

(۱) جمعیت چیست نمونه چیست (۵نمره)

جمعیت: مجموعه تمام عناصری که دارای یک یا چند ویژگی مشترک باشند
نمونه: بخشی از جمعیت میباشد

(۲) انواع روشهای مقیاس گذاری (measure) را نام ببرید هر کدام نیم خط توضیح دهید (۵نمره)

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 لیکرت شامل موافق ۳ - 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

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

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

(۴) 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

(۵) این جدول قبلا حل شده است و میانگین 13.2 و واریانس 2.26 میباشد. مجدد حل نکنید

X_i	10	12	13	14	15
f_i	1	3	9	5	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) د) جدول زیر شباهت زیادی به جدول بالا در همین سوال دارد با استفاده از جواب قسمت بالا میانگین و انحراف معیار و ضریب تغییرات این داده جدید را بدست آورید (بدون استفاده از فرمول اصلی)

X_i	90	92	93	94	95
f_i	1	3	9	5	2

پاسخ

$$\bar{x} = 13.2 \quad \sigma_x = \sqrt{2.26} = 1.50$$

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 ۸۰ واحد به آن اضافه کنیم داده جدید Y حاصل میشود
بنابراین طبق جدول فوق در خصوص Y ها میتوان نوشت که

$$y_i = x_i + 80$$

و طبق قانون میتوان نوشت

$$\bar{y} = \bar{x} + 80 = 13.2 + 80 \rightarrow \bar{y} = 93.2$$

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

$$\sigma_y = \sigma_x = \sqrt{2.26} = 1.50 \rightarrow \sigma_y = 1.50$$

چون داده های جدید همگی 80 واحد اضافه شده اند پس میانگین داده های جدید 80 واحد بیشتر از میانگین داده های قبلی میشود و انحراف معیار داده های جدید با انحراف معیار داده های قدیم هیچ فرقی نمیکند
ضریب تغییرات

$$\rho = \frac{\sigma}{\mu} = \frac{1.50}{93.2} = 0.016 = 1.6\%$$

الان $\rho = 0.016$ عدد خیلی کوچکی است یعنی پراکندگی داده‌ها کم است

۶) نمرات دانشجویان بشرح زیر است. الف) مد(نما) داده‌ها چه مقدار است (چرا) ب). میانه داده‌ها چه مقدار است؟ ج) صدک ۱۳ 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 (۴۰نمره)

نمرات: ۱۲ و ۱۱ و ۱۳ و ۱۳ و ۱۶ و ۱۲ و ۱۵ و ۱۵ و ۱۳ و ۱۶

نمره = x	11	12	13	15	16
تعداد = f	1	2	3	2	2
F	1	3	6	8	10
r_i	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{2}{10}$
R_i	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{10}{10}$

مد داده ای بمقدار 13 میباشد زیرا بیشترین تعداد را دارد

$$M=14$$

میانه داده‌ها

$$Q \sum f_i = \frac{1}{2}(1 + 2 + 3 + 2 + 2) = \frac{10}{2} = 5 \rightarrow \text{در سطر } F_i \rightarrow 5+ \rightarrow F_i=6 \rightarrow m=13$$

یعنی نصف داده‌ها 13 یا کمتر از 13 هستند

صدک 13 داده‌ها

$$Q \sum f_i = \frac{13}{100}(1 + 2 + 3 + 2 + 2) = \frac{13}{100} = 1.3 \rightarrow \text{در سطر } F_i \rightarrow 1.3+ \rightarrow F_i=3 \rightarrow x=12$$

یعنی 13% داده‌ها 12 یا کمتر از 12 هستند

=====

$$\bar{x} = \mu = \frac{\sum x_i * f_i}{\sum f_i} = \frac{(11 * 1) + (12 * 2) + (13 * 3) + (15 * 2) + (16 * 2)}{1 + 2 + 3 + 2 + 2} = \frac{136}{10} = 13.6$$

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

$$\sigma^2 = \frac{(11 - 13.6)^2 * 1 + (12 - 13.6)^2 * 2 + (13 - 13.6)^2 * 3 + (15 - 13.6)^2 * 2 + (16 - 13.6)^2 * 2}{1 + 2 + 3 + 2 + 2} = 2.84 \text{ واریانس}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.84} = 1.69 \text{ انحراف معیار}$$

سه تا از داده‌ها ۱۳ میباشد، ۶ تا از داده‌ها ۱۳ یا کمتر از ۱۳ میباشد سه دهم یا ۳۰٪ داده‌ها ۱۳ میباشد و شش دهم یا ۶۰٪ داده‌ها ۱۳ یا کمتر از ۱۳ میباشد

۷) 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 (۲۰نمره)

سال=X	۱۳۹۷	۱۳۹۸	۱۳۹۹	۱۴۰۰	۱۴۰۱	۱۴۰۲	۱۴۰۳
خون	۲	۴	۴	۴	۵	??	??

