

Homework #3

Name _____

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

- 1) A _____ association between two quantitative variables x and y is present if y tends to go down as x goes up. 1) _____
- 2) The _____ is a summary measure that describes the strength of the linear association between two quantitative variables. 2) _____

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

- 3) Suppose you were to collect data for the pair of given variables in order to make a scatterplot. For the variables cloudy days and rainy days, which is more naturally the response variable and which is the explanatory variable? 3) _____
- A) Cloudy days: explanatory variable
Rainy days: response variable
- B) Cloudy days: response variable
Rainy days: explanatory variable

In 2006, the General Social Survey asked "Do you strongly agree, agree, disagree or strongly disagree that it is sometimes necessary to discipline a child with a good, hard spanking?" The responses are cross-tabulated with the respondent's gender in the table below.

Gender	Favor Spanking?				Total
	Strongly Agree	Agree	Disagree	Strongly Disagree	
Male	238	419	165	41	863
Female	294	474	237	102	1107
Total	532	893	402	143	1970

- 4) Which variable is the explanatory variable? 4) _____
- A) Gender
B) Response to whether the respondent favors spanking
- 5) What proportion of males responded that they agree with spanking as a means of disciplining a child? 5) _____
- A) 0.28 B) 0.21 C) 0.49 D) 0.12 E) 0.76

Select the most appropriate answer.

- 6) The _____ is the outcome variable on which comparisons are made. 6) _____
- A) predictor variable
B) Both B and C
C) response variable
D) explanatory variable
E) lurking variable

In 2006, the General Social Survey asked "Do you strongly agree, agree, disagree or strongly disagree that it is sometimes necessary to discipline a child with a good, hard spanking?" The responses are cross-tabulated with the respondent's gender in the table below.

Gender	Favor Spanking?				Total
	Strongly Agree	Agree	Disagree	Strongly Disagree	
Male	238	419	165	41	863
Female	294	474	237	102	1107
Total	532	893	402	143	1970

7) Which gender is more likely to respond that they agree that it is sometimes necessary to discipline a child with spanking? 7) _____

- A) Neither, they are both equally likely to offer this response
- B) Men
- C) Women

8) Overall, what proportion of people responded that they strongly disagree with spanking as a form of discipline? 8) _____

- A) 0.05
- B) 0.07
- C) 0.09
- D) 0.02

Answer true or false.

9) The advantage of a side-by-side bar graph is that it allows for easy comparison of the explanatory variable groups with respect to values on the response variable. 9) _____

- A) False
- B) True

The following table summarizes the responses of 1255 adults when asked by the 2006 General Social Survey whether they had ever taken the drug Prozac.

	Male	Female	Total
Yes	36	96	132
No	495	628	1123
Total	531	724	1255

10) Given that the respondent is male, what is the probability that he responded yes? 10) _____

- A) 0.11
- B) 0.27
- C) 0.03
- D) 0.07

11) Given that the respondent answered "Yes", what is the probability that the respondent was female? 11) _____

- A) 0.08
- B) 0.58
- C) 0.13
- D) 0.73

Provide an appropriate response.

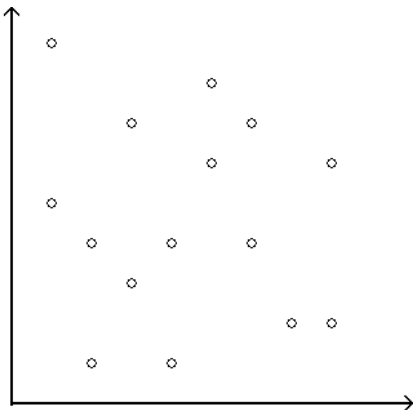
- 12) The relationship between the number of games won by a minor league baseball team and the average attendance at their home games is analyzed. A regression to predict the average attendance from the number of games won has an $r = 0.73$. Interpret this statistic. 12) _____
- A) Positive, fairly strong linear relationship. 53.29% of the variation in average attendance is explained by the number of games won.
 - B) No association
 - C) Negative, fairly strong linear relationship. 53.29% of the variation in average attendance is explained by the number of games won.
 - D) Positive, weak linear relationship. 7.29% of the variation in average attendance is explained by the number of games won.
 - E) Positive, fairly strong linear relationship. 73% of the variation in average attendance is explained by the number of games won.

- 13) A random sample of records of electricity usage of homes gives the amount of electricity used in July and size (in square feet) of 135 homes. A regression was performed to predict the amount of electricity used (in kilowatt-hours) based on size. The residuals plot indicated that a linear model is appropriate. Do you think the slope is positive or negative? Why? 13) _____
- A) Positive. Larger homes should use more electricity.
 - B) Positive. More square feet indicates more houses.
 - C) Negative. Smaller homes should use less electricity.
 - D) Positive. The larger the number of houses the more electricity used.
 - E) Negative. Larger homes should use less electricity.

- 14) Would you expect the following pair of variables measured for 200 individuals aged 18-32 to have a positive association, negative association, or no association: amount of time spent exercising per week; height? 14) _____
- A) no association B) negative association C) positive association

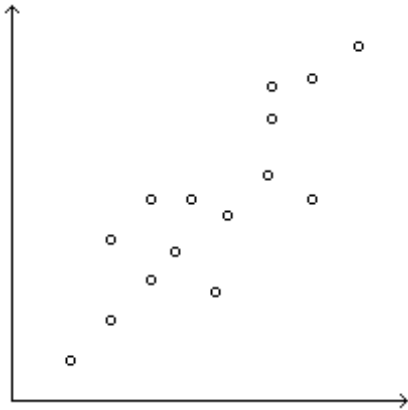
Determine the type of association apparent in the following scatterplot.

