سوالات نمونه سوالات امتحان آمار استنباطی و پاسخ برای سال ۱۴۰۲   (با سوالات فاصله اطمینان و فرض صفر)

۱) نام دو نرم افزار در آمار که در این ترم تحصیلی آموزش دیدید را ذکر کنید(۲نمره)

Excel   SPSS (Statistical package for social science)

۲) آمار توصیفی و آمار استنباطی را توضیح دهید(۲نمره)

بدست آوردن شاخصهای آماری (مثل مود میانه میانگین و ...) در خصوص دادههایی که داریم چه نمونه باشد چه جمعیت ، را آمار توصیفی گویند

بررسی اینکه آیا شاخص های آماری دادههای بدست آمده در نمونه، در آمار توصیفی،   برای جمعیت تعمیم (عمومیت) دارد را آمار استنباطی گویند

۳) جمعیت چیست نمونه چیست (۲نمره)

جمعیت :   مجموعه تمام عناصری که دارای یک یا چند ویژگی مشترک باشند

نمونه: بخشی از جمعیت میباشد

۴) انواع روشهای مقیاس گذاری (measure) را نام ببرید هر کدام نیم خط توضیح دهید(۲نمره)

پاسخ : مقیاس اسمی (NOMINAL)- کد گذاری عددی – اعداد نسبت داده شده ارجحیت و برتری ندارند (مثل جنسیت - زن ۲   مرد ۱)

مقیاس رتبهای(ORDINAL) (ترتیبی) - با کد گذاری عددی – اعداد نسبت داده شده ارجحیت دارند ولی تناسب ندارند (مثل پرسشنامه لینکرت شامل موافق۳ – بی نظر ۲ – مخالف ۱

مقیاس نسبتی(SCALE) (وزنی) با کد گذاری عددی – اعداد نسبت داده شده ارجحیت دارند و تناسب هم دارند (مثل سن و مثل وزن - مثلا وزن احمد ۳۵ و وزن عباس ۷۰ کیلوگرم یعنی عباس بیشتر از احمد است و به نسبت دو برابر هم میباشد)

مقیاس فاصلهای - مثلا به یک بیمار بگوییم اگر حداکثر درد مثل یک خط کش عدد ۱۰۰ باشد درد شما چقدر است و بیمار بگوید ۶۵

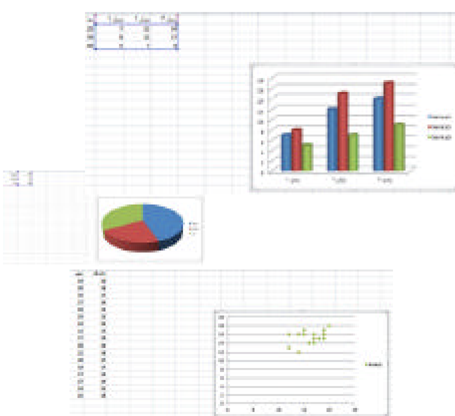
۵) سه مورد واریانس و انحراف معیار و ضریب تغییرات ، هر یک مورد بیانگر چه چیزی است معدل اعداد میانه اعداد یا پراکندگی اعداد

پاسخ: هر سه مورد سه مورد واریانس و انحراف معیار و ضریب تغییرات بیانگر پراکندگی اعداد است

۶) اگر بخواهیم بجای زرد عدد ۳ و بجای آبی عدد ۲ و بجای قرمز عدد ۵ را بگذاریم این مقیاس گذاری ordinal یا nominal یا scale کدامیک است

پاسخ: اسمی Nominal

۷) چهار مورد از انواع نمودارها نام ببرید و هر یک نیم خط توضیح دهید و شکل نمودار را ترسیم کنید (۲نمره)



نمودار میله ای Bar Chart: برای مشاهده فراوانی داده ها و مقایسه داده ها نسبت به هم