پاسخ:

همه اعدادی که به نام X داریم ۱۳۹۹ تا کم میکنیم

سال=y	-2	-1	0	+1	+2	+3	+4
خون	۲	۴	۴	۴	۵	??	??

$$y = a + bx$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i * \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$= \frac{(-2 * 2) + (-1 * 4) + (0 * 4) + (1 * 4) + (2 * 5) - \frac{(-2 - 1 + 0 + 1 + 2)(2 + 4 + 4 + 4 + 5)}{5}}{((-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2) - \frac{(-2 - 1 + 0 + 1 + 2)^2}{5}} = \frac{6}{10}$$

$$= 0.6$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{-2 - 1 + 0 + 1 + 2}{5} = 0$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2 + 4 + 4 + 4 + 5}{5} = \frac{19}{5} = 3.8$$

$$\bar{y} = a + b\bar{x} \quad 3.8 = a + (0.6 * 0) \quad a = 3.8$$

$$y = 3.8 + 0.6x$$

$$Xm = 1399 \rightarrow$$

$$y = 1402 - 1399 = 3 \rightarrow y = 3.8 + (0.6 * 3) = 5.6$$

$$y = 1403 - 1399 = 4 \rightarrow y = 3.8 + (0.6 * 4) = 6.2$$

۸) از سوابق یک آزمایشگاه روی حیوانات مشخص شده که ۸۰٪ آزمایشها صحیح است و بقیه غلط است. هفته آینده قرار است ۲۵ آزمایش با توزیع نرمال انجام شود. الف: میانگین، واریانس، انحراف معیار، آزمایش صحیح در هفته آینده چقدر است؟

ب) میانگین با انحراف معیار جمع و تفریق کنید و یکبار دیگر میانگین با دو برابر انحراف معیار جمع و تفریق کنید و برای داده فوق که نرمال است هر دو جواب را کامل توضیح دهید ..

ج: احتمال اینکه هفته آینده حداکثر ۱۷ آزمایش صحیح باشد چقدر است -- د) احتمال اینکه هفته آینده بیشتر از ۲۱ آزمایش صحیح باشد چقدر است -- ه) احتمال اینکه هفته آینده بیشتر از ۱۹ آزمایش صحیح باشد چقدر است

جدول توزیع نرمال

$$P(z \leq 2) = 0.98, \quad P(z \leq 1.5) = 0.93, \quad P(z \leq 1) = 0.85, \quad P(z \leq 0.5) = 0.7, \\ P(z \leq 0) = 0.5, \quad P(z \leq -0.5) = 0.3, \quad P(z \leq -1) = 0.15, \\ P(z \leq -1.5) = 0.07, \quad P(z \leq -2) = 0.02$$

پاسخ:

$$\text{شکست} = q = 1 - p = 1 - 0.8 = 0.2 = \text{شکست} \rightarrow p = 0.8 = \text{موفقیت}$$

$$\bar{x} = \mu = n * p = 25 * 0.8 = 20 \text{ میانگین}$$

$$\sigma^2 = n * p * q = 25 * 0.8 * 0.2 = 4 \text{ واریانس}$$

$$\sigma = \sqrt{4} = 2 = \text{انحراف معیار}$$

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

$$\mu \pm \sigma = 20 \pm 2 = 18 \text{ تا } 22$$

یعنی بیش از ۶۸٪ داده ها بین ۱۸ تا ۲۰ میباشند

$$\mu \pm 2\sigma = 20 \pm (2 * 2) = 20 \pm 4 = 16 \text{ تا } 24$$

یعنی بیش از ۹۶٪ داده ها بین ۱۶ تا ۲۴ میباشند

$$P(x \leq 17) = P\left(\frac{x - \mu}{\sigma} \leq \frac{17 - 20}{2}\right) = P(z \leq -1.5) = 0.07$$

$$P(x > 21) = 1 - P(x \leq 21) = 1 - P\left(\frac{x - \mu}{\sigma} \leq \frac{21 - 20}{2}\right) = 1 - P(z \leq +0.5) = 1 - 0.7 = 0.3$$

$$P(x > 19) = 1 - P(x \leq 19) = 1 - P\left(\frac{x - \mu}{\sigma} \leq \frac{19 - 20}{2}\right) = 1 - P(z \leq -0.5) = 1 - 0.3 = 0.7$$

سوال اضافه : در همین تمرین - احتمال اینکه هفته آینده دقیقاً ۲۱ آزمایش صحیح باشد چقدر است

$$P(x = 21) = P(x \leq 21) - P(x \leq 20) = P\left(\frac{x - \mu}{\sigma} \leq \frac{21 - 20}{2}\right) - P\left(\frac{x - \mu}{\sigma} \leq \frac{20 - 20}{2}\right) = P(z \leq 0.5) - P(z \leq 0) = 0.7 - 0.5 = 0.2$$

۹) سوال پایان ترم قبل، 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

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

							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		

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

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

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
Rx	1	2	3	4	5	6
Rx	1	2.5	2.5	5	5	5

داده سری اول

رتبه سری اول

رتبه سری اول نرم شده

Y	1	2	2	2	3	3
rY	1	2	3	4	5	6
Ry	1	3	3	3	5.5	5.5

داده سری دوم

رتبه سری دوم

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