AP Statistics – Unit 2 (Chapters 7-10) Warm-Ups: Part 1

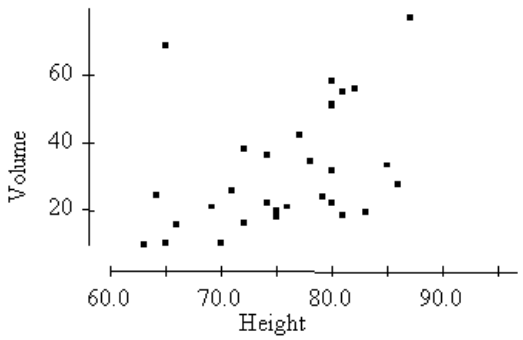
2. A researcher is interested in determining if one could predict the score on a statistics exam from the amount of time spent studying for the exam. In this study, the explanatory variable is

A) the researcher B) the amount of time spent studying for the exam
C) the score on the exam D) the fact that this is a statistics exam

A researcher measures the height (in feet) and volume of usable lumber (in cubic feet) of 32 cherry trees. The goal is to determine if volume of usable lumber can be estimated from the height of a tree. The results are plotted below.

4. In the study above, the response variable is

A) Height B) Volume
C) height or volume; it doesn't matter which is considered the response
D) neither height nor volume; the measuring instrument used to measure height is the response variable



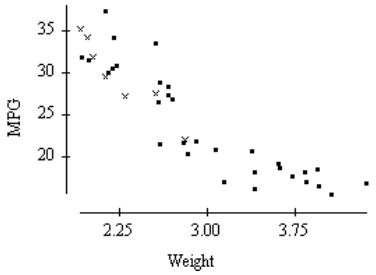
5. The scatterplot above suggests that

A) there is a positive association between height and volume
B) there is an outlier in the plot
C) both a and b D) neither a nor b

6. At a large university, the office responsible for scheduling classes notices that demand is low for classes that meet before 10:00 AM or after 3:00 PM and is high for classes that meet between 10:00 AM and 3:00 PM. Which of the following may we conclude about the relationship between demand for classes and the time the classes meet?

A) there's an association B) there's a positive association C) there's a negative association D) there's no association

7. The graph below plots the gas mileage (miles per gallon, or MPG) of various 1978 model cars versus the weight of these cars in thousands of pounds.



In the graph, the points denoted by the plotting symbol x correspond to cars made in Japan. From this plot, we may conclude that

A) in 1978 there was little difference between Japanese cars and cars made in other countries
B) in 1978 Japanese cars tended to be lighter in weight than other cars
C) in 1978 Japanese cars tended to get poorer gas mileage than other cars
D) the plot is invalid. A scatterplot is used to represent quantitative variables, and the country that makes a car is a qualitative variable

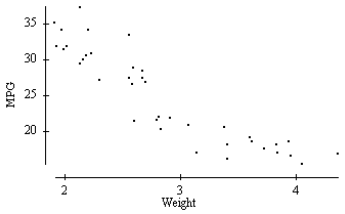
9. A school guidance counselor examines the number of extracurricular activities of students and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero." A correct interpretation of this statement would be that

A) active students tend to be students with poor grades, and vice versa
B) students with good grades tend to be students that are not involved in many activities, and vice versa
C) students involved in many extracurricular activities are just as likely to get good grades as bad grades. The same is true for students involved in few extracurricular activities
D) involvement in many extracurricular activities and good grades go hand in hand

11. Which of the following statements is true?
A) The correlation coefficient equals the proportion of times two variables lie on a straight line
B) The correlation coefficient will be +1.0 only if all the data lie on a perfectly horizontal straight line
C) The correlation coefficient measures the fraction of outliers that appear in a scatterplot
D) The correlation coefficient has no unit of measurement and must always lie between –1.0 and +1.0, inclusive

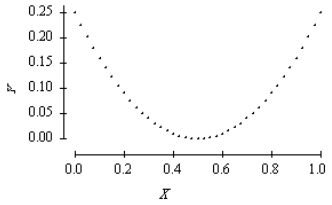
12. A study found a correlation of $r = –0.61$ between the gender of a worker and his or her income. You may correctly conclude that
A) women earn more than men on the average
B) women earn less than men on the average
C) an arithmetic mistake was made. Correlation must be positive
D) this is incorrect because r makes no sense here

13. Consider the scatterplot below.



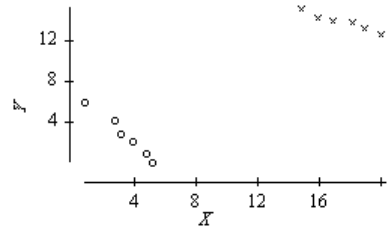
According to the scatterplot, which of the following is a plausible value for the correlation coefficient between weight and MPG?
A) +0.2 B) –0.9 C) +0.7 D) –1.0