نمودار دایره ایPie Chart : برای مشاهده سهم هر مورد از داده ها

نمودار هیستوگرام Histogram: برای مشاهده نحوه توزیع داده ها بصورت پیوسته

نمودار خطی Line Chart: برای نشان دادن رابطه بین دوسری اعداد مثلا خط رگرسیون

نمودار پراکندگی Scatter Plot : برای نشان دادن پراکندگی بین دوسری اعداد مثلا در رگرسیون

۸) وقتی دو سری داده مثلا سال و مصرف اینترنت در سالهای گذشته داریم با استفاده از فرمول رگرسیون یا همبستگی خطی چه دستاوردی خواهیم داشت؟ (۲ نمره)

پیش بینی آینده با داشتن دادههای سالهای قبل

۹) در یک منطقه عدد آزمایش قند افراد در چند سال گذشته بشرح ذیل میباشد معادله خط رگرسیون را نوشته پیش بینی سال ۱۴۰۲ و ۱۴۰۳ چقدر میباشد (۴نمره)

سال=x	۱۳۹۷	۱۳۹۸	۱۳۹۹	۱۴۰۰	۱۴۰۱	۱۴۰۲	۱۴۰۳
خون	۳	۳	۴	۴	۵	؟؟	؟؟

پاسخ:

از همه اعدادی که به نام x داریم ۱۳۹۹ تا کم میکنیم

سال = x	-2	-1	0	+1	+2	+3	+4
Y=خون	۳	۳	۴	۴	۵	؟؟	??

$y = a + bx$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i * \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{(-2*3)+(-1*3)+(0*4)+(1*4)+(2*5) - \frac{(-2-1+0+1+2)(3+3+4+4+5)}{5}}{((-2)^2+(-1)^2+(0)^2+(1)^2+(2)^2) - \frac{(-2-1+0+1+2)^2}{5}} = \frac{5}{10} = 0.5$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{-2-1+0+1+2}{5} = 0 \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{3+3+4+4+5}{5} = \frac{19}{5} = 3.8$$

$\bar{y} = a + b\bar{x} \qquad 3.8 = a + (0.5*0) \qquad a = 3.8$

$y = 3.8 + 0.5x$

$Xm = 1399 \quad \rightarrow$

$y = 1402 - 1399 = 3 \quad \rightarrow y = 3.8 + (0.5*3) = 5.3$

$y = 1403 - 1399 = 4 \quad \rightarrow y = 3.8 + (0.5*4) = 5.8$

۱۰) این جدول قبلا حل شده است و میانگین 13.7 و واریانس 4.61 میباشد . مجدد حل نکنید

X_i	10	12	13	15	17
f_i	1	2	3	2	2

الف) انحراف معیار را حساب کنید ب) بدون حل کردن اگر هنگام درج فرمول واریانس توان دوم را فراموش کنیم نتیجه کسر حاصل چه خواهد شد(محاسبه نکنید فقط جواب بنویسید) ج) جدول زیر شباهت زیادی به جدول بالا در همین سوال دارد با استفاده از جواب قسمت بالا میانگین و انحراف معیار و و ضریب تغییرات این داده جدید را بدست آورید(بدون استفاده از فرمول اصلی)

Y_i	90	92	93	95	97
f_i	1	2	3	2	2

پاسخ

$\bar{x} = 13.7 \qquad \sigma_x = \sqrt{4.61} = 2.15$

ب) در محاسبه کسر واریانس اگر توان دوم فراموش کنیم جواب صفر میشود و به این محاسبه، واریانس نمیگویند

ج) با مشاهده داده های قدیم x ملاحظه میکنیم اگر 80 واحد به آن اضافه کنیم داده جدید به نام y حاصل شده

بنابراین طبق جدول فوق در خصوص y ها میتوان نوشت که

$y_i = x_i + 80$

و طبق قانون میتوان نوشت چون داده های جدید همگی80 واحد اضافه شده اند پس میانگین داده های جدید80 واحد بیشتر از میانگین داده های قبلی میشود و انحراف معیار داده های جدید با انحراف معیار داده های قدیم هیچ فرقی نمیکند

