Chapter 8: Correlation

> There is strong relationship between age and the birthdays. The people who celebrate the most birthdays were found to be those with the highest ages.
> ~ author unknown

Learning Objectives

Upon completion of this chapter, students should know

- The nature and types of correlation.
- How to generate and interpret a scatterplot.
- How to compute and interpret correlation coefficients.
- How to determine statistical significance of correlation coefficients.
- How to compute and interpret the coefficient of determination (effect size).
- The distinction between correlation and causation.

Key Terms

Correlation is a relationship between two variables whereby a change in one variable is associated with a congruent change in the other.

Scatterplots are a type of graph in which two sets of data are displayed, one along the abscissa and the other along the ordinate. The points are plotted where corresponding X and Y values intersect and show the direction and degree of correlation between two variables.

Correlation coefficient is a measure that indicates the relationship between two variables. The Pearson product moment correlation coefficient symbol is *r* and range from –1.00 to +1.00. The sign of the coefficient indicates the direction of the relationship and the number indicates the size of the relationship.
A positive correlation represents the situation where both variables vary in the same direction and a negative correlation indicates a situation where the variables vary in opposite directions. A correlation coefficient near +1.00 or –1.00 is a large correlation, whereas a correlation close to 0 is small.

$$r = \frac{\mathrm{cov}_{XY}}{S_X \cdot S_Y} \qquad r = \frac{(n \cdot \sum XY) - (\sum X \cdot \sum Y)}{\sqrt{[(n \cdot \sum X^2) - (\sum X)^2] \cdot [(n \cdot \sum Y^2) - (\sum Y)^2]}}$$

Covariance is basically a number that represents the degree to which two different variables change together.

$$\text{cov}_{XY} = \frac{\sum (X - \bar{X}) \cdot (Y - \bar{Y})}{n-1} \qquad \text{cov}_{XY} = \frac{\sum XY - \frac{\sum X \cdot \sum Y}{n}}{n-1}$$

Critical value refers to a numerical value, according to the value of the **degrees of freedom** that is used as a decision point. In statistical analyses, critical values are based on the probability of a certain outcome occurring merely by chance. When the absolute value of the correlation coefficient is greater than the critical value is said to be statistically significant.

$df = n - 2$

Coefficient of determination is the effect size. It is the part of the variance of one variable that can be explained or attributed to the variance of a related variable. Coefficient of determination = r^2

Lecture and Demonstration Ideas

The relationship between two variables is a concept students say they understand. However, the notion of correlation and causation seems to be confusing, especially when the correlation is strong. For example, for ethical reasons human smoking behavior and the incidence of cancer is correlational research, yet causation is assumed. Thus, if the finding seems sensible, causality is often assumed. Yet, on the other hand, there are times when correlation and causation are easily called absurd, such as the number of electrical appliances people own and increases in birth rates. While the logic of applying causality sometimes and not others is perplexing, it is important they understand correlation is not causation. Discussion on cause and effect directionality and the third variable problem help clarify the reason why correlation cannot be causation.

Lecture and Demonstration Ideas

Math Scores and Calculators. Introduce students to correlation using Transparencies 8-1 to 8-4. The example of the results found in the 2000 Mathematics Assessment (2001) study shown on Transparency 8-7, may be a useful aid to your lecture. The results indicated a relationship between student-reported calculator use for class work and mathematics performance. Interestingly, calculator use and math performance were negatively related in grade 4 and positively related in grades 8 and 12 (see Transparency 8-7). Discuss these findings with students and what other factors might account for these relationships. During this discussion, causal conclusions may arise. If so, this will be an opportune time to discuss why correlation is not correlation and the third variable problem.

Cause? Effect? Third-Variable Problem?. Transparency 8-5 may help students visualize the three possible causes explanations for correlational findings. There are many examples to use during this discussion. Here are several: 1) positive relationship between the number of bathing suits sold and the incidence of skin cancer, 2) positive relationship between the number of years married and life satisfaction, 3) positive relationship between length of arms and reasoning ability, and 4) negative relationship between cigarettes smoked and grades in school.

Computing *r* and *r*²*.* You may find the data in Handout 1-A helpful as a computation demonstration. The scatterplot and complete solution are shown on Transparencies 8-8 and 8-9. Interpreting coefficients are sometimes difficult for students. Use Transparency 8-6, an illustration of the continuum of correlation coefficients, as a lecture aid in this discussion.

