

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 regression equation is obtained for a set of data. After examining a scatter diagram, the researcher notices a data point that is potentially an influential point. How could she confirm that this data point is indeed an influential point?

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) $r = 0.434$, $n = 15$

- A) Critical values: $r = \pm 0.514$, significant linear correlation
 B) Critical values: $r = \pm 0.532$, no significant linear correlation
 C) Critical values: $r = 0.514$, no significant linear correlation
 D) Critical values: $r = \pm 0.514$, no significant linear correlation

Find the value of the linear correlation coefficient r .

- 3) The paired data below consist of the temperatures on randomly chosen days and the amount a certain kind of plant grew (in millimeters):

Temp	62	76	50	51	71	46	51	44	79
Growth	36	39	50	13	33	33	17	6	16

- A) -0.210 B) 0 C) 0.256 D) 0.196

Describe the error in the stated conclusion.

- 4) Given: There is a significant linear correlation between the number of homicides in a town and the number of movie theaters in a town.

Conclusion: Building more movie theaters will cause the homicide rate to rise.

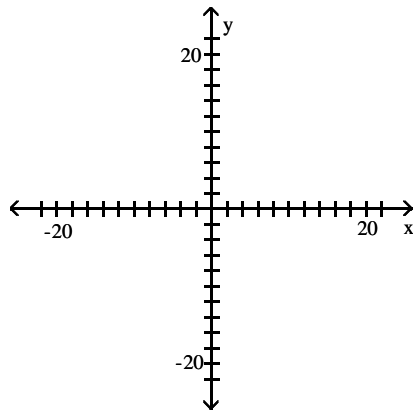
Use the given data to find the best predicted value of the response variable.

- 5) Four pairs of data yield $r = 0.942$ and the regression equation $\hat{y} = 3x$. Also, $\bar{y} = 12.75$.
 What is the best predicted value of y for $x = 4.8$?
- A) 2.826 B) 12.75 C) 0.942 D) 14.4

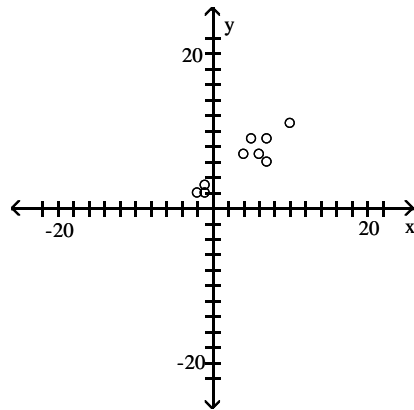
Construct a scatterplot for the given data.

6)

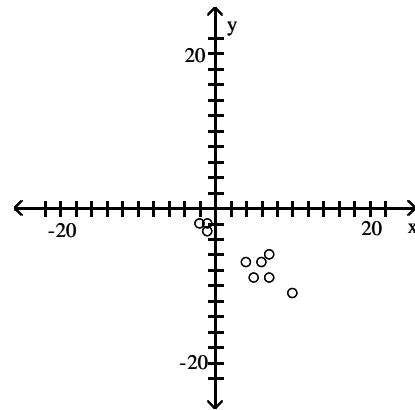
x	-2	5	7	7	10	6	4	-1	-1
y	2	9	6	9	11	7	7	3	2



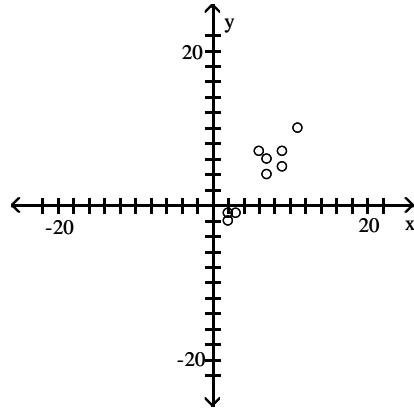
A)