$\bar{y} = \bar{x} + 80 = 13.7 + 80 \rightarrow \bar{y} = 93.7$

$\sigma_y^2 = \sigma_x^2 \rightarrow$

$\sigma_y = \sigma_x = \sqrt{4.61} = 2.15 \rightarrow \sigma_y = 2.15$

ضریب تغییرات

$$\rho = \frac{\sigma}{\mu} = \frac{2.15}{93.7} = 0.022 = 2.2\%$$

الان $\rho = 0.022$ عدد خیلی کوچکی است یعنی پراکندگی دادهها کم است

۱۱) نمرات دانشجویان بشرح زیر است. الف) مد(نما) داده ها چه مقدار است (چرا) ب). میانه داده ها چه مقدار است؟ ج) صدک 24 را محاسبه کنید. د) میانگین و واریانس و انحراف معیار نمرات دانشجویان را بدست آورید ه) جدول فراوانی و فراوانی تجمعی و فراوانی نسبی و فراوانی نسبی تجمعی را تکمیل کنید و در این جدول در خصوص همه اعداد زیر داده ۱۳ یک به یک توضیح دهید (۴ نمره)

نمرات : ۱۶ و ۱۲ و ۱۱ و ۱۸ و ۱۸ و ۱۳ و ۱۶ و ۱۶ و ۱۱ و ۱۳ (۴نمره)

نمره = x	11	12	13	16	18
تعداد =f	2	1	2	3	2
F	2	3	5	8	10
r_i	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{2}{10}$
R_i	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	$\frac{8}{10}$	$\frac{10}{10}$

مد داده ای بمقدار 16 میباشد زیرا بیشترین تعداد را دارد

M=16

میانه داده ها

$Q\sum f_i = \frac{1}{2}(2+1+2+3+2) = \frac{10}{2} = 5$ ➔ در سطر Fi ➔ 5+ ➔ Fi=8 ➔ $x_i = 16 = m$ میانه

یعنی نصف داده ها 16 یا کمتر از 16 هستند

صدک24 داده ها

$Q\sum f_i = \frac{24}{100}(2+1+2+3+2) = \frac{24}{100} = 2.4$ ➔ در سطر Fi ➔ 2.4+ ➔ Fi=3 ➔ x=12 صدک ۲۴

یعنی 24% داده ها 12 یا کمتر از 12 هستند

============

$$\bar{x} = \mu = \frac{\sum x_i * f_i}{\sum f_i} = \frac{(11*2)+(12*1)+(13*2)+(16*3)+(18*2)}{2+1+2+3+2} = \frac{144}{10} = 14.4$$

$$\sigma^2 = \frac{\sum (x-\bar{x})^2 f_i}{\sum f_i}$$

$$\sigma^2 = \frac{(11-14.4)^2*2+(12-14.4)^2*1+(13-14.4)^2*2+(16-14.4)^2*3+(18-14.4)^2*2}{2+1+2+3+2} = 6.64$$ واریانس

$$\sigma = \sqrt{\sigma^2} = \sqrt{6.64} = 2.58$$ انحراف معیار

دو تا از داده ها ۱۳ میباشد ، ۵ تا از داده ها ۱۳ یا کمتر از ۱۳ میباشد دو دهم یا ۲۰٪ داده‌ها ۱۳ میباشد و پنج دهم یا ۵۰٪ داده ها ۱۳ یا کمتر از ۱۳ میباشد

مثال احتمال برای جمعیت(نرمال) :

۱۲) یک دانشجو ۱۱۰ واحد درسی را با توزیع نرمال و میانگین نمرات 17 و واریانس 4 گذرانده است، در یک درس در ترم بعدی. الف) احتمال اینکه نمره این دانشجو حداکثر 15 شود چقدر است؟ ب) احتمال اینکه نمره این دانشجو بیشتر از 19 شود چقدر است ج) احتمال اینکه نمره دقیقا 18 شود؟ د) جمع میانگین با انحراف معیار را تحلیل کنید

