«فصل هشتم» «برآورد میانگین جامعه» «نکات و سؤالات مهم»

«در این فصل توضیح می‌دهیم که چگونه پارامترهای جامعه را از روی داده‌های نمونه تصادفی ساده برآورد کنیم.»

میانگین و انحراف معیار دو پارامتر مهم جامعه هستند که در صورت نامعلوم بودن مقادیرشان باید آن‌ها را برآورد کنیم.

برآورد:
- نقطه‌ای => هرگاه مشخصه‌ای از جامعه با یک عدد برآورد شود.
- فاصله‌ای => با استفاده از برآورد نقطه‌ای حدودی برای جامعه در نظر می‌گیریم.

برآوردکننده آماره‌ای است که برای برآورد کردن مشخصه‌ای از جامعه به کار می‌رود.

مثلاً اگر پارامتر جامعه θ باشد، برآوردکننده‌ی آن $\hat{\theta}$ می‌باشد.

«ویژگی‌های یک برآوردکننده‌ی خوب نقطه‌ای»

الف) نااریب بودن

۱- آماره‌ی $\hat{\theta}$ را یک برآوردکننده‌ی نااریب پارامتر θ گوییم هرگاه میانگین توزیع نمونه‌ای آن برابر با پارامتر θ باشد یعنی:

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

اگر رابطه‌ی بالا برقرار نباشد $\hat{\theta}$ برآوردکننده‌ی اریب θ است.

۲- آماره‌ی $\bar{x}$ یک برآوردکننده‌ی نااریب برای μ است هرگاه: $E(\bar{x})=\mu$

۳- واریانس نمونه S^2 وقتی جامعه نامتناهی باشد، برآوردکننده نااریب σ^2 است هرگاه $E(S^2)=\sigma^2$

نکته => S یک برآوردکننده اریب σ می‌باشد => $E(S)\neq\sigma$

ب) کارایی

هرگاه دو برآوردکننده نااریب برای پارامتر θ موجود باشند، برآوردکننده‌ای که واریانس کوچکتری دارد، کارایی نسبی بیشتری نسبت به دیگری دارد.

ج) سازگاری

برآوردکننده‌ای مانند $\hat{\theta}$ را یک برآوردکننده‌ی سازگار برای پارامتر θ می‌نامیم هرگاه با افزایش n، $\hat{\theta}$ با احتمال بیشتری به θ نزدیک شود.

برآورد فاصله‌ای:

در این روش با استفاده از برآوردکننده نقطه‌ای پارامتر جامعه، حدودی برای پارامتر مورد نظر پیدا می‌کنیم، این حدود یک بازه یا فاصله را مشخص می‌کند.

حد پایین ← $L < \theta < U$ → حد بالا

پارامتر مجهول

ضریب اطمینان:

با c نمایش داده می‌شود، مقدارش برابر است با احتمال اینکه فاصله‌ی اطمینان شامل مقدار واقعی پارامتر برآورد شده باشد.

$$c = P(L < \theta < U)$$

برآورد نقطه‌ای میانگین جامعه:

بهترین برآوردکننده میانگین جامعه، $\bar{x}$، میانگین نمونه‌ای است که در صورت بزرگ بودن اندازه‌ی نمونه، جامعه دارای توزیع تقریبی نرمال با میانگین μ و انحراف معیار $\frac{\sigma}{\sqrt{n}}$ است.

خطای برآورد:

اگر خطای برآورد را با d نشان دهیم مقدار آن از رابطه‌ی زیر بدست می‌آید:

$$d = \left| \text{مقدار واقعی پارامتر} - \text{مقدار برآورد پارامتر} \right|$$

نکته ⇐ خطای برآورد μ برای نمونه‌های بزرگ به صورت روبه‌رو است: $d = |\bar{x} - \mu|$

چون در رابطه‌ی بالا مقدار μ نامعلوم است ⇐ خطای برآورد نیز نامعلوم می‌گردد اما اگر اندازه‌ی نمونه بزرگ باشد می‌توانیم یک عبارت احتمالی به صورت زیر درباره‌ی مقدار خطا بنویسیم:

$$P(d < Z_{\frac{\alpha}{2}} \sigma_{\bar{x}}) = 1 - \alpha$$

✓ $(1-\alpha)$ ٪۱۰۰ اطمینان می‌دهد که خطای برآورد کمتر از $Z_{\frac{\alpha}{2}} \sigma_{\bar{x}}$ است.

$$\begin{cases} \%90 \rightarrow Z_{/05} = 1/645 \\ \%95 \rightarrow Z_{/025} = 1/96 \\ \%99 \rightarrow Z_{/005} = 2/575 \end{cases}$$

مثال <= جامعه‌ای بزرگ دارای انحراف معیار $\sigma = 21$ و میانگین نامعلوم μ است. برای برآورد μ، نمونه‌ای به اندازه $n = 100$ از جامعه انتخاب کرده و مقدار میانگین نمونه‌ای، $\bar{x} = 871$ مشاهده شده است. اگر بخواهیم با اطمینان ٪۹۵ قضاوت کنیم، حداکثر خطای برآورد میانگین چقدر است؟

$$P(d < Z_{\frac{\alpha}{2}}\sigma_{\bar{x}}) \Rightarrow ?$$

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow \boxed{Z_{0.025} = 1.96}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{21}{10} \Rightarrow \boxed{\sigma_{\bar{x}} = 2.1}$$

$$\Rightarrow P(d < 1.96 \times 2.1) \Rightarrow P(d < 4.116)$$

<= حداکثر خطای برآورد میانگین $d = 4.116$ می‌باشد.

برآورد نقطه‌ای واریانس جامعه:

بهترین برآوردکننده σ^2 اگر σ^2 معلوم نباشد، S^2، واریانس نمونه‌ای آماره‌ای است که ویژگی‌های یک برآوردکننده خوب را دارد. بنابراین نسبت به سایر برآوردکننده‌ها σ^2 بهترین است.

مثال <= می‌خواهیم میانگین روزانه‌ی تعداد قطعات تولید شده توسط یک ماشین را برآورد کنیم. تعداد قطعات تولید شده در $n = 50$ روز را ثبت کرده، میانگین و انحراف معیار نمونه‌ای را محاسبه کرده‌ایم <= $\bar{x} = 871$ ، $S = 21$

احتمال اینکه خطای برآورد میانگین کمتر از ۵ باشد چقدر است؟

$$P(|\bar{x} - \mu| < 5) \Rightarrow P(-5 < \bar{x} - \mu < 5) \Rightarrow P(-5 + \mu < \bar{x} < 5 + \mu)$$

مقدار استاندارد شده
$$= P\left(\frac{-5 + \mu - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{5 + \mu - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{-5}{\frac{21}{\sqrt{50}}} < Z < \frac{5}{\frac{21}{\sqrt{50}}}\right)$$

$$= P(-1.68 < Z < 1.68) = 2P(0 < Z < 1.68) = 2 \times 0.4535 = 0.907$$

مثال <= نمونه‌ای تصادفی به اندازه‌ی $n = 100$ از جامعه‌ای نامتناهی با میانگین μ و واریانس σ^2 استخراج شده است. اگر بخواهیم با ۹۵ درصد اطمینان قضاوت کنیم حداکثر خطای برآورد μ برای مقدار $\sigma = 100$ داده شده چقدر است؟

$$1-\alpha = 0.95 \Rightarrow Z_{\frac{\alpha}{2}} = 1.96$$

$$d = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow d = 1.96 \times \frac{100}{10} \Rightarrow \boxed{d = 19.6}$$

$$d = |x - \mu| \Rightarrow \mu = ?$$

★ هرچه قدر σ افزایش یابد، خطای برآورد نیز افزایش می‌یابد و بالعکس. اما هرچه n افزایش یابد d کاهش می‌یابد و بالعکس.

www.pnu-m-s.com

مثال =) برای برآورد μ ، میانگین تعداد افرادی که در پرواز ساعت ۱۶ به مقصدی حاضر نمی‌شوند، نمونه‌ای مرکب از ۱۰۰ روز را بررسی و داده‌های حاصل را در جدول زیر خلاصه کردیم.

الف) مقدار μ و σ را بدست آورید [برآورد کنید؟]

ب) احتمال اینکه خطای برآورد میانگین کمتر از ۰/۲ باشد چقدر است؟

تعداد غایبین X	۰	۱	۲	۳	۴	۵	۶	
f_i تعداد روزها	۲۰	۳۷	۲۳	۱۵	۴	۰	۱	
$x_i f_i$	۰	۳۷	۴۶	۴۵	۱۶	۰	۶	$\sum x_i f_i = 150$
$f_i x_i^2$	۰	۳۷	۹۲	۱۳۵	۶۴	۰	۳۶	$\sum f_i x_i^2 = 364$

حل الف)

$$\bar{x} = \mu = \frac{\sum x_i f_i}{n} = \frac{150}{100} = 1.5 \qquad \boxed{\bar{x} = \mu = 1.5}$$

واریانس
$$S^2 = \sigma^2 = \frac{\sum f_i x_i^2 - \frac{\left(\sum (x_i f_i)\right)^2}{n}}{n-1} = \frac{364 - \frac{(150)^2}{100}}{99} = 1.40$$

$$\boxed{S = \sigma = 1.18}$$

انحراف معیار $= \sqrt{1.4} = 1.18$

حل ب)

$$P(|\bar{x} - \mu| < 0.2) = P(-0.2 < \bar{x} - \mu < 0.2) = P(-0.2 + \mu < \bar{x} < 0.2 + \mu)$$

استاندارد سازی $\Rightarrow$
$$P\left(\frac{-0.2 + \mu - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{0.2 + \mu - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{-0.2}{\frac{1.18}{10}} < Z < \frac{0.2}{\frac{1.18}{10}}\right)$$

$$= P(-1.69 < Z < 1.69) = 2(0 < P < 1.69) = 2 \times 0.4545 = 0.909$$

برآورد فاصله‌ای μ برای نمونه‌ی بزرگ:

الف) اگر انحراف معیار جامعه معلوم باشد داریم:

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = (L, U)$$

ب) اگر انحراف معیار جامعه معلوم نباشد داریم:

$$(L, U) = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

مثال => مبالغ هزینه‌ی پستی، برای نمونه‌ای به اندازه‌ی $n=400$ بسته که در روزی خاص به وسیله‌ی اداره‌ی پست جابه‌جا شده‌اند ثبت شده است. میانگین و انحراف معیار نمونه عبارتند از: $\bar{x} = 364.7$ $\quad S = 259$

الف) انحراف معیار $\bar{x}$ را برآورد کنید؟

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{259}{20} = 12.95$$

ب) یک فاصله‌ی اطمینان ۹۰٪ برای μ پیدا کنید؟

$$(L, U) = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 364.7 \pm \underbrace{(1.645 \times 12.95)}_{21.30} = (343.4, 386)$$

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

ج) حال اگر انحراف معیار جامعه معلوم باشد و برابر ۳۰۰ باشد آن‌گاه یک فاصله‌ی اطمینان با ضریب ۰/۹۵ برای μ بدست آورید؟

$$(L, U) = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 364.7 \pm (1.96 \times \cancel{12.95}) = (339, 390.1)$$

۳۳۵٫۳ ، ۳۹۴٫۱

$$\frac{300}{20} = 15$$

$$Z_{\frac{\alpha}{2}} = 1.96$$

نکته => هرچه قدر ضریب اطمینان $(1-\alpha)$ افزایش یابد => محدوده‌ی اعداد بزرگتر شده => از دقت برآورد کاسته می‌شود.

$1-\alpha$	$Z_{\frac{\alpha}{2}}$
٪۹۰	1.645
٪۹۵	1.96
٪۹۹	2.575

برآورد فاصله‌ای μ برای نمونه‌های کوچک :

★ الف) حدود اطمینان μ در جامعه‌ی نرمال (σ نامعلوم) :

$$(L,U) = \bar{x} \mp t_{(\frac{\alpha}{۲}, n-۱)} \frac{S}{\sqrt{n}}$$

مثال <= نمونه‌ای مرکب از ۵ قوطی رب به تصادف از خط تولید انتخاب می‌شود و محتوای ماده‌ی هر قوطی وزن می‌شود تجربه‌ی گذشته حاکی از این است که توزیع وزن ماده‌ی هر قوطی نرمال است و یک فاصله‌ی اطمینان ۹۹ درصد برای μ بسازید؟ نتایج نمونه در زیر آمده‌اند.

x_i : ۲۳۰ ۲۳۵ ۲۳۵ ۲۵۰ ۲۴۵

$$\bar{x} = \frac{\sum x_i}{n} \Rightarrow \bar{x} = \frac{۱۱۹۵}{۵} = ۲۳۹$$

$$S^۲ = \frac{\sum (x - x_i)^۲}{n-۱} = \frac{۸۱ + ۱۶ + ۱۶ + ۱۲۱ + ۳۶}{۴} = \frac{۲۷۰}{۴} = ۶۷/۵$$

$$S = \sqrt{۶۷/۵} = ۸/۲۱۶$$

$$۱ - \alpha = ۰/۹۹ \Rightarrow \alpha = ۰/۰۱ \Rightarrow \boxed{\frac{\alpha}{۲} = ۰/۰۰۵}$$

① نمونه کوچک است ② جامعه نرمال است ③ σ جامعه نامعلوم است

$$(L,U) = \bar{x} \pm t_{(\frac{\alpha}{۲}, n-۱)} \frac{S}{\sqrt{n}} = ۲۳۹ \pm t_{(۰/۰۰۵, ۴)} \frac{۸/۲۱۶}{۲/۲۳۶}$$

$$= ۲۳۹ \pm (۴/۶۰۴)(۳/۶۷۴) = (۲۲۲ ; ۲۵۶)$$

★ب) حدود اطمینان μ در جامعه نرمال (σ معلوم)

$$(L,U) = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

مثال => در مثال قبل اگر انحراف معیار جامعه برابر ۷ گرم باشد، یک فاصله‌ی اطمینان ۹۵ درصدی برای μ بدست آورید:

① نمونه کوچک است امّا انحراف معیار جامعه معلوم ② نرمال ③ σ معلوم است

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \quad , \quad Z_{\frac{\alpha}{2}} = 1.96$$

$$(L,U) = 239 \pm \underbrace{\left(1.96 \times \frac{7}{2.236}\right)}_{6.136} = (232.86, 245.14)$$

★ج) حدود اطمینان μ در جامعه غیرنرمال

$$(L,U) = \bar{x} \pm t_{(\frac{\alpha}{2}, n-1)} \frac{\sigma}{\sqrt{n}} \iff \sigma \text{ معلوم}$$

$$(L,U) = \bar{x} \pm t_{(\frac{\alpha}{2}, n-1)} \frac{S}{\sqrt{n}} \iff \sigma \text{ نامعلوم}$$

سؤال => پژوهشگری تأثیر روش خاصی را بر میانگین زمان انجام کاری مشخص، بررسی می‌کند، نمونه‌ای تصادفی از ۱۶ گروه، این کار را بطور تصادفی متوسط در زمان ۲۵/۹ دقیقه و با انحراف معیار (نمونه) نه جامعه ۳/۶ دقیقه انجام داده‌اند. فرض کنید که زمان‌های انجام کار از توزیعی پیروی کنند که تقریباً نرمال باشد. برای میانگین زمان انجام کار با این روش یک فاصله‌ی اطمینان ۹۵٪ بنا کنید و برآورد فاصله‌ای را تفسیر کنید.

① تقریباً نرمال ② نمونه هم کوچک است ③ σ جامعه هم نامعلوم

$$(L,U) = \bar{x} \pm t_{(\frac{\alpha}{2}, n-1)} \frac{S}{\sqrt{n}} \Rightarrow (L,U) = 25.9 \pm \underbrace{t_{(0.025, 15)}}_{2.120} \frac{3.6}{4}$$

$$\rightarrow (23.99, 27.81)$$

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

www.pnu-m-s.com

(نمونه سؤالات امتحانی)

۱- داده های ۱و۲و۳و۴و۵ از یک نمونه می باشند برآورد واریانس جامعه چند است؟

$$\bar{x} = \frac{\sum(x_i)}{n} \Rightarrow \bar{x} = \frac{15}{5} = 3$$

$$S^2 = \frac{(5-3)^2+(4-3)^2+(0)+(2-3)^2+(1-3)^2}{4} \Rightarrow \boxed{S^2 = \frac{10}{4}}$$

۲- درصورتیکه مجموع ۵ عدد ۵۰ و مجموع توان دوم آن ها ۶۰۰ باشد، برآورد واریانس چند است؟

$$S^2 = \frac{\sum x_i^2 - \frac{\sum(x_i)^2}{n}}{n-1} = \frac{600 - \frac{(50)^2}{5}}{4} = 25$$

۳- نمونه ای تصادفی با اندازه ی n=100 از جامعه ی نامتناهی با میانگین μ و واریانس ۱۰۰ انتخاب می کنیم اگر بخواهیم با ۹۵% اطمینان تفاوت کنیم حداکثر خطای برآورد μ چقدر است؟

$\boxed{\text{راه اول}}$

$$P(|\bar{x}-\mu| < d) = 0.95 \Rightarrow P(-d < \bar{x}-\mu < d) = 0.95$$

$$\Rightarrow P\left(\frac{-d}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{d}{\frac{\sigma}{\sqrt{n}}}\right) = 0.95 \Rightarrow P(-d < Z < d) = 0.95$$

$$\Rightarrow 2P(0<Z<d) = 0.95 \Rightarrow P(0<Z<d) = 0.475 \Rightarrow \boxed{d = 1.96}$$

راه دوم $\Rightarrow P\left(d < Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) \Rightarrow P(d < 1.96 \times 1) \Rightarrow P(d < 1.96)$

← حداکثر خطا ۱٫۹۶ می باشد

۴- جامعه ای دارای انحراف معیار $\sigma=21$ برای برآورد میانگین، نمونه ای به حجم n=100 اختیار می کنیم، اگر بخواهیم با ۹۵% اطمینان تفاوت کنیم حداکثر خطای برآورد میانگین چقدر است؟

$$d = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow d = 1.96 \times \frac{21}{10} = 4.116$$

نکته ⇐ با افزایش حجم نمونه طول فاصله‌ی اطمینان کاهش می‌یابد. (طبق رابطه‌ی زیر)

$$\bar{x} \pm Z_{\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n} \rightarrow \uparrow}$$

۵- برای نمونه‌ای تصادفی به اندازه n از جامعه‌ای نرمال مقادیر زیر حاصل شده است:

۱٫۹   ۲   ۳٫۲   ۲٫۱   ۱٫۸   ۱٫۴   ۳٫۶   ۲٫۳   ۲٫۸   ۲٫۱   ۱٫۹   ۲٫۳

یک فاصله‌ی اطمینان ۹۹ درصد برای میانگین واقعی بدست آورید؟

$$\bar{x} = \frac{۲۷٫۴}{۱۲} = ۲٫۲۸$$

$$S^۲ = \frac{۰/۰۰۱۴ + ۰/۱۴۴۴ + ۰/۰۳۲۴ + ۰/۲۷۰۴ + ۰/۰۰۰۴ + ۱٫۷۴۲۴ + ۰/۷۷۴۴ + ۰/۲۳۰۴ + ۰/۰۳۲۴ + ۰/۸۴۶۴ + ۰/۰۷۸۴ + ۰/۱۴۴۴}{۱۱}$$

$$S^۲ = \frac{۴٫۵۸۸۴}{۱۱} = ۰/۴۲ \qquad \boxed{S = ۰/۶۴}$$

① نرمال ② n کوچک ③ σ نامعلوم

$$\bar{x} \pm t(\tfrac{\alpha}{۲}, n-1) \frac{S}{\sqrt{n}} = ۲٫۲۸ \pm (۰/۰۰۵ , ۱۱) \frac{۰/۶۴}{۳٫۴۶} = \boxed{(۱٫۷۲۱ , ۲٫۸۳۹)}$$

$$1 - \alpha = ۰/۹۹ \Rightarrow \alpha = ۰/۰۱ \Rightarrow \frac{\alpha}{۲} = ۰/۰۰۵$$

با اطمینان ۹۹ درصد میانگین واقعی جامعه در بازه‌ی بالا قرار می‌گیرد.

