AP Statistics Practice Test
Unit Three – Exploring Relationships Between Variables

Name_____Period_____Date_____

True or False:

1. Correlation and regression require explanatory and response variables.   1._____

2. Every least squares regression line passes through $(\overline{x}, \overline{y})$.   2._____

Short Answer:

3. In a simple, linear regression model, the variable that is being predicted may be called which of the following: (Check all that apply)

    _   independent variable   _   regression variable
    _   dependent variable   _   response variable
    _   X variable   _   Y variable

4. Which of the following correlation values represents a perfect linear relationship between two quantitative variables? (Check all that apply)

    _   0   _   1
   _   .9   _   .5
    _   -1

5. A positive correlation between two variables X and Y means: If you increase the X value, this will cause the value of Y to increase. Circle the correct answer and briefly explain.

    a.   This is always true.

    b.   This is sometimes true.

    c.   This is never true.

6. Describe a situation in which you would expect a fairly strong association between two variables that do not have a cause-and-effect relationship.

7. Think of a set of bivariate, quantitative data. When the correlation coefficient for these data is close to 1, must the slope also be close to 1? Explain.

8. In the scatterplot of y versus x shown at the right, **circle** the point with the largest residual and place a **square** around the point that would be considered an influential observation.


1
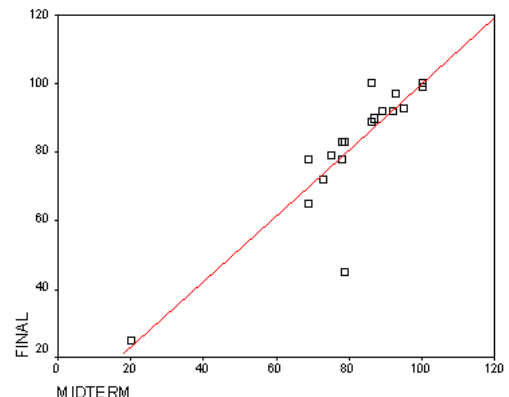
Multiple-Choice: Circle the best answer.

9. Which of these statements are false?

(A) There is a strong linear relationship between gender and height because we found a correlation of .65.
(B) Plant height and leaf height were found to be negatively correlated because the correlation coefficient is -1.41.
(C) Since the correlation between X and Y is 0, this means there is no relationship whatsoever between these two variables.
(D) All of the above.
(E) None of the above.

10. Two measures x and y were taken on 18 subjects. The first of two regressions, Regression I, yielded $\hat{y} = 24.5 + 16.1x$ and had the following residual plot.
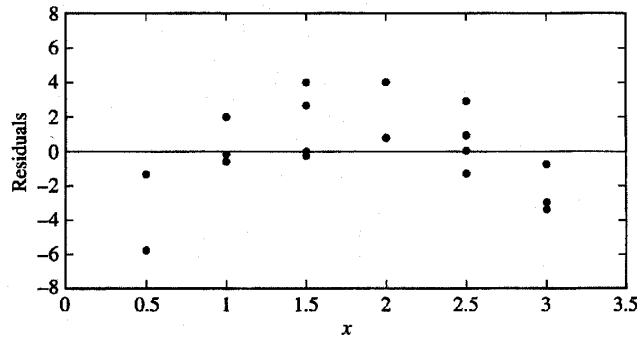
The second regression, Regression II, yielded $\log(\hat{y}) = 1.6 + 0.51 \log(x)$ and had the following residual plot.
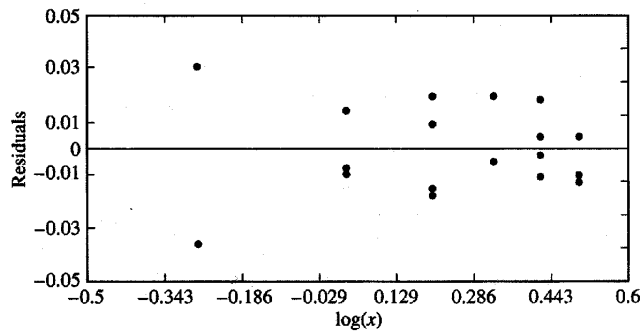
Which of the following conclusions is best supported by the evidence above?