(جدول توزیع نرمال استاندارد بشرح زیر در نظر بگیرید ۴نمره)

$P(z \leq 2) = 0.98$, $P(z \leq 1.96) = 0.975$, $P(z \leq 1.645) = 0.95$, $P(z \leq 1.5) = 0.93$,
$P(z \leq 1) = 0.85$, $P(z \leq 0.5) = 0.7$, $P(z \leq 0.25) = 0.6$,
$P(z \leq 0) = 0.5$, $P(z \leq -0.25) = 0.4$, $P(z \leq -0.5) = 0.3$, $P(z \leq -1) = 0.15$,
$P(z \leq -1.5) = 0.07$, $P(z \leq -2) = 0.02$

حل:

$$P(x \leq 15) = P\left(\frac{x-\mu}{\sigma} \leq \frac{15-17}{2}\right) = P(z \leq -1) = 0.15$$

$$P(x > 19) = 1 - P(x \leq 19) = 1 - P\left(\frac{x-\mu}{\sigma} \leq \frac{19-17}{2}\right) = 1 - P(z \leq +1) = 1 - 0.85 = 0.15$$

$$P(x = 18) = P(x \leq 18) - P(x \leq 17) = P\left(\frac{x-\mu}{\sigma} \leq \frac{18-17}{2}\right) - P\left(\frac{x-\mu}{\sigma} \leq \frac{17-17}{2}\right) = P(z \leq +0.5) - P(z \leq 0)$$
$$= 07 - 0.5 = 0.2$$

تحلیل $\mu \pm \sigma$ $\quad$ $\mu \pm 2\sigma$

$\mu \pm \sigma = 17 \pm 2 = 15$ تا 19

یعنی ترم بعد بیش از ۶۸٪ نمره بین 18 تا 20 میشود

$\mu \pm 2\sigma = 17 \pm (2 * 2) = 17 \pm 4 = 13$ تا 21

یعنی بیش از ۹۸٪ نمره بین *13* تا *20* میشود

مثال احتمال برای جمعیت(نرمال) :

۱۳) در کل یک دانشگاه میانگین ١٤ و واریانس ٣,٦ میباشد احتمال اینکه نمره یک دانشجو درکل این دانشگاه بیشتز از ١٦ باشد چقدر است؟

(قبلا در مورد توزیع نرمال اشاره شده بود)

$$X \sim N(\mu = 14, \sigma^2 = 3.6) \qquad \sigma = \sqrt{3.6} = 1.89$$

$$\boldsymbol{p(x > 16) = 1 - p(x \leq 16) = 1 - p\left(\frac{x-\mu}{\sigma} \leq \frac{16-\mu}{\sigma}\right) = 1 - p\left(z \leq \frac{16-14}{1.89}\right) = 1 - p(z \leq +1.058)}$$
$$\boldsymbol{= 1 - 0.8531 = 0.145}$$

به احتمال ۱۴٪ نمره یک دانشجو بیشتر از ۱۶ میشود

در اطمینان ۹۵٪ یک طرفه (یک سمت) این احتمال را از توزیع نرمال استاندارد نباید فراموش کنیم

$$P(z \leq 1.645) = 0.95$$

در فاصله اطمینان۹۵٪ دو طرفه (دو سمت) این احتمال را از توزیع نرمال استاندارد نباید فراموش کنیم

$$P(z \leq 1.96) = 0.975$$

مثال فاصله اطمینان برای جمعیت(نرمال) :

۱۴) در یک کلاس میانگین ۱٤ و واریانس ۳,٦ میباشد با اطمینان ۹۵٪($X_{0.05}$) بیشترین نمره کلاس را بدست آورید

