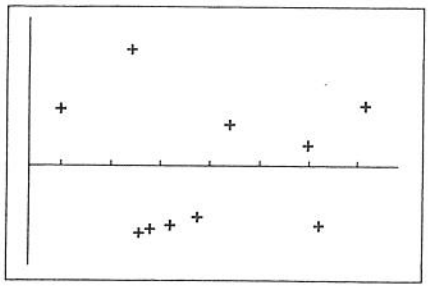
Name _____ Date _____

Mr. Bannon - AP Statistics Period _____

# Unit 3 - Review

## Part I

1) Consider the following residual plot:



Which of the following statements is (are) true?

I. The residual plot indicates that a line is a reasonable model for the data.
II. The residual plot indicates that there is no relationship between the data.
III. The correlation between the variables is probably non-zero.

(a) I only
(b) II only
(c) I and III only
(d) II and III only
(e) I and II only

2) Which of the following are true about a point that is an outlier with respect to the response variable but is not an outlier with respect to the explanatory variable?

I. Significantly affects the slope of the regression line.
II. Significantly affects the correlation.
III. Significantly affects the y-intercept of the regression line.

**A.** I only
**B.** II only
**C.** I and II
**D.** II and III
**E.** I, II, and III

3) One of the points in a set of points is (4, 8). What is the residual for this point if the equation of the regression line for these points is $\hat{y} = 3.6 + 1.8x$?

**A.** –1.8
**B.** –2.4
**C.** –2.8
**D.** –3.2
**E.** –3.6

Use the following information for problems 9 and 10:

The following table shows the salaries of 10 employees (in $1000's) along with the purchase price of the cars they drive (in $1000's).

| Salary | 27 | 62 | 42 | 18 | 35 | 20 | 55 | 25 | 27 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|
| Car Price | 21 | 39 | 28 | 11 | 24 | 13 | 36 | 18 | 20 | 16 |

4) Using the information in the table, what would be the expected value of a car driven by an employee with a salary of $30,000?

A. $20,300
B. $20,500
C. $20,700
D. $20,900
E. $21,100

5) Using the information in the table, what would be the expected value of the salary for an employee who drives a car valued at $30,000?

A. $43,550
B. $44,400
C. $45,650
D. $47,800
E. $49,850

6) A data set with 5 points has a correlation of 0.8 and a linear regression line with a slope of 0.8. Which of the following statements are true?

I. A sixth point can be added so that the resulting slope is 1.
II. A sixth point can be added so that the resulting correlation is 1.
III. A sixth point can be added so that the resulting slope is –1.
IV. A sixth point can be added so that the resulting correlation is –1.

A. I and III
B. II and IV
C. I and II
D. III and IV
E. I, II, III, and IV

7) A study found a correlation of $r = -0.58$ between hours per week spent watching television and hours per week spent exercising. That is, the more hours spent watching television, the less hours spent exercising per week. Which of the following statements is most accurate?

(a) About one-third of the variation in hours spent exercising can be explained by hours spent watching television.
(b) A person who watches less television will exercise more.
(c) For each hour spent watching television, the predicted decrease in hours spent exercising is 0.58 hrs.
(d) There is a cause-and-effect relationship between hours spent watching television and a decline in hours spent exercising.
(e) 58% of the hours spent exercising can be explained by the number of hours watching television.

| $x$ | 23 | 15 | 26 | 24 | 22 | 29 | 32 | 40 | 41 | 46 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 19 | 18 | 22 | 20 | 27 | 25 | 32 | 38 | 35 | 45 |

8) The regression line for the two-variable dataset given above is $\hat{y} = 2.35 + 0.86x$. What is the value of the residual for the point whose $x$-value is 29?

(a) 1.71
(b) −1.71
(c) 2.29
(d) 5.15
(e) −2.29

9) Given a set of ordered pairs $(x, y)$ so that $s_x = 1.6$, $s_y = 0.75$, $r = 0.55$. What is the slope of the least-square regression line for these data?

(a) 1.82
(b) 1.17
(c) 2.18
(d) 0.26
(e) 0.78

10) Suppose the LSRL for predicting Weight (in pounds) from Height (in inches) is given by *Weight* = −115 + 3.6 (*Height*). Which of the following statements is correct?

I. A person who is 61 inches tall will weigh 104.6 pounds.
II. For each additional inch of Height, Weight will increase on average by 3.6 pounds.
III. There is a strong positive linear relationship between Height and Weight.

