# Regression Review Self-Test

1. What do variance and standard deviation represent?

2. What is a sampling distribution?

3. What is the difference between the standard deviation and standard error?

4. What role does the standard error play in the confidence interval?

   CI = b +/- (SE$_b$)(t-value)

5. What are the definitions of covariance and correlation?

6. What does it mean to say that a regression line is a conditional average?

7. Explain the "intuitive" regression slope as an input, output function.

   Slope = cov(x,y) / var(x)

8. How does the regression line split the variance of Y into the explained and unexplained portion?
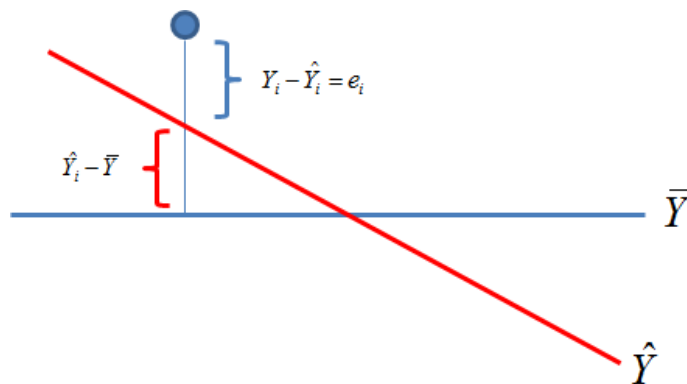
9. What are the three parts of the standard error of the slope?

$$SE_b = \sqrt{\frac{residual}{sampSize \cdot var(x)}}$$

Looking Ahead:

1. In program evaluation, what does it mean to say that a regression slope is an input-output function? What is the input, and what is the output in this metaphor?

2. We have two types of treatments – binary (you either do or do not receive the treatment) and levels (subjects receive different dosages).

   a) What does the regression coefficient represent if subjects receive different dosages?

   b) What does the regression coefficient represent if subjects are in treatment or control groups?

3. What is the null hypothesis in every regression model?

4. The "p" in p-value stands for "probability". The probability of what?

5. What does it mean for program impact (the slope) to be statistically significant? How do we interpret the case when it is NOT statistically significant?

6. What are two ways that omitted variable bias can make a slope appear statistically significant when it is not?