Instructional Video. *Against All Odds: Inside Statistics***.** Program Nine, "Causality" discusses correlational relationships between baseball players' salaries and home run statistics. Another example involves a study comparing identical twins raised together and apart. These videos are produced by the Consortium for Mathematics and Its Applications and Chedd-Angier (1989) and available through Annenberg/CPB.

Active-Learning Activities

Practice Direction. To give students practice identifying the direction of relationships, have students complete the activity in Handout 8-B. Also, ask students to generate their own examples of positive and negative relationships.

Additional Assignments

Critique the Popular Press. The purpose of this assignment is to give students practice critical evaluation and challenge assumptions reported in the popular press. Ask students to review articles written in popular magazines, newspapers, and the Internet to find two correlational studies reported in causal terms. Have students make copies of these articles (with citations) and include a critique of the problem. In addition, challenge students to think of several alternative explanations for these findings.

References:

National Center for Education Statistics, National Assessment of Educational Progress (2001). 2000 Mathematics Assessment.

Handout 8-A.

Computing Correlations

Student	Hours TV Per Week	Quiz Scores
1	16	64
2	12	82
3	23	56
4	24	58
5	30	60
6	24	75
7	25	72
8	15	83
9	10	90
10	7	95

1. Prepare a scatterplot.
2. Compute the correlation coefficient.
3. Compute the coefficient of determination.
4. Compute the degrees of freedom.
5. Determine the critical value.
6. Results

Handout 8-B.

Identification Exercise

	Positive Correlation	Negative Correlation	Zero Correlation
1. Age of cars and the price of cars.			
2. Years of driving experience and car accidents.			
3. Age of children and monthly allowance.			
4. Length of hair and intelligence.			
5. Hours playing video games and number of friends.			
6. Stressful events and depression.			
7. Age of wine and price of wine.			

Other Examples:

Transparency 8-1.

Correlation

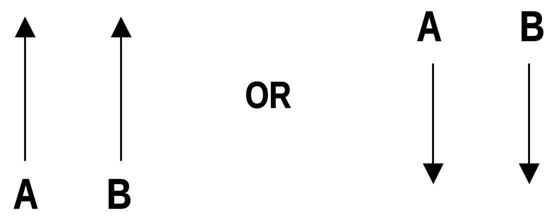
- ✓ The relationship between two variables.
- ✓ A change in one variable is associated with a concurrent change in the other.
- ✓ The direction and degree of the association (relationship) is described by a correlation coefficient.
  - ➢ Coefficients range from + 1.00 to –1.00
    - ♦ The sign of the coefficient + or – indicates the direction of the relationship.
    - ♦ The value of the coefficient 0 to 1 indicates the size of the relationship.

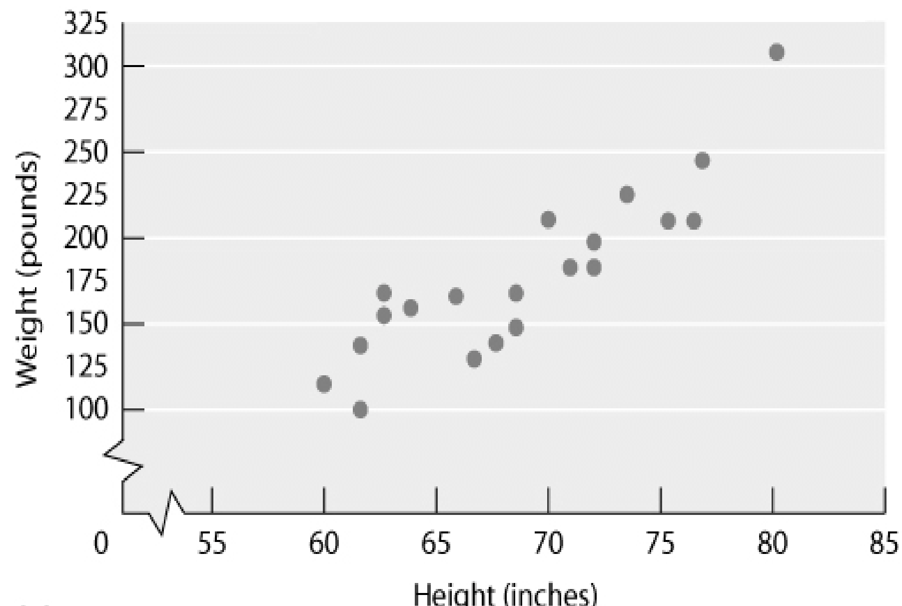
Coefficient of Determination

The portion of the variance of one variable that can be explained by the variance of the other.

