Math 222 - Project 5 **Due Friday, March 2**

1. In 1995, the National Highway System Designation Act abolished the federal mandate of 55 miles per hour maximum speed limit and allowed states to establish their own limits. Of the 50 states (plus District of Columbia), 32 increased their speed limits in 1996. The data in TrafficFatalities.txt shows the percentage change in interstate highway traffic fatalities from 1995 to 1996 and whether or not the state increased their speed limit. (Data from the National Highway Traffic Safety Administration as reported in Ramsey and Schafer, 2002.)
   (a) Identify the individuals and response variable of interest. Is this a randomized experiment or an observational study?
   (b) Produce numerical and graphical summaries of these data and describe how the two groups compare. Does it look like raising the speed limit caused more people to die?
   (c) Are the technical conditions for a two-sample t-test met for this study? Explain.
   (d) Carry out a two-sample t-test to determine whether the average percentage change in interstate highway traffic fatalities is significantly higher in states that increased their speed limit. If you find a significant difference, estimate its magnitude with a confidence interval.
   (e) Discuss what the p-value in (d) measures.

2. Repeat the significance test in the previous problem, but this time use a permutation test instead of a two-sample t-test. Be sure to explain how your permutation test simulation works. How close are the p-values for the two methods?

3. How powerful was the t-test in problem 1? Compute its statistical power if the true difference in population means was only 10 percentage points (more fatalities with the higher speed limits). Note: for a 2-sample t-test, computing the power requires you to use a non-central t-distribution with non-centrality parameter:

$$\delta = \frac{\mu_1 - \mu_2}{\sqrt{\left(\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right)}}.$$

   Since we cannot know the true values of σ_1 and σ_2, use the sample standard deviations as a reasonable guess.