در اینجا اطمینان از یک سمت خواسته است (بیشترین نمره را خواسته) (یک طرفه) و بازه بین کمترین و بشترین نمره را نخواسته)

$$\sigma = \sqrt{3.6} = 1.89 \qquad \mathbf{X \sim N(\mu = 14, \sigma^2 = 3.6)}$$

$$p(x \leq k) = 0.95$$

$$p\left(\frac{x-\mu}{\sigma} \leq \frac{k-\mu}{\sigma}\right) = 0.95$$

$$p\left(z \leq \frac{k-14}{1.89}\right) = 0.95$$

$$p\left(z \leq \frac{k-14}{1.89}\right) = 0.95 = \text{از جدول} = 0.95 = p(z \leq 1.645)$$

$$\frac{k-14}{1.89} = 1.645 \quad \rightarrow \quad k - 14 = 1.89 * 1.645$$

$$\rightarrow \quad k = 17.1$$

یعنی به احتمال ۹۵٪ نمره دانشجو حداکثر ۱۷,۱ خواهد بود یا ۵٪ احتمال دارد نمره دانشجویی بیشتر از ۱۷,۱ شود

در فاصله اطمینان۹۵٪ دو طرفه (دو سمت) این احتمال را از توزیع نرمال استاندارد نباید فراموش کنیم

$$P(z \le 1.96) = 0.975$$

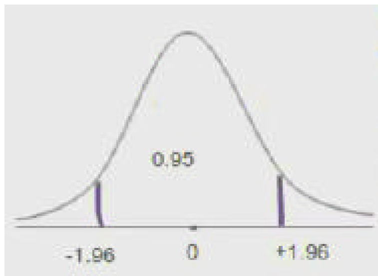
مثال فاصله اطمینان برای جمعیت(نرمال)

۱۵) : در کل یک کلاس میانگین ۱٤ و واریانس ۳,٦ میباشد   با اطمینان ۹۵٪($X_{0.05}$) محدوده نمرات را بدست آورید

$\mathbf{X \sim N(\mu = 14, \sigma^2 = 3.6)}$ $\qquad$ $\boldsymbol{\sigma = \sqrt{3.6} = 1.89}$

در اینجا فاصله اطمینان خواسته است (دو طرفه) (اطمینان از یک سمت نخواسته است   و   بازه بین کمترین و بشترین نمره را خواسته)

برای بدست آوردن اطمینان ۹۵٪ یعنی محدوده بالا و پایین نمرات از هر طرف ۲.۵٪ جمعا به احتمال ۹۵٪ میتوان نوشت



حل بطور خلاصه

$$z = \frac{k-\mu}{\sigma} \qquad -1.96 < \frac{k-\mu}{\sigma} < +1.96$$

$$Low = k - 1.96 * \sigma = 14 - 1.96 * 1.89 = 14 - 1.96 * 1.89 = 10.29$$

$$Up = k + 1.96 * \sigma = 14 + 1.96 * 1.89 = 14 + 1.96 * 1.89 = 17.7$$

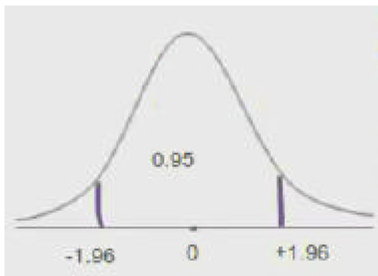
یعنی با اطمینان ۹۵٪ میتوان گفت میانگین جمعیت بین ۱۰,۲۹ تا ۱۷,۷ میباشد

مثال فاصله اطمینان برای نمونه بیش از ۳۰ نفر (نرمال):

۱۶) از نمونه ۲۰۰ نفری دانشجویان میانگین نمرات ۱۳ و انحراف معیار ۱,۵ میباشد فاصله اطمینان ۹۵٪ برای میانگین را محاسبه کنید و توضیح بنویسید

در اینجا فاصله اطمینان خواسته است (دو طرفه) (اطمینان از یک سمت نخواسته است و بازه بین کمترین و بشترین نمره را خواسته)