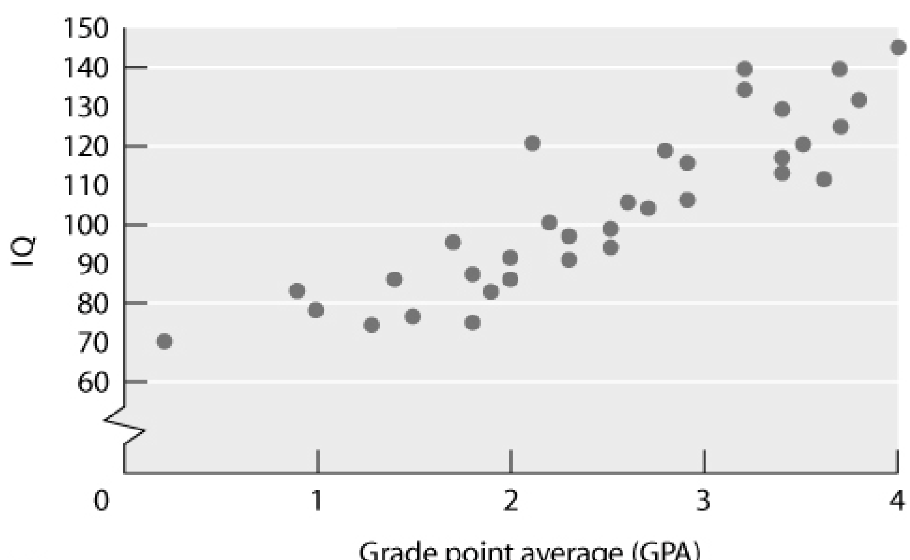
Transparency 8-2.

Positive Relationship = (+) Coefficient





(a)



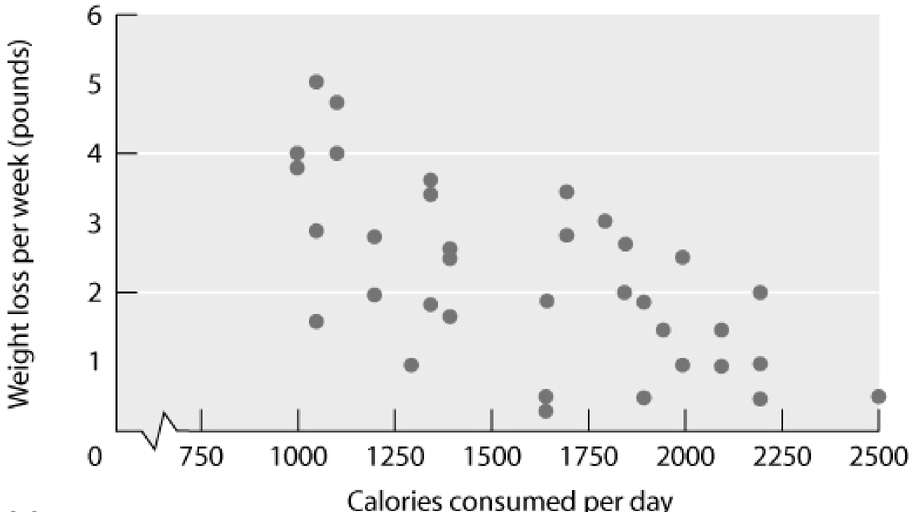
(b)

Transparency 8-3.