★ ۶- توزیع نمونه‌ای $\bar{x}$ دارای واریانس ۹ می‌باشد. اگر انحراف معیار جامعه‌ی آماری ۸۱ باشد، مقدار n را بیابید؟

$$var(\bar{x}) = \sigma^۲_{\bar{x}} = \frac{\sigma^۲_x}{n} = ۹ = \frac{۶۵۶۱}{n} \Rightarrow ۹n = ۶۵۶۱$$

$$\Rightarrow n = \frac{۶۵۶۱}{۹} = ۷۲۹$$

۷- جامعه‌ای دارای ۵ عنصر می‌باشد که دارای میانگین ۱۲ و واریانس ۱۸ می‌باشد. اگر نمونه‌ای تصادفی به اندازه ۳ از این جامعه استخراج کنیم، واریانس $\bar{x}$ چقدر است؟

$$\left.\begin{array}{l} N=5 \\ \mu=12 \\ \sigma^2=18 \\ n=3 \\ \sigma^2_{\bar{x}}=? \end{array}\right\} \Rightarrow var(\bar{x}) = \sigma^2_{\bar{x}} = \frac{\sigma^2_x}{n} = \frac{18}{3} = 6$$

۸- توزیع نمونه‌گیری $\bar{X}$ دارای میانگین ۱۰ می‌باشد، میانگین واقعی جامعه چقدر است؟ ۵۰ $\boxed{۱۰}$ ۲۰ ۳۰

۹- برای اینکه طول بازه‌ی اطمینان برای میانگین جامعه نصف شود. اگر انحراف معیار و سطح معنی ثابت باشد مقدار n چه تغییری باید بکند؟ دو برابر شود $\boxed{\text{۴ برابر شود}}$

ثابت ← $Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}$ ؛ σ_x → ثابت ؛ $\sqrt{n}$ → x $\xrightarrow[\text{شدن باید}]{\text{برای نصف}}$ n ۴ برابر گردد

۱۰- به طور کلی اگر بخواهیم طول یک فاصله‌ی اطمینان را کوتاه‌تر کنیم به طوری که ضریب اطمینان فاصله کاهش نیابد باید اندازه‌ی نمونه را افزایش دهیم. و بالعکس.

۱۱- همچنین با افزایش ضریب اطمینان در صورت ثابت بودن دیگر متغیرها ← محدوده‌ی اعداد بزرگتر می‌گردد.

۱۲- هرگاه جامعه بزرگ بوده و حجم نمونه از ۳۰ کمتر باشد و انحراف معیار جامعه نامعلوم باشد، مناسب‌ترین توزیع برای تخمین فاصله‌ای میانگین جامعه از طریق میانگین نمونه توزیع t می‌باشد.

۱۳- برای برآورد میانگین جامعه ای با انحراف معیار ۱٫۷۵، نمونه‌ای ۴۹ تایی انتخاب کرده‌ایم. احتمال اینکه خطای برآورد میانگین حداکثر ۰٫۴۱ شود برابر است با: ۰٫۱ ۰٫۰۵ ۰٫۹۵ $\boxed{0/9}$

$$P(|\bar{x}-\mu| \leq 0.41) = P(-0.41 < \bar{x}-\mu < 0.41) = P\left(\frac{-0.41}{\frac{1.75}{7}} < Z < \frac{0.41}{\frac{1.75}{7}}\right)$$

$$= P(-1.64 < Z < 1.64) = 2P(0 < Z < 1.64) = 2 \times 0.4495 = 0.9$$

۱۴- نمرات دانش‌آموزان کلاس پنجم در یک آزمون هوش (x)، دارای میانگین ۲۰۰ امتیاز و انحراف معیار ۳۶ امتیاز بصورت نرمال توزیع شده است. احتمال اینکه در یک نمونه ۹ تایی $\bar{x}$ حداکثر در فاصله ۰٫۲۵ از میانگین جامعه قرار بگیرد کدام است؟ ۰٫۰۰۸ ۰٫۹۹۲ $\boxed{0/016}$ ۰٫۰۸

$$P(|\bar{x}-\mu| \leq 0.25) = P(-0.25 \leq x-\mu \leq 0.25) = P\left(\frac{-0.25}{\frac{36}{3}} < Z < \frac{0.25}{\frac{36}{3}}\right)$$

$$= P(-0.02 < Z < 0.02) = 2P(0 < Z < 0.02) = 2(0.0080) = 0.016$$

۱۵- در بررسی اثرات تغذیه بر کارایی کارگران، از کارگرانی که نمونه تصادفی به اندازه ۱۶ را تشکیل می‌دهند، خواسته شده که از یک رژیم غذایی خاص پیروی کنند و سطح قند خون (x) هر یک از کارگران متعلق به نمونه، دو ساعت بعد از خوردن صبحانه، اندازه‌گیری شده است. نتایج عبارتند از: $\bar{x} = 128$ $S = 9.6$

اگر مقادیر x بصورت نرمال توزیع شده باشند برای میانگین سطح قند تحت این رژیم غذایی یک فاصله اطمینان ۹۵ درصد بسازید $(t_{0.025,15}) = 2.131$.

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

$$\bar{x} \pm t\left(\frac{\alpha}{2}, n-1\right)\frac{S}{\sqrt{n}} = 128 \pm (2.131)\left(\frac{9.6}{4}\right) \Rightarrow (L, U) = 128 \pm (5.1144)$$

$$(122.8856 \,,\, 133.1144)$$

۱۶- نمونه‌ای تصادفی شامل ۵ وسیله‌ی نقلیه موتوری خاصی را انتخاب و هزینه استفاده از هر کدام را ثبت کرده‌ایم ۲۳۲ – ۲۴۵ – ۲۷۰ – ۲۸۸ – ۳۱۵. اگر هزینه‌ها دارای توزیع نرمال باشند. الف) یک فاصله‌ی اطمینان ۹۸ درصد برای متوسط هزینه‌ها بدست آورید.

$$\bar{x} = 270$$

$$\alpha = 0.02 \longrightarrow t(0.01, 4) = 3.747$$

$$S^2 = \frac{2025 + 324 + 625 + 1444}{4} = 1104.5 \Rightarrow \boxed{S = 33.23}$$

$$(L, U) = 270 \pm (3.747 \times 14.8) = (214.32,$$

«فصل نهم» «برآورد نسبت و واریانس جامعه» نکات و سؤالات مهم

«در این فصل برآورد نقطه‌ای و برآورد فاصله‌ای P ، نسبت جامعه را برای نمونه‌های بزرگ بدست می‌آوریم»

اگر x در فواست‌های کامل در یک نمونه nتایی باشد، $\bar{P}$ نسبت نمونه‌ای از رابطه‌ی زیر بدست می‌آید.

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

نکته => نسبت نمونه‌ای $\bar{P} = \frac{x}{n}$ ، بهترین برآوردکننده‌ی P نسبت به سایر برآوردکننده‌هاست. $E(\bar{P}) = P$

میانگین و واریانس $\bar{P}$:

میانگین و واریانس آماره‌ی $\bar{P}$ از رابطه‌ی زیر بدست می‌آید :

$$E(\bar{P}) = P \qquad \sigma^2_{\bar{P}} = \frac{P(1-P)}{n} \qquad \sigma_{\bar{P}} = \sqrt{\frac{P(1-P)}{n}}$$

برآوردکننده $\sigma^2_{\bar{P}}$:

اگر نسبت جامعه (P) نامعلوم باشد، $\sigma^2_{\bar{P}}$ هم نامعلوم باشد، برآورد آن به صورت زیر بدست می‌آید :

$$S^2_{\bar{P}} = \frac{\bar{P}(1-\bar{P})}{n} \qquad S_{\bar{P}} = \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$$

خطای برآورد P برای نمونه‌های بزرگ :

$$d = |\bar{P} - P|$$

برای نمونه‌های بزرگ، عبارت احتمال زیر را درباره‌ی مقدار خطا می‌نویسیم.

$$P(d < Z_{\frac{\alpha}{2}} S_{\bar{P}}) \approx 1 - \alpha$$

بنابراین با $100(1-\alpha)$ درصد اطمینان می‌گوییم که حداکثر خطای برآورد P ، برابر با $(Z_{\frac{\alpha}{2}} S_{\bar{P}})$ است.

مثال => در یک نمونه تصادفی با اندازه‌ی $n=100$ از کارگران یک کارخانه، تعداد ۲۰ نفر بی سواد هستند. نسبت بی سوادان را در این کارخانه برآورد کنید. با احتمال ۹۵٪ حداکثر خطای این برآورد چقدر است؟

$$\bar{p} = \frac{\text{تعداد بی‌سوادان در نمونه}}{\text{اندازه‌ی نمونه}} = \frac{20}{100} \Rightarrow \boxed{\bar{p} = 0/2}$$

$$\boxed{Z_{\frac{\alpha}{2}} = 1/96} \qquad d = Z_{\frac{\alpha}{2}} \cdot S_{\bar{p}} \Rightarrow d = Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \Rightarrow d = 1/96 \times \sqrt{\frac{0/2 \times 0/8}{100}}$$

$$\Rightarrow d = 1/96 \times \sqrt{\frac{0/16}{100}} \Rightarrow d = 1/96 \times \frac{0/4}{10} \Rightarrow \boxed{d = 0/0784}$$

حداکثر خطای برآورد با احتمال ۹۵ درصد برابر با $d = 0/0784$ می باشد.

برآورد فاصله‌ای p برای نمونه‌های بزرگ:

برای برآورد فاصله‌ای رابطه‌ی زیر برقرار است:

$$(L,U) = \bar{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

مثال => از کفش هایی که تحت فرآیند معینی تولید شده اند، نمونه‌ای به اندازه‌ی ۴۰۰ جفت انتخاب شده و ۴۱ جفت آن ها در رده معیوب قرار گرفته اند. یک فاصله اطمینان ۹۵ درصدی برای p، نسبت کفش های معیوبی که تحت این فرآیند تولید شده‌اند بدست آورید.

$$\left.\begin{array}{l} Z_{0/025} = 1/96 \\ \bar{p} = \frac{41}{400} = 0/1025 \end{array}\right\} \Rightarrow (L,U) = \bar{p} \pm Z_{0/025} \sqrt{\frac{p(1-p)}{n}} = 0/1025 \pm 1/96 (0/01516)$$

$$\Rightarrow (L,U) = (0/0727 , 0/1323)$$

مثال => یک اتحادیه صنفی، نمونه‌ای از ۱۰۰ شرکت عضو را برای تعیین برآورد p، نسبت شرکت هایی که کارکنان نیمه وقت ندارند، انتخاب می‌کند، اگر در این نمونه، ۱۰ شرکت دارای کارکنان نیمه وقت باشند.

الف) مقدار p را برآورد کنید. ب) یک فاصله‌ی اطمینان ۹۵ درصد برای p بدست آورید.

$$\bar{p} = \frac{90}{100} = 0/9 \Rightarrow \boxed{\bar{p} = 0/9} \Rightarrow$$ نسبت شرکت هایی که کارکنان نیمه وقت ندارند.

$$Z_{0/025} = 1/96 \qquad (L,U) = \bar{p} \pm Z_{0/025} \sqrt{\frac{p(1-p)}{n}} = 0/9 \pm 1/96 \sqrt{\frac{0/9 \times 0/1}{100}}$$

$$\Rightarrow (L,U) = (0/8412 , 0/9588)$$

(برآورد واریانس جامعه)

برآورد فاصله‌ای σ^2 در جامعه‌ی نرمال:

بهترین برآوردکننده نقطه‌ای σ^2 واریانس نمونه‌ای است، بصورت زیر:

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

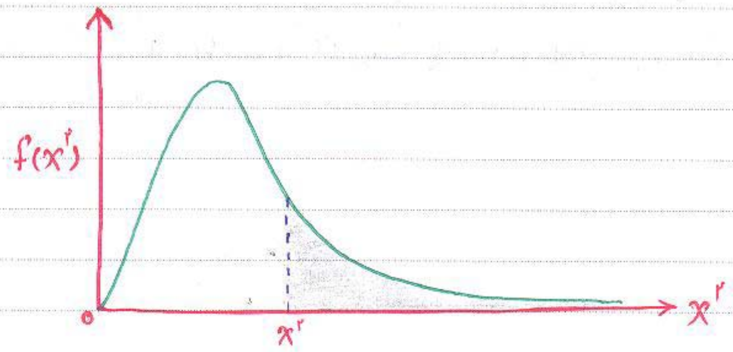
برای پیدا کردن حدود اطمینان σ، از نقاط درصد توزیع جدیدی به نام توزیع χ^2 (خی‌دو) استفاده می‌کنیم.
شکل توزیع خی‌دو با r که همان درجه‌ی آزادی است بستگی دارد.

حال اگر نمونه‌ای n تایی بطور تصادفی با واریانس S^2 از جامعه‌ای نرمال با واریانس σ^2 استخراج کنیم، در این صورت حدود فاصله‌ی اطمینان با ضریب $(1-\alpha)$ برای σ^2 بصورت زیر است:

$$L = \frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2}, n-1)} \qquad u = \frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2}, n-1)}$$

نکته <= توزیع خی‌دو از جدول ضمیمه کتاب بدست می‌آید.

نکته <= برخلاف توزیع نرمال یا t، منحنی توزیع احتمال χ^2، منحنی نامتقارنی است که به طرف راست کشیده شده است. [چوله به راست است]



مثال => یک شرکت دارویی قرص‌هایی تولید می‌کند که تصور می‌شود شامل مقداری مناسب از ماده‌ای مؤثر است. نمونه‌ای تصادفی شامل ۴۱ قرص که از فرآیند تولید انتخاب شده، دارای انحراف معیار $S=1.09$ میلیگرم از مقدار ماده‌ی مؤثر در هر قرص است. فرض کنید توزیع مقدار ماده‌ی مؤثر در هر قرص دارای توزیع نرمال است.

الف) یک فاصله‌ی اطمینان ۹۵% درصدی برای واریانس مقدار ماده‌ی مؤثر قرص‌های کل فرآیند بسازید.

ب) برآورد فاصله‌ای قسمت الف را به فاصله‌ی اطمینان ۹۵% برای انحراف معیار تبدیل کنید.

(الف)

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \boxed{\frac{\alpha}{2} = 0.025} \quad \text{و} \quad \boxed{1-\frac{\alpha}{2} = 0.975}$$

$$L = \frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2}, n-1)} = \frac{40 \times (1.09)^2}{\chi^2(0.025, 40)} = \frac{47.524}{59.3417} = 0.80$$

$$U = \frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2}, n-1)} = \frac{40 \times (1.09)^2}{\chi^2(0.975, 40)} = \frac{47.524}{24.4331} = 1.95$$

$$0.80 < \sigma^2 < 1.95$$

(ب) چون انحراف معیار جذر واریانس است => فاصله‌ی اطمینان ۹۵% برای انحراف معیار به صورت زیر است.

$$\sqrt{0.80} < \sigma < \sqrt{1.95} \Rightarrow \boxed{0.89 < \sigma < 1.40}$$

مثال => انحراف معیار طول ۲۵ پیچ که توسط دستگاهی تولید شده‌اند برابر $S=4.35$ سانتی‌متر است. طول پیچ‌ها دارای توزیع نرمال است. یک فاصله‌ی اطمینان ۹۹ درصد برای واریانس طول پیچ‌ها بسازید.

$$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \boxed{\frac{\alpha}{2} = 0.005} \qquad \boxed{1-\frac{\alpha}{2} = 0.995}$$

$$\left.\begin{array}{l} L = \dfrac{(24)(4.35)^2}{\chi^2(0.005, 24)} = 9.968 \\[2em] U = \dfrac{(24)(4.35)^2}{\chi^2(0.995, 24)} = 45.937 \end{array}\right\} \Rightarrow (L, U) = (9.968, 45.137)$$

مثال ⇐

پژوهشگری درجه حرارتی را که برای ۵ فلز مورد آزمایش بدست آورده است را در جدول زیر قرار داده است. پژوهشگر می‌خواهد یک فاصله‌ی اطمینان برای σ (انحراف معیار جامعه) بدست آورد. فرض کنید که ۵ درجه‌ی حرارت مشاهده شده نمونه‌ی تصادفی از جامعه‌ای نرمال هستند. یک فاصله‌ی اطمینان ۹۹ درصدی برای σ بسازید؟

فلز	۱	۲	۳	۴	۵	
درجه‌ی حرارت	۲۱۶	۲۲۱	۲۲۰	۲۱۸	۲۲۵	۱۱۰۰=
x_i^2	۴۶۶۵۶	۴۸۸۴۱	۴۸۴۰۰	۴۷۵۲۴	۵۰۶۲۵	۲۴۲۰۴۶

$$S^2=\frac{\sum x_i^2-\frac{\sum (x_i)^2}{n}}{n-1}=\frac{۲۴۲۰۴۶-\frac{(۱۱۰۰)^2}{۵}}{۴}=\frac{۲۴۲۰۴۶-۲۴۲۰۰۰}{۴}$$

$$\Rightarrow \boxed{S^2=۱۱/۵}$$

$$1-\alpha=۰/۹۹ \Rightarrow \alpha=۰/۰۱ \Rightarrow \boxed{\frac{\alpha}{۲}=۰/۰۰۵} \qquad \boxed{۱-\frac{\alpha}{۲}=۰/۹۹۵}$$

$$L=\frac{(۴)(۱۱/۵)}{\chi^2(۰/۰۰۵,۴)}=۳/۰۹۶$$

$$U=\frac{(۴)(۱۱/۵)}{\chi^2(۰/۹۹۵,۴)}=۲۲۲/۲۳$$

$$\Rightarrow (L,U)=(۳/۰۹۶,\ ۲۲۲/۲۳)$$

$$\Rightarrow \sqrt{۳/۰۹۶}<\sigma<\sqrt{۲۲۲/۲۳}$$

$$\Rightarrow \boxed{۱/۷۵۹<\sigma_x<۱۴/۹۰۷}$$

مثال ⇐ صفحه‌های پلاستیکی که توسط یک ماشین تولید می‌شوند به طور متناوب مورد بازبینی قرار می‌گیرد تا تفاوت‌های ضخامت آن‌ها بررسی گردد. ناهمگنی در غلظت ماده‌ای که به کار می‌رود وجود تفاوت‌هایی در ضخامت‌های صفحه‌ها را غیر قابل اجتناب می‌سازد. در صفحه‌ی تولید شده در یک نوبت کاری اندازه‌ی ضخامت برحسب میلی‌متر به قرار زیر بوده است. یک فاصله‌ی اطمینان ۹۵ درصدی برای انحراف معیار واقعی ضخامت صفحه‌های تولید شده بسازید؟