(A) There is a linear relationship between x and y, and Regression I yields a better fit.
(B) There is a linear relationship between x and y, and Regression II yields a better fit.
(C) There is a negative correlation between x and y.
(D) There is a nonlinear relationship between x and y, and Regression I yields a better fit.
(E) There is a nonlinear relationship between x and y, and Regression II yields a better fit.

11. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least squares fit of some data collected by a biologist gives the model

$$\hat{y} = 25.2 + 3.3x, \qquad 9 < x < 25,$$

where x is the number of chirps per minute and $\hat{y}$ is the estimated temperature in degrees Fahrenheit. What is the estimated increase in temperature that corresponds to an increase of 5 chirps per minute?

(A)   3.3° F
(B) 16.5° F
(C) 25.2° F

(D) 28.5° F
(E) 41.7° F

12. The equation of the least squares regression line for the points on the scatterplot at the right is $\hat{y} = 1.3 + .73x$. What is the residual for the point (4, 7)?

(A) 2.78
(B) 3.00
(C) 4.00
(D) 4.22
(E) 7.00
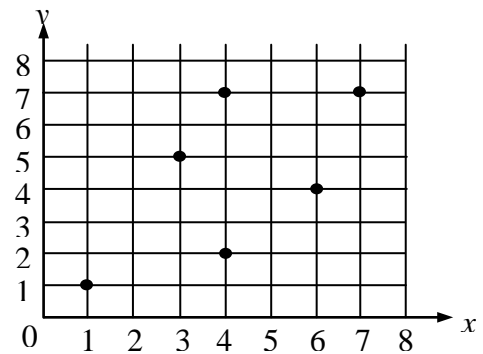
13. What does the pattern of the residuals tell you about the linear model?

(A) The evidence is inconclusive.
(B) The residual plot clearly contradicts the linearity of the data.
(C) The residual plot confirms the linearity of the fuel economy data.
(D) None of the above.
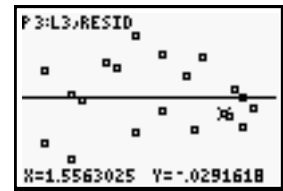
Free Response:

Exercises 14-18 relate to the following.
Joey read in his biology book that fish activity increases with water temperature, and he decided to investigate this issue by conducting an experiment. On nine successive days, he measures fish activity and water temperature in his aquarium. Larger values of his measure of fish activity denote more activity. The figure below presents the scatterplot of his data.
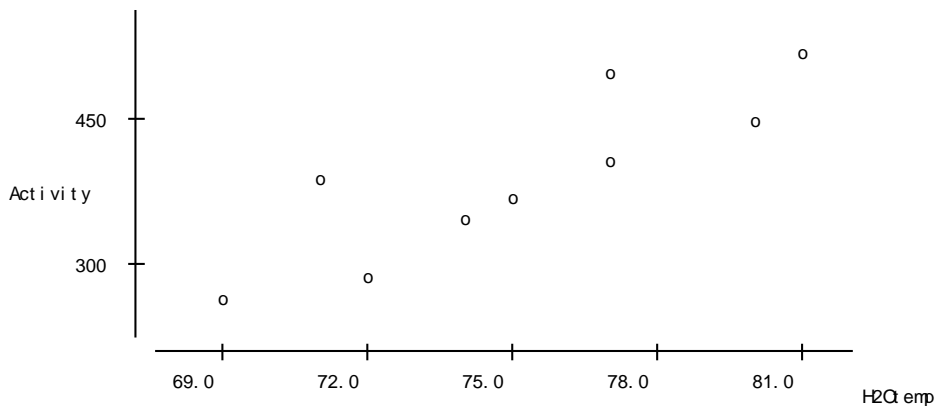
14. What does the scatterplot reveal? State in plain language as if you were explaining to a friend who knows very little statistics.