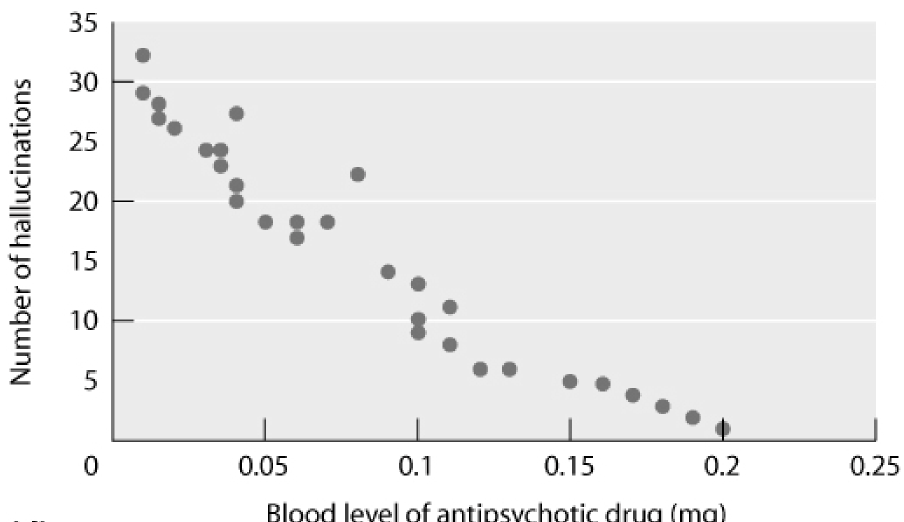
Negative Correlation = (-) Coefficient

↓ ↑

A B



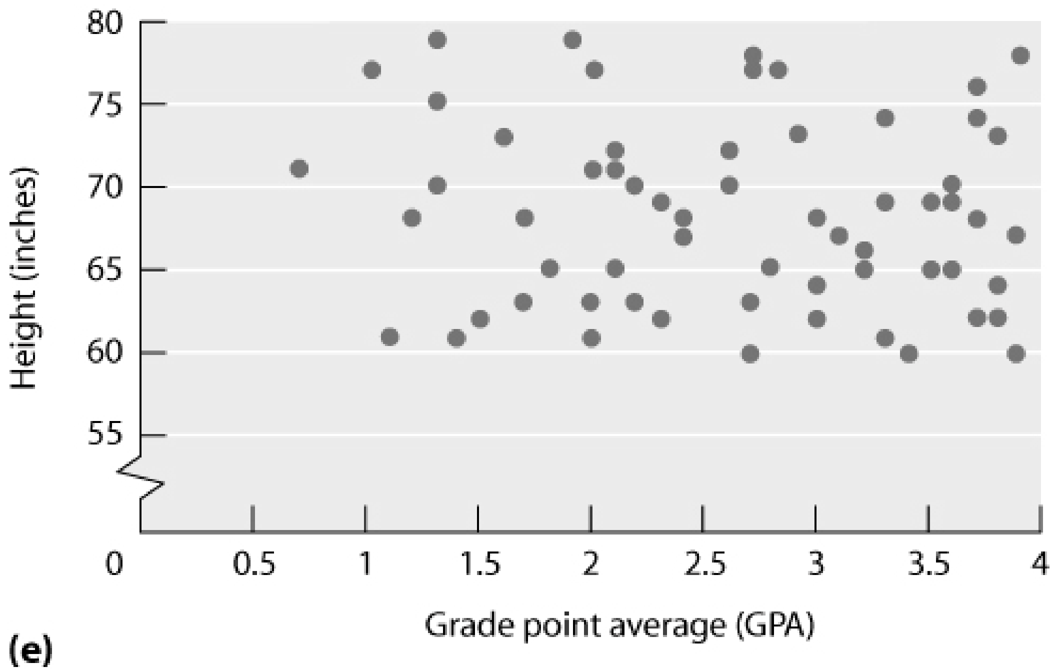
(c)



(d)

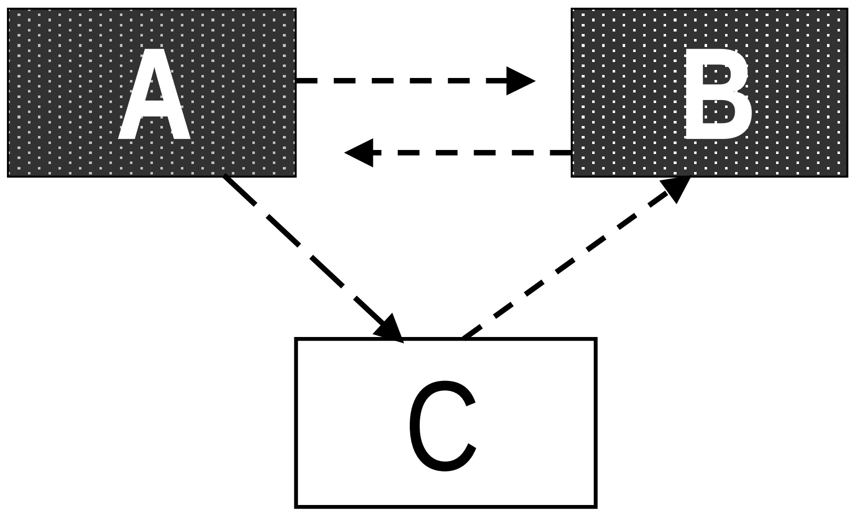
Transparency 8-4.