۲۲۶، ۲۲۸، ۲۲۶، ۲۳۰، ۲۲۵، ۲۲۹، ۲۲۷، ۲۲۸، ۲۳۲، ۲۲۵

$$\left.\begin{array}{l}\sum(x_i) = 2276 \\ \sum(x_i^2) = 518064\end{array}\right\} \Rightarrow S^2 = \frac{518064 - 518017.6}{9} = 5.16$$

$$\Rightarrow \boxed{S^2 = 5.16} \qquad 1-\alpha = 0.95 \Rightarrow \boxed{\frac{\alpha}{2} = 0.025} \qquad \boxed{1-\frac{\alpha}{2} = 0.975}$$

$$L = \frac{(9)(5.16)}{\chi^2(0.025, 9)} = 2.44$$

(نمونه سؤالات امتحانی)

۱- در یک نمونه تصادفی ۴۰۰ تایی از محصولات یک تولیدی، ۱۰٪ محصولات معیوب هستند. حد پایین فاصله‌ی اطمینان ۹۰٪ برای P، نسبت کل محصولات معیوب کدام است؟

$P = 0.1$

$$1-\alpha = 0.90 \Rightarrow \alpha = 0.1 \Rightarrow \boxed{\frac{\alpha}{2} = 0.05} \Rightarrow \boxed{Z_{0.05} = 1.645}$$

$$L = 0.1 - \left(1.645 \times \sqrt{\frac{0.1 \times 0.9}{400}}\right) \Rightarrow L = 0.1 - (0.0246) \Rightarrow \boxed{L = 0.075}$$

۲- در یک نمونه تصادفی با حجم $n=5$ ، $\sum_{i=1}^{5} x_i = 20$ می‌باشد و $\sum_{i=1}^{5} x_i^2 = 120$ می‌باشند. برآورد فاصله‌ای انحراف معیار با ضریب اطمینان ۰/۹۵ را بدست آورید؟

$$S^2 = \frac{120 - \frac{(20)^2}{5}}{4} = 10$$

$$(L) = \frac{(4)(10)}{\chi^2(0.025, 4)} = \frac{40}{11.1433} = 3.59$$

$$\Rightarrow 3.59 < \sigma^2 < 82.57$$

$$u = \frac{(4)(10)}{\chi^2(0.975, 4)} = \frac{40}{0.484419} = 82.57 \qquad 1.90 < \sigma < 9$$

۳- مطالعه‌ای به منظور تعیین سبک تصمیم‌گیری مشارکتی مدیران در جهت برنامه‌ریزی است. تحقیقات مشابه نشان می‌دهد که این نسبت بزرگتر از ۶۰ درصد نیست. دقت برآورد را ۰/۰۴ در نظر بگیرید و حجم نمونه را در سطح ۰/۰۱ تعیین کنید. ۹۹۵

$\alpha = 0/01 \qquad Z_{\frac{\alpha}{2}} = \boxed{Z_{0.005} = 2{,}575}$

$$d = Z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \Rightarrow d^2 = (Z_{\frac{\alpha}{2}})^2\frac{pq}{n} \Rightarrow n = \frac{(Z_{\frac{\alpha}{2}})^2 pq}{d^2} \Rightarrow \boxed{n = 995}$$

۴- در یک بسته ۵۰ تایی از لامپ‌ها، ۵ لامپ خراب است. انحراف معیار نسبت نمونه‌ای لامپ‌های خراب برابر است با:

۰/۰۰۱۸ ۰/۳ $\boxed{0/0424}$ ۰/۰۹

$$\left.\begin{array}{l} S_p = \sqrt{\dfrac{P(1-P_0)}{n}} \\ \bar{P} = \dfrac{5}{50} = 0/1 \end{array}\right\} \Rightarrow S_{\bar{P}} = \sqrt{\frac{0/1 \times 0/9}{50}} = 0/0424$$

۵- برای یافتن فاصله‌ی اطمینان $(1-\alpha)$ ۱۰۰% برای واریانس جامعه از چه توزیعی استفاده می‌گردد؟

توزیع خی‌دو با $n-1$ درجه‌ی آزادی

۶- در نمونه‌ای ۱۰ تایی، اگر واریانس نمونه‌ای ۴ باشد، فاصله‌ی اطمینان ۹۹ درصدی برای انحراف معیار جامعه عبارت است از:

$1-\alpha = 0/99 \Rightarrow \alpha = 0/01 \Rightarrow \frac{\alpha}{2} = 0/005$

$$(L,U) = \left(\frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2}, n-1)}, \frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2}, n-1)}\right) = \left(\frac{(9)(4)}{\chi^2(0/005, 9)}, \frac{(9)(4)}{\chi^2(0/995, 9)}\right) = \left(\frac{36}{23{,}59}, \frac{36}{1{,}735}\right)$$

$$\Rightarrow (L,U) = (1{,}526\ ,\ 20{,}75)$$

۷- نمونه‌ای تصادفی شامل ۵ وسیله‌ی نقلیه‌ی موتوری خاصی را انتخاب و هزینه‌ی استفاده از هر کدام را ثبت کرده‌ایم. (۲۳۲، ۲۴۵، ۲۷۰، ۲۸۸، ۳۱۵) اگر هزینه‌ها دارای توزیع نرمال باشند، در سطح $\alpha = 0/01$ فرض بزرگتر بودن انحراف معیار از ۳۴ را آزمون کنید؟

$$\begin{cases} H_0 : \sigma \leq 34 \\ H_1 : \sigma > 34 \end{cases} \qquad S = 33{,}234$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \Rightarrow \chi^2 = \frac{(4)(1104{,}49)}{1156} = 3{,}821$$

$$\chi^2(\alpha, n-1) = \chi^2(0/01, 4) = 13{,}276$$

فرض H_0 قبول می‌گردد.

۸- اگر از ۱۰۰ دانشجوی یک دانشگاه، ۲۴ نفر اتومبیل داشته باشند. یک فاصله‌ی اطمینان ۹۶ درصد برای نسبت دانشجویان دارای اتومبیل بدست آورید.

$$\bar{p} = \frac{۲۴}{۱۰۰} = ۰/۲۴$$

$$۱ - \alpha = ۰/۹۶ \Rightarrow \alpha = ۰/۰۴ \Rightarrow \boxed{\frac{\alpha}{۲} = ۰/۰۲} \qquad Z_{۰/۰۲} = ۰/۵ - ۰/۰۲ = ۰/۴۸$$

$$\Rightarrow \frac{۲/۰۶ + ۲/۰۵}{۲} = ۲/۰۵۵ \Rightarrow \boxed{Z_{۰/۰۲} = ۲/۰۵۵}$$

$$(L, U) = ۰/۲۴ \pm \sqrt{\frac{۰/۲۴ \times ۰/۷۶}{۱۰۰}} = (۰/۱۵۲ , ۰/۳۲۷)$$

تجربه گذشته نشان می‌دهد که سطوح pH در آب دریاچه تقریباً بصورت نرمال توزیع می‌شوند. اگر نتایج نمونه تصادفی استخراج شده بصورت زیر باشد برای انجام آزمون فرض روی میانگین توزیع آماره آزمون کدام است؟

$$n = ۱۸ \qquad \bar{x} = ۶/۸ \qquad S = ۰/۱$$

$$\chi^2(۱۶) \qquad Z \qquad \boxed{t(۱۷)} \qquad F(۱,۱)$$

۱۰- حداکثر خطای برآورد p با (۱-α) ۱۰۰٪ اطمینان برای نمونه‌های بزرگ $Z_{\frac{\alpha}{2}} S_{\bar{p}}$ است.

۱۱- طول یک لوله‌ی سلفنی دارای توزیع نرمال با میانگین μ، و واریانس σ^2 است. یک نمونه‌ی تصادفی ۲۵ تایی از لوله‌ها جمع آوری شده است و مقادیر $\sum_{i=1}^{۲۵} x_i = ۲۵۲/۷$ و $\sum_{i=1}^{۲۵} x_i^2 = ۲۵۷۹/۷$ حاصل شده است یک برآورد نقطه‌ای و یک فاصله‌ی اطمینان ۹۰ درصد برای واریانس جامعه بدست آورید.

$$S^2 = \hat{\sigma}^2 = \frac{۲۵۷۹/۷ - \frac{(۲۵۲/۷)^2}{۲۵}}{۲۴} = ۱/۰۶$$

$$\left.\begin{array}{l} \chi^2(۰/۰۵ , ۲۴) = ۳۶/۴ \Rightarrow L = \frac{(۲۴)(۱/۰۶)}{۳۶/۴} \\ \chi^2(۰/۹۵ , ۲۴) = ۱۳/۸ \Rightarrow U = \frac{(۲۴)(۱/۰۶)}{۱۳/۸} \end{array}\right\} \Rightarrow (L, U) = (۰/۶۹۹ , ۱/۸۴۳)$$

در یک نمونه تصادفی به حجم ۱۵ از جامعه‌ای نرمال با واریانس ۷۰ مقدار واریانس نمونه‌ای برابر ۸۵ می‌باشد. مقدار آماره‌ی خی دو (χ^2) برای آزمون واریانس کدام است؟

۱۲/۳۵۲   ۱۱/۵۲۹   ۱۸/۲۱۴   ۱۷

«فصل دهم» «آزمون هایی درباره‌ی میانگین جامعه» «نکات و سؤالات مهم»

«در این فصل نوع خاصی از استنباط آماری با نام آزمون فرض های آماری را معرفی می کنیم.»

آزمون فرض: هر حکمی درباره‌ی توزیع جامعه یا پارامتر جامعه را یک فرض آماری می نامند و ممکن است درست یا نادرست باشد، درست یا نادرست بودن یک فرض، باید بر مبنای اطلاعات حاصل از نمونه گیری از جامعه بررسی شود. این عمل را آزمون فرض می نامیم.

چون ادعا ممکن است صحیح یا غلط باشد بنابراین دو فرض مکمل در ذهن بوجود می آید:

الف) فرض اول => ادعا صحیح است   ب) فرض دوم => ادعا غلط است

(فرض صفر و فرض مقابل).

الف) فرض صفر (H_0): به فرضی که باید آن را اثبات کنیم و در صدد رد یا عدم رد آن می باشیم و در آن علامت = وجود دارد.

ب) فرض مقابل (H_1): به فرض مخالف H_0 گویند که در صورت عدم اثبات H_0 پذیرفته می شود.

★ همواره باید فرض H_0 را اثبات کنیم و در صورت رد آن H_1 را بپذیریم.

★ ادعا ممکن است فرض صفر H_0 یا فرض مقابل H_1 شود.

«انواع آزمون های آماری»

آزمون یک طرفه چپ $\begin{cases} \mu \geq \mu_0 \\ \mu < \mu_0 \end{cases}$ آزمون یک طرفه راست $\begin{cases} \mu \leq \mu_0 \\ \mu > \mu_0 \end{cases}$ آزمون دو طرفه $\begin{cases} \mu = \mu_0 \\ \mu \neq \mu_0 \end{cases}$

مثال => ادعاهای داده شده را بخوانید و H_0 و H_1 آن را تعیین کنید:

الف) میانگین سن استادان دانشگاه بیش تر از ۳۰ سال است ← $\begin{cases} H_0: \mu \leq 30 \\ H_1: \mu > 30 \end{cases}$ آزمون یک طرفه راست

ب) میانگین قد دانشجویان دانشگاه حداقل ۱۶۰ می باشد => $\begin{cases} H_0: \mu \geq 160 \\ H_1: \mu < 160 \end{cases}$ آزمون یک طرفه چپ می باشد

ج) میانگین هزینه ماهانه نگهداری هواپیما ۳۲۰۰ دلار است => $\begin{cases} H_0: \mu = 3200 \\ H_1: \mu \neq 3200 \end{cases}$ آزمون دو طرفه است

ناحیه‌ی قبول و رد: ناحیه‌ی رد یک آزمون در واقع ناحیه‌ی رد فرض H_0 می‌باشد و ناحیه‌ی قبول یک آزمون در واقع ناحیه‌ی قبول فرض H_0 است. در هر قاعده‌ی تصمیم آماری، برد مقادیری از آماره‌ی آزمون که به ازای آن فرض H_0 نتیجه گرفته می‌شود به ناحیه‌ی قبول موسوم است. برد مقادیری که به ازای آن فرض H_1 را نتیجه می‌گیریم به $\boxed{\text{ناحیه‌ی رد}}$ یا $\boxed{\text{ناحیه‌ی بحرانی موسوم}}$ است.

(خطای نوع اول و نوع دوم):

الف) خطای نوع اول (α) ⇐ احتمال رد کردن H_0 وقتی H_0 درست است یا احتمال رد کردن H_0 وقتی H_1 غلط است به عبارتی:

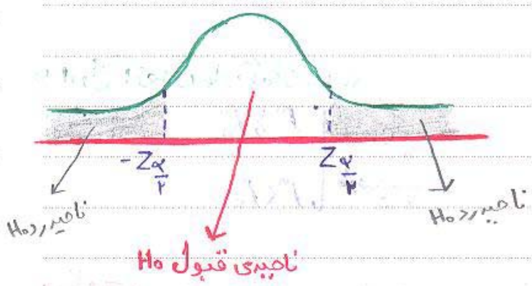
$\alpha = P($ خطای نوع اول $) = P($ رد کردن H_0 | درست بودن H_0 $) = P($ رد کردن H_0 | غلط بودن H_1 $)$

ب) خطای نوع دوم (β) ⇐ احتمال قبول کردن H_0 وقتی H_0 غلط است یا احتمال قبول کردن H_0 وقتی H_1 درست است به عبارتی:

$\beta = P($ خطای نوع دوم $) = P($ قبول کردن H_0 | غلط بودن H_0 $) = P($ قبول کردن H_0 | درست بودن H_1 $)$

حالت اول

آزمون فرض آماری میانگین یک جامعه $(\mu_{\bar{x}})$
- نوع آزمون ⇐ دوطرفه
- نوع جامعه ⇐ نرمال یا $n > 30$
- انحراف معیار جامعه ⇐ معلوم یا نامعلوم



در صورت معلوم بودن انحراف معیار: $Z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$

در صورت معلوم نبودن انحراف معیار: $Z = \dfrac{\bar{x} - \mu_0}{S_{\bar{x}}}$

رد فرض H_0 ⇐ $Z < -Z_{\frac{\alpha}{2}}$ یا $Z > Z_{\frac{\alpha}{2}}$

مثال ⇐ نمونه‌ای تصادفی به اندازه $n=49$ از جامعه‌ای، دارای میانگین ۳۰ و انحراف معیار ۴، است. براساس اطلاعات حاصل از این نمونه، آزمون آماری زیر را در سطح معنی‌دار بودن $\alpha = 0.05$ انجام دهید.

$$\begin{cases} H_0 : \mu = 33 \\ H_1 : \mu \neq 33 \end{cases}$$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \Rightarrow \boxed{Z_{0.025} = 1.96}$$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{30 - 33}{\frac{4}{7}} = -5.25 \Rightarrow \boxed{Z = -5.25}$$

$$\boxed{Z_{\frac{\alpha}{2}} = 1.96}$$

قاعده اول ⇐ $Z > Z_{\frac{\alpha}{2}} \Rightarrow -5.25 > 1.96$ برقرار نیست ✗

قاعده دوم ⇐ $Z < -Z_{\frac{\alpha}{2}} \Rightarrow -5.25 < -1.96$ برقرار است ✓ اما

⇐ فرض H_0 را رد می‌کنیم و H_1 را می‌پذیریم.

مثال ⇐ یک سازنده‌ی داروهای دامی، قرص‌های هورمون رشد را برای گاوها در بسته‌های بزرگ تولید می‌کند. مشخصه‌های تولید ایجاب می‌کند که میانگین محتوای هورمونی قرص‌ها معادل $\mu_0 = 1280$ میلی‌گرم برای هر قرص باشد. از هر بسته نمونه‌ای تصادفی مرکب از ۸۰ قرص انتخاب می‌کنند تا آزمون نمایند که هر بسته دارای این مشخصه هست یا نه. از روی تجربه می‌دانند که انحراف معیار $\sigma = 110$ است. اگر میانگین نمونه‌ای ۱۲۹۰ میلی‌گرم باشد آزمون را در سطح معنی‌دار بودن $\alpha = 0.05$ انجام دهید.

$$\begin{cases} H_0 : \mu = 1280 \\ H_1 : \mu \neq 1280 \end{cases}$$

$$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow \boxed{Z_{\frac{\alpha}{2}} = 1.96}$$

$$Z = \frac{1290 - 1280}{\frac{110}{\sqrt{80}}} = \frac{10}{\frac{110}{9}} \Rightarrow \boxed{Z = 0.81}$$

$Z > Z_{\frac{\alpha}{2}} \Rightarrow$ فرض H_0 رد می‌گردد $\Rightarrow 0.81 \ngtr 1.96 \Rightarrow$ فرض H_0 قبول می‌گردد ✓

$Z < -Z_{\frac{\alpha}{2}} \Rightarrow$ فرض H_0 رد می‌گردد $\Rightarrow 0.81 \nless -1.96 \Rightarrow$ فرض H_0 قبول می‌گردد ✓

⇐ فرض H_0 قبول می‌گردد

www.pnu-m-s.com

حالت دوم

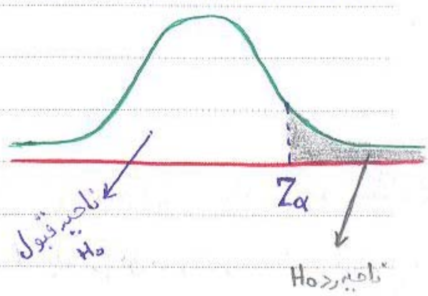
آزمون فرض آماری میانگین یک جامعه $\mu(x)$

- نوع آزمون <= یک طرفه
- نوع جامعه <= نرمال یا $n>30$
- انحراف معیار جامعه <= معلوم یا نامعلوم

- در صورت معلوم بودن انحراف معیار <= $Z = \frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}}$

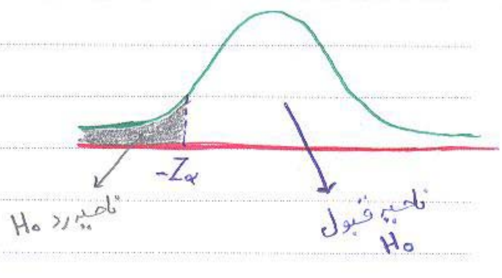
- در صورت معلوم نبودن انحراف معیار <= $Z = \frac{\bar{x}-\mu_0}{S_{\bar{x}}}$

