Assignment 8

Due Wednesday May 6

1. A pastor expects hi parishioners to spend some time each week in a Bible study. He randomly selects 30 parishioners from among those coming to church one Sunday and asks each the number of minutes spent reading the bible during the pervious week. Using this data, the pastor's secretary, who once took a college statistics class, calculates the following:

$$n = 30, \quad \overline{x} = 31.2, \quad s = 13.2, \quad \text{Min} = 13.5$$

$$Q_1 = 18.7, \quad \text{Med} = 27, \quad Q_3 = 41.2, \quad \text{Max} = 57$$

The pastor asks the secretary to compute a confidence interval estimate of the mean number of minutes parishioners read the bible weekly. Comment on this study, including all necessary assumptions and how well they are me. (You are not asked to calculate the confidence interval).

- There is a real problem with possible response bias in that parishoners may not answer the pastor honestly.
- The assumption of a random sample from the population of all parishoners is not met as the sample was taken from those coming to church one Sunday, and they may not be representative of all parishoners.
- for normality the sample is n=30, which is fairly large for CLT. However, it should be noted that the sample distribution does not appear symmetric.

2. A high school math department conducts a study to determine whether a classroom with windows leads to higher exam scores than a classroom without any windows. Two algebra classes are scheduled, each with 25 students. It is randomly decided which class will use which classroom. During the year, each teacher administers the same exams. At the end of the academic year, overall exam grades are compared.

(a) Identify the response variable, the treatments, and the experimental units.

The response variable is the exam scores. The treatments are the two classrooms. The experimental units are the two algebra classes.
(not the students themselves!)

(b) Was randomization properly used? Explain.

· Yes, the two classrooms (the treatments) were randomly assigned to the two algebra classes.

(c) was replication properly used? Explain

No, there was no replication in this study. Each treatment was applied to only one experimental unit.

(d) Teacher is a confounding variable. Explain.

If the difference in exam scores is noted between the two classes, it is not known if this is due to the difference in classrooms or the difference in teachers. That is, the treatments (classrooms) are confounded with teachers.

3. A popular office chain has 11,500 stores in the U.S. and 13,500 stores outside the U.S. Each store has approximately 10 full-time employees. Suppose the company would like to survey 250 of its employees about new coffee drinks under consideration. Under discussion are three sampling methods.

   1. A simple random sample
   2. Randomly pick 125 of their 115000 U.S. full-time employees and randomly pick 125 of their 135000 non-U.S. full-time employees
   3. Randomly pick 25 of their 25000 stores and pick all full-time employees at these 25 stores

(a) Give a design for carrying out method 1, and give a disadvantage of using method 1.

1. Number employees 1 through 250 000.
2. Use random number generator to pick integers between 1 and 250 000, throwing out repeats, until 250 unique integers have been picked.
3. The sample will be the 250 employees whose numbers correspond to the picked integers. A disadvantage would be the difficulty of listing and numbering 250000 employees.

(b) What is method 2 called? What is an advantage of method 2 in this context?

Stratefied Random Sampling

- Stratefied sample where U.S. and non-U.S. employees are two strata.
- An advantage of Stratefied sampling here is this method garuntees good numbers of both U.S. and Non-U.S. employees will be in the sample. We would like both groups to be well represented because they might have different views about new coffee drinks that their clientele will enjoy and purchase.

(c) What is method 3 called? What is an advantage of method 3 in this context?

- Method 3 is a cluster random sample, where the Stores are the 25 clusters.
- An advantage of cluster random sampling here is the ease and low cost in implimenting the method. Employees in the company do not have to be listed. The research team only has to pick the 25 stores randomly and then simply use all full-time employees in those stores for their sample.

4. A reading specialist plans a study to determine if high school students read faster from a physical book or an iPad. The reading specialist randomly selects 60 high school students for the study.

   (a) Describe a randomization process and an inference procedure for the study to be conducted with a completely randomized design.

1. Number the students from 1 to 60.
2. Use a random number generator to pick numbers between 1 and 60, throwing out repeats, until a set of 30 unique numbers has been selected.
3. The students corresponding to these numbers read from a book, while the remaining group of 30 students read the same passages on an iPad for the same predetermined length of time.
4. The mean number of words read by each of the two groups can be compared using a two-sample t-test for the difference between two means or by using a confidence interval for the difference between two means.

   (b) Describe a randomization process and an inference procedure for the study to be conducted with a matched pairs design.

- Randomly assign the 60 students do two treatment groups as above. The first group reads from a physical book, while the remaining 30 students reads the same passage on an iPad for the same predetermined length of time.
- Next, the opposite treatment should be applied to each student.
- Again all students read the same set of passages for the same predetermined length of time.
- The difference in number of words between treatments is recorded for each student.
- Finally, use a one-sample t-test on this set of differences. Alternatively, determine a confidence interval of this set of differences and see if 0 is in the interval.

note:
$H_0: M_d = 0$
$H_a: M_d \neq 0$