- 15) 15) _____



- A) Negative association, moderately strong association
- B) Little or no association
- C) Positive association, moderately strong association
- D) Positive association, linear association
- E) Negative association, linear association

16)



- A) Linear association, moderately strong association
 B) Positive association, moderately strong association
 C) Positive association, moderately strong association, linear association
 D) Positive association
 E) Linear association

16) _____

Answer true or false.

- 17) A scatterplot is a graphical display for two quantitative variables.

- A) True
 B) False

17) _____

Provide an appropriate response.

- 18) For the 14 teams in baseball's American league, the correlation with number of wins in the 2007 regular season is 0.51 for shutouts, 0.61 for hits made, -.70 for runs allowed and -0.56 for homeruns allowed. Which variable has the weakest linear association with number of wins?

- A) homeruns allowed
 B) runs allowed
 C) shutouts
 D) hits made

18) _____

Answer true or false.

- 19) The closer
- r
- is to 0, the weaker is the linear association between the variables.

- A) True
 B) False

19) _____

- 20) The value of the correlation is always between 0 and 1.

- A) True
 B) False

20) _____

Provide an appropriate response.

- 21) Based on findings from the Health and Nutrition Examination Survey conducted by the National Center for Health Statistics from April 1971 to June 1974, the regression equation predicting the average weight of a male aged 18-24 (
- y
-) based on his height (
- x
-) is given by
- $\hat{y} = -172.63 + 4.842x$
- . (www.cdc.gov/nchs/data/ad/ad014acc.pdf) What is the best prediction for the weight of a male aged 18-24 who is 70 inches tall? Round your answer to the nearest pound.

- A) \bar{y}
 B) 166
 C) 339
 D) 70

21) _____

- 22) Nine data points of data yield
- $r = 0.867$
- and the regression equation
- $\hat{y} = 19.4 + 0.93x$
- . Also,
- $\bar{y} = 64.7$
- . What is the best predicted value of
- y
- for
- $x = 40$
- ?

- A) 37.2
 B) 57.8
 C) 64.7
 D) 56.6
 E) 79.6

22) _____

Select the most appropriate answer.

- 23) The slope is the 23) _____
- A) predicted value of y .
 - B) predicted value of y when $x = 0$.
 - C) change in the predicted value of y per unit increase in x .
 - D) point where the regression line crosses the y -axis.
 - E) smallest value for the residual sum of squares.

Provide an appropriate response.

- 24) A regression line for predicting the selling prices of homes in Chicago is 24) _____
 $\hat{y} = 168 + 102x$, where x is the square footage of the house. A house with 1500 square feet recently sold for \$140,000. What is the residual for this observation?
- A) 13,168 B) 1316.80 C) 13,000 D) -13,168 E) -13,000

- 25) A regression line for predicting Internet usage (%) for 39 countries is $\hat{y} = -3.61 + 1.55x$, where x is the per capita GDP, in thousands of dollars, and y is Internet usage. Interpret the residual for one of the 39 countries with per capita GDP of \$15,000 and actual Internet use of 20 percent. 25) _____
- A) The actual Internet usage for this country is 3.6% higher than expected from the regression equation.
 - B) The actual Internet usage for this country is 0.36% lower than expected from the regression equation.
 - C) The actual Internet usage for this country is 3.25% lower than expected from the regression equation.
 - D) The actual Internet usage for this country is 0.36% higher than expected from the regression equation.
 - E) The actual Internet usage for this country is 3.25% higher than expected from the regression equation.

Select the most appropriate answer.

- 26) The prediction error for an observation, which is the difference between the actual value and the predicted value of the response variable, is called _____ 26) _____.
- A) an extrapolation
 - B) an intercept
 - C) a correlation
 - D) an outlier
 - E) a residual

Fill in the missing information.

- 27)

\bar{x}	s_x	\bar{y}	s_y	r	$\hat{y} = a + bx$
40	20	8	11	8	$\hat{y} = ?$

27) _____
- A) $-154 + 4.4x$
 - B) $-168 + 4.4x$
 - C) $175 + 2.2x$
 - D) $-38 + 4.4x$
 - E) $-168 + 2.2x$

28)	\bar{x}	s_x	\bar{y}	s_y	r	$\hat{y} = a + bx$	28)	_____
	?	?	18	4	-0.5	$\hat{y} = 30 - 4x$		

A) $\bar{x} = 3; s_x = 0.50$
 B) $\bar{x} = 12; s_x = 2.00$
 C) $\bar{x} = 12; s_x = 1.00$
 D) $\bar{x} = 3; s_x = 1.00$
 E) $\bar{x} = 48; s_x = -18.00$

Provide an appropriate response.

- 29) Which of the following has the potential for affecting the relationship between the response and explanatory variable, but was not measured by the study? 29) _____
- A) None of these
 B) All of these
 C) A confounding variable
 D) Outliers
 E) A lurking variable
- 30) A study of consumer behavior finds a positive correlation between sales of ice cream and sales of soda. What might explain the strong correlation? 30) _____
- A) No lurking variable
 B) Ice cream creates a thirst for soda
 C) People generally have ice cream for dessert if they have drunk soda with a meal.
 D) Arithmetic mistake
 E) Outdoor temperature