

Name _____ Course Number: _____ Section Number: _____

Directions: Answer the questions and solve the problems in the spaces provided, or attach paper. Circle the correct choice for each response set. Where allowed, use computer assistance.

Provide an appropriate response.

- 1) A set of data consists of the number of years that applicants for foreign service jobs have studied German and the grades that they received on a proficiency test. The following regression equation is obtained: $\hat{y} = 31.6 + 10.9x$, where x represents the number of years of study and y represents the grade on the test. What does the slope of the regression line represent in terms of grade on the test?

Given the linear correlation coefficient r and the sample size n , determine the critical values of r and use your finding to state whether or not the given r represents a significant linear correlation. Use a significance level of 0.05.

- 2) $r = 0.898$, $n = 9$
- A) Critical values: $r = \pm 0.666$, no significant linear correlation
 B) Critical values: $r = 0.666$, no significant linear correlation
 C) Critical values: $r = \pm 0.666$, significant linear correlation
 D) Critical values: $r = -0.666$, no significant linear correlation

Find the value of the linear correlation coefficient r .

- 3) The paired data below consist of the costs of advertising (in thousands of dollars) and the number of products sold (in thousands):

Cost	9	2	3	4	2	5	9	10
Number	85	52	55	68	67	86	83	73

- A) -0.071 B) 0.708 C) 0.246 D) 0.235

Describe the error in the stated conclusion.

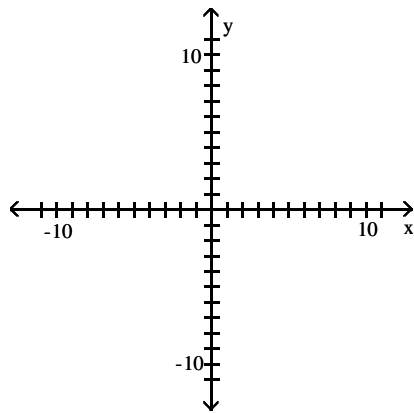
- 4) Given: Each school in a state reports the average SAT score of its students. There is a significant linear correlation between the average SAT score of a school and the average annual income in the district in which the school is located.

Conclusion: There is a significant linear correlation between individual SAT scores and family income.

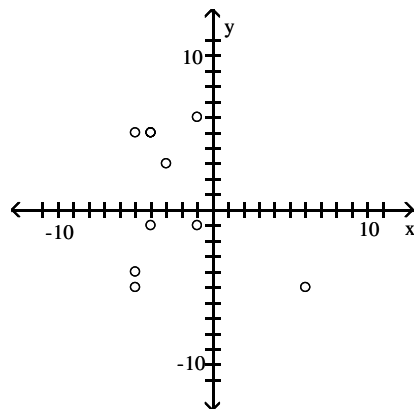
Construct a scatterplot for the given data.

5)

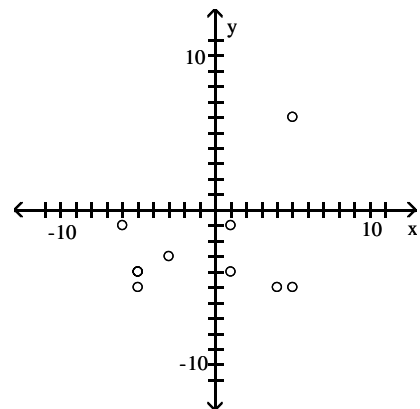
x	-3	-4	-4	-5	-5	-1	-4	6	-5	-1
y	-3	-5	-5	5	4	1	1	5	-5	-6



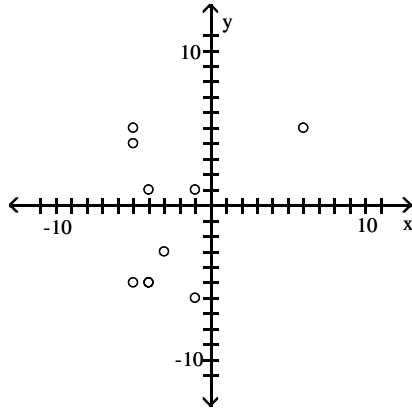
A)



