Name:_____ Date:_____

**AP Statistics Assignment 3.1 Correlation and Least Square Regression Lines**

1. Definitions: Define each of the following using your own words:

*Response variable:*

*Explanatory variable:*

*Scatterplot:*

*Interpreting a scatterplot:*

*Positive association:*

*Negative association:*

*Categorical variables in scatterplots:*

*Correlation (r):*

$r = \frac{1}{n-1}\Sigma(\frac{x_i - \bar{x}}{s_x})(\frac{y_i - \bar{y}}{s_y})$ or $r = \Sigma\frac{(Z_x)(Z_y)}{n-1}$

2. For each of the following situations, is there a clear indication of an explanatory and response variable? If yes, define which variable is which. If not, explain why. Also, indicate if the variables are quantitative or categorical:

i) Your height in inches and weight in lbs

ii) Your GPA and the number of hours you spend watching TV

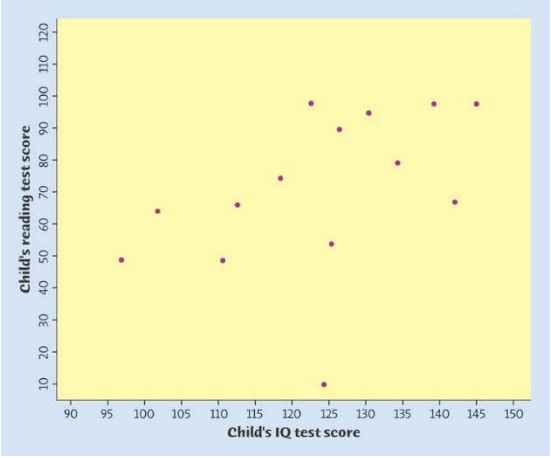
iii) Height in inches and how much hair you have on your head

iv) The type of instructor for driving school and the number car accidents in the first five years

v) The heights of both the father and son

vi) Your average salary in the first five years and the number of years in college/university

3. Here is a scatterplot of reading test scores against IQ test scores for 14 fifth-grade children.




i) Is there a clear explanatory variable and response variable in this setting? If so, tell which is which. If not, explain why not.

ii) Does the scatterplot show a positive association, negative association, or neither? Explain why this makes sense.

**iii).** How would you describe the *form* of the relationship?

**iv)** Which of the following is closest to the correlation between reading test score and IQ score for the group of 15 children? Explain.

$r = 0.9$ $r = 0.7$ $r = 0.4$ $r = -0.4$

**v.** There is one "unusual point" on the graph. Explain what is "unusual" about this student.

**vi.** What effect would removing the "unusual point" have on the correlation? Justify your answer.

4. A student wonders if tall women tend to date taller men than do short women. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (heights in inches):

| | | | | | | |
|---|---|---|---|---|---|---|
| Women | 66 | 64 | 66 | 65 | 70 | 65 |
| Men | 72 | 68 | 70 | 68 | 71 | 65 |

**i)** Is there a clear explanatory variable and response variable in this setting? If so, tell which is which. If not, explain why not.

**ii)** Make a well-labeled scatterplot of these data.

**iii)** How would you describe the *form* of the relationship?

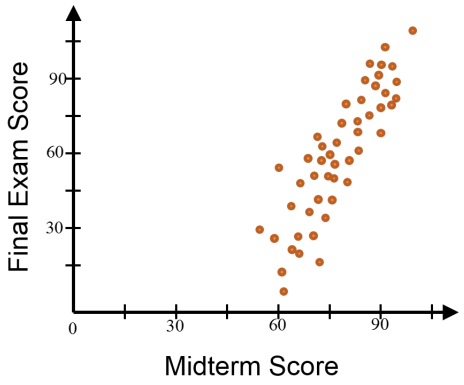
**vi)** Based on the scatterplot, do you expect the correlation to be positive or negative? Near $\pm 1$ or not? Explain.

**v).** Use your calculator to find the correlation $r$ between the heights of the men and women. Do the data show that taller women tend to date taller men? Explain.

**vi).** How would $r$ change if

- all the men were 6 inches shorter than the heights given in the table?
- heights were measured in centimeters rather than inches? (There are 2.54 centimeters in an inch.)

5. The scatter diagram shows scores on the midterm and final exam of a 1st year calculus course.




i) Approximate the average midterm score:

ii) Approximate the average final exam score:

iii) Was there more spread in the midterm or final scores?

iv) Is there a strong positive association between midterm and final scores?

6. For a representative sample of cars, would the correlation between the age of the car and its gasoline economy (miles per gallon) be positive or negative?

7. The correlation between gasoline economy and income of owner turns out to be positive. How do you account for this positive association?

8. Given the table of values, calculate "r":

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y$ | 2 | 1 | 4 | 6 | 5 | 8 | 10 |

9. Match each scatterplot with the corresponding correlations:
-1.00, -0.80, -0.40, 0.30, 0.70, 0.90