Stat 500 **Lab 4**

1. For a school statistics poster competition in 2006, students timed 15 randomly selected teenagers from the school and 15 randomly selected staff from the school over the age of 30 on how long it took each person to text the following sentence on their phone: "the quick brown fox jumps over the lazy dog." Each subject had the sentence in front of them while they were typing. The text message had to be typed with no errors, no abbreviations, and no use of the phone directory. Time was measured using a stop watch to within 0.01 seconds. Participants were timed using two phones – their own phone and a "control" phone, which was the same for all participants. The data are located in the file **smsspeed.csv** and the SAS code for the analyses below can be found in the file **smsspeed.sas**.

We would like to determine if teenagers were faster "texters", on average, than adults.

a. Using the formula from the notes, calculate a 95% confidence interval for the difference in the two treatment means. Use $t_{28,0.025} = 2.0484$. Verify your calculation with the values on the SAS output.

b. Give an interpretation of the confidence interval you calculated in part (f).

2. In lecture notes, we have an example of a randomized experiment to determine which of two treatments was the most effective at reducing bone loss in elderly women. In this experiment, we will assume equal sample sizes, equal population variances, and normally distributed response variables in both samples. We will also assume an estimate of the pooled sample variance for the response variable is available from previous studies, denoted as S_p^2.

a. Suppose our research question is to determine whether or not the two treatment means are different. We will use a Type I error rate of α and will want the power to detect a difference of δ units between the treatment means to be $1-\beta$.

In lecture, our example set $\alpha = 0.05, 1-\beta = 0.8, \delta = 4, S_p^2 = 25$. The code that produced the result of 26 subjects in each treatment group is given below. Copy and paste this code into a SAS program window and run the code.

```
proc power;
      twosamplemeans
      meandiff = 4
      alpha = 0.05
      power = 0.80;
      stddev= 5
      npergroup = .
run;
```

Then, use SAS to help you determine the effect of changes to the values of α, $1-\beta$, δ, and S_p^2 on the sample size. Write a summary of your findings. To make it easier to study these changes, you can modify the SAS code to study the sample size for multiple values of an input value at the same time. For example, to study the effect of increasing power $1-\beta$, you can change the power command to

```
power = 0.80 to 0.95 by 0.05;
```

or you can list values to study, like

```
power = 0.80, 0.9, 0.95, 0.99;
```

b. You can also calculate power for multiple values of sample sizes. For example, to study the effect of increasing sample size, you can change the power command to

```
proc power;
        twosamplemeans
        meandiff=4
        power= .
        stddev=5
        npergroup= 10 to 100 by 5 ;
        plot x = n;    /* plot y = power vs x = sample size */
run;
```

c. Instead of the analysis above, suppose our research question is to estimate the difference between the two treatment group means. We would like to estimate this difference with $100(1-\alpha)\%$ confidence with width of no more than δ units.

   In lecture, our example set $\alpha = 0.05, \delta = 4, S_p^2 = 25$. From the calculation, we obtained a sample size of 50 from each sample. The code that produced the result of 50 subjects in each treatment group is given below. Copy and paste this code into a SAS program window and run the code.

```
proc power;
       twosamplemeans
       alpha = 0.05
       meandiff = 4.0
       stddev= 5
       npergroup = .
       power = 0.975;
run;
```

   Why should we use power = 0.975?

   Use SAS to determine the effect of changes to the values of α, δ, and S_p^2 on the sample size. Note: the value of power should be set to $1 - {}^{\alpha}/_{2}$, why? Write a summary of your findings.

3. In homework 1, we look at data taken from an experiment used to determine the best variety of corn for a particular field. In this experiment, two corn varieties were randomly assigned to 36 plots (18 plots per variety) in order to determine which variety produced the highest mean yield. The pooled sample standard deviation from our experimental data is $S_p = 9.18$.

a. Use the SAS code below to calculate the power of this hypothesis test to detect a difference in the mean yields between the two treatments of 12 units.

```
proc power;
       twosamplemeans
       alpha = 0.05
       meandiff = 12.0
       stddev= 9.18
       npergroup = 18
       power = .;
run;
```

b. Change the code from part (a) to determine the power for this hypothesis test when the difference in the mean yields between the two treatments decreases from 12 to 10.
c. Repeat part (b), except use a significance level for the test of $\alpha = 0.1$.
d. Repeat part (b), except use a significance level for the test of $\alpha = 0.01$.
e. Write a statement summarizing your findings from parts b, c, and d.