# HW 1

1. What is the difference between a randomized experiment and a random sample? Under what type of study/sample can a causal inference be made?
   **In a randomized experiment, a random mechanism is used to allocate the available subjects to treatment groups.**
   **In a random sample, a random mechanism is used to select subjects from the population of interest.**
   **A causal inference can be made in a randomized experiment study.**

2. In 1936, the *Literary Digest* polled 1 out of every 4 Americans and concluded that Alfred Landon would win the presidential election in a landon-slide. Of course, history turned out dramatically different (see http://historymatters.gmu.edu/d/5168/ for further details). The magazine combined three sampling sources: subscribers to its magazine, phone number records, and automobile registration records. Comment on the desired population of interest of the survey and what population the magazine actually drew from.
   **The Literary Digest used a sample in which one in every five voters was polled, but only used subscribers to their magazine. Their subscribers were more affluent than the average American since they owned cars and telephones , so they didn't get the lower income people who couldn't afford subscriptions to their magazine so it excluded them from this poll.**

**3.** Suppose we have developed a new fertilizer that is supposed to help corn yields. This fertilizer is so potent that a small vial of it sprayed over an entire field is a sufficient dose. We find that the new fertilizer results in an average yield of 60 more bushels over the old fertilizer with a p-value of 0.0001. Write up a scope of inference under the following study designs that generated this data.

   **a.** We offer the new fertilizer at a discount to customers who have purchased the old fertilizer along with a survey for them to fill out. Some farmers send in the survey after the growing season, reporting their crop yield. From our records, we know which of these farmers used the new fertilizer and which used the old one.
   **Is it a Random Sample?**
   **No since the farmers are not randomly selected to get the new fertilizer, they self-select into the study when they fill out the survey and hence no inference to the population and results are confined only to this group.**
   **Is this Randomly assignment?**
   **No Here we don't randomize the assignment of 'old' and 'new' fertilizer and hence no casual inference can be made. Hence casual inference can't be established and results can't be generalized.**

   b. When a customer makes an order, we randomly send them either the old or new fertilizer. At the end of the season, some of the farmers send us a report of their yield. Again, from our records, we know which of these farmers used the new fertilizer and which used the old.
   **Is this a random sample?**

**Yes since the old or new fertilizers are sent randomly to customer who placed the orders these results can be generalized (or inferences) to the population (All farmers).**

**Is this a random assignment?**

**No since only some of the farmers send the yield report, it is self-selecting themselves for the study.Hence casual inference can't be made and but the results can be generalized**.

**c.** When a customer makes an order, we randomly send them either the old or new fertilizer. At the end of the season, we sub-select from the fertilizer orders and send a team out to count those farmers' crop yields.

**Is this a random sample?**

**Yes since the old or new fertilizers are sent randomly to customer who place the order these results can be generalized (or inferences) to the population (All farmers).**

**Is this a random assignment?**

**Yes Since we sub-select from the orders we are randomly assigning the fertilizer groups to the farmers. Hence casual inference can be made and results can be generalized.**

**d.** We offer the new fertilizer at a discount to customers who have purchased the old fertilizer. At the end of the season, we sub-select from the fertilizer orders and send a team out to count those farmers' crop yields. From our records, we know which of these farmers used the new fertilizer and which used the old one.

**Is this a random sample?**

**No since the farmers are not randomly selected to get the new fertilizer.**

**Is this a random sample**

**Yes Since we sub-select from the orders we are randomly assigning the fertilizer groups to the farmers. the results can be generalized, but causality can't be established**

4. A Business Stats class here at SMU was polled, and students were asked how much money (cash) they had in their pockets at that very moment. The idea was to see if there was evidence that those in charge of the vending machines should include the expensive bill / coin acceptor or if the machines should just have the credit card reader. Also, a professor from Seattle University polled her class last year with the same question. Below are the results of the polls.

**SMU**

34, 1200, 23, 50, 60, 50, 0, 0, 30, 89, 0, 300, 400, 20, 10, 0

**Seattle U**

20, 10, 5, 0, 30, 50, 0, 100, 110, 0, 40, 10, 3, 0

a. Use SAS to make a histogram of the amount of money in a student's pocket from each school. Does it appear there is any difference in ***population*** means? What evidence do you have? Discuss your thoughts.