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \Longrightarrow$$ آزمون یک طرفه راست



رد فرض H_0 <= $Z > Z_\alpha$ یا $Z < -Z_\alpha$

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \Longrightarrow$$ آزمون یک طرفه چپ



رد فرض H_0 <= $Z > -Z_\alpha$

www.pnu-m-s.com

مثال => ادعا شده است که میانگین برق مصرفی فروردین ماه یک ناحیه تهران دست کم ۱۳۰۰ کیلووات ساعت بوده است. برای این منظور یک نمونه تصادفی به تعداد ۴۰ خانوار را از منطقه انتخاب شده که میانگین و انحراف معیار برق مصرفی آن‌ها به ترتیب ۱۲۵۲ و ۲۵۷ می‌باشد. سطح خطای ۰/۰۱ را در نظر گرفته و صحت ادعا را بررسی کنید.

$$\begin{cases} H_0 : \mu \geq 1300 \\ H_1 : \mu < 1300 \end{cases}$$

$$\alpha = 0.01 \Rightarrow \boxed{Z_\alpha = 2.325}$$

$$Z = \frac{1252 - 1300}{\frac{257}{40}} \Rightarrow \boxed{Z = -3.73}$$

برای رد فرض H_0 : $Z > -Z_\alpha \Rightarrow 3.73 > -2.35 \Longrightarrow$ لذا فرض H_0 رد می‌گردد و H_1 را می‌پذیریم

مثال => فرض کنید که مقدار ماده‌ی اولیه‌ای را در یک کارگاه تولیدی در یک روز مصرف می‌شود را با متغیر x نشان می‌دهیم. مقدار x را در $n = 50$ روز ثبت کرده‌ایم و نتایج زیر بدست آمده. $S = 21$ $\bar{x} = 871$

آیا در سطح معنی‌دار $\alpha = 0.05$ می‌توان این ادعا را پذیرفت که میانگین مصرف روزانه‌ی این ماده کمتر از ۸۸۰ کیلوگرم

$$\begin{cases} H_0 : \mu \geq 880 \\ H_1 : \mu < 880 \end{cases}$$

$$\alpha = 0.05 \Rightarrow \boxed{Z_\alpha = 1.96}$$

$$Z = \frac{871 - 880}{\frac{21}{\sqrt{50}}} = \frac{-9}{\frac{21}{7}} \Rightarrow \boxed{Z = -3.03}$$

$Z > -Z_\alpha$ یا $Z < Z_\alpha$ $\quad -3.03 < 1.96 \Rightarrow$ فرض H_1 را رد می‌کنیم

مثال => از روی تجربه‌ای طولانی می‌دانند که میانگین تعداد افرادی که در پروازهای ساعت معینی حاضر نمی‌شوند ۱٫۳۲ است. در نمونه‌ای مرکب از ۱۰۰ روز، نتایج زیر بدست آمده‌اند. $S = 1.185$ $\bar{x} = 1.5$

آزمون کنید که آیا میانگین تعداد این گونه افراد در پرواز از ۱٫۳۲ تجاوز می‌کند یا نه؟ $\alpha = 0.05$

$$\begin{cases} H_0 : \mu \leq 1.32 \\ H_1 : \mu > 1.32 \end{cases}$$

$$\boxed{Z_\alpha = 1.645}$$

$$Z = \frac{1.5 - 1.32}{\frac{1.185}{10}} = 1.52 \Rightarrow \boxed{Z = 1.52}$$

$Z > Z_\alpha \Rightarrow 1.52 \not> 1.645 \Rightarrow$ فرض H_0 قبول می‌گردد

مثال ⇐ در یک داروی تعیین مزه‌ی غذا، افراد نمونه‌ای تصادفی مرکب از ۱۰۰ نفر که در جامعه مصرف کنندگان مورد نظر است یک نوع غذا را چشیده‌اند و هرکس نمره‌ای از ۱ تا خوب به این غذا داده است. اگر میانگین نمونه ۶/۵ و انحراف معیار نمونه ۱/۷۶ باشد، فرض $H_0: \mu \leq 6$ را در برابر $H_1: \mu > 6$ در سطح $\alpha = 0/1$ را آزمون کنید.

$$\begin{cases} H_0: \mu \leq 6 \\ H_1: \mu > 6 \end{cases} \qquad \boxed{Z_{0/1} = 1/285} \qquad Z = \frac{6/5 - 6}{\frac{1/76}{10}} \Rightarrow \boxed{Z = 2/84}$$

$$\boxed{Z_\alpha = 1/285}$$

$Z > Z_\alpha \Rightarrow 2/84 > 1/285 \Rightarrow$ فرض H_0 رد می‌گردد

مثال ⇐ در تابستان‌های گذشته در یک کتابخانه‌ی بزرگ، میانگین تعداد کتاب‌هایی که هر عضو به امانت گرفته ۸/۵ بوده است. مدیریت کتابخانه مایل است آزمون کند که میانگین تعداد کتاب‌هایی که در این تابستان، تحت مقررات اصلاح شده امانت دادن به هر عضو، به امانت گرفته شده با سطح تابستان‌های گذشته تفاوت دارد یا نه؟ نمونه‌ای تصادفی مرکب از ۱۰۰ عضو، نتایج نمونه‌ای زیر را نشان می‌دهد. $\bar{x} = 9/34$ $S = 3/31$

آزمون را در سطح $\alpha = 0/05$ انجام دهید.

$$\begin{cases} H_0: \mu = 8/5 \\ H_1: \mu \neq 8/5 \end{cases} \qquad Z_{\frac{\alpha}{2}} = Z_{0/025} = 1/96 \Rightarrow \boxed{Z_{\frac{\alpha}{2}} = 1/96}$$

$$Z = \frac{9/34 - 8/5}{\frac{3/31}{10}} = 2/54 \Rightarrow \boxed{Z = 2/54}$$

$Z > Z_{\frac{\alpha}{2}} \quad 2/54 > 1/96 \Longrightarrow$ فرض H_0 رد می‌گردد

حالت سوم

آزمون فرض آماری میانگین یک جامعه $\mu(x)$ $\left\{ \begin{array}{l} \text{نوع جامعه نرمال و } n<30 \\ \text{انحراف معیار جامعه معلوم است} \end{array} \right.$

در این حالت آماره‌ی آزمون عبارت است:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

در این حالت ناحیه‌ی رد یا قبول آزمون مانند آزمون های بزرگ نمونه‌ای می‌باشد.

حالت چهارم

آزمون فرض آماری میانگین یک جامعه $\mu(x)$ $\left\{ \begin{array}{l} \text{نوع جامعه نرمال و } n<30 \\ \text{انحراف معیار جامعه نامعلوم است.} \end{array} \right.$

در این حالت آماره‌ی آزمون عبارت است از:

$$T = \frac{\bar{x} - \mu_0}{S_{\bar{x}}}$$

توزیع t کمی از توزیع Z کوتاهتر است لذا پراکندگی بیش‌تری نسبت به توزیع Z دارد.

مثال ⇐ فرضیه‌ای به این صورت توسط یک دانشجوی مدیریت طراحی شده است $\left\{ \begin{array}{l} H_0: \mu = 55 \\ H_1: \mu \neq 55 \end{array} \right.$ به منظور بررسی فوق، دانشجو از بین مدیران سازمان (الف) یک نمونه تصادفی ۱۲ نفره انتخاب کرده که میانگین و انحراف معیار آن به ترتیب ۶۰ و ۱۵ می‌باشد. فرض کنید فرضیه بالا دارای توزیع نرمال است. در سطح اطمینان ۹۹ درصد صحت فرضیه‌ی فوق را بررسی کنید.

$$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005 \Rightarrow t_{(0.005,\,11)} = 3.106$$

$$t = \frac{60 - 55}{\frac{15}{\sqrt{12}}} = 1.15 \Rightarrow \boxed{t = 1.15}$$

$$\boxed{t_{\frac{\alpha}{2}} = 3.106}$$

$\left. \begin{array}{l} t > t_{\frac{\alpha}{2}} \ \times \\ t < -t_{\frac{\alpha}{2}} \ \times \end{array} \right\} \Rightarrow$ فرض H_0 قبول می‌گردد.

مثال ⇐ اداره‌ی بهداشت یک شهر، می‌خواهد تعیین کند که آیا میانگین تعداد باکتری‌ها در واحد حجم آب شهر از سطح ایمنی یعنی ۲۰۰ کمتر است یا نه. پژوهشگران ۱۰ نمونه از آب را گردآوری کرده و دیده‌اند که تعداد باکتری‌ها عبارتند از:

۱۷۵ , ۱۹۰ , ۲۱۵ , ۱۹۸ , ۱۸۴ , ۲۰۷ , ۲۱۰ , ۱۹۳ , ۱۹۶ , ۱۸۰

اگر تعداد باکتری‌ها در واحد حجم آب شهر از توزیع نرمال پیروی کند آزمون را در سطح $\alpha = ۰/۰۱$ انجام دهید.

$$\begin{cases} H_0 : \mu \geq ۲۰۰ \\ H_1 : \mu < ۲۰۰ \end{cases} \qquad t(۰/۰۱ , ۹) = ۲/۸۲۱ \Rightarrow \boxed{t_\alpha = ۲/۸۲۱} \quad ① \overset{\text{مولد}}{\underset{\text{چپ}}{\Rightarrow}} \quad t_\alpha = -۲/۸۲۱$$

$$S^2 = \frac{۳۸۱۰۲۴ - ۳۷۹۴۷۰۷/۴}{۹} = ۱۷۲/۶ \Rightarrow \boxed{S = ۱۳/۱۴}$$

$$\bar{x} = \frac{۱۹۴۸}{۱۰} = ۱۹۴/۸ \Rightarrow \boxed{\bar{x} = ۱۹۴/۸}$$

$$t = \frac{۱۹۴/۸ - ۲۰۰}{\frac{۱۳/۱۴}{\sqrt{۱۰}}} = -۱/۲۵ \Rightarrow \boxed{t = -۱/۲۵} \quad ②$$

$t > -t_\alpha$ یا $t < t_\alpha \Rightarrow -۱/۲۵ < -۲/۸۲۱ \Rightarrow$ فرض H_0 را قبول می‌کنیم

(میانگین های دو جامعه):

حالت اول

وقتی نمونه های تصادفی مستقل و با اندازه های بزرگ هستند. ($n_1>30$ و $n_2>30$)

در این حالت آماره های آزمون عبارتند از:

$$Z = \frac{\bar{x}_2 - \bar{x}_1}{\sigma(\bar{x}_2 - \bar{x}_1)} \qquad \sigma(\bar{x}_2-\bar{x}_1) \longrightarrow \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

اگر σ_1 و σ_2 معلوم نباشند به جای آن ها از S_1 و S_2 استفاده می کنیم.

مثال=> برای بررسی اینکه تفاوتی بین میانگین های دو گروه A و B وجود دارد یا نه نتایج به شرح زیر بدست آمده است. آزمون برابری میانگین دو گروه $\alpha = 0{/}01$ است.

$A:\ n_1 = 200 \qquad \bar{x}_1 = 15530 \qquad S_1 = 5160$

$B:\ n_2 = 250 \qquad \bar{x}_2 = 16910 \qquad S_2 = 5840$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad \boxed{Z_{0{/}005} = 2{/}575} \qquad Z = \frac{16910 - 15530}{\sqrt{\frac{(5160)^2}{200} + \frac{(5840)^2}{250}}}$$

$$\Rightarrow Z = \frac{+1380}{519{/}5} = +2{/}66 \Rightarrow \boxed{Z = +2{/}66}$$

فرض H_0 رد می گردد و H_1 را می پذیریم

مثال => در مطالعه‌ی حقوق کارکنان یک شرکت بزرگ، نمونه‌های تصادفی مرکب از ۱۵۰ کارشناس به طور مستقل از دو بخش بزرگ شرکت انتخاب شده‌اند و نتایج زیر بدست آمده است.

آیا در سطح معنی‌دار $\alpha = 0/05$ می‌توان پذیرفت که میانگین حقوق کارکنان بخش ۲ بیش‌تر از میانگین حقوق کارکنان بخش ۱ است.

بخش ۲	بخش ۱
$n_2 = 150$	$n_1 = 150$
$\bar{x}_2 = 39212$	$\bar{x} = 37250$
$S_2 = 5356$	$S_1 = 5541$

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

① $\boxed{Z_\alpha = 1/645}$

$$Z = \frac{39212 - 37250}{\sqrt{\frac{(5541)^2}{150} + \frac{(5356)^2}{150}}} = \frac{+1962}{\sqrt{204684/54 + 191245}}$$

$$\Rightarrow Z = \frac{+1962}{\sqrt{395929/54}} = \frac{+1962}{629/22} \Rightarrow \boxed{Z = +3/118}$$ ②

فرض H_0 رد می‌گردد و H_1 را می‌پذیریم.

www.pnu-m-s.com

حالت دوم

$n_1<30$ و $n_2<30$ ، واریانس جامعه ها معلوم و جامعه ها نرمال باشند آنگاه:

همانند نمونه های بزرگ آماره آزمون عبارتست از:

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

حالت سوم

$n_1<30$ و $n_2<30$ ، واریانس جامعه ها نامعلوم و جامعه ها نرمال باشند آنگاه:

آماره ی آزمون عبارت است از:

$$T = \frac{\bar{x}_2 - \bar{x}_1}{S_{(\bar{x}_2 - x_1)}} \qquad \begin{cases} S_{(\bar{x}_2 - \bar{x}_1)} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ S_p = \sqrt{\frac{(n-1)S_1^2 + (n-1)S_2^2}{n_1 + n_2 - 2}} \end{cases}$$

آماره ی t دارای توزیع با درجه ی آزادی $(n_1 + n_2 - 2)$ است.

مثال <= تحلیل گری می خواهد میانگین طول عمر عاج یک نوع لاستیک اتومبیل را در حالتی که فشار باد لاستیک بصورت استاندارد است و در حالتی که فشار باد لاستیک بیش تر از حد استاندارد است با هم مقایسه کند. او دو نمونه تصادفی مستقل مرکب از ۱۵ لاستیک را از خط تولید انتخاب کرده است. لاستیک نمونه ۱ را با فشار باد استاندارد و لاستیکها نمونه ۲ را با فشار باد بیش از حد استاندارد تنظیم کرده، عمر عاج لاستیکها را مورد آزمایش قرار داده و نتایج زیر بدست آمده است.

نمونه ۱	نمونه ۲
$n_1 = 15$	$n_2 = 15$
$\bar{x}_1 = 43$	$\bar{x}_2 = 40.7$

اگر هر دو جامعه دارای توزیع نرمال با واریانس مشترک $\sigma = \sigma_1 = \sigma_2 = 1.2$ باشند آیا در سطح معنی‌دار بودن $\alpha = \%1$ می‌توان وجود اختلاف بین میانگین طول عمر لاستیک‌ها را در شرایط گفته شده پذیرفت؟

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \qquad \boxed{Z_{\alpha/2} = 2.576}$$

$$Z = \frac{40.7 - 43}{\sqrt{\frac{(1.2)^2}{15} + \frac{(1.2)^2}{15}}} = \frac{-2.3}{0.438} \Longrightarrow \boxed{Z = -5.25}$$

فرض H_0 رد می‌گردد ⇐ فرض H_1 را می‌پذیریم

مثال ⇐ داده‌های دو نمونهٔ تصادفی مستقل که از دو جامعهٔ نرمال با واریانس‌های مساوی‌ها استخراج شده‌اند در جدول زیر آمده است. با توجه به این داده‌ها آیا در سطح $\alpha = 0.01$ می‌توان نتیجه گرفت که میانگین جامعهٔ ۱ از ۲ بیشتر است؟

نمونه ۱	۲۲٫۵	۲۵	۳۰	۲۷٫۵	۲۰
نمونه ۲	۲۱	۱۷٫۵	۱۷	۲۰	

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

① $\bar{x}_1 = \frac{\sum x_i}{n} = \frac{125}{5} = 25$ ② $\sum x_1^2 = 3187.5$

③ $\bar{x}_2 = \frac{75.5}{4} = 18.875$ ④ $\sum x_2^2 = 1436.25$

$$S_1^2 = \frac{(3187.5) - \frac{(125)^2}{5}}{4} = 15.625$$

$$S_2^2 = \frac{1436.25 - \frac{(75.5)^2}{4}}{3} = 3.73$$

$$S_p = \sqrt{\frac{62.5 + 11.1875}{7}} = 3.24$$

$$t = \frac{25 - 18.875}{3.24\sqrt{\frac{1}{5} + \frac{1}{4}}} = 2.82 \Longrightarrow \boxed{t = 2.82}$$

$$t_{(0.01,7)} = 2.998 \Longrightarrow \boxed{t_7 = 2.998}$$

⇐ فرض H_0 را رد نمی‌کنیم و آن را قبول می‌کنیم

نکته: خطای آزمون/ در یک آزمون آماری باید قاعده‌ی تصمیم را طوری انتخاب کنیم که α، احتمال خطای I و β، احتمال خطای نوع II، در سطحی قابل قبول کنترل شوند.

وقتی اندازه‌ی نمونه از قبل معین نشده باشد، می‌توان اندازه‌ی نمونه را آنقدر بزرگ گرفت که مقدار α و β به اندازه‌ی کافی کوچک شوند، ولی در حالت معمولی که حجم نمونه از پیش تعیین شده، فقط یکی از این دو نوع خطا را می‌توان کنترل کرد. در این حالت معمولاً α را در سطح پایین کنترل می‌کنیم و مقدار آن را از قبل مشخص می‌کنیم.

«فصل یازدهم» «آزمون‌های آماری درباره‌ی نسبت و واریانس جامعه» «نکات و سؤالات مهم»

(آزمون فرض نسبت موفقیت در جامعه):

آزمون‌های مربوط به نسبت جامعه، p، وقتی اندازه‌ی نمونه بزرگ باشند ($np \geq 5$ و $nq > 5$) به همان روش آزمون‌های مربوط به میانگین جامعه انجام می‌شوند.

$$\text{آماره آزمون} \Rightarrow Z = \frac{\bar{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

① $P_0 \Rightarrow$ نسبت مورد آزمون و موجود در مسأله.

② $\bar{P}$ را با استفاده از فرمول $\frac{x}{n}$ بدست می‌آوریم.

مثال $\Rightarrow$ در محموله‌ای از ۱۴۰ گوی شیشه‌ای تزیینی که از فروشنده‌ای جدید دریافت شده است، ۲ گوی ترک دارند. فرض کنید که گوی‌های محموله نمونه‌ای تصادفی از فرآیند تولید جاری فروشنده باشد. آیا با اطلاعات داده شده می‌توان این ادعا را پذیرفت که نسبت گوی‌های شیشه‌ای معیوب کمتر از ۰/۰۲ است؟ ($\alpha = 0.05$)

$$\begin{cases} H_0: P \geq 0.02 \\ H_1: P < 0.02 \end{cases}$$

$$\bar{P} = \frac{2}{140} = 0.0143$$

$$Z = \frac{0.0143 - 0.02}{\sqrt{\frac{(0.02)(0.98)}{140}}} = \frac{-0.0057}{0.0118} = -0.4830$$

$$Z_\alpha = Z_{0.05} = 1.645$$

فرض H_0 پذیرفته می‌گردد و ادعای H_1 رد می‌گردد.

