**7/22/15**

Final Project Announcements:

--Can search for a data set online. Search “multivariate regression data set”.

--Technical report suggested length is 5 pages. Can have charts and stuff but explanation can still be around 5 pages. Nontechnical report is 1-page. In-class ppt.

--At least 10 observations per variable

--Create a training set and test set. 50/50 or 80/20 for training/test split.

Compromise in regression models:

--Needs to be **accurate**

--Needs to be **parsimonious** (talk little with a lot of meaning)

Collinearity

--Means highly correlated variables…**implies you can get rid of one of them/both of them?**

**Review Exercises**

What is a standard normal density?

--Special case when mean is 0 and variance is 1. AKA **N(0,1)**

**More Notes**

Six types of statistical test:

1. One-sample z-test – n >= 30
2. One-sample t-test – n < 30 and sample is normal
3. Paired sample z-test (will not cover)
4. Paired sample t-test
5. Independent two-sample z-test (will not cover)
6. Independent two-sample t-test

\*R and SAS only do t-tests. Something about fatter tails so the p-value is bigger?

**Sample Problem (Textbook p.45)**

--Step 1: State null and alternative hypothesis

--Step 2: Compute the test statistic.

Z = (x-bar – mu)/SEofmean

We use the Standard Error of the mean for this z-statistic formula because we are doing a t-test

--Step 3: Compute confidence interval for Z [-1.96, 1.96]

--Step 4: Reject null hypothesis

--Step 5: P-value is the probability of obtaining a test statistic as extreme or more extreme than the one you actually obtained given that the null hypothesis is true. If the null hypothesis is not true the p-value doesn’t matter? If p-value is less than level of test, reject H0. The closer you are to the p-value then you barely rejected the null hypothesis.

**\*\*How do I do a normal plot in R?**