۵ در صد رد نمی شود یعنی داده ها از توزیع نرمال پیروی می کنند.

آزمون گردش (Runstest):

روش های ناپارامتری زیادی برای این آزمون که آیا داده ها به صورت تصادفی جمع آوری شده اند وجود دارد. روش مورد نظر بر پایه گردش ها استوار است یک گردش در واقع دنبال ای از داده ها است که ویژگی های همانندی را نشان می دهدت و قبل و بعد از آن داده ی متفاوتی بوده یا اصلا داده ای نباشد.

مثال : فرض کنید دنباله ای از حروف nوd به شکل زیر داریم:

nnnnndddnnnnnnnnnnddnndddddnddnn

یک گردش مرکب از ۵ تا nو۴ تا d و ۱۰ تا n و ۲ تا d و ۲ تا n و ۴ تا d و ۱ تاn و ۲ تا d و ۲ تا n داریم که در مجموع تعداد گردش ها ۹ می باشد.

تذکر : تعداد گردش های خیلی زیاد یا خیلی کم می تواند نشانه ای از عدم تصادفی بودن باشد.آزمون گردش بر اساس ترتیبی است که داده ها اتفاق می افتند نه بر اساس فراوانی داده ها،فرضیه صفر آزمون مورد نظر این است که ترتیب تصادفی است و فرضیه مقابل عدم تصادفی بودن ترتیب است.اگر n_1 تعداد اجزای دنباله ای دارای ویژگی های همانند باشد و n_2 را برابر تعداد اجزای دنباله ای که دارای ویژگی همانند دیگری است در نظر بگیریم،وقتی n_2,n_1 کوچک باشند،آزمون فرضیه تصادفی بودن با توجه به جدوال خاصی انجام می شود که در کتاب های آمار موجود است.

اگر n_2,n_1هر دو بزرگتر از ۲۰ باشند، توزیع نمونه گیری تعداد گردش ها که آنرا با آمارهG نشان می دهیم را می توان با یک توزیع نرمال تقریب زد و مقادیر بحرانی را از جداول توزیع نرمال استاندارد استخراج نمود.میانگین و واریانس تعدادا گردش ها عبارتند از :

$$\mu_G = \frac{2n_1n_2}{n_1+n_2}+1$$

$$\sigma_G^2 = \frac{(2n_1n_2)(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

و آماره آزمون $Z = \frac{G - \mu_G}{\sigma_G}$ می باشد.

مثال : ماشینی قطعات معیوب A و قابل قبول B را به صورت دنباله زیر تولید می کند.در سطح خطای پنج در صد این ادعا را آزمون کنید که دنباله تولید محصولات تصادفی است.

B,B,B,B,A,B,B,B,B,A,B,B,B,B,A,B,B,B,B,A,B,B,B,B,A,B,B,B,B,A

حل :

$n_1 = 6$ تعدادکالامعیوب $\begin{cases} H_0: \text{دنباله تصادفی است} \\ H_1: \text{دنباله تصادفی نیست} \end{cases}$ و تعداد گردش ها G=12

$n_2 = 24$ تعداد کالای قابل قبول

$$\mu_G = \frac{2 \times 6 \times 24}{6 + 24} + 1 = 10/6$$

$$\sigma_G = \sqrt{\frac{(2 \times 6 \times 24)(2 \times 6 \times 24 - 6 - 24)}{(6 + 24)^2(6 + 24 - 1)}} = 1/69$$

$$Z = \frac{G - \mu_G}{\sigma_G} = \frac{12 - 10/6}{1/69} = 0/83$$

$$\alpha = 0/05 \rightarrow Z_{1-\frac{\alpha}{2}} = Z_{0/975} = 1/96$$

چون $1/96 \not> 0/83$ پس فرض صفر رد نمی شود

آزمون علامت :

در این آزمون به جای مقادیر عددی مشاهده شده از علائم مثبت و منفی به عنوان یافته های مورد نظر استفاده می شود.این آزمون به خصوص در مواردی که کاربرد مقادیر عددی امکان ندارد یا اصولا مناسب باشد ولی می توان یافته ها را به صورت رتبه بندی در آورده کاربرد دارد.آزمون علامت دارای دو حالت است :

حالت اول : حالت تک نمونه ای که برای مقایسه میانه جامعه با یک ثابت استفاده می شود.

حالت دوم : حالت زوج نمونه ای که برای مقایسه میانه دو جامعه وابسته استفاده می شود.

