Name _____   STT 3850 Test 2

Directions: Answer each of the following problems directly and completely.   If the question is not asking for an explanation, don't put an explanation.   Try not to spend more than 2 minutes per question, on average and not more than 30 minutes for Part A.

Part A. (60 points) A researcher randomly assigns four 3 ft. by 3 ft. agricultural plots to two conditions (with sunlight and without sunlight), planted 10 seeds of a certain plant species on each of the plots, then after a month counted how many seeds germinated. He obtained the following data on each plot and the count of seeds that germinated on each plot:

Plots With Sunlight: Plot 1   Plot 4

                                    10   8

Plots Without Sunlight: Plot 2   Plot 3

                                     4   6

1. a.   Is this study an observational study?

   b.   Based on the number and type of variables used in the study, what design was used in this study?
        Based on how the subjects were assigned to the treatments, what design was used in this study?

   c.   Identify exactly the experimental units in this study and how many?

   d.   Identify the factor and the response variable in the study.

   e.   Suppose the soil type is a potential confounding variable in this study, what design (based on how the subjects are assigned to the treatments) should the researcher use to eliminate the confounding effect of this variable on the results of the study?

2. To determine if sunlight have a positive effect on seed germination in the study on the previous page, the researcher needs to conduct a hypothesis testing procedure.

a. What are the null and alternative hypotheses for this study?

Ho:

Ha:

b. Calculate the value of the appropriate test statistic for this study.

c. Obtain the exact permutation reference distribution of the test statistic in part b if there is really no difference between the two treatments by completing the table below. How many possible treatment allocations can be made if 2 plots will be assigned to each treatment?

With Sunlight	Without Sunlight	Test Statistic Value
10 8	4 6	

d. Calculate the p-value of the test based on the exact reference distribution under H_0.

e. If the level of significance is set at $\alpha = 0.2$, interpret the p-value and state the conclusion of the test in context of the study.

3. Another researcher realized that the sample sizes in the study on the first page were too small. He repeated the study but this time used 200 plots, recorded if more than half of the seeds planted on each plot have germinated (Y) or not (N) and obtained the following data summary:

   With Sunlight: 86 out of 100 plots have more than half of the seed planted that have germinated
   Without Sunlight: 64 out of 100 plots have more than half of the seed planted that have germinated

   a. Calculate the observed value of the appropriate test statistic for testing if sunlight have positive effect on seed germination in this study.

   b. What are the values of the mean and standard deviation of the approximate reference distribution of the test statistic in part a if there is really no difference between the two treatments?

   c. If there is no difference between the two treatments, what is the probability that the test statistic is equal to or more extreme than the observed value in part a using the reference distribution in part b ?

   d. If the level of significance is set at $\alpha = 0.1$, what is the conclusion of the study based on the probability obtained in part c?

Part B. (40 points) Start R Studio and enter the command **attach(mtcars)** on the R console. This will allow you to access the contents of the data frame **mtcars**. Enter the command **help(mtcars)** to get a description of this data frame then answer the following questions:

1. Using the indexing method, write the R command(s) that will find all possible combinations of 3 cars from the data frame (no need to write the result of executing the commands). How many possible combinations can be made?

2. Using the indexing method, write the R command(s) needed to randomly assign each of the cars to one of two treatments (no need to write the result of executing the commands). How many cars were assigned to each of the two treatments?

3. Using the **mpg** data, perform an appropriate test procedure using the approximate reference distribution of the test statistic under H_0 to test if the mean miles per gallon (**mpg**) of cars is higher than 18. Calculate the value of the appropriate test statistic, state and interpret the p-value of the test and state the conclusion of the test procedure if the level of significance is 0.1.

4. What is the critical value of the test statistic in #3 if the level of significance is 0.1?