2028: Basic Statistical Methods
Homework 6

This homework is due Monday, Nov 11th in class **BEFORE class starts**. Late papers will not be accepted. Please do not turn in any papers to any mailbox.

- Please remember to staple if you turn in more than one page.
- Please make sure to **SHOW ALL WORK** in order to receive full credit.

I Comparing Two Population Means

1 Textbook 10-30 (a-d) page 372-373

2 Textbook 10-77 page 396

3 A random sample of $n = 16$ 1-kilogram sugar packets of brand A have weights with a sample mean $\bar{x} = 1.053$ kg and a sample standard deviation $s_x = 0.058$ kg. In addition, a random sample of $m = 16$ 1-kilogram sugar packets of brand B have weights with a sample mean $\bar{y} = 1.071$ kg and a sample standard deviation $s_y = 0.062$ kg. It is safe to conclude that brand B sugar packets weigh slightly more on average than brand A sugar packets?

II Comparing Two Proportion Parameters

1 Textbook 10-70 page 394-395

2 Textbook 10-82 page 397

III Computer Problem: Friday of 13th

Is Friday the 13th an unusually unlucky day, or is this just superstition? How do superstitions affect people's behavior? These questions were addressed by researchers Scanlon, et al. (1993) in a study that examined the relationship between behavior and superstition in the United Kingdom. They analyzed shopping and traffic patterns, as well as the numbers and types of accidents that occurred on past Friday the 13th's. The study, conducted in England, focused on two questions: 1) How do superstitions regarding Friday the 13th affect human behavior?, and 2) Is Friday the 13th more unlucky than other Fridays?

The first question was addressed by comparing traffic patterns and shopping behaviors on all Friday the 13th's between July of 1990 and November of 1992 to the patterns and behaviors on the Friday of the week before. The second question was addressed in a similar fashion using the numbers of emergency admissions to hospitals due to transport accidents from October of 1989 to November of 1992.

The variable names in the data file are:

1. Dataset: Identifies source dataset (traffic, shopping, or accident)
2. Dates: year and month in which the Friday the 13th occurred

3. 6th: Number of cars passing through junction (traffic dataset), shoppers for each supermarket (shopping dataset), or admissions due to transport accidents (accident dataset) on Friday the 6th

4. 13th: Number of cars passing through junction (traffic dataset), shoppers for each supermarket (shopping dataset), or admissions due to transport accidents (accident dataset) on Friday the 13th

5. Location: Motorway junction (traffic dataset), supermarket location (shopping dataset) or hospital (accident dataset) to which the data correspond

Reference: Scanlon, T.J., Luben, R.N., Scanlon, F.L., Singleton, N. (1993), "Is Friday the 13th Bad For Your Health?," *BMJ*, 307, 1584-1586.

Instructions for reading the data. The data are available on T-square (Assigments/Homeework 6) and the text file name is `Friday13.txt`. To read the data in R, save the file in your working directory (make sure you have changed the directory if different from the R working directory) and read the data using the R function `read.table`.

```
data = read.table("Friday13.txt",header=TRUE,sep="\t")
```

In order to test whether superstitions regarding Friday the 13th affect traffic and shopping patterns, we will compare the traffic data and shopping data for the two days when these data were recorded: Friday of 6th and Friday of 13th. We will have to write these data in separate vectors for analysis.

```
source = as.character(data[,1])
traffic.data.6th = as.numeric(data[source=="traffic",3])
traffic.data.13th = as.numeric(data[source=="traffic",4])
```

Using similar R commands, extract the shopping data for each of the two days (6th and 13th) and define the two vectors as `shopping.data.6th` and `shopping.data.13th`.

Question 1. Does the traffic data and shopping data look normally distributed? You may check normality using the histogram and boxplot graphical displays.

Question 2. In order to compare the traffic and shopping patterns for the two days, we will use a t-test using the `t.test` function in R. However, this approach should be used with caution. Why?

The `t.test` function in R needs the following inputs:

```
t.test(x,y,alternative,paired,conf.level)
```

where x and y are the vectors of data that are compared (for example, the traffic data from 6th will be x and the traffic data from 13th will be y); `alternative` with three different options, "two.sided", "less" or "greater"; `paired` which is TRUE when the test is paired and FALSE when unpaired; and `conf.level` which could be 0.95 (the default), 0.99, etc.

Do these data support the hypothesis that there is less traffic on Friday 13th than on Friday 6th? Do these data support the hypothesis that there is less intensive shopping on Friday 13th than on Friday 6th?

Note: Make sure you write down the command line that you used to derive p-values and provide the R output.

Question 3. Are the data representative for all Fridays 13th? If not, how would you sample to obtain a representative sample?

Question 4. What are your final conclusions?