ابتدا حالت دوم را توضیح می دهیم و سپس اشاره می کنیم که چگونه می توان از آن برای انجام آزمون تک نمونه ای استفاده کرده فرضیه صفری که به وسیله آزمون علامت سنجیده می شود بر پایه احتمالهای زیر استوار است :

$$P(X_A > X_B) = P(X_A < X_B) = 0/5$$

که در آن X_A نمره یا قضاوت تحت یکی از شرایط و X_B نمره یا قضاوت تحت یکی دیگر از شرایط است. یا فرض آماری این است که میانه تفاوت ها برابر صفر است.

اگر بین علائم مثبت و منفی از نظر تعداد تفاوتی باشد،.H را رد می کنیم.

مثال : تحقیقی برای بررسی اثر هیپنوتیسم روی کاهش درد انجام شده و نتایج برای افرادی که به تصادف انتخاب شده اند مطابق جدول زیر است. در سطح معنی داری ۵ در صد این ادعا را که اندازه های حساسیت پس از انجام هیپنوتیسم کمتر شده است آزمون کنید.

افراد	A	B	C	D	E	F	G	H
قبل هیپنوتیسم	۶/۶	۶/۵	۹	۱۰/۳	۱۱/۳	۸/۱	۶/۳	۱۱/۶
بعد هیپنوتیسم	۶/۸	۲/۴	۷/۴	۸/۵	۸/۱	۶/۱	۳/۴	۲

حل : ابتدا اختلاف ها را بدست می آوریم:

	A	B	C	D	E	F	G	H
Dتفاوت	۰/۲	-۴/۱	-۱/۶	-۱/۸	-۳/۲	-۲	-۲/۹	-۹/۶
علامت	+	-	-	-	-	-	-	-

حال اگر M_e میانه تفاضل افراد جامعه در مورد درد بعد از هیپنوتیسم منهای قبل از آن باشد آنگاه آزمون به صورت $\begin{cases} H_0: M_e = 0 \\ H_1: M_e < 0 \end{cases}$ می باشد یعنی میانه منفی منجر به رد .H می شود.

طبق جدولv علامت منفی و یک علامت مثبت است پس :

P (علامت مثبت) = P (علامت منفی) = ۰/۵

و چون n<30 پس از توزیع دو جمله ای استفاده می کنیم تعداد علامت های مثبت را با γ^+ نشان می دهیم و مقدار P- مقدار را می یابیم.اگر P- مقدار کمتر از α باشد فرضیه آماری را رد می کنیم.آماره آزمون $\gamma = \gamma^+$ است اگر $H_1: M_e > 0$ باشد آنگاه:

$$P-value = P_{H_0}(X \geq \gamma = \gamma^+) = \sum_{x=\gamma^+}^{n} \binom{n}{x} (0/5)^x \times (0/5)^{n-x} = \sum_{x=\gamma^+}^{n} \binom{n}{x} (0/5)^x$$

و اگر $H_1 : M_e < 0$ آنگاه:

$$P-value = P_{H_0}(X \leq \gamma = \gamma^+) = \sum_{x=0}^{\gamma^+} \binom{n}{x} (0/5)^x \times (0/5)^{n-x} = \sum_{x=0}^{\gamma^+} \binom{n}{x} (0/5)^n$$

و اگر $H_1 : M_e \neq 0$ باشد آنگاه آماره آزمون $\gamma = \min\ (\gamma^+ \text{ و } \gamma^-)$ است پس:

$$P-value = 2P_{H_0}(X \leq \gamma = \min(\gamma^+ \text{ و} \gamma^-)) = 2 \sum_{x=0}^{\min(\gamma^+ \text{ و}\gamma^-)} \binom{n}{x} (0/5)^n$$

چون آزمون فوق یک طرفه چپ است پس :

$$P-value = P_{H_0}(X \leq \gamma^+ = 1) = \sum_{x=0}^{1} \binom{n}{x} (0/5)^x \times (0/5)^{n-x} = \sum_{x=\gamma^+}^{1} \binom{n}{x} (0/5)^n = 0/035$$

چون مقدار کمتر از ∝=۰/۰۵ است پس در سطح خطای ۵ درصد فرضیه آماری رد می شود یعنی به طور متوسط درد بعد از هیپنوتیسم کاهش یافته است.