مثال $\Rightarrow$ در یک کلاس ۲۵ نفری ۱۰ نفر از درس آمار نمره‌ی قبولی دریافت نکرده‌اند، آیا می‌توانیم فرض ادعا مبنی بر آنکه نسبت قبولی درس آمار با معیار استاندارد ۸۵٪ مغایرت دارد؟

$$\begin{cases} H_0: P = 0.85 \\ H_1: P \neq 0.85 \end{cases}$$

$$\bar{P} = \frac{15}{25} = 0.6 \quad \text{نسبت قبولی}$$

$$Z = \frac{0.6 - 0.85}{\sqrt{\frac{(0.85)(0.15)}{25}}} = \frac{-0.25}{0.0714} = -3.50$$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

فرض H_0 رد می‌گردد $\Rightarrow$ ادعا (فرض H_1) پذیرفته می‌گردد.

مثال => فرض کنید p، نسبت مشتریانی باشد که محصول نوع A را به محصول نوع B ترجیح می‌دهند. مدیر فروشگاهی تصمیم می‌گیرد که اگر بیش از ۵۰ درصد مشتریان، محصول A را ترجیح دهند، فقط این نوع محصول را در فروشگاه خود عرضه کند. او نمونه‌ای تصادفی از ۲۶۵ مشتری را انتخاب و از آن‌ها سؤال می‌کند که کدام محصول را ترجیح می‌دهند. اگر تعداد مشتریانی که محصول نوع A را ترجیح می‌دهند ۱۴۴ نفر باشد، در سطح $\alpha = 0/05$ آیا مدیر شرکت فقط باید محصول نوع A را در فروشگاه عرضه کند؟

$$\begin{cases} H_0: p \leq 0/5 \\ H_1: p > 0/5 \end{cases} \qquad \bar{P} = \frac{144}{265} = 0/543$$

$$Z = \frac{0/543 - 0/5}{\sqrt{\frac{(0/5)(0/5)}{265}}} = \frac{0/043}{0/0307} = 1/40$$

$$Z_\alpha = Z_{0/05} = 1/645$$

فرض H_0 را می‌پذیریم => ادعا رد می‌گردد. بنابراین مدیر فروشگاه نباید محصول A را عرضه کند.

نکته‌ی اول => برای نسبت، معیار همواره Z می‌باشد.

نکته‌ی دوم => $\begin{cases} H_0: P = P_0 \\ H_1: P \neq P_0 \end{cases}$ => آزمون دوطرفه $\xleftarrow{\text{ناحیه رد}}$ $|Z| > Z_{\frac{\alpha}{2}}$ ، $Z < -Z_\alpha$, $Z_\alpha < Z$

نکته‌ی سوم => $\begin{cases} H_0: P \leq P_0 \\ H_1: P > P_0 \end{cases}$ => آزمون یک طرفه راست $\xleftarrow{\text{ناحیه رد}}$ $Z_\alpha < Z$

نکته‌ی چهارم => $\begin{cases} H_0: P \geq P_0 \\ H_1: P < P_0 \end{cases}$ => آزمون یک طرفه چپ $\xleftarrow{\text{ناحیه رد}}$ $Z < -Z_\alpha$

آزمون‌های فرض آماری برای واریانس جامعه:

آزمون‌های مربوط به واریانس جامعه براساس آماره‌ی آزمون زیر انجام می‌گیرند:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

قاعده‌ی تصمیم‌گیری:

وقتی جامعه‌ی موردنظر نرمال یا تقریباً نرمال باشد، آماره‌ی χ^2 دارای توزیع χ^2 با $(n-1)$ درجه‌ی آزادی است و بنابراین ناحیه‌ی رد آزمون‌های مطرح شده در سطح α عبارتند از:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases} \Rightarrow \text{آزمون دوطرفه} \xrightarrow{\text{ناحیه رد}} \chi^2 > \chi^2_{(\frac{\alpha}{2}, n-1)} \quad \chi^2 < \chi^2_{(1-\frac{\alpha}{2}, n-1)}$$

$$\begin{cases} H_0 : \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases} \Rightarrow \text{آزمون یک طرفه راست} \xrightarrow{\text{ناحیه رد}} \chi^2 > \chi^2(\alpha, n-1)$$

$$\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases} \Rightarrow \text{آزمون یک طرفه چپ} \xrightarrow{\text{ناحیه رد}} \chi^2 < \chi^2(1-\alpha, n-1)$$

مثال $\Leftarrow$ آزمایشگری معتقد است که واریانس اندازه‌هایی که در طول آزمایش ثبت می‌کند، کوچکتر از ۲ می‌باشد. در یک آزمایش او اندازه‌های ۴٫۱ و ۵٫۲ و ۱۰٫۲۵ را ثبت کرده است. اگر اندازه‌ها دارای توزیع نرمال باشند آیا می‌توان ادعای آزمایشگر را در سطح $\alpha = ۰٫۱$ پذیرفت یا نه؟

$$\begin{cases} H_0 : \sigma^2 \geq ۲ \\ H_1 : \sigma^2 < ۲ \end{cases} \qquad \bar{x} = \frac{۱۹٫۵۵}{۳} = ۶٫۵$$

$$S^2 = \frac{(۱۰٫۲۵ - ۶٫۵)^2 + (۵٫۲ - ۶٫۵)^2 + (۴٫۱ - ۶٫۵)^2}{۲} = ۱۰٫۵۷$$

$$\chi^2 = \frac{(۲)(۱۰٫۵۷)}{۲} \Rightarrow \chi^2 = ۱۰٫۵۷$$

$$\chi^2(1-\alpha, n-1) = \chi^2(۰٫۹۹, ۲) = ۰٫۰۲۰$$

فرض H_0 پذیرفته می‌شود

مثال <= کارخانه‌ای ادعا می‌کند که واریانس تولیدات آن برابر ۲ می‌باشد. مسئول کنترل کیفیت نمونه‌ای ۱۵ تایی را استخراج و واریانس نمونه را محاسبه کرده است که برابر ۴ شده است. آیا می‌توانیم در سطح ۹۵٪ تفاوت معنی‌دار بین یافته‌های مسئول کنترل کیفیت و ادعای کارخانه را بپذیریم؟

$$\begin{cases} H_0: \sigma^2 = 2 \\ H_1: \sigma^2 \neq 2 \end{cases} \qquad \chi^2 = \frac{(14)4}{2} = 28$$

$$\left.\begin{array}{l} \chi^2(1-\frac{\alpha}{2}, 14) = \chi^2(0.975, 14) = 5.63 \\ \chi^2(\frac{\alpha}{2}, 14) = \chi^2(0.025, 14) = 26.12 \end{array}\right\} \Rightarrow 5.63 < 28 < 26.12$$

فرض H_0 رد می‌گردد <= فرض H_1 را می‌پذیریم

کنترل کیفیت:

یکی از روش‌های استنباط آماری، کنترل کیفیت آماری است.

در این مبحث به دو روش مهم در کنترل کیفیت یعنی نمونه‌گیری برای پذیرش و نمودارهای کنترل می‌پردازیم.

۱- نمونه‌گیری برای پذیرش:

این روش برای پذیرش یا رد یک محموله مورد بررسی قرار می‌گیرد. فرضیه H_0 فرضیه‌ی پذیرش محموله و فرضیه H_1 فرضیه رد آن می‌باشد. اصطلاحات رایج این روش عبارتند از:

الف) سطح کیفیت قابل پذیرش (P_0) <= حداکثر نسبت اقلام معیوبی که منجر به پذیرش محموله می‌شود.

ب) عدد پذیرش <= حداکثر تعداد اقلام معیوبی که منجر به پذیرش محموله می‌شود.

ج) عدد رد <= حداقل تعداد اقلام معیوبی که منجر به رد محموله می‌شود.

د) مخاطره‌ی تولید (عرضه کننده): همان خطای α می‌باشد که حداکثر احتمال رد یک محموله‌ی قابل پذیرش است.

هـ) مخاطره‌ی خریدار: همان خطای β می‌باشد که احتمال پذیرش یک محموله‌ی غیرقابل پذیرش است

www.pnu-m-s.com

معمولاً α و β در سطح قابل قبولی کنترل می‌شوند که اگر قاعده‌ی تصمیم منجر به افزایش α از حد قابل قبول باشد باید عدد پذیرش را زیاد کنیم.

و اگر β بیش از حد قابل قبول باشد باید عدد پذیرش را کم کنیم.

مثال => یک تولید کننده کنسرو ماهی، تولید خود را در محموله‌های ۴ خرده‌فروش‌ها عرضه می‌کند. این تولید کننده طرح بازرسی خود را بدین صورت انجام می‌دهد که براساس نمونه‌ی ۴۰ تایی از محموله‌ها، اگر حداکثر ۲ قوطی معیوب باشد محموله بازرده می‌شود.

الف) عدد پذیرش و عدد رد را مشخص کنید.

ب) قاعده‌ی تصمیم را براساس $\bar{P}$ بیان کنید.

ج) اگر سطح کیفیت قابل پذیرش برابر ۰/۰۴ باشد، مخاطره‌ی عرضه کننده چقدر است؟

د) مخاطره‌ی تأیید کردن یک محموله برای آن $P = ٠/٠٦$ چقدر است؟ این مخاطره مربوط به عرضه کننده است یا خریدار؟

(الف)
همان‌طور که می‌دانیم حداکثر تعداد اقلام معیوب را عدد پذیرش گویند => عدد پذیرش = ۲

" " حداقل " " " را عدد رد گویند => عدد رد = ۳

ب) $\bar{P} = \frac{x}{n} = \frac{2}{40} = 0/05$

$$\begin{cases} H_0: P \leqslant 0/05 & (\text{پذیرش محموله}) \\ H_1: P > 0/05 & (\text{رد محموله}) \end{cases}$$

ج) مخاطره‌ی عرضه کننده همان α است و چون نمونه بزرگ است، $\bar{P}$ دارای توزیع نرمال است => $\bar{P} = \frac{P(1-P)}{n}$

$$\alpha = P(\text{رد محموله} \mid \text{محموله قابل پذیرش است}) = P\left(\bar{P} > 0/05 \mid P = 0/04\right)$$

$$P\left(\frac{\bar{P} - 0/04}{\sqrt{\frac{0/04 \times 0/96}{40}}} > \frac{0/05 - 0/04}{\sqrt{\frac{0/04 \times 0/96}{40}}}\right) = P(Z > 0/32) = 0/3745$$

(د)

$$\beta = P(\text{محموله قابل پذیرش نیست} \mid \text{پذیرش محموله}) = P(\bar{P} \leq 0.05 \mid P = 0.06)$$

$$P\left(\frac{\bar{P} - 0.06}{\sqrt{\frac{0.06 \times 0.94}{40}}} \leq \frac{0.05 - 0.06}{\sqrt{\frac{0.06 \times 0.94}{40}}}\right) = P(Z \leq -0.270) = 0.3936$$

۲- نمودارهای کنترل:

از این نمودارها که به روش زیر برای مشخصه ای مانند θ ساخته می‌شوند برای کنترل کردن کیفیت فرآیندهایی مثل فرآیندهای تولیدی و خدماتی و .... استفاده می‌شود:

الف) فاصله‌ی اطمینان (L و U) با ضریب اطمینان ۰/۹۹۷ را برای θ بدست می‌آوریم.

ب) محور افقی نمودار را برحسب زمان یا شماره‌ی نمونه‌ی انتخاب شده و محور عمودی را برحسب مقدار مشخصه θ در نمونه، درجه‌بندی می‌کنیم.

ج) میانگین فرآیند θ و حدود کنترل را بصورت خطوط افقی رسم می‌کنیم:

$$L = \theta - 3\sigma_{\bar{x}}$$

$$U = \theta + 3\sigma_{\bar{x}}$$

د) برای هر نمونه انتخاب شده مقدار مشخصه‌ی θ را محاسبه کرده و آن را روی نمودار بصورت یک نقطه مشخص می‌کنیم. اگر نقطه‌ای خارج از حدود کنترل قرار بگیرد نتیجه می‌گیریم فرآیند در آن دوره‌ی زمانی خارج از کنترل بوده و در صدد پیدا کردن علت آن برمی‌آئیم.

مثال $\Leftarrow$ وزن بسته‌های فرآورده‌ایی که با ماشین بسته‌بندی می‌شوند دارای توزیع نرمال با انحراف معیار $\sigma = 1.5$ گرم می‌باشد. میانگین مطلوب وزن هر بسته $\theta_0 = 375$ گرم است. یک نمونه تصادفی شامل ۶ بسته در هر ساعت در طی مدت عمل پر کردن انتخاب و میانگین نمونه‌ای $\bar{X}$ محاسبه شده است.

الف) یک نمودار کنترل برای میانگین فرآیند بنا کنید.

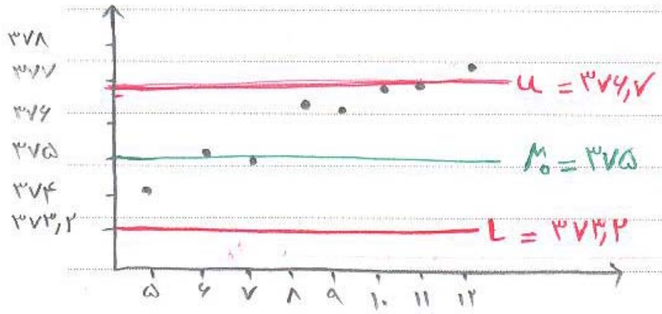
ب) میانگین های نمونه‌ای ساعات ۵ تا ۱۲ به شرح زیرند.

ساعت	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
$\bar{X}$	۳۷۴٫۲	۳۷۵٫۱	۳۷۴٫۹	۳۷۶٫۳	۳۷۶٫۲	۳۷۶٫۶	۳۷۶٫۷	۳۷۷٫۲

$$L = \mu_0 - 3\sigma_{\bar{x}} = \mu_0 - 3\frac{\sigma_x}{\sqrt{n}} = 375 - 3\,\frac{1/5}{\sqrt{6}} \simeq 373/2$$

$$U = \mu_0 + 3\sigma_{\bar{x}} = \mu_0 + 3\frac{\sigma_x}{\sqrt{n}} = 375 + 3\,\frac{1/5}{\sqrt{6}} \simeq 376/7$$

نمودار کنترل به شرح زیر است:



(فقط در ساعت ۱۲ میانگین خارج از کنترل بوده است)

مثال => شرکتی صورتحساب های دارندگان کارت اعتباری را بصورت ماهانه ارسال می کند. تجربه گذشته نشان داده است که ۴٪ درصد صورت حساب ها دارای یک یا چند اشتباه اند. نمونه هایی تصادفی به اندازه ۲۰۰ صورتحساب در ماه انتخاب و اشتباه صورت حساب ها کنترل می شود.

الف) یک نمودار کنترل برای نسبت فرآیند صورتحساب های شامل یک یا چند اشتباه بسازید.
مقادیر P_0 و L و U را روی نمودار مشخص کنید.

ب) تعداد صورت حساب هایی که یک یا چند اشتباه برای نمونه های ۱۲ ماه گذشته دارند به ترتیب زیر است

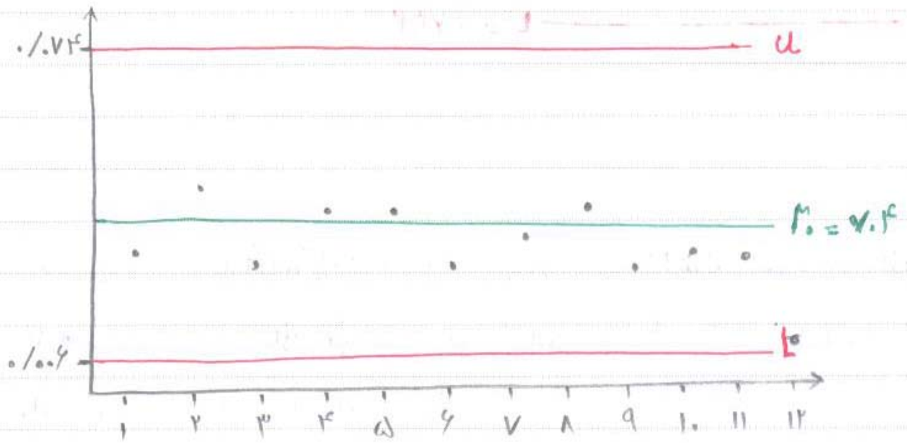
ماه	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
X تعداد	۱۰	۱۵	۱۰	۱۳	۱۳	۱۰	۱۱	۱۳	۱۴	۸	۹	۳

نسبت نمونه ای $\bar{P}$ را برای ۱۲ ماه گذشته حساب و در نمودار مشخص کنید. آیا کیفیت صورت حساب ها در ۱۲ ماه گذشته در کنترل بوده است؟

(الف):

$$\left.\begin{array}{l} P_0 = 0/04 \\ n = 300 \\ q_0 = 0/96 \end{array}\right\} \begin{array}{l} \Rightarrow L = P_0 - 3\sigma_{\bar{p}} = 0/04 - 3\sqrt{\dfrac{0/04 \times 0/96}{300}} = 0/006 \\ \Rightarrow U = P_0 + 3\sigma_{\bar{p}} = 0/04 + 3\sqrt{\dfrac{0/04 \times 0/96}{300}} = 0/074 \end{array}$$

ماه	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
$\bar{P} = \frac{x}{n}$	۰/۰۳۳	۰/۰۵۰	۰/۰۳۳	۰/۰۴۳	۰/۰۴۳	۰/۰۳۳	۰/۰۳۷	۰/۰۴۳	۰/۰۴۲	۰/۰۲۷	۰/۰۳	۰/۰۱۰



«سؤالات امتحانی»

اگر حدود کنترل برای نمونه‌های ۱۶ تایی (۳ و ۳-) باشد، میانگین و واریانس فرآیند عبارتند از:

$\mu = 0 \quad \sigma^2 = 4$ $\boxed{\mu = 0 \quad \sigma^2 = 16}$

$\mu = \sigma^2 = 0$ $\mu = 1 \quad \sigma^2 = 4$

$$\begin{cases} \mu + 3\sigma_{\bar{x}} = 3 \\ \mu - 3\sigma_{\bar{x}} = -3 \end{cases}$$

$$2\mu = 0 \Rightarrow \boxed{\mu = 0} \longrightarrow 0 + 3\sigma_{\bar{x}} = 3 \Rightarrow \boxed{\sigma_{\bar{x}} = 1}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = 1 \Rightarrow \frac{\sigma_{\bar{x}}}{4} = 1 \Rightarrow \sigma_{\bar{x}} = 4 \Rightarrow \boxed{\sigma_x^2 = 16}$$

نمونه‌هایی برابر از دو جامعه مستقل نرمال با واریانس‌های برابر گرفته‌ایم. اگر انحراف معیارهای نمونه‌ای ۲ و ۶ باشند S_p^2 کدام است؟

نمی‌توان حساب کرد. ۴ $\boxed{2\sqrt{5}}$ ۲۰

$$S_p = \sqrt{\frac{(n-1)S_1^2 + (n-1)S_2^2}{n_1+n_2-2}} \xRightarrow[\text{مسئله}]{\text{فرض}} S_p = \sqrt{\frac{(n-1)(S_1^2+S_2^2)}{2n-2}} = \sqrt{\frac{(n-1)(4+36)}{2(n-1)}}$$

$$\Rightarrow S_p = \sqrt{\frac{40}{2}} \Rightarrow S_p = \sqrt{20} = 2\sqrt{5}$$

چند نکته دیگر:

الف) ناحیه‌ی بحرانی همان ناحیه‌ی رد است.

