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?
   1. Random experiments can lead to causal statements. Observational studies may include confounding variables and thus cannot lead to causal statements. Randomization allows for impersonal chance and allows for multiple treatment groups to be studied.

1. 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.
   1. This survey did not choose a group that was representative of the population of the US in 1936. Only the upper class would have had cars and a telephone and subscribed to the magazine. Given this the sample set chosen was skewed. Just because a large sampling of over 2 million subscribers was chosen is not a sufficient argument that this poll was conclusive if it is not chosen at random from the full population. Most polls can be much more accurate with far less data collected.
2. 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.
   1. 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.

The Farmers volunteer for the study, so this study cannot be generalized and therefore are only representative of this sample set. So, it is not a Random Sample, nor is it Randomly Assigned. (Random Sample = No, Random Assignment = No)

* 1. 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.

The farmer that gets the new fertilizer in this study has been randomly assigned, but he is not randomly assigned since they will be selected from the current list. (Random Sample = Yes, Random Assignment = No)

* 1. 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.

The farmer will be selected randomly and then at the end the participants will be grouped according to old and new formula and then the results will be verified by an independent group. This study will provide the best comparisons of the old and new formula while minimizing any confounding variables. (Random Sample = Yes, Random Assignment = Yes)

* 1. 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.

The farmer will volunteer and then at the end the participants will be grouped according to old and new formula and then the results will be verified by an independent group. This study will be of no use since it will not be representative of the population. (Random Sample = No, Random Assignment = Yes)

1. 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

* + - * 1. 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.

|  |
| --- |
| SMU |
|  |
| Seattle University |
|  |

Just looking at the histograms alone, it would appear that the median is approx. equal between the SMU and Seattle Univ students.

* + - * 1. 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)***

|  |
| --- |
| SMU |
|  |
| Seattle University |
|  |

* + - * 1. 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.)

|  |
| --- |
| Distribution of the mean |
| ../../../Screen%20Shot%202018-05-13%20at%202.40.10%20PM.png |
|  |
| ../../../Screen%20Shot%202018-05-13%20at%202.41.25%20PM.png |
| … |
| ../../../Screen%20Shot%202018-05-13%20at%202.41.57%20PM.png |
| There is not a strong evidence that the mean scores of students between these two schools are different in any statistical meaning. With a p-value of 0.18. is way larger than an expected 95% or 1-.95 = 0.05. This shows that having outliers in the data sampling may disturb the mean and stdev between the two sample groups, that the permutation test can indeed help researchers determine a proper probability of causation. |