آزمون من – ویتنی (آزمون رتبه ایU):

برای برابری میانه در دو جامعه مستقل از آزمون من – ویتنی استفاده می کنیم.فرض صفر در این آزمون این است که توزیع ها یکسان نیستند.

اگر w_1 مجموع رتبه های مقادیر نمونه اول و w_2 مجموع رتبه های مقادیر نمونه دوم باشدآنگاه $w_2+ w_1$ مجموع $n_1 + n_2$عدد صحیح مثبت است.یعنی برای هر جفت w_2 و w_1داریم:

$$w_1 + w_2 = \frac{(n_1+n_2)(n_1+n_2+1)}{2}$$

و از دو آماره زیر نیز می توان استفاده کرد:

$U_1 = w_1 - \frac{n_1(n_1+1)}{2}$ و$U_2 = w_2 - \frac{n_2(n_2+1)}{2}$

میانگین و واریانس U_1و U_2به ترتیب عبارتند از :

$$\mu_{U_1} = \mu_{U_2} = \frac{n_1\, n_2}{2} \text{و} \sigma_{v_1}{}^2 = \sigma_{v_2}{}^2 = \frac{n_1\, n_2(n_1 + n_2 + 1)}{12}$$

و آماره آزمون با داشتن میانگین و واریانس و توزیع نرمال استاندارد به صورت $Z = \frac{U_1 - \mu_{U_1}}{\sigma_{v_1}}$

می باشد و در صورتیکه $|Z| > Z_{1-\frac{\alpha}{2}}$آنگاه فرضیه آماری در سطح خطای $\propto$ رد می شود.

مثال: با توجه به داده های جدول زیر که مربوط به طول عمر دو چراغ روشنایی است، در سطح تشخیص ۵ درصد این فرضیه را که دو نمونه از جامعه یکسان آمدند در مقابل فرضیه جانشین یعنی متوسط طول عمر چراغ نوع A کمتر از متوسط طول عمر چراغ B است را آزمون کنید.

نوع A	۱۶/۹	۱۳	۱۵/۴	۱۴/۱	۱۷	۱۶/۶	۱۳/۲	۱۱/۲	۱۴/۹	
نوع B	۱۹/۴	۱۵/۳	۱۲/۲	۱۸/۹	۲۱/۲	۱۶/۲	۱۸/۳	۱۴/۷	۱۹/۸	۱۵/۲

حل :

$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_B < \mu_B \end{cases}$ اگر داده ها را روی هم ریخته و از کوچک به بزرگ ردیف کنیم و به آنها رتبه های ۱و۲و...و۱۹ دهیم مقادیر نوع A رتبه های ۱۴و۱۳و۱۲و۱۰و۷و۵و۴و۳و۱ می گیرند پس:

$$W_1 = 1+3+4+5+7+10+12+13+14=69$$

و مقادیر نوع B رتبه های ۱۹و۱۸و۱۷و۱۶و۱۵و۱۱و۹و۸و۶و۲ پس

$2+6+8+9+11+15+16+17+17+18+19\ W_2=$ و $n_1= 9$ و$n_2= 10$

$$\mu_{U_1} = \frac{9\times10}{2} = U_1 U_1 45 \sigma_{v_1}{}^2 = \frac{9\times10\,(9+10+1)}{12} = 150 = 69 - \frac{9\times10}{2} = 24$$

$$Z = \frac{U_1 - \mu_{U_1}}{\sigma_{v_1}} = \frac{24-45}{\sqrt{150}} = -1/71$$

$$\propto = 05/0 \rightarrow Z_{\propto} = -1/645$$

-۱/۷۱ > -۱/۶۴۵ پس فرض H_0 رد می شود یعنی طول عمر چراغ نوع A از طول عمر چراغ نوع B کمتر است.

آزمون ویلکاکسون:

برای برابری میانه های دو گروه وابسته از آزمون ویلکاکسون استفاده می کنیم.آماره این آزمون برای $n \geq 30$ به شکل $Z=\frac{T-\frac{n\,(n+1)}{4}}{\sqrt{\frac{n\,(n+1)\,(\,2n+1)}{24}}}$ و اگر مقدار z کمتر یا مساوی ناحیه بحرانی شود آنگاه H_0 را رد می کنیم.