**The class wasn't randomly asked about how much money was in their pockets since they were polled instead. The means difference between the two colleges is significant but can't be applied to the population of the students at each university.**

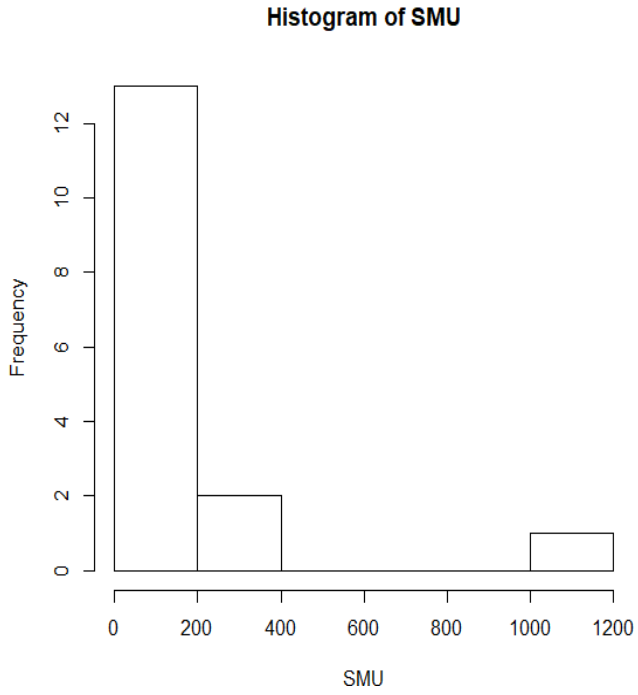
b. Use the following R code to reproduce your histograms. Simply cut and paste the histograms into your HW.

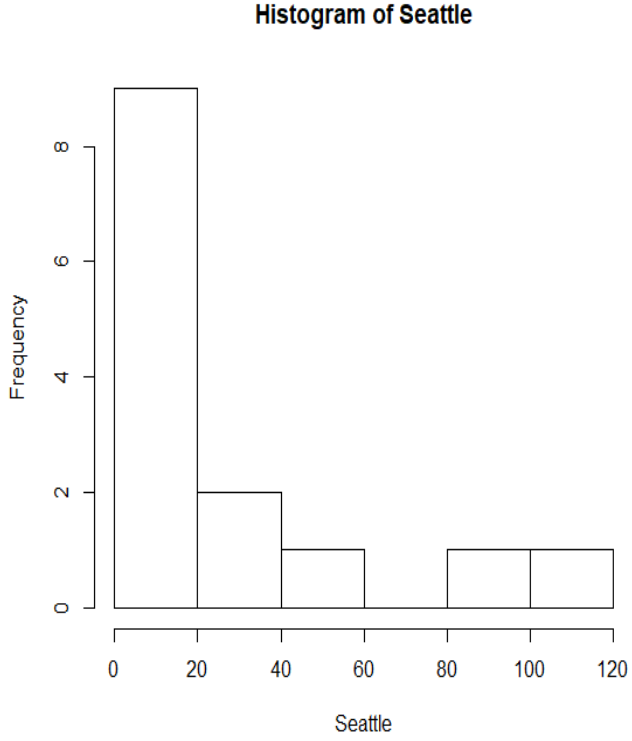
***SMU = c(34, 1200, 23, 50, 60, 50, 0, 0, 30, 89, 0, 300, 400, 20, 10, 0)***

***Seattle = c(20, 10, 5, 0, 30, 50, 0, 100, 110, 0, 40, 10, 3, 0)***

***hist(SMU)***

***hist(Seattle)***




c. Run a permutation test to test if the mean amount of pocket cash from students at SMU is different than that of students from Seattle University. Write up a statistical conclusion and scope of inference (similar to the one from the PowerPoint). (This should include identifying the Ho and Ha as well as the p-value.)

