§12.1: Inferences for Regression

Skills

- Construct a confidence interval for the slope of the regression line
- Conduct a test of significance for the slope of the regression line
The Idea

It begins with a question: what line describes the true relationship between these variables?

If we could gather all the data, we'd be able to construct this line...i.e., find the two numbers (true slope, true intercept) needed to write the equation.
Since we can't measure the population, we'll have to settle for a sample.

Every sample is going to produce a different line. Every sample is going to produce two statistics (sample slope, sample intercept). ...so there are really two sampling distributions involved in this!

What's more important: slope, or intercept?
The Parameters: $\alpha, \beta$

The Statistics: $a, b$

The Sampling Distribution of $b$:

Mean: $\mu_b = \beta$

Standard Deviation: $SE_b = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{(n-2)\sum (x_i - \bar{x})^2}}$

(let's not get into its shape just yet)
Under certain circumstances, 

\[ t = \frac{b - \beta}{SE_b} \]

has a \( t \) distribution with \( n - 2 \) degrees of freedom.

About those conditions...
The Conditions

- The data must have a linear relationship
- All observations are independent of one another
- The response variable varies normally about the least squares regression line
- The variation in the response variable is the same at any point along the least squares regression line
Checking the conditions

- Linearity? Look at the data
- Independence? Random Sample
- Normality: Histogram of the residuals
- Constant variation: Residual plot
The Interval

statistic $\pm$ (critical value)(measure of variation)

$b \pm t^* SE_b$

...and all of the usual stuff from there.
The Test

$H_0: \beta = 0$

$H_a: \beta \neq 0$

This calls for a $t$-test for slope.

...and then all the usual stuff.
More about $SE_b$...

- Finding it in computer output
- Tricking the calculator into giving it to you
- The meaning of $s$