**STA 6166 – Project 2 – Fall 2016**

**Due Wednesday, October 5**

# Part 1: Airline Demand Effects of 9/11

The dataset **airq4.dat** **(airq4.xls)** contains data on average fare and average weekly passengers for the fourth quarters of years 2000 and 2001 for a **population** of 4177 markets. The file **airq4.txt** contains variable descriptions and file layout. Complete the following steps and write up a brief, but informative summary of results.

* Generate the percent difference (2001q4-2000q4) in total revenue for each market. Note that to obtain total revenue for each quarter, you multiply average fare by average weekly passengers by 13 weeks for each quarter.
* Obtain a histogram of percent difference in total revenues and comment on its shape
* Obtain the mean and variance of this population.
* What size of sample would be needed to estimate this population mean percent within plus or minus 4% with 95% confidence?
* What is the power of testing H0: = 0 versus HA:  < 0 for the true value of  for sample sizes of: n=25,36,64? Use =0.05 significance level
* Generate a random sample of n=50 markets and complete the following parts (use the middle 4 digits of UFID as a seed if prompted):
* Obtain a 95% Confidence Interval for 
* Test H0: = 0 versus HA:  < 0 at =0.05 significance level

# Part 2: Comparison of 2 Yields

A firm wants to compare the yield of two types of animal feed. They wish to determine whether the true mean weight gain from the higher price type is higher than that for the lower price type. The response measured is weight gain (in kilograms) in a 3-month period on the diet. Complete the following steps and write up a brief, but informative summary of results:

* Based on a small pilot study, they estimate the standard deviations of weight gains to be 12 kg. How many animals should be assigned to each diet if they wish the power to detect a difference in means of 4 kg to be .90 (with =0.05)?
* How many animals should be assigned to each diet if they wish the margin of error to be less than or equal to 0.5 kg with 95% confidence?
* The dataset **animfeed.dat (animfeed.xls)** contains simulated data from an experiment with nH=nL=12 animals that were randomly assigned to the 2 diets.
* Obtain a 95% confidence interval for the difference in true mean weight gains (high price – low price). Note that the labels for diet are 1=Low, 2=High.
* Conduct a test of whether the population variances are equal.
* Based on your results of your previous part, use the appropriate t-test to determine whether the true mean weight gain is higher for the high price formulation (=0.05).
* Use the Wilcoxon Rank-Sum Test to test whether the high price formulation is more effective in weight gain (=0.05).

**Part 3 – Comparing Means & Medians – Dependent Samples – LPGA 2003**

The LPGA 2003 dataset has a “sample” of n=27 professional women golfers who completed all 4 rounds of the U.S. Women’s Open and Women’s British Open in 2003. Complete the following parts and determine whether true mean and median scores tended to differ under the two course/environmental conditions. Conduct all tests at  = 0.05 significance level.

* Conduct the paired t-test, clearly stating null and alternative hypotheses, test statistic, rejection region, and P-value.
* Conduct the Wilcoxon Signed Rank-Test, clearly stating null and alternative hypotheses, and P-Values (different software packages use different “normalizations” of the test statistic).
* Plot the scores of the U.S. Open versus British Open. Is there evidence of an association between scores.