

(۱) سوالات توضیحی ( دو سوال دو نمره)

A. نام دو نرم افزار در آمار که در این ترم تحصیلی آموزش دیدید را ذکر کنید

SPSS (Statistical package for social science)

Excel

B. جمعیت چیست نمونه چیست

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

C. آمار توصیفی چیست آمار استنباطی چیست

در آمار توصیفی از ویژگی های نمونه آماری تحقیق به همراه شاخص های گرایش به مرکز و یا شاخص های گرایش به پراکندگی می باشد. در حالی که در آمار استنباطی و یا تحلیلی می توان نتایج و یافته های به دست آمده از نمونه آماری را به کل جامعه آماری تحقیق تعمیم داد. به عبارتی، مفهوم کانونی آمار استنباطی، تعمیم پذیری است.

D. انواع روشهای مقیاس گذاری (measure) را نام ببرید هر کدام نیم خط توضیح دهید

مقیاس اسمی (NOMINAL) - کد گذاری عددی - اعداد نسبت داده شده ارجحیت و برتری ندارند (مثل جنسیت - زن ۲ - مرد ۱)  
مقیاس رتبه ای (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 فاصله ای - مثلا به یک بیمار بگوییم اگر حداکثر درد مثل یک خط کش عدد ۱۰۰ باشد درد شما چقدر است و بیمار بگوید ۶۵

E. چهار مورد از انواع نمودارها نام ببرید و هر یک نیم خط

توضیح دهید و شکل نمودار را ترسیم کنید

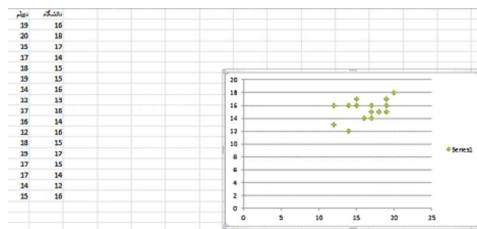
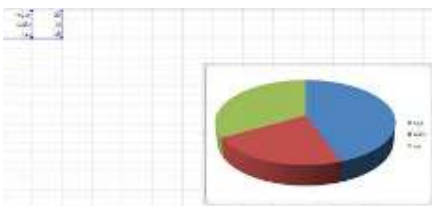
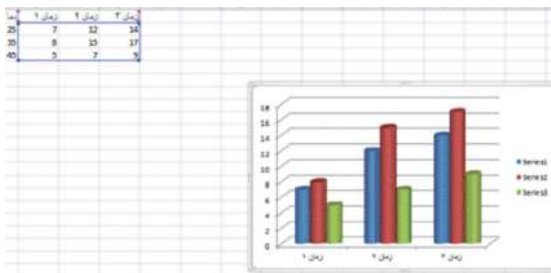
نمودار میله ای Bar Chart: برای مشاهده فراوانی داده ها و مقایسه داده ها نسبت به هم

نمودار دایره ای Pie 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



F. بدون حل کردن اگر هنگام محاسبه واریانس طبق فرمول مربوطه، توان دوم را فراموش کنیم نتیجه کسر حاصل چه خواهد شد. جواب: صفر

G. شاخصهای مرکزی را نام ببرید؟ میانگین میانه مود

H. شاخصی پراکندگی را نام ببرید؟ واریانس، انحراف معیار، ضریب تغییرات

I. از رگرسیون خطی چه استفاده ای میشود؟ پیش بینی آینده با استفاده از داده های گذشته

۲) جدول سوال زیر قبلا حل شده است و میانگین 13.3 و واریانس 4.01 میباشد. مجدد حل نکنید

$X_i$	10	11	12	15	16
$f_i$	1	1	3	4	1

الف) انحراف معیار را حساب کنید ج) 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. د) جدول زیر شباهت زیادی به جدول بالا در همین سوال دارد با استفاده از جواب قسمت بالا میانگین و واریانس و انحراف معیار داده جدید را بدست آورید (بدون استفاده از فرمول اصلی)، و ضریب تغییرات این داده جدید را بدست آورید (۴ نمره)

$X_i$	110	111	112	115	116
$f_i$	1	1	3	4	1

$$\bar{x} = 13.3 \quad \sigma_x = \sqrt{4.01} = 2.0025$$

$$\mu \pm 2\sigma \quad \mu \pm \sigma \quad \text{تحلیل}$$

$$\mu \pm \sigma = 13.3 \pm 2.0025 = 11.2975 \quad \text{تا} \quad 15.3025$$

یعنی بیش از ۶۸٪ داده ها بین 11.297 تا 15.302 میباشند

$$\mu \pm 2\sigma = 13.3 \pm (2 * 2.0025) = 13.3 \pm 4.005 = 9.295 \quad \text{تا} \quad 17.305$$