15. What is your best guess for the correlation coefficient between water temperature and fish activity.

_____

16. If we reversed the variables and made a scatterplot of fish activity vs. water temperature, what would the new correlation coefficient be? *Briefly* explain your reasoning.

17. Suppose we were to record temperature in Celsius degrees instead of Fahrenheit. How would the correlation change? Why?

18. Suppose a new point at (84, 300), i.e., water temperature = 84° F and fish activity = 300, is added to the plot. Describe the effect, if any, that this new point will have on the correlation coefficient of water temperature vs. fish activity and how it will affect the slope of the line. Why does it have this effect?

The table shows the number of live births per 1000 women aged 15-44 years in the United States starting in 1965. When you enter the data, enter year as 65, 70, 75, 80, 85, etc.:
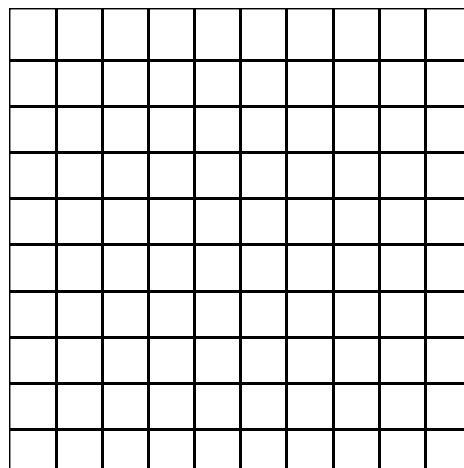
YEAR	1965	1970	1975	1980	1985	1990	1995	1999
Rate	19.4	18.4	16.3	15.9	15.6	16.4	14.8	14.5

19. Which is the explanatory variable?_____

20. Make a scatterplot of the data <u>by hand</u>.

21. Determine the equation of the LSRL for this data using your TI-84.

    _____

    Define your variables:

22. Plot the line on your scatterplot. Plot it on your TI-84 and then just copy it from the screen.

23. What is the slope? **Define the slope in the context of this problem**.

24. What is the correlation between x and y? **Interpret this value in the context of this problem.**

25. What is the coefficient of determination? **Interpret this number.**

4

26. Predict what the birth rate will be in 2010. Comment on your faith in this prediction.

27. Find the residual for the <u>first point</u> in the data set. Show your work below for credit.

28. Construct a residual plot in your calculator and sketch it below.

29.   Use the residual plot to answer this question:
      Is a linear model appropriate for these two
      variables?   Explain.

A consumer organization has reported test data for 50 car models. We will examine the association between the weight of the car (in thousands of pounds) and the fuel efficiency (in miles per gallon).

```
Predictor   Coef    StDev    T      P
Constant    48.739  1.976   24.7   0.000
Weight     -8.2136  0.6738 -12.2   0.000

S = 2.413   R-Sq = 75.6%   R-Sq(adj) = 71.2%
```

30. Using the printout above, find the LSRL.   30._____

31. Find the correlation between weight of a car and its gas mileage.   31._____

32. What is your estimate of fuel efficiency (MPG) for a car that weighs 3000 lbs?

                                                                            32._____

Review:

33. The number of new projects started each month at an advertising agency for the last six months is:
       2   5   3   3   6   3
    The interquartile range for the above data is

   (A) 1.0
   (B) 2.0
   (C) 3.0
   (D) 4.0
   (E) 5.0   33._____

34. The weight of a randomly selected can of a new soft drink is known to have a normal distribution with a mean of 8.3 ounces and a standard deviation of 0.2 ounces. The weight that should be stamped on each can so that only 2% of all cans are underweight is

(A) 7.89 ounces.
(B) 8.13 ounces.
(C) 8.26 ounces.
(D) 8.71 ounces.
(E) 8.02 ounces.   34._____

35. Birth weights at a local hospital have a normal distribution with a mean of 110 ounces and a standard deviation of 15 ounces. The proportion of infants with birth weights under 95 ounces is

(A) 0.341
(B) 0.500
(C) 0.159
(D) 0.682
(E) 0.841   35._____