در جمعیت انحراف معیار $\boldsymbol{\sigma}$ بود که که در حالت نمونه انحراف معیار را $\boldsymbol{SD}$ مینامیم. با توجه به صورت مسئله آماره ها $\boldsymbol{\bar{x} = 13}$ و $\boldsymbol{SD = 1.5}$ میباشد برآورد نقطه ای میانگین جمعیت 13 میباشد اما این برآورد از نمونه به جمعیت مقداری خطا دارد و طبق شکل زیر 95% در وسط به مفهوم دوتا 2.5% از دو طرف میشود



$$\boldsymbol{(z \le 0.975) = 1.96}$$

برای جمعیت $\boldsymbol{z = \dfrac{\bar{x} - \mu}{\sigma}}$ $\rightarrow$ برای نمونه $\boldsymbol{z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$ $\rightarrow$ در این مسئله $\boldsymbol{-1.96 < \dfrac{\bar{x} - \mu}{\frac{SD}{\sqrt{n}}} < +1.96}$

$$\boldsymbol{Low = \bar{x} - 1.96 * \frac{SD}{\sqrt{n}} = 13 - 1.96 * \frac{SD}{\sqrt{n}} = 13 - 1.96 * \frac{1.5}{\sqrt{200}} = 12.78}$$

$$\boldsymbol{Up = \bar{x} + 1.96 * \frac{SD}{\sqrt{n}} = 13 + 1.96 * \frac{SD}{\sqrt{n}} = 13 + 1.96 * \frac{1.5}{\sqrt{200}} = 13.22}$$

یعنی با اطمینان ۹۵٪ میتوان گفت میانگین جمعیت بین ۱۲,۷۸ تا ۱۳,۲۲ میباشد

در فرض صفر، اطمینان از یک طرف بخواهد این احتمال را از توزیع نرمال استاندارد نباید فراموش کنیم

$$P(z \leq 1.645) = 0.95$$

مثال فرض صفر برای نمونه بیشتر از ۳۰ تا :

۱۷) از تحقیقات قبلی مشخص میشود که در ایران هوش عاطفی زنانی که تحصیلات عالی دارند دارای توزیع نرمال با میانگین ٦٥ و انحراف معیار ۱۲ میباشد این تحقیقات ادعا دارد گه در استان آذربایجان میانگین هوش عاطفی زنان بیشتر از دیگر استانها است . یک نمونه ۱۰۰ نفری در آذربایجان انتخاب کردیم و مشخص شد که میانگین ٦٨ است با سطح اطمینان ۹۵٪ ادعای تحقیقات را آزمون کنید

آزمونی یک دامنه است(یک سمت)

$$\begin{Bmatrix} H_0:\ \mu \leq \mu_0 \\ H_1:\ \mu > \mu_0 \end{Bmatrix} \qquad \begin{Bmatrix} H_0:\ \mu \leq 65 \\ H_1:\ \mu > 65 \end{Bmatrix}$$

obs یعنی مشاهده = observe

در جمعیت اینگونه $\frac{\bar{x}-\mu_0}{\sigma}$ مینوشتیم در حال بررسی در نمونه میباشیم پس به صورت $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ مینویسیم

روش اول از طریق Z (یعنی نگاه به بیرون جدول نرمال استاندارد)

اکثر اوقات با ۹۵٪ ادعا را میخواهیم اثبات یا رد کنیم

$P = 0.95 \quad \rightarrow$ طبق جدول $\rightarrow \ 0.95 = P(Z \leq 1.645)$ رابطه اول

حال طبق صورت مسئله

$P\left(z \leq \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) = P\left(z \leq \frac{68-65}{\frac{12}{\sqrt{100}}}\right) = P\left(z \leq \frac{3}{1.2}\right) = P(Z \leq 2.5) = P(Z \leq Z_{obs}) \rightarrow Z_{obs} = 2.5$ رابطه دوم