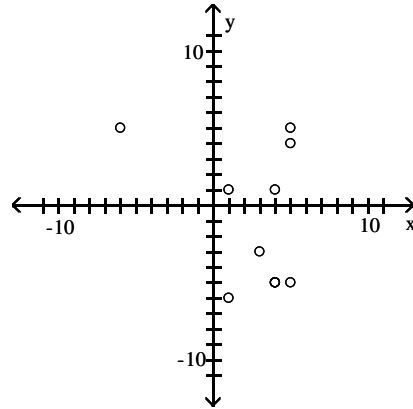
B)



C)



D)



Use the given data to find the best predicted value of the response variable.

6) Six pairs of data yield $r = 0.444$ and the regression equation $\hat{y} = 5x + 2$. Also, $\bar{y} = 18.3$. What is the best predicted value of y for $x = 5$?

- A) 27 B) 93.5 C) 18.3 D) 4.22

Use the given data to find the equation of the regression line. Round the final values to three significant digits, if necessary.

7)

x	6	8	20	28	36
y	2	4	13	20	30

- A) $\hat{y} = -3.79 + 0.897x$ B) $\hat{y} = -3.79 + 0.801x$
 C) $\hat{y} = -2.79 + 0.897x$ D) $\hat{y} = -2.79 + 0.950x$

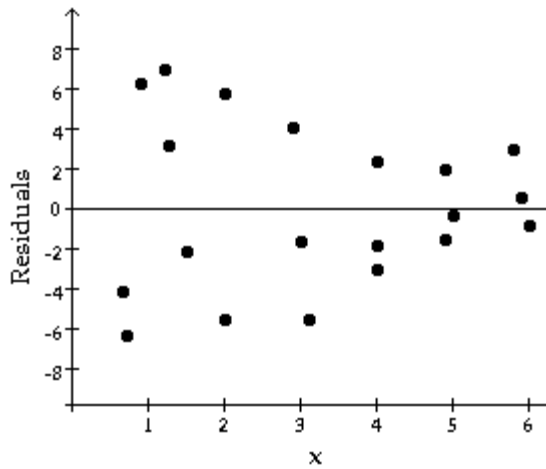
Is the data point, P, an outlier, an influential point, both, or neither?

8) The regression equation for a set of paired data is $\hat{y} = 38.4 + -0.9x$. The values of x run from 100 to 400. A new data point, P(375, -299.1), is added to the set.

- A) Neither B) Influential point
 C) Outlier D) Both

Provide an appropriate response.

- 9) The following residual plot is obtained after a regression equation is determined for a set of data. Does the residual plot suggest that the regression equation is a bad model? Why or why not?



Use the given information to find the coefficient of determination.

- 10) A regression equation is obtained for a collection of paired data. It is found that the total variation is 128.6, the explained variation is 84, and the unexplained variation is 44.6. Find the coefficient of determination.

A) 0.653 B) 0.531 C) 1.531 D) 0.347

Find the explained variation for the paired data.

- 11) The paired data below consists of test scores and hours of preparation for 5 randomly selected students. The equation of the regression line is $\hat{y} = 44.8447 + 3.52427x$. Find the explained variation.

x Hours of preparation	5	2	9	6	10
y Test of score	64	48	72	73	80

A) 599.2 B) 511.724 C) 87.4757 D) 498.103

Use the computer display to answer the question.

- 12) A collection of paired data consists of the number of years that students have studied Spanish and their scores on a Spanish language proficiency test. A computer program was used to obtain the least squares linear regression line and the computer output is shown below. Along with the paired sample data, the program was also given an x value of 2 (years of study) to be used for predicting test score.

The regression equation is

$$\text{Score} = 31.55 + 10.90 \text{ Years.}$$

Predictor	Coef	StDev	T	P
Constant	31.55	6.360	4.96	0.000
Years	10.90	1.744	6.25	0.000

$$S = 5.651 \quad R\text{-Sq} = 83.0\% \quad R\text{-Sq (Adj)} = 82.7\%$$

Predicted values

Fit	StDev Fit	95.0% CI	95.0% PI
53.35	3.168	(42.72, 63.98)	(31.61, 75.09)

If a person studies 4.5 years, what is the single value that is the best predicted test score? Assume that there is a significant linear correlation between years of study and test score.