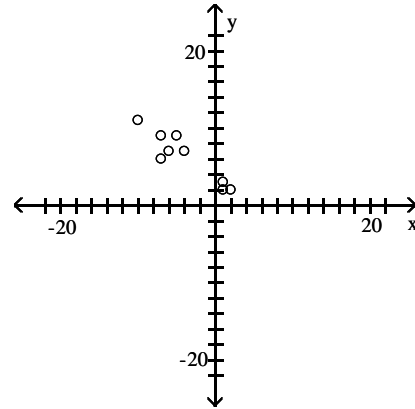
B)



C)



D)



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

7)

x	0	3	4	5	12
y	8	2	6	9	12

A) $\hat{y} = 4.98 + 0.425x$

C) $\hat{y} = 4.88 + 0.625x$

B) $\hat{y} = 4.88 + 0.525x$

D) $\hat{y} = 4.98 + 0.725x$

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

8) The regression equation for a set of paired data is $\hat{y} = 1.18 + 1.90x$. The values of x run from 3 to 20. A new data point, P(30, 60), is added to the set.

A) Influential point

B) Outlier

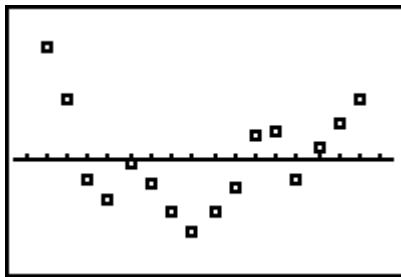
C) Both

D) Neither

Provide an appropriate response.

- 9) The following table gives the US domestic oil production rates (excluding Alaska) over the past few years. A regression equation was fit to the data and the residual plot is shown below.

Year	Millions of barrels per day	Year	Millions of barrels per day
1987	6.39	1995	5.08
1988	6.12	1996	5.07
1989	5.74	1997	5.16
1990	5.58	1998	5.08
1991	5.62	1999	4.83
1992	5.46	2000	4.85
1993	5.26	2001	4.84
1994	5.10	2002	4.83



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) Find the coefficient of determination, given that the value of the linear correlation coefficient, r , is -0.346 .
- A) 0.120 B) 0.346 C) 0.654 D) 0.880

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

- 11) The equation of the regression line for the paired data below is $\hat{y} = 6.18286 + 4.33937x$. Find the standard error of estimate.

x	9	7	2	3	4	22	17
y	43	35	16	21	23	102	81

- A) 3.1270 B) 2.6959 C) 0.8275 D) 1.6419

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)

What percentage of the total variation in test scores is unexplained by the linear relationship between years of study and test scores?

- A) 82.7% B) 8.9% C) 17.0% D) 83.0%

Find the unexplained variation for the paired data.

- 13) 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 unexplained variation.

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

- A) 87.4757 B) 96.103 C) 511.724 D) 599.2

Construct the indicated prediction interval for an individual y .

- 14) The equation of the regression line for the paired data below is $\hat{y} = 3x$ and the standard error of estimate is $s_e = 2.2361$. Find the 90% prediction interval of y for $x = 3$.

x	2	4	5	6
y	7	11	13	20

- A) $7.1 < y < 10.9$ B) $4.5 < y < 13.5$ C) $1.2 < y < 16.8$ D) $6.8 < y < 11.2$

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

- 15) A wildlife analyst gathered the data in the table to develop an equation to predict the weights of bears. He used WEIGHT as the dependent variable and CHEST, LENGTH, and SEX as the independent variables. For SEX, he used male=1 and female=2.

WEIGHT	CHEST	LENGTH	SEX
344	45.0	67.5	1
416	54.0	72.0	1
220	41.0	70.0	2
360	49.0	68.5	1
332	44.0	73.0	1
140	32.0	63.0	2
436	48.0	72.0	1
132	33.0	61.0	2
356	48.0	64.0	2
150	35.0	59.0	1
202	40.0	63.0	2
365	50.0	70.5	1