یعنی بیش از ۹۶٪ داده ها بین 9.295 تا 17.305 میباشند

طبق جدول فوق در خصوص  $y$  ها میتوان نوشت که

$$y_i = x_i + 100$$

و طبق قانون میتوان نوشت

$$\bar{y} = \bar{x} + 100 = 13.3 + 100 \rightarrow \bar{y} = 113.3$$

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

$$\sigma_y = \sigma_x = \sqrt{4.01} = 2.0025 \rightarrow \sigma_y = 2.0025$$

چون داده های جدید همگی 100 واحد اضافه شده اند پس میانگین داده های جدید 100 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

$$\rho = \frac{\sigma}{\mu} = \frac{2.0025}{113.3} = 0.0177 = 1.7\%$$

الآن  $\rho = 0.0177$  عدد خیلی کوچکی است یعنی پراکندگی داده ها کم است

همیشه  $0 \leq \rho \leq 1$  است. اگر  $\rho$  صفر شود یعنی داده ها اصلا پراکنده نیستند داده ها بسیار بهم نزدیک هستند

اگر  $\rho$  یک شود یعنی داده ها خیلی پراکنده هستند

۳) نمرات دانشجویان بشرح زیر است. الف) مد(نما) داده ها چه مقدار است (چرا) ب). میانه داده ها چه مقدار است؟ ج) صدک ۲۵ را محاسبه کنید. د) اعداد زیر یک ستون از داده ها را توضیح دهید ۵) میانگین و واریانس و انحراف معیار نمرات دانشجویان را بدست آورید (۴ نمره)

12,17,10,15,14,14,15,14,12,17

حل : ابتداء داده ها مرتب میکنیم

x = نمره	10	12	14	15	17
f = تعداد	1	2	3	2	2
F (فروانی تجمعی) (انباشته)	1	3	6	8	10
f <sub>i</sub> فروانی نسبی	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{2}{10}$
R <sub>i</sub> فراوانی نسبی تجمعی	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{10}{10}$

مد داده ای بمقدار 14 میباشد زیرا بیشترین تعداد را دارد

$$M=14$$

میانه داده ها

$$Q \sum f_i = \frac{1}{2}(1 + 2 + 3 + 2 + 2) = \frac{10}{2} = 5 \rightarrow \text{در سطر } F_i \rightarrow 5+ \rightarrow F_i=6 \rightarrow m=14$$

یعنی نصف داده ها 14 یا کمتر از 14 هستند

صدک ۲۵ داده ها

$$Q \sum f_i = \frac{25}{100}(1 + 2 + 3 + 2 + 2) = \frac{25}{100} = 2.5 \rightarrow \text{در سطر } F_i \rightarrow 2.5+ \rightarrow F_i=3 \rightarrow x=12$$

یعنی 25% داده ها 12 یا کمتر از 12 هستند

در یک ستون مثلا اعداد زیر ۱۲ توضیح داده میشود

۲ تا از داده ها ۱۲ هستند

۳ تا از داده ها ۱۲ یا کمتر از ۱۲ هستند

$\frac{2}{10}$  تا از داده ها ۱۲ هستند

$\frac{3}{10}$  تا از داده ها ۱۲ یا کمتر از ۱۲ هستند

$$\bar{x} = \mu = \frac{\sum x_i * f_i}{\sum f_i} = \frac{(10 * 1) + (12 * 2) + (14 * 3) + (15 * 2) + (17 * 2)}{1 + 2 + 3 + 2 + 2} = \frac{140}{10} = 14$$

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

$$\sigma^2 = \frac{(10 - 14)^2 * 1 + (12 - 14)^2 * 2 + (14 - 14)^2 * 3 + (15 - 14)^2 * 2 + (17 - 14)^2 * 2}{1 + 2 + 3 + 2 + 2} = 4.4 \text{ واریانس}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{4.4} = 2.1 \text{ انحراف معیار}$$