- A) 49.1 B) 80.6 C) 83.0 D) 53.35

Construct the indicated prediction interval for an individual y .

- 13) The regression equation for the given paired data is $\hat{y} = 73.012 + 0.8859x$ and the standard error of estimate is $s_e = 2.78807$. Find the 95% prediction interval of \hat{y} for $x = 37$.

x	25	26	36	36	40	48
y	95	95	102	109	110	114

- A) $73.0 < y < 105.8$ B) $97.4 < y < 10.8$
 C) $73.0 < y < 114.2$ D) $97.4 < y < 114.2$

Find the standard error of estimate for the given paired data.

- 14) The data set given below has $\hat{y} = 3x - 5$ for its regression equation. Find the standard error of estimate.

x	3	2	5	8
y	4	1	10	19

- A) 2.06 B) 1.29 C) 0 D) 0.17

Use computer software to find the multiple regression equation. Can the equation be used for prediction?

- 15) An anti-smoking group used data in the table to relate the carbon monoxide of various brands of cigarettes to their tar and nicotine content.

CO	TAR	NIC
15	1.2	16
15	1.2	16
17	1.0	16
6	0.8	9
1	0.1	1
8	0.8	8
10	0.8	10
17	1.0	16
15	1.2	15
11	0.7	9
18	1.4	18
16	1.0	15
10	0.8	9
7	0.5	5
18	1.1	16

- A) $CO = 1.3 + 5.5TAR - 1.3NIC$; Yes, because the adjusted R^2 is high
 B) $CO = 1.37 - 5.53TAR + 1.33NIC$; Yes, because the R^2 is high
 C) $CO = 1.37 + 5.50TAR - 1.38NIC$; Yes, because the P-value is high
 D) $CO = 1.25 + 1.55TAR - 5.79NIC$; Yes, because the P-value is too low

Use computer software to obtain the multiple regression equation and identify R^2 , adjusted R^2 , and the P-value.

- 16) A study of food consumption in the country related the level of food consumed to an index of food prices and an index of personal disposable income.

FOODCONS	INCOME	PRICE
98.6	87.4	108.5
101.2	97.6	110.1
102.4	96.7	110.4
100.9	98.2	104.3
102.3	99.8	107.2
101.5	100.5	105.8
101.6	103.2	107.8
101.6	107.8	103.4
99.8	96.6	102.7
100.3	88.9	104.1
97.6	75.1	99.2
97.2	76.9	99.7
97.3	84.6	102.0
96.0	90.6	94.3
99.2	103.1	97.7
100.3	105.1	101.1
100.3	96.4	102.3
104.1	104.4	104.4
105.3	110.7	108.5
107.6	127.1	111.3

A) 0.912, 0.901, 0.010

B) 0.843, 0.799, 0.005

C) 0.855, 0.844, 0.002

D) 0.867, 0.852, 0.000

Use computer software to obtain the multiple regression equation. Use the estimated equation to find the predicted value.

- 17) A visitor to Yellowstone National Park sat down one day and observed Old Faithful, which faithfully spurts throughout the day, day in and day out. He surmised that the height, in feet, of a given spurt was caused by the pressure build-up during the interval between spurts and by the momentum build-up during the duration of the spurt. He wrote down the data to test his hypothesis. He then used the regression equation he obtained to predict the height of the geyser if the interval is 95 seconds and the duration is 218 seconds.

HEIGHT	INTERVAL	DURATION
150	86	240
154	86	237
140	62	122
140	104	267
160	62	113
140	95	258
150	79	232
150	62	105
160	94	276
155	79	248
125	86	243
136	85	241
140	86	214
155	58	114
130	89	272
125	79	227
125	83	237
139	82	238
125	84	203
140	82	270
140	82	270
140	78	218
135	87	270
140	70	241
100	56	102
105	81	271

A) 144 ft

B) 132 ft

C) 156 ft

D) 151 ft

Find the indicated multiple regression equation.