- A) $\text{WEIGHT} = -320 + 10.6\text{CHEST} + 7.3\text{LENGTH} - 10.7\text{SEX}$; Yes, because the P-value is high
- B) $\text{WEIGHT} = -442.6 + 12.1\text{CHEST} + 3.6\text{LENGTH} - 23.8\text{SEX}$; Yes, because the adjusted R^2 is high
- C) $\text{WEIGHT} = 196 + 2.35\text{CHEST} + 3.40\text{LENGTH} + 25\text{SEX}$; Yes, because the R^2 is high
- D) $\text{WEIGHT} = 442.6 + 12.1\text{CHEST} + 4.2\text{LENGTH} - 21\text{SEX}$; Yes, because the P-value is low

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

- 16) 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) 0.931, 0.902, 0.000

B) 0.943, 0.934, 0.000

C) 0.861, 0.900, 0.015

D) 0.976, 0.921, 0.002

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

- 17) 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. Next year, the income index number is expected to be 104.4, and the price index is expected to be 105.8. These numbers would indicate a predicted value for food consumption.

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) 94.43 B) 105.83 C) 99.33 D) 102.23

Find the indicated multiple regression equation.

- 18) Below are the productivity, dexterity, and job satisfaction ratings of ten randomly selected employees.

Productivity	23	25	28	21	21	25	26	30	34	36
Dexterity	49	53	59	42	47	53	55	63	67	75
Job satisfaction	56	58	60	50	54	61	59	63	67	69

Find the multiple regression equation that expresses the job satisfaction scores in terms of the productivity and dexterity scores.

- A) $\hat{S} = 28.28 + 0.237P + 0.728D$ B) $\hat{S} = 28.28 + 0.517P + 0.0860D$
 C) $\hat{S} = 28.28 + 0.0860P + 0.517D$ D) $\hat{S} = 28.28 + 0.011P + 0.437D$

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	X ₃		
	344	45.0	67.5	1	CORRELATION COEFFICIENTS	
	416	54.0	72.0	1		
	220	41.0	70.0	2		Y/ X ₁ = 0.951
	360	49.0	68.5	1		Y/ X ₂ = 0.789
	332	44.0	73.0	1		Y/ X ₃ = -0.616
	140	32.0	63.0	2	COEFFICIENTS OF DETERMINATION	
	436	48.0	72.0	1		
	132	33.0	61.0	2		
	356	48.0	64.0	2		Y/ X ₁ = 0.905
	150	35.0	59.0	1		Y/ X ₁ , X ₂ = 0.919
	202	40.0	63.0	2		Y/ X ₁ , X ₂ , X ₃ = 0.927
	365	50.0	70.5	1		
	A) $\hat{Y} = -442 + 12.1 X_1 + 3.58 X_2 - 23.8 X_3$					
	B) $\hat{Y} = -543 + 12.8 X_1 + 4.15 X_2$					
	C) $\hat{Y} = -412 + 13.6 X_1 + 3.15 X_2$					
	D) $\hat{Y} = -355 + 14.9 X_1$					

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 R².

20)		
	$\begin{array}{c ccccc} x & 1 & 2 & 3 & 4 & 5 \\ \hline y & 3 & 9 & 17 & 30 & 40 \end{array}$	
	A) $y = -8.70 + 9.50 x$	B) $y = 2.96 x^{1.628}$
	C) $y = 2.07 e^{0.638x}$	D) $y = -1.59 + 22.3 \ln x$

Answer Key

Testname: CHAPTER 10 FORM C

- 1) The researcher should first graph the regression line resulting from the data with the point included. She should then remove that data point, recalculate the regression equation, and graph the resulting regression line. If the graph changes considerably, the researcher can conclude that the point is an influential observation. Influential points are outliers that are horizontally far away from other data points.
- 2) D
- 3) D
- 4) Significant correlation does not imply causality. Both variables are affected by a third variable (a lurking variable), namely the population of the town.
- 5) B
- 6) A
- 7) B
- 8) B
- 9) Yes, the residual plot suggests that the regression equation is a bad model. The residual plot has an obvious pattern that is not a straight line. This suggests that a scatterplot of the sample data is not a straight line.
- 10) A
- 11) D
- 12) C
- 13) A
- 14) C
- 15) B
- 16) B
- 17) D
- 18) C
- 19) D
- 20) B