(a) I only
(b) II only
(c) III only
(d) II and III only
(e) I and II only

11) The correlation between two variables X and Y is –0.26. A new set of scores, X* and Y*, is constructed by letting X* = –X and Y* = Y +12. The correlation between X* and Y* is

a. – 0.26
b. 0.26
c. 0
d. 0.52
e. – 0.52

12) A study of stopping distances found that the least squares regression line for predicting mileage (in miles per gallon) from the weight of the vehicle (in hundreds of pounds) was MPG = 32.50 – 0.45(Weight). The mean weight for the vehicles in the study was 2980 pounds. What was the mean MPG in the study?

a. 19.09
b. 15.27
c. –1308.5
d. 18.65
e. 20.33

13) The linear correlation coefficient of a set of points is –0.75. If the $x$-coordinate of each point is doubled in value and the y-coordinate of each point were halved in value, which of the following statements most accurately describes what will happen to the correlation coefficient?

**A.** It will increase by an unknown amount.
**B.** It will decrease by an unknown amount.
**C.** It will stay the same.
**D.** It will change by a fixed amount.
**E.** The answer cannot be determined from the information given.

14) The correlation of a set of 25 points is positive. If two points are chosen at random and a line is drawn through these two points, which of the following could be true about the slope of that line?

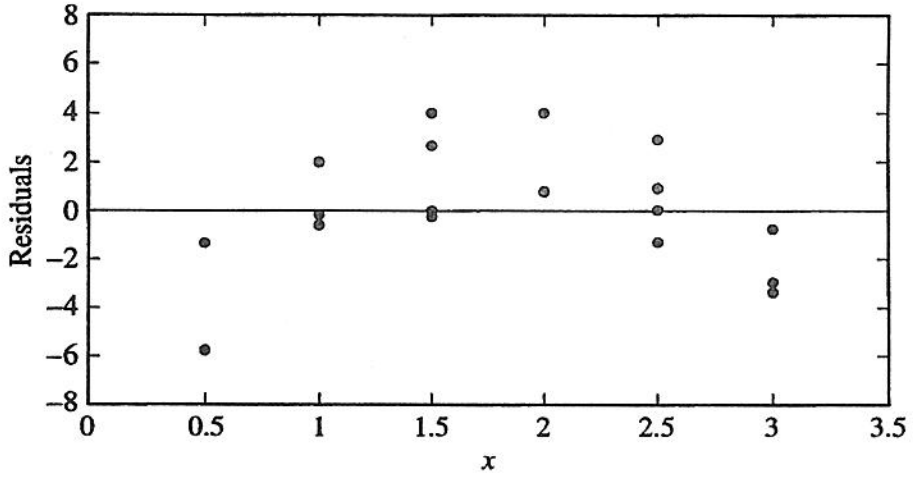
I. The slope is positive.
II. The slope is negative.
III. The slope is zero.
IV. The slope is undefined.

**A.** None of the above
**B.** One of the above
**C.** Two of the above
**D.** Three of the above
**E.** All four of the above

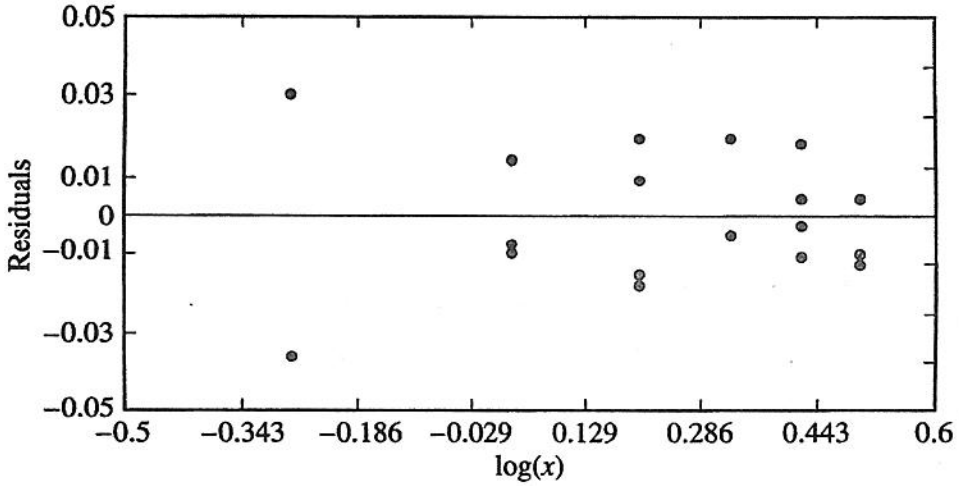
15) A response variable appears to be exponentially related to the explanatory variable. The natural logarithm of each $y$-value is taken and the least-squares regression line is found to be $ln(y) = 1.64 - 0.88x$. Rounded to two decimal places, what is the predicted value of $y$ when $x = 3.1$?