18) Below are performance and attitude ratings of employees.

Performance	59	63	65	69	58	77	76	69	70	64
Attitude	72	67	78	82	75	87	92	83	87	78

Managers also rate the same employees according to adaptability, and below are the results that correspond to those given above.

Adaptability : 50 52 54 60 46 67 66 59 62 55

Find the multiple regression equation that expresses performance in terms of attitude and adaptability.

- A) $\hat{P} = 14.09 + 0.014(\text{Att.}) + 0.907(\text{Adapt.})$
- B) $\hat{P} = 14.09 + 0.213(\text{Att.}) + 0.895(\text{Adapt.})$
- C) $\hat{P} = 14.09 + 0.907(\text{Att.}) + 0.014(\text{Adapt.})$
- D) $\hat{P} = 14.09 + 0.895(\text{Att.}) + 0.213(\text{Adapt.})$

Use computer software to find the best multiple regression equation to explain the variation in the dependent variable, Y, in terms of the independent variables, X₁, X₂, X₃.

19)

Y	X ₁	X ₂
98.6	87.4	108.5
101.2	97.6	110.1
102.4	96.7	110.4
100.9	98.2	104.3
102.3	99.8	107.2
101.5	100.5	105.8
101.6	103.2	107.8
101.6	107.8	103.4
99.8	96.6	102.7
100.3	88.9	104.1
97.6	75.1	99.2
97.2	76.9	99.7
97.3	84.6	102.0
96.0	90.6	94.3
99.2	103.1	97.7
100.3	105.1	101.1
100.3	96.4	102.3
104.1	104.4	104.4
105.3	110.7	108.5
107.6	127.1	111.3

CORRELATION COEFFICIENTS

Y/ X₁ = 0.850
 Y/ X₂ = 0.742

COEFFICIENT OF DETERMINATION

Y/ X₁ = 0.723
 Y/ X₂ = 0.550
 Y/ X₁, X₂ = 0.867

- A) $\hat{Y} = 57.7 + 0.153 X_1 + 0.270 X_2$
- B) $\hat{Y} = 48.0 + 0.398 X_1 + 0.228 X_2$
- C) $\hat{Y} = 58.9 + 0.612 X_1$
- D) $\hat{Y} = 52.6 + 0.462 X_2$

Construct a scatterplot and identify the mathematical model that best fits the data. Assume that the model is to be used only for the scope of the given data and consider only linear, quadratic, logarithmic, exponential, and power models. Use a calculator or computer to obtain the regression equation of the model that best fits the data. You may need to fit several models and compare the values of R^2 .

- 20) The table below shows the population of a city (in millions) in each year during the period 1990 – 1995. Using the number of years since 1990 as the independent variable, find the regression equation of the best model.

x	1990	1991	1992	1993	1994	1995
y	1.08	1.37	1.68	2.19	2.73	3.34

A) $y = 0.930 + 0.454x$

B) $y = 1.27x^{0.550}$

C) $y = 1.08e^{0.228x}$

D) $y = 0.05x^2 + 0.27x + 1.06$

Answer Key

Testname: CHAPTER 10 FORM B

- 1) The grade on the test increases by an estimated 10.9 points for each additional year of study.
- 2) C
- 3) B
- 4) Averages suppress individual variation and tend to inflate the correlation coefficient. The fact that there is significant linear correlation between average SAT scores and average incomes in the district does not necessarily imply that there is significant linear correlation between individual SAT scores and family incomes.
- 5) C
- 6) C
- 7) A
- 8) A
- 9) Yes, the residual plot suggests that the regression equation is a bad model. The residual plot becomes thinner when viewed from left to right. This suggests that for different fixed values of x , the distributions of the corresponding y -values have different standard deviations.
- 10) A
- 11) B
- 12) B
- 13) D
- 14) C
- 15) B
- 16) D
- 17) A
- 18) A
- 19) A
- 20) C