Math 244 Lecture Notes

Chapter 25 Day One: Linear Regression Day 2

Overview: Linear Regression is used for examining the relationship between two variables–an **explanatory variable**, x, and a **response variable**, y.

When done correctly, it allows us to make predictions about what should happen given a limited amount of information. Here's what we know so far.

- Our approximation for the line of best fit is $\hat{y} = b_1(x) + b_0$ where b_1 is our slope and b_0 is our y-intercept.
- We calculate the slope using the formula $b_1 = r\frac{s_y}{s_x}$.
- We found the slope by plugging in the point $(\bar{x}, \bar{y})$ and solving for b_0.
- r and R^2 were both calculated using technology. Recall that R^2 is the percentage of variation accounted for by our model.
- Residuals are the error in our predictions. We find these by $e = y - \hat{y} =$ data value $-$ line value.

Much to our surprise, we found that under ideal situations the residuals followed a t-distribution

$$e \approx t_{k-2}(0, s_e)$$

where $s_e^2 = \Sigma\frac{(e-0)^2}{k-2}$.

Unfortunately, error is unavoidable. We are making two big predictions that cause this. We predict a value for the slope b_1 when the actual slope is β_1. We predict a value for the starting value b_0 when the actual starting value is β_0. As a result, we "lost" two degrees of freedom due to the unknown factors. This gave us $df = k - 2$. Both of these predictions are affected by our residual distribution, so both of these unknowns can be analyzed with the t-distribution. The assumptions for these tests include...

- The usual: Random, Independence.
- Linearity: The data should look linear when plotted. Of course.
- Equal Spread: Much like ANOVA. This means that the data stays clumped near the line.
- Normal Population for Errors: We want the errors to look normal. Everything else is based off this. We can either do a histogram of the errors to check for nearly normal OR do a normality plot.

Let's do one final experiment for the term! I want to know if the drop height of a ball in centimeters affects the bounce height of a ball.

Drop Height	Bounce Height
10	
15	
20	
25	
30	
35	
40	
45	
50	
55	
60	
65	
70	
75	
80	
85	

(a) State the Explanatory and Response Variable.

(b) Determine the line of best fit. Using technology is okay.

(c) Calculate R^2. How good is the fit?

(d) **Test for Linearity** Using technology, perform the test for linearity! Recall that this refers to β_1.

(e) Assumption Check:
  (a) The usual:
  (b) Linearity:
  (c) Equal Spread:
  (d) Normal Population for Errors:

We will use a printout from CrunchIt <http://crunchit2.bfwpub.com/crunchit2/ips5e/?section_id=> for the last two components. Fill in the table first. You may be expected to read such tables for the final.

Estimates	t-Stat	P-Value	CI Low	CI High
(Intercept)				
Drop Height				

(f) **Test for y-intercept, β_0.** Using the printout, perform a test for a y-intercept. State your hypotheses, degrees of freedom, model, P-value, and conclusion.

(g) State the 95% CI for the slope, β_1.

(h) State the 95% CI for the y-intercept.

(i) Create the 95% CI for predicted bounce height if we drop the ball from 100 cm.