Zero Correlation



Transparency 8-5.

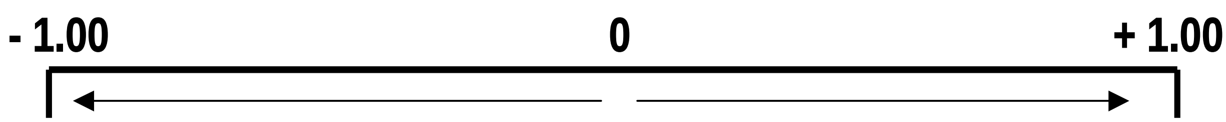
Correlation is not causation. Why not?



Transparency 8-6.

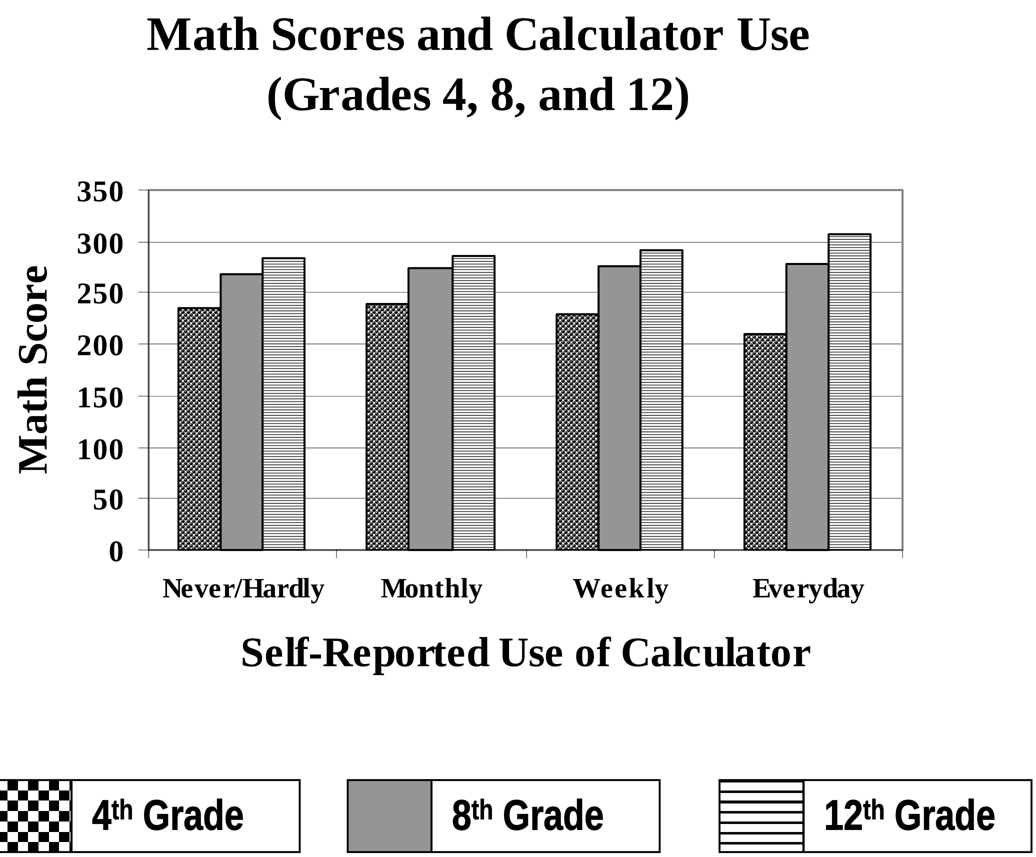
Interpreting Correlation Coefficients

The strength of the association increases in both directions from 0.



Transparency 8-7.