حال چون رابطه زیر برقرار است

$$\mathbf{1.645 \leq 2.5}$$

در نتیجه با اطمینان ۹۵٪ میتوان فرض صفر H0 را رد نمود و پذیرای فرض H1 شد یعنی میانگین هوش عاطفی زنان با تحصیلات دانشگاهی از ۶۵ بیشتر است

روش دوم از طریق P (یعنی نگاه به داخل جدول نرمال استاندارد)

ادعای سوال را مینویسیم

$$Sig = P(Z > Z_{obs}) = 1 - P(Z \leq Z_{obs}) = 1 - P\left(z \leq \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) = 1 - P\left(z \leq \frac{68-65}{\frac{12}{\sqrt{100}}}\right) = 1 - P\left(z \leq \frac{3}{1.2}\right)$$

$$= 1 - P(Z \leq 2.5) = \text{از جدول} = 1 - 0.9938 = 0.0062$$

حال چون رابطه زیر برقرار است

$$\boldsymbol{Sig = 0.0062 \leq 0.05}$$

در نتیجه با اطمینان ۹۵٪ میتوان فرض صفر H0 را رد نمود و پذیرای فرض H1 شد یعنی میانگین هوش عاطفی زنان با تحصیلات دانشگاهی از ۶۵ بیشتر است

مثال فرض صفر برای نمونه بیش از ۳۰ نفر (نرمال):

۱۸) لامپهای ساخت کارخانه ای با انحراف معیار ۱۴۲ ساعت در یک نمونه ۱۰۰ تایی میانگین عمرشان ۱۲۸۰ ساعت شد صاحب کارخانه ادعا میکند که با همین اطلاعات میانگین عمر لامپهای ای کارخانه از ۱۲۰۰ ساعت بیشتر است آیا میتوان ادعای او را قبول یا رد کرد ؟ (هر وقت آلفا مشخص ندادند مقدارش 0.05 در نظر میگیریم = اطمینان ٪۹۵)

آزمونی یک دامنه است(یک سمت)

$$\begin{Bmatrix} H_0:\ \mu \le \mu_0 \\ H_1:\ \mu > \mu_0 \end{Bmatrix} \qquad \begin{Bmatrix} H_0:\ \mu \le 1200 \\ H_1:\ \mu > 1200 \end{Bmatrix}$$

obs یعنی مشاهده = observe

در جمعیت اینگونه $\frac{\bar{x}-\mu_0}{\sigma}$ مینوشتیم در حال بررسی در نمونه میباشیم پس به صورت $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ مینویسیم

روش اول از طریق Z (یعنی نگاه به بیرون جدول نرمال استاندارد)

اکثر اوقات با ٪۹۵ ادعا را میخواهیم اثبات یا رد کنیم

$P = 0.95 \quad \rightarrow$ طبق جدول $\rightarrow \ 0.95 = P(Z \le 1.645)$ رابطه اول

حال طبق صورت مسئله

$$\boldsymbol{P}\left(z \le \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) = \boldsymbol{P}\left(z \le \frac{1280-1200}{\frac{142}{\sqrt{100}}}\right) = \boldsymbol{P(Z \le 5.6)} = P(Z \le \boldsymbol{Z_{obs}}) \rightarrow \boldsymbol{Z_{obs}} = 5.6$$ رابطه دوم

با مقایسه رابطه اول و دوم ، چون رابطه زیر برقرار است

$$\mathbf{1.645 \le 5.6}$$

پس در نتیجه با اطمینان ٪۹۵ میتوان فرض صفر H0 را رد نمود و پذیرای فرض H1 شد یعنی ادعای صاحب کارخانه را که عمر لامپهایش بیشتر از ۱۲۰۰ ساعت است، میپذیریم

روش دوم از طریق P (یعنی نگاه به داخل جدول نرمال استاندارد)