ب) قاعده‌ی تصمیم => در یک آزمون آماری باید تصمیم‌گیری در مورد قبول یا رد هر یک از فرض‌های H_0 و H_1 بنابر قاعده‌ی معینی انجام شود. این قاعده را قاعده‌ی تصمیم می‌نامیم.

ج) آماره‌ی آزمون => آماره‌ای است که قاعده‌ی تصمیم را براساس مقدار آن طرح ریزی می‌کنیم

خطای α => وقتی H_0 صحیح است، تصمیم نادرستی که ممکن است اتخاذ شود قبول H_1 است.

خطای β => وقتی H_1 صحیح است، تصمیم نادرستی که ممکن است اتخاذ شود قبول H_0 است.

(فصل دوازدهم) «آنالیز واریانس» «خلاصه درس»

در این فصل قصد داریم مقایسه‌ی همزمان میانگین های چند جامعه‌ی نرمال را با استفاده از آنالیز واریانس انجام دهیم.

نمای کلی از جدول آنالیز واریانس:

منبع متغییر	مجموع توان دوم SS	درجه‌ی آزادی d.f	واریانس MS	F
بین گروهها	SSR	$V_1 = k-1$	MSR	$\frac{MSR}{MSE}$
درون گروهها	SSE	$V_2 = N-k$	MSE	
کل	SST	$N-1$	MST	

$$N = nk$$

$k \longrightarrow$ تعداد جامعه

$$SSR = \sum_{i=1}^{k} \frac{T_i^2}{n} - \frac{T^2}{N} \qquad SSR = MSR(k-1)$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{T^2}{N} \qquad SSR + SSE = SST$$

روش چگونگی بدست آوردن مقدار T در مثالها بیان می گردد.

نکته مهم: اگر بخواهیم آزمون را در سطح معنی داربودن α انجام دهیم باید مقدار F محاسبه شده را با $F_{(v_1, v_2, \alpha)}$ که از جدول توزیع F بدست می آید مقایسه کنیم:

$F > F_{(v_1, v_2, \alpha)} \Longrightarrow$ فرض H_0 را رد می کنیم.

$F \leq F_{(v_1, v_2, \alpha)} \Longrightarrow$ فرض H_0 را می پذیریم.

مثال <= برای مقایسه‌ی میانگین های سه جامعه‌ی نرمال با واریانس‌های مشترک σ^2، نمونه‌های تصادفی مستقل به اندازه‌های $n_1 = n_2 = n_3 = ۵$ استخراج کرده‌ایم. اگر واریانس کل داده‌ها برابر با ۹۵/۴۶ و واریانس بین گروه‌ها برابر با ۲۳۷/۵ باشد، ابتدا جدول آنالیز واریانس را تشکیل دهید و سپس فرض‌ها برابری میانگین های سه جامعه را در سطح معنادار بودن $\alpha = ۰/۰۵$ آزمون کنید.

منبع متغییر	SS	d.f	MS	F
بین گروه‌ها	$SSR = ۴۷۵$	$V_1 = ۲$	$MSR = ۲۳۷/۵$	$F = ۳,۳۱$
درون گروه‌ها	$SSE = ۸۶۱,۴۴$	$V_2 = ۱۲$	$MSE = ۷۱/۷۹$	
کل	$SST = ۱۳۳۶/۴۴$	$N - 1 = ۱۴$	$MST = ۹۵/۴۶$	

$$\left.\begin{matrix} n = ۵ \\ k = ۳ \end{matrix}\right\} \Rightarrow N = nk \Rightarrow \boxed{N = ۱۵}$$

$$MST = ۹۵,۴۱ \Rightarrow SST = MST(N-1) = ۹۵,۴۶ \times ۱۴ = ۱۳۳۶,۴۴$$

$$MSR = ۲۳۷/۵ \Rightarrow SSR = (MSR)(V_1) = ۴۷۵$$

$$SSE = SST - SSR \Rightarrow SSE = ۱۳۳۶,۴۴ - ۴۷۵ = ۸۶۱,۴۴$$

$$(MSE)(V_2) = SSE \Rightarrow MSE = \frac{SSE}{V_2} \Rightarrow MSE = \frac{۸۶۱,۴۴}{۱۲} = ۷۱,۷۹$$

$$\left.\begin{matrix} F = \dfrac{MSR}{MSE} = ۳,۳۱ \\ F_{(۲, ۱۲, ۰/۰۵)} = ۳,۸۹ \end{matrix}\right\} \Rightarrow$$

فرض H_0 را که فرض اختلاف در میانگین‌ها ست را می‌پذیریم

مثال ⇐ در یک آزمایش، نمونه‌های تصادفی به اندازه‌ی $n_1 = n_2 = n_3 = n_4 = ۵$ از چهار جامعه‌ی نرمال با واریانس مشترک استخراج شده اند. جدول آنالیز واریانس زیر را تکمیل کنید و فرض برابری میانگین جامعه‌ها را در سطح $\alpha = ۰/۰۵$ آزمون کنید؟

منبع متغییر	SS	d.f	MS	F
بین گروه‌ها	۱۶۹٫۲	۳	۵۶٫۴	۸٫۷۸
درون گروه‌ها	۱۰۲٫۷	۱۶	۶٫۴۲	

$$\left.\begin{matrix} n = ۵ \\ k = ۴ \end{matrix}\right\} \Rightarrow N = nk = ۲۰$$

$$\left.\begin{matrix} F_{(۳, ۱۶, ۰/۰۵)} = ۳٫۲۴ \\ F = ۸٫۷۸ \end{matrix}\right\} \Rightarrow$$ فرض برابری میانگین‌ها رد می‌گردد.

مثال: فرض کنید می‌خواهیم تعداد ضایعات سه ماشین را با هم مقایسه کنیم، ضایعات هر کدام از ماشین‌ها در پنج روز به صورت زیر بوده است. می‌خواهیم بدانیم آیا تفاوت معناداری بین آن‌ها وجود دارد یا نه؟ $\alpha = ۰/۰۵$

ماشین اول	۸۶	۷۹	۸۱	۷۰	۸۴
ماشین دوم	۸۹	۸۲	۸۸	۷۶	۹۰
ماشین سوم	۸۲	۶۸	۷۳	۷۱	۸۱

$$\left.\begin{matrix} k = ۳ \\ n = ۵ \end{matrix}\right\} \Rightarrow \boxed{N = ۱۵}$$

$$\left.\begin{matrix} T_1 = ۸۶ + ۷۹ + ۸۱ + ۷۰ + ۸۴ = ۴۰۰ \\ T_2 = ۸۹ + ۸۲ + ۸۸ + ۷۶ + ۹۰ = ۴۲۵ \\ T_3 = ۸۲ + ۶۸ + ۷۳ + ۷۱ + ۸۱ = ۳۷۵ \end{matrix}\right\} \Rightarrow T = ۱۲۰۰$$

Subject:
Year: Month: Date:

$$SSR = \sum_{i=1}^{k} \frac{T_i^2}{n} - \frac{T^2}{N} = \frac{(۴۰۰)^2 + (۴۲۵)^2 + (۳۷۵)^2}{۵} - \frac{(۱۲۰۰)^2}{۱۵} = ۲۵۰$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}^2 - \frac{T^2}{N} = (۸۶^2 + ۷۹^2 + \ldots + ۸۱^2) - \frac{(۱۲۰۰)^2}{۱۵} = ۶۹۸$$

$$SSE = ۶۹۸ - ۲۵۰ = ۴۴۸$$

$$\left.\begin{array}{l} (MSR \times \nu_1) = SSR \Rightarrow MSR = ۱۲۵ \\ (MSE \times \nu_2) = SSE \Rightarrow MSE = ۳۷٫۳ \end{array}\right\} \Rightarrow \boxed{F = ۳٫۳۳}$$

$$MST \times (N-1) = SST \Rightarrow SST = ۴۹٫۸۶$$

$$\boxed{F(۲, ۱۲, ۰٫۰۵) = ۳٫۸۸}$$

فرض H_0 را می‌پذیریم و اختلافی $\Leftarrow$
در میانگین‌ها وجود ندارد.

مثال $\Leftarrow$ جدول آنالیز واریانس زیر برای مقایسه‌ی میانگین‌های چهار جامعه‌ی نرمال با واریانس مشترک تشکیل شده است، جدول را تکمیل کنید و فرض برابری میانگین‌ها را در $\alpha = ۰٫۱$ آزمون کنید.

منبع تغییرات	SS	d.f	MS	F
بین گروه‌ها	۱۲۳٫۳	۳	۴۱٫۱	$F = ۱٫۱۷۱$
درون گروه‌ها	۴۲۱٫۲	۱۲	۳۵٫۱	
کل	۵۴۴٫۵	۱۵	۳۶٫۳	

$$\left.\begin{array}{l} \boxed{k = ۴} \\ N - 1 = ۱۵ \Rightarrow \boxed{N = ۱۶} \end{array}\right\} nk = N \Rightarrow ۴n = ۱۶ \Rightarrow \boxed{n = ۴}$$

$$\left.\begin{array}{l} F(۳, ۱۲, ۰٫۱) = ۲٫۶۱ \\ F = ۱٫۱۷۱ \end{array}\right\} \Rightarrow$$

فرض H_0 را می‌پذیریم و H_1 را رد می‌کنیم

Subject:
Year: Month: Date:

مثال => آزمایشی برای مقایسه قیمت کالا در چهار محله یک شهر طرح ریزی شده ، شش فروشگاه از هر محله به طور تصادفی انتخاب شده و قیمت کالا در هر فروشگاه در جدول زیر ثبت شده است . با فرض اینکه قیمت کالا در هر محله دارای توزیع نرمال با واریانس σ^2 است ، آیا می‌توان فرض برابری قیمت کالا را در این چهار محله پذیرفت یا نه ؟ $\alpha = ٪۵$

محله	قیمت کالا					
۱	۱۳۹	۱۴۳	۱۴۵	۱۴۱	۱۳۸	۱۴۴
۲	۱۳۸	۱۴۱	۱۴۴	۱۳۷	۱۴۰	۱۴۳
۳	۱۳۴	۱۳۹	۱۳۵	۱۳۹	۱۳۶	۱۳۸
۴	۱۴۹	۱۵۰	۱۴۸	۱۴۶	۱۵۱	۱۵۰

$\left.\begin{matrix} k=۴ \\ n=۶ \end{matrix}\right\} \Rightarrow N = ۲۴$ $\quad T_1 = ۸۵۰ \quad T_2 = ۸۴۳ \quad T_3 = ۸۲۱ \quad T_4 = ۸۹۴$

$T = ۳۴۰۸$

$$SSR = \sum_{i=1}^{۴} \frac{T_i^2}{n} - \frac{T^2}{N} = \frac{(۸۵۰)^2 + (۸۴۳)^2 + (۸۲۱)^2 + (۸۹۴)^2}{۶} - \frac{(۳۴۰۸)^2}{۲۴}$$

$$\Rightarrow SSR = ۴۶۸٫۳$$

$$SST = \sum_{i=1}^{۴} \sum_{j=1}^{۶} x_{ij}^2 - \frac{T^2}{N} = ۴۸۴۵۲۰ - \frac{(۳۴۰۸)^2}{۲۴} = ۵۸۴$$

$$SSE = ۵۸۴ - ۴۶۸٫۳ = ۱۱۵٫۷$$

$MSR = ۱۵۶٫۱ \qquad MSE = ۵٫۷۸۵ \qquad MST = ۲۲٫۸۳$

$\left.\begin{matrix} F = ۲۶٫۹۸ \\ F(۳, ۲۰, ۰٫۰۵) = ۳٫۱۰ \end{matrix}\right\} \Longrightarrow$ فرض برابری قیمت کالا رد می‌گردد.

Subject:
Year: Month: Date:

مثال ⇐ چندین دسته از حشره های پردار میوه را به سه نوع حشره‌کش سمپاشی و درصد مرگ و میر هر دسته را ثبت کرده‌ایم.

داده‌های زیر بدست آمده‌اند. آیا این سه حشره‌کش اثرهای یکسانی دارند؟ $\boxed{\alpha = ٠/٠٢٥}$

$n = ٧$
$k = ٣$

حشره‌کش اول	حشره‌کش دوم	حشره‌کش سوم
۴۰	۳۸	۶۸
۲۸	۴۹	۵۱
۳۱	۵۶	۴۵
۳۸	۲۵	۷۵
۴۳	۳۷	۷۵
۴۶	۳۰	۶۹
۲۹	۴۱	۶۰

$$\left.\begin{array}{l} T_1 = ٢٥٥ \\ T_2 = ٢٧٦ \\ T_3 = ٤٤٣ \end{array}\right\} \Rightarrow T = ٩٧٤$$

$$SSR = \frac{(٢٥٥)^2 + (٢٧٦)^2 + (٤٤٣)^2}{٧} - \frac{(٩٧٤)^2}{٢١}$$

$$SSR = \frac{٦٥٠٢٥ + ٧٦١٧٦ + ١٩٦٢٤٩}{٧} - ٤٥١٧٥$$

$$\Rightarrow SSR = ٣٠٣٢/١٤$$

$$SST = (٤٠)^2 + (٢٨)^2 + \dots (٦٠)^2 - \frac{(٩٧٤)^2}{٢١} = ٥٠٠١٢ - ٤٥١٧٥ = ٤٨٣٧$$

$$SSE = ١٨٠٤/٨٦$$

$$MST = ٢٤١/٨٥$$

$$\left.\begin{array}{l} MSR = ١٥١٦/٠٧ \\ MSE = ١٠٠/٢٧ \end{array}\right\} = F = ١٥/١٢$$

$$F(٢, ١٨, ٠/٠٢٥) = ٤/٥٦$$

⇐ فرض H_0 رد می‌گردد.

Subject:
Year: Month: Date:

مثال => نتایج زیر از جدول آنالیز واریانس برای مقایسه میانگین های ۴ جامعه نرمال با واریانس مشترک تشکیل شده است.
جدول را تکمیل کنید و فرض برابری میانگین ها را آزمون کنید.

$F_{۳,۱۲,۰/۱} = ۲/۶۱$

$d.f_t = ۱۵$ $MST = ۳۶/۳$ $MSE = ۳۵/۱$ $\alpha = ۰/۱$

منبع متغیر	SS	d.f	MS	F
بین گروهها	۱۲۳/۳	۳	۴۱/۱	$F = ۱/۱۷۱$
درون گروهها	۴۲۱/۲	۱۲	۳۵/۱	
کل	۵۴۴/۵	۱۵	۳۶/۳	

$N - 1 = ۱۵ \Rightarrow N = ۱۶$

$F_{(۳,۱۲,۰/۱)} > F \Rightarrow$ فرض برابری پذیرفته می گردد.

نکته => با استفاده از آنالیز واریانس بطور همزمان میانگین های چند جامعه نرمال را با هم مقایسه می کنیم

★ در بحث آنالیز واریانس برای مقایسه میانگین k جامعه چه موقع فرض برابری میانگین ها رد می گردد؟

$\boxed{\text{الف}}$ اگر برآورد σ^2 از طریق واریانس ادغام شده خیلی بزرگتر از برآورد σ^2 از طریق واریانس $\bar{x}$ باشد.

ب) اگر برآورد σ^2 از طریق واریانس ادغام شده و از طریق واریانس $\bar{x}$ خیلی با هم اختلاف داشته باشند.

ج) وقتی برآورد σ^2 از طریق واریانس $\bar{x}$ خیلی بزرگتر از برآورد σ^2 از طریق واریانس ادغام شده باشد.

د) وقتی که برآورد σ^2 از طریق واریانس ادغام شده و از طریق واریانس $\bar{x}$ باهم برابر باشند

★ برای مقایسه ی ۴ جامعه، از هر جامعه، نمونه ای ۵ تایی استخراج می کنیم. اگر $SSR = ۸۸۰$ و $SST = ۱۸۳۱$ باشد مقدار میانگین تغییرات درون گروهها چقدر می باشد؟

۳۱۷ ۲۹/۳۳ $\boxed{۵۹/۴۳}$ ۲۴۵

۸۸۰ ۳ ۲۹۳/۳

۱۶

۹۵۰

۱۸۳۱ ۱۵

مثال <= چهار جامعه ی نرمال با واریانس مشترک σ^2 داریم . از هر یک نمونه ای ۵ تایی گرفته ایم . نتایج در جدول زیر خلاصه شده اند .

جدول آنالیز واریانس را برای این داده ها تشکیل دهید و فرض تساوی میانگین ها را در سطح $\alpha = 0/01$ آزمون کنید

$$F(3,16,0/01) = 5/29$$

جامعه اول	جامعه دوم	جامعه سوم	جامعه چهارم	
۳۸۵	۳۸۰	۳۴۹	۴۴۱	T_i : مجموع داده ها
۷۷	۷۶	۶۹/۸	۸۸/۲	$\bar{x}_i$: میانگین داده ها
۲۹۹۲۵	۲۹۰۲۰	۲۴۷۸۷	۳۹۰۰۱	$\sum x_i^2$: مجموع توان دوم داده ها

منبع متغیر	SS	d.f	MS	F
بین گروهها	۸۸۰/۱۵	۳	۲۹۳/۳۹	F = ۴/۹۳۴
درون گروهها	۹۵۱/۶	۱۶	۵۹/۴۷۵	
کل	۱۸۳۱/۷۵	۱۹	۹۶/۴۱	

$$\left.\begin{array}{l} k-1=3 \Rightarrow k=4 \\ N-k=16 \Rightarrow N=20 \end{array}\right\} \Rightarrow n=5$$

$$N-1=19$$

$$SSR = \frac{(441)^2+(349)^2+(380)^2+(385)^2}{5} - \frac{(1555)^2}{20}$$

$$SSR = \frac{194481+121801+144400+148225}{5} - \frac{2418025}{20}$$

$$SSR = 121781/4 - 120901/25 = 880/15$$

$$SST = \underbrace{122733}_{\sum x_i^2} - \frac{(1555)^2}{20} = 1831/75$$

فرض تساوی میانگین ها پذیرفته می شود .

Subject:
Year: Month: Date:

بخش مهم از فصل ۱۲

مثال => برای مقایسه‌ی میانگین‌های سه جامعه‌ی نرمال با واریانس‌های مشترک، نمونه‌های تصادفی مستقل انتخاب کرده و داده‌های زیر را بدست آورده‌ایم.

