

LAST NAME (Print): **KEY**

Statistics 111 **Quiz 19**

1. You do a regression analysis. The response variable is the average income at age 30 for graduates of a university, and the explanatory variable is the average math SAT score for that university. You sample 16 universities, for which the average math SAT score is 500 with standard deviation 50. STATA says the estimated intercept is \$80K, the estimated slope is -0.02, the coefficient of determination is 0.8, and the standard deviation of the residuals is \$15K.

Hint: Recall that $SS_x = \sum(x_i - \bar{x})^2$ and relate this to the variance.

-0.89 What is the value of the correlation coefficient?

Square root of the coefficient of determination, with sign matching the slope.

True True or false: This is an ecological correlation.

Since each point in the scatterplot is an average for the entire university. We don't know if the high SAT score students are getting the large salaries or the low salaries.

\$66K What is the estimated average income for Duke students, whose average math SAT score is 700?

80 - 0.02 * 700 = 66.

\$104.54K Set a 95% upper bound on the average income for students at Duke.

$$U = 66 + \hat{\sigma}_\epsilon \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_x}} * t_{n-2, 0.95}$$

which is $66 + 15\sqrt{1 + \frac{1}{16} + \frac{200^2}{37500}} * 1.761$.

\$94.07K Set a 95% lower confidence bound on the average income for students at schools whose mean math SAT score is 700.

This is the same formula as above, but without the 1 in the square root.

2. Write the equation of the line that contains (5, 6) and (-1, -2).

High school. Rise over run. $Y = -\frac{2}{3} + \frac{4}{3}X$.

3. What does the coefficient of determination tell you?

The proportion of variance in Y explained by knowing X .

5. Which of the following assumptions are needed for regression? **A, B, C**

- A: The relationship between X and Y is linear.
- B: The errors are independent.
- C: The errors have mean zero.
- D: The X values are independent.
- E: The Y values are measured without error.
- F: The scatterplot is football shaped.