ادعای سوال را مینویسیم

$$\boldsymbol{Sig = P(Z > Z_{obs}) = 1 - P(Z \le Z_{obs}) = 1 - P}\left(z \le \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) = \boldsymbol{1-P}\left(z \le \frac{1280-1200}{\frac{142}{\sqrt{100}}}\right) = \boldsymbol{1-P}\left(z \le \frac{80}{14.2}\right)$$

$$= \boldsymbol{1 - P(Z \le 5.6)} = \text{از جدول} = \mathbf{1 - 0.9999 = 0.0001}$$

حال چون رابطه زیر برقرار است

$$\boldsymbol{Sig = 0.0001 \le 0.05}$$

پس در نتیجه با اطمینان ٪۹۵ میتوان فرض صفر H0 را رد نمود و پذیرای فرض H1 شد یعنی ادعای صاحب کارخانه را که عمر لامپهایش بیشتر از ۱۲۰۰ ساعت است، میپذیریم

مثال ضریب همبستگی پیرسون و اسپیرمن (رتبه‌ای)

۱۹) تعدادی افراد بالای ۷۰ سال میزان فشارخون آنها را اندازه گرفتیم جدول زیر حاصل شد

X= میزان سن بالای ۷۰ سال	3	4	4	5	5	5
Y= فشارخون	1	2	2	2	3	3

ضریب همبستگی پیرسون و اسپیرمن (رتبه‌ای) بین دو سری عدد x و y را بدست آورید و در خصوص علامت این ضریب و مقدار آن توضیح بنویسید (۲۰ نمره)

پاسخ : حل به روش ضریب همبستگی پیرسون

پیرسون برای x و y و x*y از این سه فرمول میانگین و واریانس و انحراف معیار x و y و x*y را بدست میاوریم

$\bar{x} = \mu = \frac{\sum x_i * f_i}{\sum f_i}$ میانگین $\sigma^2 = \frac{\sum (x-\bar{x})^2 f_i}{\sum f_i}$ واریانس $\sigma = \sqrt{\sigma^2}$ انحراف معیار

									میانگین	واریانس	انحراف معیار
x=فاصله سن	3	4	4	5	5	5			4.3333	0.5556	0.7454
y=فشار خون	1	2	2	2	3	3			2.1667	0.4722	0.6872
x*y	3	8	8	10	15	15			9.8333		

$$\rho(x,y) = Corr(x,y) = r = \frac{Cov(x,y)}{\sigma_x \, \sigma_y} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \, \sigma_y}$$

$$\rho(x,y) = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \, \sigma_y} = \frac{9.8333 - (4.3333 * 2.1667)}{0.7454 * 0.6872} = 0.8677$$

ضریب همبستگی پیرسون 0.86+ شد یعنی دو سری داده وابستگی زیاد در جهت مستقیم دارند

حل به روش ضریب همبستگی اسپیرمن

X	3	4	4	5	5	5	داده سری اول
Rx	1	2	3	4	5	6	رتبه سری اول
Rx	1	2.5	2.5	5	5	5	رتبه سری اول نرم شده

Y	1	2	2	2	3	3	داده سری دوم
rY	1	2	3	4	5	6	رتبه سری دوم
Ry	1	3	3	3	5.5	5.5	رتبه سری دوم نرم شده

Rx-Ry	0	-0.5	-0.5	2	-0.5	-0.5
(Rx-Ry)^2	0	0.25	0.25	4	0.25	0.25

$$r_s = 1 - \frac{6\sum_{i=1}^{n} D^2}{n(n^2-1)} = 1 - \frac{6 * \{0^2 + (-0.5)^2 + (-0.5)^2 + (2)^2 + (-0.5)^2 + (-0.5)^2\}}{6(6^2-1)} = 1 - \frac{6 * 5}{210} = 1 - 0.1429 = +0.8571$$

ضریب همبستگی اسپیرمن 0.85+ شد یعنی دو سری داده وابستگی زیاد در جهت مستقیم دارند