$\bar{x}_1 = 81.06$ $\bar{x}_2 = 78.56$ $\bar{x}_3 = 87.81$

$S_1 = 17.05$ $S_2 = 15.43$ $S_3 = 14.36$ $n = 16$, $k = 3$

$n_1 = 16$ $n_2 = 16$ $n_3 = 16$

فرض برابری میانگین‌های سه جامعه را آزمون کنید؟ $(\alpha = 0.05)$

$$F = \frac{n S_{\bar{x}}^2}{S_p^2}$$

$$S_p^2 = \frac{(17.05)^2 + (15.43)^2 + (14.36)^2}{3} = \frac{735}{3} = 245 \Rightarrow \boxed{S_p^2 = 245}$$

$$\bar{x} = \frac{81.06 + 78.56 + 87.8}{3} = 82.48$$

$$S_{\bar{x}}^2 = \frac{\overset{2.0164}{(81.06 - 82.48)^2} + \overset{15.3664}{(78.56 - 82.48)^2} + \overset{28.3024}{(87.8 - 82.48)^2}}{3} = 22.9 \Rightarrow \boxed{S_{\bar{x}}^2 = 22.9}$$

$$\left.\begin{array}{l} F = \dfrac{(16+16+16)(22.9)}{245} = 4.49 \\ F_{(2, 45, 0.05)} = 3.32 \end{array}\right\} \Rightarrow$$ فرض برابری میانگین‌ها رد می‌گردد.

نکته مهم:

$$\frac{n S_{\bar{x}}^2}{S_p^2} = \frac{MSR}{MSE} = F \qquad MSR = \frac{SSR}{k-1} \qquad MSE = \frac{SSE}{N-k}$$

★مثال => اگر $SSR = 12$ و $n S_{\bar{x}}^2 = 3$ آنگاه تعداد گروه‌ها چند تاست؟ $\boxed{5}$ ۴ ۳ ۶

$$n S_{\bar{x}}^2 = MSR = 3 \Rightarrow (3)(k-1) = 12 \Rightarrow 3k - 3 = 12 \Rightarrow 3k = 15 \Rightarrow \boxed{k = 5}$$

مثال ⇐ فرض کنید می‌خواهیم تعداد ضایعات سه ماشین را با هم مقایسه کنیم. ضایعات هر کدام از ماشین‌ها در پنج روز به این صورت بوده است.

ماشین اول	86	79	81	70	84
″ دوم	89	82	88	76	90
″ سوم	82	68	73	71	81

$$\bar{x}_1 = \frac{86+79+81+70+84}{5} = 80$$

$$x_3 = \frac{82+68+73+71+81}{5} = 75$$

$$\bar{x}_2 = \frac{89+82+88+76+90}{5} = 85$$

$$\bar{x} = \frac{80+85+75}{3} = 80$$

$$s^2_{\bar{x}} = \frac{(80-80)^2+(85-80)^2+(75-80)^2}{2} = 25 \Rightarrow \boxed{s^2_{\bar{x}} = 25}$$

$$\Rightarrow \boxed{n s^2_{\bar{x}} = 5 \times 25 = 125} \star$$

$$① \ s_1^2 = \frac{(86-80)^2+(79-80)^2+(81-80)^2+(70-80)^2+(84-80)^2}{4} = 38.5$$

$$② \ s_2^2 = 35$$

$$③ \ s_3^2 = 38.5$$

$$\left.\begin{matrix} ① \\ ② \\ ③ \end{matrix}\right\} \Rightarrow \frac{38.5+35+38.5}{3} = \boxed{s_p^2 = 37.3} \star$$

$$\left.\begin{matrix} F = \dfrac{125}{37.3} = 3.35 \\ F(2, 12, 0.05) = 3.88 \end{matrix}\right\} \Rightarrow$$

★ فرض برابری میانگین‌ها پذیرفته می‌شود.

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

$$s_{\bar{x}}^2 = \frac{\quad}{k-1}$$

Subject:
Year: Month: Date:

★ مثال => از ۴ کارخانه ۴ نمونه ۵ تایی با انحراف معیار نمونه‌ای ۳ ۳ ۱ ۱ بدست آمده است، MSE را بدست آورید

۲۰ ۸ ۵ ۴

$$k=۴,\ n=۵ \Rightarrow N=۲۰$$

$$MSE = S_p^2 = \frac{9+9+1+1}{۴} = ۵$$

★ مثال => برای مقایسه متوسط ضایعات لگنا در ۴ استان، از هر استان ۵ مزرعه انتخاب و متوسط ضایعات در نمونه‌های هر استان عبارت است از: ۱۱ - ۱۱ - ۸ - ۱۰

اگر $SST = ۴۰$ باشد. جدول را تشکیل داده و فرض برابری میانگین‌های استان‌ها را در سطح $\alpha = ۰/۱$ آزمون کنید؟

$n=۵$, $k=۴$, $N=۲۰$

$$\bar{x} = \frac{۱۰+۸+۱۱+۱۱}{۴} = ۱۰$$

$$S_{\bar{x}}^2 = \frac{۰+۴+۱+۱}{۳} \Rightarrow \boxed{S_{\bar{x}}^2 = ۲}$$

$$nS_{\bar{x}}^2 = MSR \Rightarrow (۵)(۲) = MSR \Rightarrow \boxed{MSR = ۱۰}$$

$$\left.\begin{array}{l} SSR = ۱۰ \times ۳ = ۳۰ \\ SST = ۴۰ \end{array}\right\} \Rightarrow \boxed{SSE = ۱۰}$$

فرض برابری میانگین‌ها رد می‌گردد.

	SS	d.f	MS	F
تیمار	۳۰	۳	۱۰	
خطا	$\boxed{۱۰}$	۱۶	۰/۶۲۵	F=۱۶
کل	۴۰	۱۹		

$$\left.\begin{array}{l} F > F_{\alpha} \\ ۱۶ > ۲,۴۶ \end{array}\right\} \Rightarrow$$ فرض H_0 رد می‌گردد.

Subject:
Year: Month: Date:

از ۳ کارخانه، ۳ نمونه ۵ تایی انتخاب کرده ایم و داریم $SSR = ۸$ و $SST = ۲۰$ مقدار آماره آزمون برابری میانگین ها چند است؟

$$\left.\begin{array}{l} k = ۳ \\ n = ۵ \end{array}\right\} \Rightarrow N = ۱۵$$

$$\left.\begin{array}{l} \left.\begin{array}{l} SSR = ۸ \\ k-1 = ۲ \end{array}\right\} \Rightarrow MSR = ۴ \\ \\ \left.\begin{array}{l} SSE = ۱۲ \\ N-k = ۱۲ \end{array}\right\} \Rightarrow MSE = ۱ \end{array}\right\} \Rightarrow F = ۴$$

Subject:
Year: Month: Date:

«فصل سیزدهم» (ضریب همبستگی و رگرسیون) «خلاصه درس»

← در این فصل حالت ساده‌ای بررسی می‌شود که تعیین رابطه‌ی بین دو متغیر x و y و ضریب همبستگی این دو متغیر مورد نظر است.

(مفاهیم پایه‌ای)

متغیر مستقل یا پیشین => متغیری که توسط آزمایشگر کنترل می‌شود و آن را با x نشان می‌دهیم.

متغیر وابسته یا پاسخ => متغیری که مقدار آن به x بستگی دارد و آن را با y نشان می‌دهیم و آن را متغیر اثر یا پاسخ می‌نامیم.

مثال => در هریک از حالات زیر متغیر مستقل x و متغیر پاسخ y را تعیین کنید:

الف) رابطه‌ی بین نرخ رشد یک گیاه قارچی و میزان رطوبت محیط آن.

متغیر مستقل (x): میزان رطوبت محیط یک گیاه قارچی

متغیر وابسته (y): نرخ رشد یک گیاه قارچی

ب) رابطه‌ی بین کاهش زمان خشک شدن یک رنگ و غلظت ماده‌ی شیمیایی که بر آن می‌افزاید.

متغیر مستقل (x): غلظت ماده‌ی شیمیایی افزوده شده بر رنگ

متغیر وابسته (y): کاهش زمان خشک شدن رنگ.

(نمودار پراکنش)

در مطالعه‌ی رابطه‌ی بین دو متغیر اولین قدم رسم داده‌ها به صورت نقاطی بر روی یک نمودار است. نمودار حاصل نمودار پراکنش داده‌ها نامیده می‌شود.

(مدل رگرسیون خطی)

مدلی که برای رابطه‌ی بین x و y در نظر می‌گیریم یک رابطه‌ی خطی به شکل زیر است:

$$y_i = \alpha + \beta x_i + e_i \qquad i = 1, 2, \dots, n$$

e_i → خطای تصادفی

β → شیب خط

α → محل تلاقی با محور yها (عرض از مبدأ)

روش کمترین مربعات (روش حداقل توان های دوم) :

(پارامترهای α و β نامعلومند و باید آنها را با استفاده از داده ها برآورد کنیم. روش کمترین مربعات، روشی برای برآورد پارامترهای رگرسیون است.)

$$\hat{y} = \alpha + \hat{\beta} x$$

① $\bar{x} = \dfrac{\sum x_i}{n}$ ② $\bar{y} = \dfrac{\sum y_i}{n}$

③ $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^۲ = \sum x_i^۲ - n\bar{x}^۲$

④ $S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^۲ = \sum y_i^۲ - n\bar{y}^۲$

⑤ $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$

⑥ $\beta = \dfrac{S_{xy}}{S_{xx}}$ ⑦ $\alpha = \bar{y} - \beta\bar{x}$

⑧ $SSE = S_{yy} - \hat{\beta}^۲ S_{xx} = S_{yy} - \dfrac{(S_{xy})^۲ \cancel{(S_{xx})}}{(S_{xx})^۲} \Rightarrow SSE = S_{yy} - \dfrac{(S_{xy})^۲}{S_{xx}}$

مثال <= براساس داده های زیر خط رگرسیون را بدست آورید.

قیمت x	۱۵	۱۹	۲۱	۱۶	۱۸
تقاضا y	۱۲۳	۵۵	۲۰	۸۸	۷۶
xy	۱۸۴۵	۱۰۴۵	۴۲۰	۱۴۰۸	۱۳۶۸
$x^۲$	۲۲۵	۳۶۱	۴۴۱	۲۵۶	۳۲۴

$\boxed{\bar{x} = \dfrac{۸۹}{۵} = ۱۷٫۸}$ $\boxed{\bar{y} = \dfrac{۳۶۲}{۵} = ۷۲٫۴}$

$\boxed{\sum xy = ۶۰۸۶}$ $\sum x_i^۲ = ۱۶۰۷$

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{(۶۰۸۶) - (۵)(۱۷٫۸)(۷۲٫۴)}{۱۶۰۷ - (۵)(۱۷٫۸)^۲} = \frac{۶۰۸۶ - ۶۴۴۴}{۱۶۰۷ - ۱۵۸۴} = -۱۵٫۵۶$$

$$\alpha = \hat{y} - \beta\bar{x} \Rightarrow \alpha = ۷۲٫۴ - (-۱۵٫۵۶)(۱۷٫۸) \Rightarrow \boxed{\alpha = ۳۴۹٫۴}$$

$$\boxed{y = ۳۴۹٫۴ - ۱۵٫۵۶x}$$

Subject:
Year: Month: Date:

مثال => برای تعیین رابطه‌ی خطی بین دو متغیر x و y ، نمونه‌ای تصادفی به اندازه‌ی $n = ۱۵$ از جامعه استخراج و مقادیر دو متغیر x و y را ثبت کرده‌ایم. نتایج زیر را با استفاده از داده‌های نمونه بدست آورید؟

$\sum x_i = ۱۶۲$ $\quad \sum y_i = ۱۸۴۰٫۵$ $\quad \sum x_i^2 = ۱۸۲۰٫۲$ $\quad \sum y_i^2 = ۲۲۵۹۲۷٫۸۵$

$$\sum x_i y_i = ۱۹۹۴۵٫۷$$

معادله‌ی خط رگرسیون را پیدا کنید و مجموع مربعات مانده‌ها را بدست آورید.

$\bar{x} = ۱۰٫۸$

$\bar{y} = ۱۲۲٫۷$

$$S_{yy} = ۲۲۵۹۲۷٫۸۵ - (۱۵)(۱۲۲٫۷)^2 = ۹۸٫۵$$

$$S_{xx} = ۱۸۲۰٫۲ - (۱۵)(۱۰٫۸)^2 = ۷۰٫۶$$

$$S_{xy} = ۱۹۹۴۵٫۷ - (۱۵)(۱۰٫۸)(۱۲۲٫۷) = ۶۸٫۳$$

$$\beta = \frac{۶۸٫۳}{۷۰٫۶} \Rightarrow \boxed{\beta = ۰٫۹۶۷} \Rightarrow \alpha = ۱۲۲٫۷ - (۰٫۹۶۷)(۱۰٫۸) = ۱۱۲٫۲۵۶$$

$$\boxed{\alpha = ۱۱۲٫۲۵۶}$$

$\boxed{y = ۱۱۲٫۲۵۶ + ۰٫۹۶۷x} \Rightarrow$ معادله‌ی خط رگرسیون

$$SSE = ۹۸٫۵ - \frac{(۶۸٫۳)^2}{۷۰٫۶} = ۳۲٫۴۲۵ \Rightarrow \boxed{SSE = ۳۲٫۴۲۵}$$

مثال => برای تعیین رابطه‌ی خطی بین دو متغیر x و y ، نمونه‌ای تصادفی به اندازه $n = ۲۰$ از جامعه استخراج و مقادیر x و y را ثبت کرده‌ایم ، نتایج زیر بدست آمده‌اند:

$\sum x_i = ۳۵$ $\quad \sum y_i = ۴۸$ $\quad \sum x_i y_i = ۹۶۰$

$\sum x_i^2 = ۶۸۰$ $\quad \sum y_i^2 = ۱۳۴۸$

معادله‌ی خط رگرسیون و مجموع مربعات مانده‌ها را حساب کنید؟

$\bar{x} = ۱٫۷۵$ $\quad \bar{y} = ۲٫۴$

$$\left.\begin{array}{l} S_{xx} = ۶۸۰ - ۲۰(۱٫۷۵)^2 = ۶۱۸٫۷۵ \\ S_{xy} = ۹۶۰ - ۲۰(۱٫۷۵)(۲٫۴) = ۸۷۶ \end{array}\right\} \Rightarrow \beta = \frac{۸۷۶}{۶۱۸٫۷۵} = ۱٫۴۲$$

$$\alpha = ۲٫۴ - (۱٫۴۲)(۱٫۷۵) = -۰٫۰۸۵$$

$$\boxed{y = -۰٫۰۸۵ + ۱٫۴۲x}$$

$S_{yy} = ۱۳۴۸ - ۲۰(۲٫۴)^2 = ۱۳۰۰ \Rightarrow SSE = ۱۳۰۰ - \frac{(۸۷۶)^2}{\dots} = ۵۲٫۳۸$

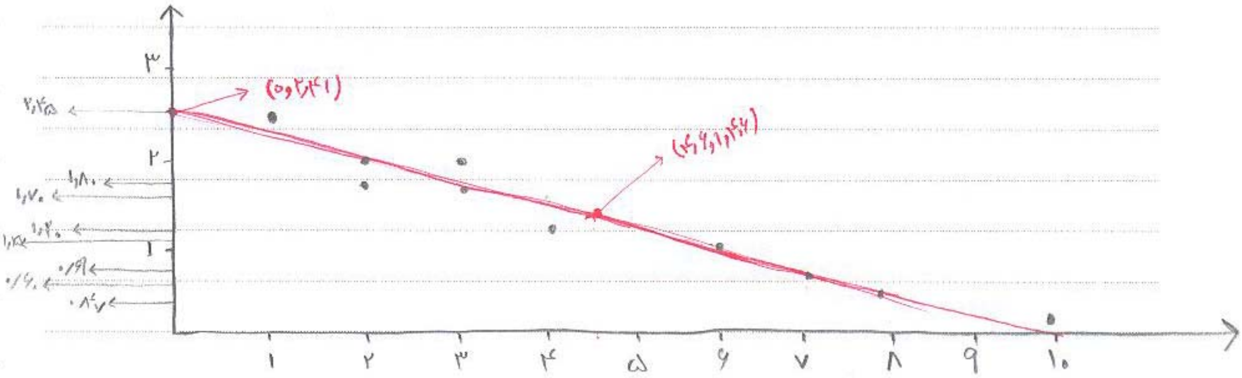
Subject:
Year: Month: Date:

مثال ⇐ قیمت ۱۰ دستگاه الکترونیکی از نوع معینی برحسب سال‌های کارکرد آن‌ها در جدول زیر آمده است.

x سال‌های کارکرد	۱	۲	۳	۳	۳	۴	۶	۷	۸	۱۰
y قیمت دستگاه	۲٫۴۵	۱٫۸۰	۲	۲	۱٫۷۰	۱٫۲۰	۱٫۱۵	۰٫۶۹	۰٫۶۰	۰٫۴۷
$x_i y_i$	۲٫۴۵	۳٫۶	۴	۶	۵٫۱	۴٫۸	۶٫۹	۴٫۸۳	۴٫۸	۴٫۷

الف) نمودار پراکنش را رسم کنید　　ب) معادلهٔ خط رگرسیون را تعیین و آن را بر روی نمودار پراکنش رسم کنید

ج) با استفاده از خط رگرسیون، مقدار پیش‌بینی شده برای متوسط قیمت یک دستگاه را که ۵ سال کارکرده تعیین کنید.



$$\bar{x} = \frac{\sum x_i}{n} = \frac{46}{10} \Rightarrow \boxed{\bar{x} = 4.6} \qquad \boxed{\bar{y} = 1.406} \qquad \boxed{\sum x_i y_i = 47.18}$$

$$S_{xx} = 292 - (10)(4.6)^2 = 80.4 \qquad S_{xy} = 47.18 - (10)(4.6)(1.406) = -17.496$$

$$\beta = \frac{-17.496}{80.4} = -0.218 \Rightarrow \alpha = \bar{y} + 0.218\bar{x} \Rightarrow \alpha = 1.406 + (0.218)(4.6) = 2.41$$

$$\boxed{y = 2.41 - 0.218x}$$

ج)

$$y = 2.41 - (0.218)(5)$$

$$\Rightarrow y = 1.32$$

x	y	xy	x^2	y^2
۱	۲٫۴۵	۲٫۴۵	۱	۳٫۰۰۲۵
۲	۱٫۸۰	۳٫۶	۴	۳٫۲۴
۳	۲	۴	۴	۴
۳	۲	۶	۹	۴
۳	۱٫۷۰	۵٫۱	۹	۲٫۸۹
۴	۱٫۲۰	۴٫۸	۱۶	۱٫۴۴
۶	۱٫۱۵	۶٫۹	۳۶	۱٫۳۲۲۵
۷	۰٫۶۹	۴٫۸۳	۴۹	۰٫۴۷۶۱
۸	۰٫۶۰	۴٫۸	۶۴	۰٫۳۶
۱۰	۰٫۴۷	۴٫۷	۱۰۰	۰٫۲۲۰۹
$\sum$ ۴۶	۱۴٫۰۶	۴۷٫۱۸	۲۹۲	۲۳٫۹۵۲

نکته: برای رسم معادلهٔ خط رگرسیون ابتدا دو نقطهٔ (۰وα) و ($\bar{y}$و$\bar{x}$) را رسم کرده و به هم وصل می‌کنیم و امتداد می‌دهیم.