مثال : به ۱۳ نفر آزمونی برای سنجش داده شده است سپس به آنها داروی آرام بخش داده شده و دوباره آزمون انجام شده است. در سطح معنی داری ۵ درصد آزمون کنید که داروی آرام بخش اثری نداشته است.داده ها در جدول زیر داده شده اند:

افراد	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳
قبل از آرام بخش	۶۷	۷۸	۸۱	۷۲	۷۵	۹۲	۸۴	۸۳	۷۷	۶۵	۷۱	۷۹	۸۰
بعد از آرام بخش	۶۸	۸۱	۸۵	۶۰	۷۵	۸۱	۷۳	۷۸	۸۴	۵۶	۶۱	۶۴	۶۳

حل: M_1: میانه میزان تفکر قبل از آرام بخش
$$\begin{cases} H_0: M_1=M_2 \ \text{آرام بخش اثر ندارد} \\ H_1: M_1\neq M_2 \ \text{آرام بخش اثر دارد} \end{cases}$$

M_2: میانه میزان تفکر بعد از ارام بخش

افراد	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳
اختلاف ها	-۱	-۳	-۴	۱۲	۰	۱۱	۱۱	۵	-۷	۹	۱۰	۱۵	۱۷
رتبه اختلاف ها	۱	۲	۳	۱۰	-	۸/۵	۸/۵	۴	۵	۶	۷	۱۱	۱۲
رتبه علامت دار اختلاف ها	-۱	-۲	-۳	۱۰	-	۸/۵	۸/۵	۴	-۵	۶	۷	۱۱	۱۲

۶۷=۱۲+۱۱+۷+۶+۴+۸/۵+۸/۵+۱۰= مجموع رتبه های مثبت

۱۱=۵+۳+۲+۱= مجموع قدرمطلق رتبه های منفی

T=min(67,11)=11

با توجه به جدول رتبه ای ویلکالسون و $\alpha = 0/05$ و $n = 12$ (نفر پنجم را کنار می گذاریم)

مقدار بحرانی ۱۴ می باشد پس چون T=11 کمتر از ۱۴ می باشد فرض صفر رد می شود. یعنی داروی آرام بخش تاثیر دارد.

آزمون کروسکال – والیس :

برای مقایسه میانه بیش از دو جامعه مستقل از آزمون کروسکال – والیس استفاده می شود آنالیز واریانس یک طرفه برای آزمون فرض بکار می رود که بخواهیم تساوی میانگین های چند نمونه را جدا از هم بررسی کنیم. آماره این آزمون دارای توزیع فیشر می باشد و جامعه های مورد نظر نرمالند با واریانسهای برابر.اگر k تعداد جامعه و $n_1,...,n_k$ نمونه هایی از آنها باشند و $n = n_1 + ... + n_k$ آنگاه آماره این آزمون در صورتیکه R_i مجموع رتبه های نمونه i م باشد به صورت زیر است :

$$H = \frac{12}{n(n+1)}(\frac{R_1^2}{n_1} + ... + \frac{R_k^2}{n_k}) - 3(n+1)$$

مثال : داده های جدول زیر نمرات امتحان نمونه هایی از سه گروه دانشجویان اند که زبان انگلیسی را به سه روش فرا گرفتند. در سطح معنی دارای ۰/۰۵ = α این ادعا را آزمون کنید که سه روش دارای تاثیر یکسانی بودند:

روش اول	۹۷	۸۷	۷۴	۹۱	۸۸	۹۴	
روش دوم	۸۰	۷۲	۶۱	۸۴	۷۹	۸۲	۸۵
روش سوم	۶۹	۷۶	۷۲	۶۷	۸۹		

حل : فرض آماری $H_0 = \mu_1 = \mu_2 = \mu_3$ را در نظر گرفته و داریم :

اگر تمام داده ها را به عنوان یک نمونه در نظر بگیریم و رتبه های آنها را بیابیم ، این رتبه ها از یک تا ۱۸ می باشد پس:

$$R_1 = 6+13+14+16+17+18 = 84$$

$$R_2 = 1+4/5+8+9+10+11+12 = 55/5$$

$$R_3 = 2+3+4/5+7+15 = 31/5$$

$$n_1 = 6,\ n_2 = 7,\ n_3 = 5 \rightarrow n = 18$$

$$\rightarrow H = \frac{12}{18(18+1)}\left(\frac{(84)^2}{6} + \frac{(55/5)^2}{7} + \frac{(31/5)^2}{5}\right) - 3(18+1) = 6/67$$

$$\chi^2_{0/95\text{و}2} = 5/991$$

$5/991 < 6/67$ پس فرض H_0 رد می شود یعنی سه روش فوق تاثیر یکسانی ندارند.