۴) یک دانشجو ۱۱۰ واحد درسی را با توزیع نرمال و میانگین نمرات ۱۷ و انحراف معیار ۲ گذرانده است، در یک درس در ترم بعدی. الف) احتمال اینکه نمره این دانشجو حداکثر ۱۵ شود چقدر است؟ ب) احتمال اینکه نمره این دانشجو بیشتر از ۱۹ شود چقدر است ج) احتمال اینکه نمره دقیقا ۱۷ شود؟ (۴نمره) (جدول توزیع نرمال استاندارد بشرح زیر در نظر

بگیرید)

$$p(z \leq 2) = 0.98, \quad p(z \leq 1.5) = 0.93, \quad p(z \leq 1) = 0.85, \quad p(z \leq 0.5) = 0.7, \quad p(z \leq 0) = 0.5$$

$$p(z \leq -2) = 0.02, \quad p(z \leq -1.5) = 0.07, \quad p(z \leq -1) = 0.15, \quad p(z \leq -0.5) = 0.3$$

حل

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

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

$$P(x = 17) = P(x \leq 17) - P(x \leq 16) = P\left(\frac{x - \mu}{\sigma} \leq \frac{17 - 17}{2}\right) - P\left(\frac{x - \mu}{\sigma} \leq \frac{16 - 17}{2}\right)$$

$$= P(z \leq 0) - P(z \leq -0.5) = 0.5 - 0.3 = 0.2$$

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

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

همه اعدادی که به نام X داریم ۱۳۹۹ تا کم میکنیم

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

$$y = a + bx$$

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

$$= \frac{(-2 * 2) + (-1 * 4) + (0 * 4) + (1 * 5) + (2 * 5) - \frac{(-2 - 1 + 0 + 1 + 2)(2 + 4 + 4 + 5 + 5)}{5}}{((-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2) - \frac{(-2 - 1 + 0 + 1 + 2)^2}{5}} = \frac{7}{10}$$

$$= 0.7$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{-2 - 1 + 0 + 1 + 2}{5} = 0 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{2 + 4 + 4 + 5 + 5}{5} = \frac{20}{5} = 4$$

$$\bar{y} = a + b\bar{x} \quad 4 = a + (0.7 * 0) \quad a = 4$$

$$y = 4 + 0.7x$$

$$X_m = 1399 \rightarrow y = 1402 - 1399 = 3 \rightarrow y = 4 + (0.7 * 3) = 6.1$$

بعد از امتحان میان ترم

۶) 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 اسپیرمن (رتبه‌ای) بین دو سری عدد X و Y را بدست آورید و در خصوص علامت این ضریب و مقدار آن توضیح بنویسید (۴نمره)

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) <sup>2</sup>	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 +0.857 شد یعنی دو سری داده وابستگی زیاد در جهت مستقیم دارند

۷) رابطه بین ساعات استفاده از اینترنت با نمره کسب شده توسط دانشجویان را بدست آورید (از روش پیرسون) (۴ نمره)

X	۱۰	۱۴	۱۵	۱۵	۱۳	۱۴	۱۳	۱۱	اعتیاد به اینترنت
Y	۲۰	۱۴	۱۲	۱۶	۱۸	۱۵	۱۷	۲۰	نمره

حل: به روش پیرسون

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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{10 + 14 + 15 + 15 + 13 + 14 + 13 + 11}{8} = 13.125$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{20 + 14 + 12 + 16 + 18 + 15 + 17 + 20}{8} = 16.5$$

x*y	200	196	180	240	234	210	221	220	نمره * اعتیاد
-----	-----	-----	-----	-----	-----	-----	-----	-----	---------------

$$\bar{xy} = \frac{\sum x_i y_i}{n} = \frac{200 + 196 + 180 + 240 + 234 + 210 + 221 + 220}{8} = 212.625$$

کوواریانس

$$\text{Cov}(x, y) = \bar{xy} - \bar{x} \bar{y} = 212.625 - (13.125 * 16.5) = -3.9375$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i} \quad \sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = 2.859 \quad \sigma_x = 1.69 \quad \sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = 7 \quad \sigma_y = 2.64$$

$$\rho(x, y) = \text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} = \frac{\bar{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} = \frac{212.625 - (13.125 * 16.5)}{1.69 * 2.64} = \frac{-3.937}{1.69 * 2.64} = -0.88$$

✓ ضریب بین ۰٫۷۰ تا ۱ نشان دهنده همبستگی قوی

عدد ۰٫۸۸ بیانگر همبستگی قوی بین X و Y است عدم منفی همبستگی در جهت معکوس است یعنی هر چه اینترنت بیشتر استفاده شود نمره دانشجو کمتر میشود Pearson Correlation = -0.88