CLASSIC

Subject:
Year: Month: Date:

★ یکی از هدف‌های مطالعه‌ی رگرسیون، استفاده از خط رگرسیون به دست آمده، برای تعیین برآورد امید ریاضی پاسخ متناظر با یک سطح معین از متغیر مستقل است.

★ (ضریب همبستگی):

روش‌های رگرسیون موقعی مناسب اند که بین متغیرهای تصادفی X و Y یک رابطه‌ی خطی قوی وجود داشته باشد.

(در این بخش با معرفی معیاری برای اندازه‌گیری شدت رابطه‌ی خطی بین دو متغیر X و Y می‌پردازیم.)

کوواریانس: متغیرهای تصادفی X و Y را با $cov(X,Y)$ نشان می‌دهیم و مقدار آن را از رابطه‌ی زیر بدست می‌آوریم:

$$cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$cov(XY) = E(XY) - E(X)E(Y)$$

ضریب همبستگی خطی: یکی از معیارهای عددی، برای تعیین شدت رابطه‌ی خطی بین X و Y ضریب همبستگی خطی است که آن را با ρ (رو) نشان می‌دهیم و مقدار آن را از رابطه‌ی زیر بدست می‌آوریم.

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

برای محاسبه‌ی ρ باید توزیع جامعه معلوم باشد.

ضریب همبستگی نمونه‌ای: آماره‌ای که برای برآورد ضریب همبستگی خطی X و Y از آن استفاده می‌کنیم را با r نشان می‌دهیم و آن را ضریب $\boxed{\text{همبستگی نمونه‌ای}}$ یا ضریب همبستگی پیرسن می‌نامیم.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

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

مثال ← ضریب همبستگی خطی بین x و y را با استفاده از ۱۰ زوج از مقادیر مشاهده‌شده‌ی x و y که در جدول آمده است برآورد کنید.

x	y	xy	x_i^2	y_i^2
۱,۲	۱۰۱	۱۲۱,۲	۱,۴۴	۱۰۲۰۱
۰/۸	۹۲	۷۳,۶	۰/۶۴	۸۴۶۴
۱	۱۱۰	۱۱۰	۱	۱۲۱۰۰
۱,۳	۱۲۰	۲۶	۱,۶۹	۱۴۴۰۰
۰/۷	۹۰	۶۳	۰/۴۹	۸۱۰۰
۰/۸	۸۲	۶۵,۶	۰/۶۴	۶۷۲۴
۱	۹۳	۹۳	۱	۸۶۴۹
۰/۶	۷۵	۴۵	۰/۳۶	۵۶۲۵
۰/۹	۹۱	۸۱,۹	۰/۸۱	۸۲۸۱
۱,۱	۱۰۵	۱۱۵,۵	۱,۲۱	۱۱۰۲۵
$\Sigma \Rightarrow$ ۹,۴	۹۵۹	۷۹۴,۸	۹,۲۸	۹۳۵۶۹

$$r = \frac{(۱۰)(۷۹۴,۸) - (۹,۴)(۹۵۹)}{\sqrt{\left[(۱۰)(۹,۲۸) - (۹,۴)^2\right]\left[۱۰(۹۳۵۶۹) - (۹۵۹)^2\right]}} = ۰/۸۷۵$$

(رابطه‌ی بین r و β :)

دو فرمول زیر را در نظر بگیرید. ملاحظه می‌شود که مقادیر r و β با هم وابسته‌اند. زیرا در فرمول‌های یاد شده در صورت کسر مقدار S_{xy} قرار دارد و اگر $S_{xy}=0$، مقادیر r و $\hat{\beta}$ هر دو برابر با صفرند.

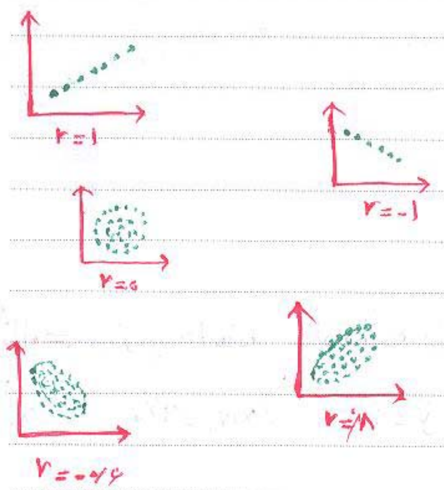
$$\beta = \frac{S_{xy}}{S_{xx}} \qquad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

نکته: پس اگر $r=0$ گردد آن گاه x و y **همبستگی خطی ندارند** و نقاط در نمودار کاملاً پراکنده‌اند.

★ (با توجه به فرمول‌هایی که برای محاسبه‌ی r مورد استفاده هستند، می‌توان نشان داد که مقدار r بین ۱- و ۱ تغییر می‌کند).

$$\boxed{-1 \le r \le 1}$$

(تفسیر r :)



همبستگی x و y مستقیم و کامل است. $r=1$

همبستگی x و y معکوس و کامل است. $r=-1$

بین x و y همبستگی خطی نیست. $r=0$

x و y دارای همبستگی مستقیم هستند. $0<r\le 1$

x و y دارای همبستگی معکوس هستند. $-1\le r<0$

Subject:
Year: Month: Date:

مثال ⇐ ضریب همبستگی نمونه‌ای را برای داده‌های جدول (۱۳-۵) زیر بدست آورید؟

x	y	xy	$x_i^۲$	$y_i^۲$
۷۶	۲٫۲	۱۶۷٫۲	۵۷۷۶	۴٫۸۴
۸۹	۲٫۴	۲۱۳٫۶	۷۹۲۱	۵٫۷۶
۸۳	۳٫۱	۲۵۷٫۳	۶۸۸۹	۹٫۶۱
۷۹	۲٫۵	۱۹۷٫۵	۶۲۴۱	۶٫۲۵
۹۱	۳٫۵	۳۱۸٫۵	۸۲۸۱	۱۲٫۲۵
۹۵	۳٫۶	۳۴۲	۹۰۲۵	۱۲٫۹۶
۸۲	۲٫۵	۲۰۵	۶۷۲۴	۶٫۲۵
۶۹	۲٫۰۰	۱۳۸	۴۷۶۱	۴
۶۶۴	۲۱٫۸	۱۸۳۹٫۱	۵۵۶۱۸	۶۱٫۸۲

$$r = \frac{\overbrace{(۸)(۱۸۳۹٫۱)}^{۱۴۷۱۲٫۸} - \overbrace{(۶۶۴)(۲۱٫۸)}^{۱۴۴۷۵٫۲}}{\sqrt{[\underbrace{۸\times(۵۵۶۱۸)}_{۴۴۴۹۴۴} - \underbrace{(۶۶۴)^۲}_{۴۴۰۸۹۶}][\underbrace{۸\times(۶۱٫۸۲)}_{۴۹۴٫۵۶} - \underbrace{(۲۱٫۸)^۲}_{۴۷۵٫۲۴}]}} = \frac{۲۳۷٫۶}{\sqrt{۴۰۴۸\times ۲۰٫۱۲}} = \frac{۲۳۷٫۶}{۲۸۵٫۳۸}$$

$$\Rightarrow \boxed{r = ۰/۸۳}$$

مثال ⇐ با توجه به اطلاعات داده شده ضریب همبستگی x و y را بدست آورده و آن را تفسیر کنید؟

$n = ۲۰ \quad \sum x_i = ۳۵ \quad \sum y_i = ۴۸ \quad \bar{x} = ۱٫۷۵ \quad \bar{y} = ۲٫۴ \quad \sum x_i y_i = ۹۶۰ \quad \sum x_i^۲ = ۶۸۰$

$\sum y_i^۲ = ۱۳۴۸$

$$\left.\begin{aligned} S_{xy} &= ۹۶۰ - (۲۰)(۲٫۴)(۱٫۷۵) = ۸۷۶ \\ S_{xx} &= ۶۸۰ - (۲۰)(۱٫۷۵)^۲ = ۶۱۸٫۷۵ \\ S_{yy} &= ۱۳۴۸ - (۲۰)(۲٫۴)^۲ = ۱۲۳۲٫۸ \end{aligned}\right\} \Rightarrow r = \frac{۸۷۶}{\sqrt{۶۱۸٫۷۵ \times ۱۲۳۲٫۸}}$$

$$\Rightarrow r = \frac{۸۷۶}{۸۷۳} \quad ??$$

Subject:
Year: Month: Date:

۱- در معادله‌ی خط رگرسیونی $y = \frac{-x}{2}$ ، پیش‌بینی امید ریاضی پاسخ به ازای $x=4$ برابر است با

$$x=4 \longrightarrow y = \frac{-4}{2} \Rightarrow \boxed{y=-2}$$

۲- اگر برآورد شیب خط رگرسیونی برابر ۲- باشد، همبستگی خطی x و y چگونه است؟
معکوس

۳- در معادله خط رگرسیونی $\hat{y} = \frac{-x}{2}$ ، اگر $\bar{y} = 1$ باشد، $\bar{x}$ کدام است؟ ۲ صفر $\boxed{-2}$ ۱

$$\alpha = \bar{y} - \beta\bar{x} \Rightarrow 0 = 1 - (-\tfrac{1}{2})(\bar{x}) \Rightarrow -1 = \tfrac{1}{2}\bar{x} \Rightarrow \boxed{\bar{x} = -2}$$

۴- اگر $\sum x_i = 15$ و $\sum y_i = 5$ و $\sum x_i y_i = 25$ و $S_{xy} = 10$ و $S_{yy} = 10$ و $S_{xx} = 10$ و تعداد داده‌ها $n=5$ باشد، ضریب همبستگی خطی بین دو متغیر x و y را بدست آورید؟

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \Rightarrow r = \frac{10}{\sqrt{100}} \Rightarrow \boxed{r=1}$$

در سؤال قبل مجموع مربعات مانده‌ها (SSE) چقدر است؟

$$SSE = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} \Rightarrow SSE = 10 - \frac{100}{10} \Rightarrow \boxed{SSE = 0}$$

در سؤال قبل معادله‌ی رگرسیون خطی کدام است؟

$$\beta = \frac{S_{xy}}{S_{xx}} \Rightarrow \boxed{\beta = 1} \qquad \boxed{\bar{x} = 3} \qquad \boxed{\bar{y} = 1}$$

$$\alpha = \bar{y} - \beta\bar{x} \Rightarrow \alpha = 1 - (1)(3) \Rightarrow \boxed{\alpha = -2}$$

$$\Rightarrow \boxed{y = -2 + x}$$

۵- فرض کنید معادله‌ی خط رگرسیون بصورت $y = 2 + 0/38x$ باشد، اگر مقدار $\bar{x} = 3/9$ و $\bar{y} = 3/5$ باشد آنگاه مقدار پیش‌بینی برای y به ازای $x = \bar{x}$ چقدر است؟ ۲ $\boxed{3/5}$ ۰/۳۸ ۳/۹

$$x = \bar{x} = 3/9 \Rightarrow y = 2 + (0/38)(3/9) \Rightarrow y = 3/482$$

$$\alpha = \bar{y} - \beta\bar{x} \quad 2$$

x	y	xy	x_i^2	y_i^2
۵	۰	۰	۲۵	۰
۱	۴	۴	۱	۱۶
۴	۲	۸	۱۶	۴
۳	۰	۰	۹	۰
۲	-۱	-۲	۴	۱
۱۵	۵	۱۰	۵۵	۲۱

۶- برای پنج زوج مشاهدات زیر r را بیابید

$$r = \frac{(5)(10) - (15)(5)}{\sqrt{\underbrace{[(5)(55) - (15)^2]}_{825}\,[(5)(21) - (5)^2]}}$$

$$\frac{50 - 75}{\sqrt{[600]\,[80]}} = \frac{-25}{219{,}089} = -0/115 = \boxed{r = -0/115}$$

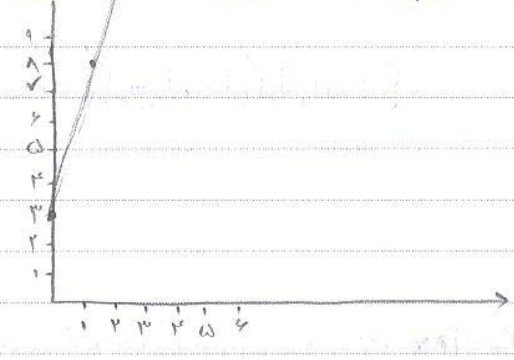
۷- در یک تحقیق می‌خواهیم معادله‌ی خط رگرسیونی بدست آوریم تا نمره‌ی امتحان دانشجویان یک دانشگاه در درس آمار را برحسب ساعات مطالعه آن درس پیش‌بینی کنیم، متغیرها را مشخص کنید.

$X \Rightarrow$ متغیر مستقل / متغیر پیش‌بین $\Rightarrow$ ساعات مطالعه‌ی درس

$Y \Rightarrow$ متغیر وابسته / متغیر پاسخ / متغیر اثر $\Rightarrow$ نمره‌ی امتحان دانشجویان

۸- فرض کنید بین متغیر x و y رابطه‌ی خطی $y = 3 + 5x$ برقرار باشد؛ آنگاه مقدار ضریب همبستگی بین x و y چقدر است؟

$r = 1$ $\quad r < 0$ $\quad r > 0$ $\quad r = 0$



$$\frac{S_{xy}}{S_{xx}} = 5$$

$$S_{xy} = 5 S_{xx}$$

$$\frac{5 S_{xx}}{\sqrt{S_{xx} S_{yy}}}$$

Subject:
Year: Month: Date:

۹- اگر مقادیر مشاهده شده برای دو متغیر x و y بصورت زیر باشد مقدار برآورد پارامتر β در کدام گزینه می باشد؟

x	۱	۲	۳
y	۳	۲	۱

(الف و ب) $\beta = ۰$ $\beta < ۰$, $\beta > ۰$

$\sum x_i = ۶$ $\bar{x} = ۲$ $\sum xy = ۱۰$ $\sum y_i^۲ = ۱۴$

$\sum y_i = ۶$ $\bar{y} = ۲$ $\sum x_i^۲ = ۱۴$

$$\beta = \frac{(۱۰) - (۳)(۲)(۲)}{(۱۴) - (۳)(۲)^۲} = \frac{۱۰ - ۱۲}{۱۴ - ۱۲} = \frac{-۲}{۲} = -۱$$

۱۰- برآورد β در معادلهٔ خط رگرسیون $y = \alpha + \beta x$ در صورتیکه $\sum x_i = ۱۶۲$ $\sum x_i^۲ = ۱۸۲۰,۲$

$n = ۱۵$ و $\sum y_i^۲ = ۲۲۵۹۲۷,۸۵$ و $\sum y_i = ۱۸۴۰,۵$ و $\sum x_i y_i = ۱۹۹۴۵,۷$ باشد چقدر است؟

$$\beta = \frac{(۱۹۹۴۵,۷) - (۱۵)(۱۰,۸)(۱۲۲,۷)}{(۱۸۲۰,۲) - (۱۵)(۱۰,۸)^۲} = \frac{۶۸,۳}{۷۰,۶} \Rightarrow \beta = ۰,۹۶۷$$

- حال مقدار α را بیابید؟

$\alpha = \bar{y} - \beta\bar{x} \Rightarrow \alpha = ۱۲۲,۷ - (۰,۹۶۷)(۱۰,۸) \Rightarrow \alpha = ۱۱۲,۲۹$

۱۱- در جدول داده‌ها مقدار ضریب زاویه خط رگرسیون را بدست آورید؟

x	y	xy	$x_i^۲$	$y_i^۲$
۳	۹	۲۷	۹	۸۱
۸	۱۳	۱۰۴	۶۴	۱۶۹
۱۰	۱۴	۱۴۰	۱۰۰	۱۹۶
۲۱	۳۶	۲۷۱	۱۷۳	۴۴۶

$$\beta = \frac{(۲۷۱) - (۳)(۷)(۱۲)}{(۱۷۳) - (۳)(۷)^۲} = \frac{۱۹}{۲۶} = ۰,۷۳$$

مقدار عرض از مبدأ را بدست آورید؟

$\alpha = (۱۲) - (۰,۷۳)(۷) \Rightarrow \boxed{\alpha = ۶,۸۸۴}$

★ در خط رگرسیون بدست آمده در سؤال بالا مقدار پیش بینی $x = ۵$ چقدر است؟

۱۱,۰۳ ۲۸,۲۹ ۶,۸۹ ۹,۸۱

CLASSIC

$y = ۶,۸۸۴ + (۰,۷۳)(۵) \Rightarrow \boxed{y = ۱۰,۵۴}$?!!

۱۲- برای داده‌های زیر برآورد خط رگرسیون را بیابید.

x_i	y_i	xy	x_i^2	y_i^2
۱	۱	۱	۱	۱
۳	۲	۶	۹	۴
۴	۴	۱۶	۱۶	۱۶
۶	۴	۲۴	۳۶	۱۶
۸	۵	۴۰	۶۴	۲۵
۹	۷	۶۳	۸۱	۴۹
۱۱	۸	۸۸	۱۲۱	۶۴
۱۴	۹	۱۲۶	۱۹۶	۸۱
۵۶	۴۰	۳۶۴	۵۲۴	۲۵۶
۷	۵			

$$\beta = \frac{(۳۶۴) - (۸)(۷)(۵)}{(۵۲۴) - (۸)(۷)^۲}$$

$$\beta = \frac{۸۴}{۱۳۲} \Rightarrow \boxed{\beta = ۰/۶۴}$$

$$\alpha = ۵ - (۰/۶۴)(۷) \Rightarrow \boxed{\alpha = ۰/۵۲}$$

$$\boxed{y = ۰/۵۲ + ۰/۶۴x}$$

۱۳- با استفاده از یک نمونه ۶ تایی از (x و y) نتایج زیر بدست آمده است:

$$\sum x_i = ۳۰ \qquad \sum y_i = ۲۴۰ \qquad \sum xy = ۱۳۲۰ \qquad \sum x_i^2 = ۱۶۶ \qquad \sum y_i^2 = ۱۰۶۰۰$$

۵ ۴۰

الف) برآورد معادله‌ی خط رگرسیون را بدست آورید.

$$\beta = \frac{(۱۳۲۰) - (۶)(۵)(۴۰)}{(۱۶۶) - (۶)(۵)^۲} = \frac{۱۲۰}{۱۶} \Rightarrow \boxed{\beta = ۷/۵}$$

$$\alpha = ۴۰ - (۷/۵)(۵) \Rightarrow \boxed{\alpha = ۲/۵}$$

$$\Rightarrow y = ۲/۵ + ۷/۵x$$

ب) ضریب همبستگی نمونه‌ای را محاسبه و آن را تفسیر کنید؟

$$S_{yy} = ۱۰۶۰۰ - (۶)(۴۰)^۲ = ۱۰۰۰$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \Rightarrow r = \frac{۱۲۰}{۱۲۶/۵} \Rightarrow \boxed{r = ۰/۹۴۸}$$

End