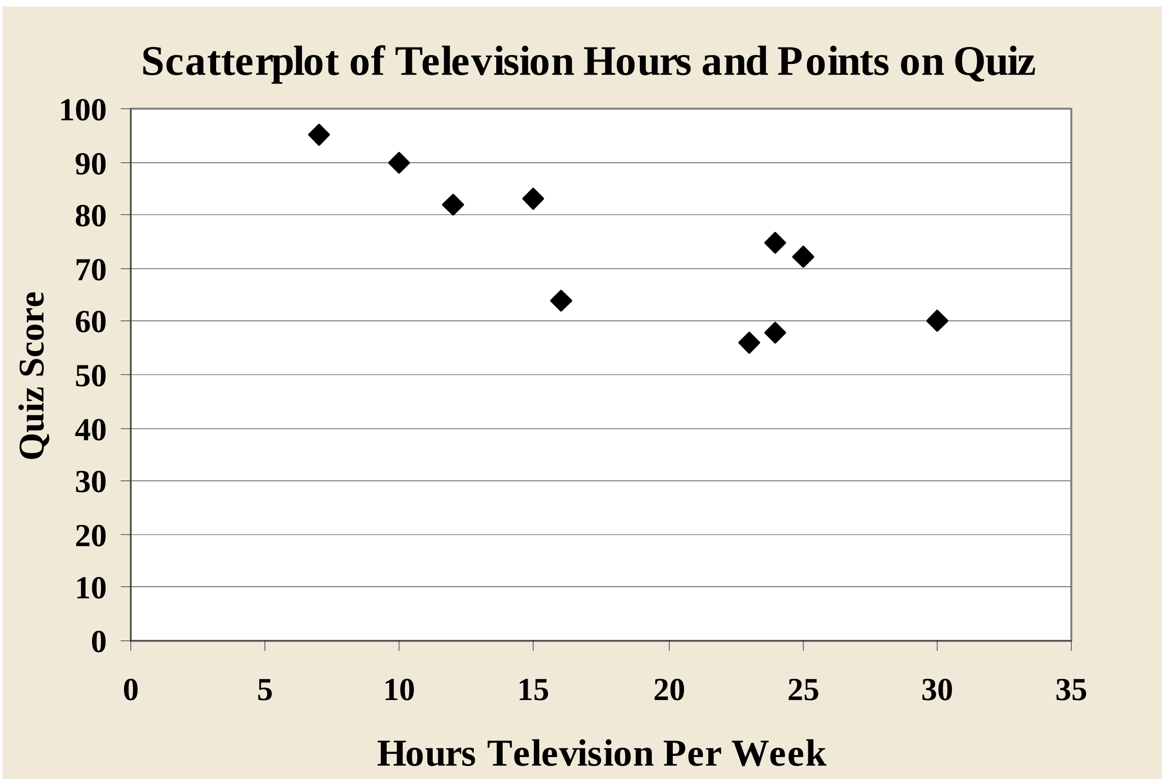
Average mathematics scores by frequency of calculator use for classwork. Grades 4, 8, and 12



	4th Grade	8th Grade	12th Grade
Never/Hardly	235	268	283
Monthly	240	275	286
Weekly	230	276	292
Everyday	210	279	308

SOURCE:
National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Mathematics Assessment.

Transparency 8-8



Transparency 8-9.

Computing Pearson Product Moment Correlation

$$r = \frac{(n \cdot \sum XY) - (\sum X \cdot \sum Y)}{\sqrt{[(n \cdot \sum X^2) - (\sum X)^2] \cdot [(n \cdot \sum Y^2) - (\sum Y)^2]}}$$

$$r = \frac{(10 \cdot 12{,}898) - (186 \cdot 735)}{\sqrt{[(10 \cdot 3{,}980) - 34{,}596] \cdot [(10 \cdot 55{,}743) - 540{,}225)]}}$$

$$r = \frac{(128{,}988) - (136{,}710)}{\sqrt{(39{,}800 - 34{,}596) \cdot (557{,}430 - 540{,}225)}}$$

$$r = \frac{-7{,}722}{\sqrt{5{,}204 \cdot 17{,}205}} = \frac{-7{,}722}{9462.284} = -.816 \qquad r^2 = .667$$

$df = n - 2 \quad = 10 - 2 = 8$