(a) –1.09
(b) –0.34
(c) 0.34
(d) 0.082
(e) 1.09

16) 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 $\widehat{\log(y)} = 1.6 + 0.51 \log(x)$ 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.

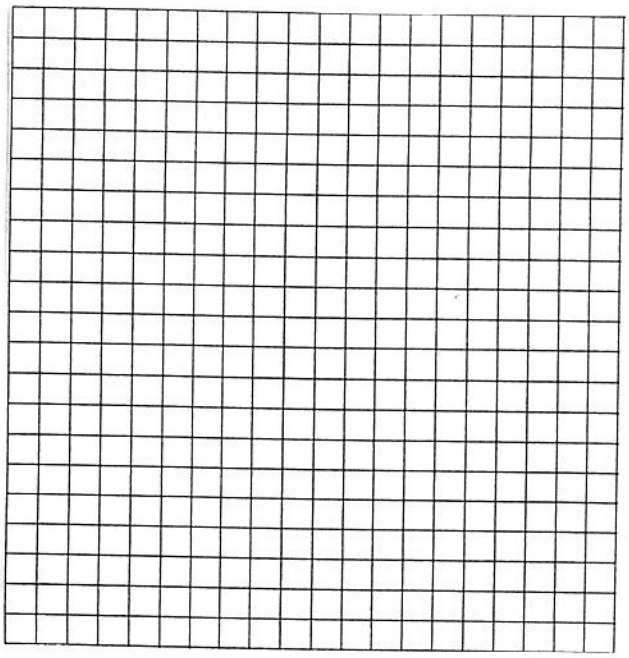
# Part II

## Free Response.

1) The data below give the first and second exam scores of 10 students in a calculus class.

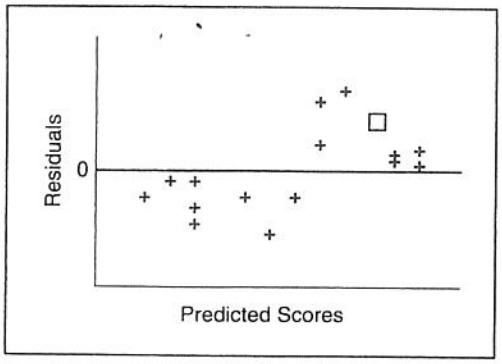
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Test 1** | 63 | 32 | 87 | 73 | 60 | 63 | 83 | 80 | 98 | 85 |
| **Test 2** | 51 | 21 | 52 | 90 | 83 | 54 | 73 | 85 | 83 | 46 |

(a) Draw a scatterplot of these data.
(b) To what extent do the scores on the two tests seem related?



2) Given a two-variable dataset such that $\bar{x} = 14.5$, $\bar{y} = 20$, $s_x = 4$, $s_y = 11$, $r = .80$, find the least-squares regression line of $y$ on $x$.

3) The following is a residual plot of a linear regression. A line would not be a good fit for these data. Why not? Is the regression equation likely to underestimate or overestimate the $y$-value of the point in the graph marked with the square?




4) The regional champion in 10 and under 100 m backstroke has had the following winning times (in seconds) over the past 8 years:

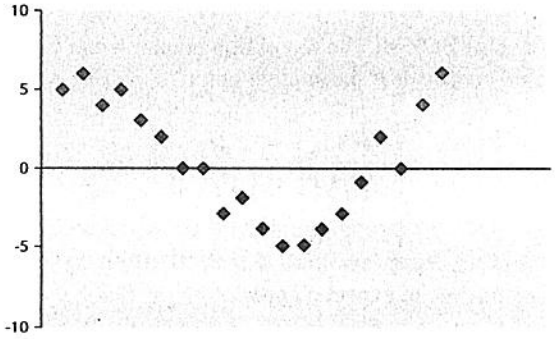
| **Year** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **Time** | 77.3 | 80.2 | 77.1 | 76.4 | 75.5 | 75.9 | 75.1 | 74.3 |

How many years until you expect the winning time to be one minute or less? What's wrong with this estimate?

5) Suppose the regression equation for predicting success on a dexterity task ($y$) from number of training sessions ($x$) is $\hat{y} = 45 + 2.7x$ and that $\frac{s_y}{s_x} = 3.33$.

What percentage of the variation in $y$ is not explained by the regression on $x$?

6) The following residual plot shows a nonlinear pattern:




If the linear regression least-squares line for the original data is $\hat{y} = 3x + 7$, discuss why the predicted value of $\hat{y} = 52$ when $x = 15$ is inaccurate. What might be done to give a better prediction?