15. Consider the scatterplot below of two variables X and Y.



We may conclude that
A) the correlation between X and Y must be close to 1 since there is a nearly perfect relation between them
B) the correlation between X and Y must be close to –1 since there is a nearly perfect relation between them, but it is not a straight-line relation
C) the correlation between X and Y is close to 0
D) the correlation between X and Y could be any number between –1 and +1. Without knowing the actual values of X and Y we can say nothing more

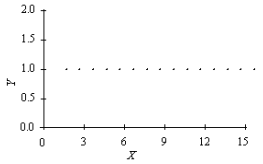
19. The scatterplot below is from a small data set.



The data were classified as either of type 1 or type 2. Those of type 1 are indicated by o's, those of type 2 by x's. The overall correlation of the data in this scatterplot is
A) positive
B) negative, since the o's display a negative trend and the x's display a negative trend
C) near 0, because the o's display a negative trend and the x's display a negative trend, but the trend from the o's to the x's is positive. The different trends cancel
D) impossible to compute for such a data set

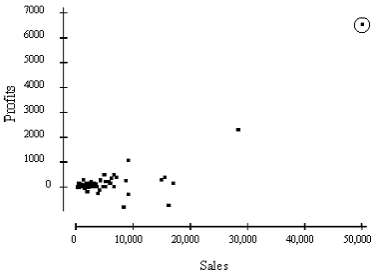
41. Which of the following statements concerning residuals is true?
A) The sum of the residuals is always 0
B) A plot of the residuals is useful for assessing the fit of the least-squares regression line
C) The value of a residual is the observed value of the response minus the value of the response that one would predict from the least-squares regression line
D) All of the above

20. A scatterplot of a variable *Y* versus a variable *X* produced the scatterplot below. The value of *Y* for all values of *X* is exactly 1.0. The correlation between *Y* and *X* is



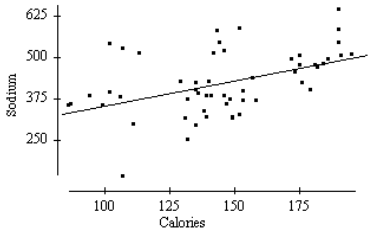
A) +1.0, because the points lie perfectly on a line
B) either +1.0 or –1.0, because the points lie perfectly on a line
C) 0, because *Y* does not change as *X* increases
D) none of the above

21. The profits (in multiples of $100,000) versus the sales (in multiples of $100,000) for a number of companies are plotted below. The correlation between profits and sales is 0.814. Suppose we removed the point that is circled from the data represented in the plot. The correlation between profits and sales would then be



A) 0.814
B) larger than 0.814
C) smaller than 0.814
D) either larger or smaller than 0.814; it is impossible to say which

24. Below is a scatterplot of the calories and sodium content of several brands of meat hot dogs. The least-squares regression line has been drawn in on the plot.



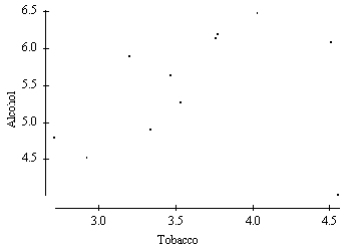
Based on the least-squares regression line in this scatterplot, one would predict that a hot dog containing 100 calories would have a sodium content of about
A) 70
B) 350
C) 400
D) 600

25. In a statistics course a linear regression equation was computed to predict the final exam score from the score on the first test. The equation of the least-squares regression line was

$$y = 10 + 0.9x$$

where *y* represents the final exam score and *x* is the score on the first exam. Suppose Joe scores a 90 on the first exam. What would be the predicted value of his score on the final exam?
A) 91
B) 89
C) 81
D) It cannot be determined from the information given. We also need to know the correlation.

27. The British government conducts regular surveys of household spending. The average weekly household spending on tobacco products and alcoholic beverages for each of 11 regions in Great Britain was recorded. A scatterplot of spending on tobacco versus spending on alcohol is given below.



Which of the following statements is true?

A) The observation in the lower right corner of the plot is influential
B) There is clear evidence of negative association between spending on alcohol and tobacco
C) The equation of the least-squares line for this plot would be approximately $y = 10 - 2x$
D) The correlation coefficient for this data is 0.99

28. John's parents recorded his height at various ages up to 66 months. Below is a record of the results.

Age (months)	36	48	54	60	66
Height (inches)	35	38	41	43	45

Which of the following is the equation of the least-squares regression line of John's height on age? (NOTE: You do not need to directly calculate the least-squares regression line to answer this question.)