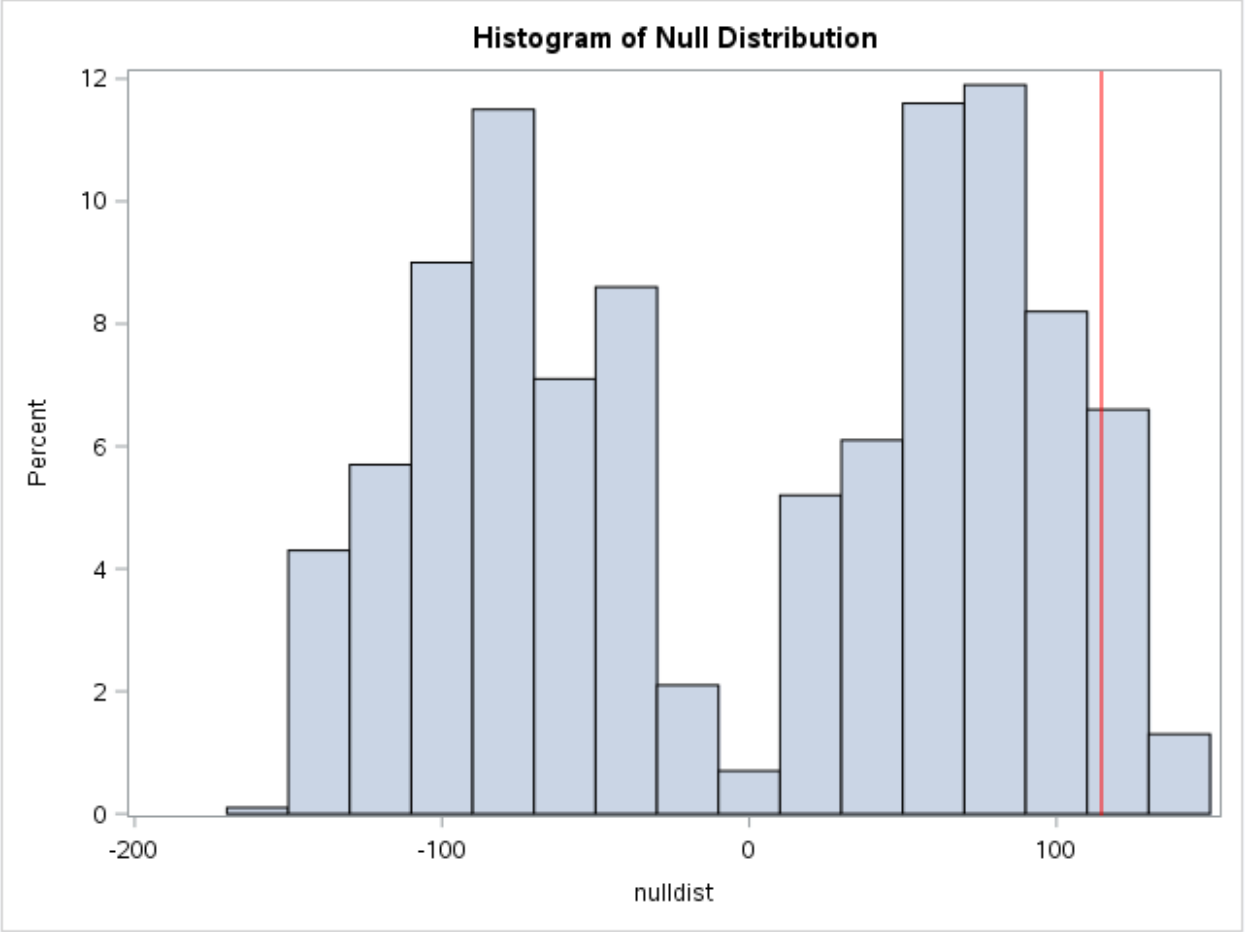
The null hypothesis is $H_o$: µSMU- µSEATTLE=0

The alternate hypothesis is $H_{a:}$ µSMU-µSEATTLE≠0

The mean for SMU was 141.6 and SEATTLE 27
Ran a program with 1000 permutations and came up with a p-value of 0.1538. We fail to reject the hypothesis. There is not sufficient evidence to suggest at the students at SMU is different than that of students from Seattle. Since this is not a random sample of the students we can't generalize this inference to all the students in the university.

Used https://blogs.sas.com/content/iml/2014/11/21/resampling-in-sas.html as a guide to run permutations.

| obsdiff |
| --- |
| 114.625 |




Histogram of Null Distribution

| pval |
| --- |
| 0.1538462 |