A) Height = 12 × (Age) B) Height = Age/12 C) Height = 60 – 0.22 × (Age) D) Height = 22.3 + 0.34 × (Age)

33. A researcher wishes to study how the average weight Y (in kilograms) of children changes during the first year of life. He plots these averages versus the age X (in months) and decides to fit a least-squares regression line to the data with X as the explanatory variable and Y as the response variable. He computes the following quantities.
r = correlation between X and Y = 0.9 J = mean of the values of X = 6.5 M = mean of the values of Y = 6.6
s_J = standard deviation of the values of X = 3.6 s_M = standard deviation of the values of Y = 1.2
The slope of the least-squares line is

A) 0.30 B) 0.88 C) 1.01 D) 3.0

35. In a study of 1991 model cars, a researcher found that the fraction of the variation in the price of cars that was explained by the least-squares regression on horsepower was about 0.64. For the cars in this study, the correlation between the price of the car and its horsepower was found to be positive. The actual value of the correlation

A) is 0.80 B) is 0.64 C) is 0.41 D) cannot be determined from the information given

38. A researcher wishes to determine whether the rate of water flow (in liters per second) over an experimental soil bed can be used to predict the amount of soil washed away (in kilograms). The researcher measures the amount of soil washed away for various flow rates and from these data calculates the least-squares regression line to be

amount of eroded soil = 0.4 + 1.3 x (flow rate)

One of the flow rates used by the researcher was 0.3 liters per second; for this flow rate the amount of eroded soil was 0.8 kilograms. These values were used in the calculation of the least-squares regression line. The residual corresponding to these values is

A) 0.01 B) –0.01 C) 0.5 D) –0.5

AP Statistics – Unit 2 (Chapters 7-10) Warm-Ups: Part 2

5. I measure a response variable Y at each of several times. A scatterplot of log$Y$ versus time of measurement looks approximately like a positively sloping straight line. We may conclude that

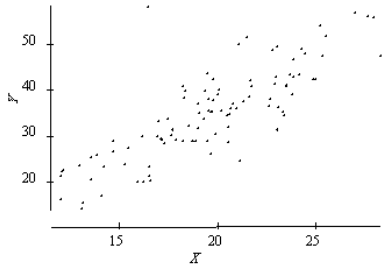
A) the correlation between time of measurement and Y is negative, since logarithms of positive fractions (such as correlations) are negative
B) the rate of growth of Y is positive, but slowing down over time
C) an exponential growth model would approximately describe the relationship between Y and time of measurement
D) a mistake has been made. It would have been better to plot Y versus the logarithm of the time of measurement

6. Using least-squares regression, I determine that the logarithm (base 10) of the population of a country is approximately described by the equation $\log(\text{population}) = -13.5 + 0.01 \times (\text{year})$
Based on this equation, the population of the country in the year 2000 should be about
A) 6.5 B) 665 C) 2,000,000 D) 3,167,277

12. Researchers studied a sample of 100 adults between the ages of 25 and 35 and found a strong negative correlation between the amount of vitamin C an individual consumed and the number of pounds the individual was overweight. Which of the following may we conclude?
A) This is strong, but not conclusive, evidence that large amounts of vitamin C inhibit weight gain
B) If the amount of vitamin C consumed and the number of pounds overweight for each individual in this study were plotted on a scatterplot, the points would lie close to a negatively sloping straight line
C) If a larger sample of adults between the ages of 25 and 35 had been studied, the correlation would have been even stronger
D) All of the above

15. When exploring very large sets of data involving many variables, which of the following is true?
A) Extrapolation is safe because it is based on a greater quantity of evidence
B) Associations will be stronger than would be seen in a much smaller subset of the data
C) A strong association is good evidence for causation because it is based on a large quantity of information
D) None of the above

20. Consider the following scatterplot.



From this plot we can conclude
A) that there is evidence of a modest cause-and-effect relation between *X* and *Y* with increases in *X* causing increases in *Y*
B) that there is an outlier in the plot C) that there is a strongly influential point in the plot D) all of the above

22. When possible, the best way to establish that an observed association is the result of a cause-and-effect relation is by means of
A) the least-squares regression line B) the correlation coefficient
C) examining *z*-scores rather than the original variables D) a well-designed experiment

Answer Key (part 1)

2.B	4.B	5.C	6.A	7.B	9.C	11.D	12.D	13.B	15.C	19.A	20.C	21.C	24.B
25.A	27.A	28.D	33.A	35.A	38.A	41.D							

Answer Key (part 2)

5.C	6.D	12.B	15.D	20.B	22.D