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?
   * ***A random sample is a collection of subjects from a population where each is chosen at random (by assigning numbers or flipping a coin). We can assume that by choosing the subjects randomly; variables that we are not testing but may influence the test (confounding variables) will be represented evenly.***
   * ***A randomized experiment uses random samples to mitigate the influence of confounding variables that may influence the outcome.***
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.
   * ***If I were conducting this poll the population of interest would be the entire voting population of the United States.***
   * ***By using subscribers, the poll may under represent people that do not read magazines. Phone records and automobile registrations may under represent lower income voters. These confounding variables could be mitigated by choosing a random sample representative of the entire population.***

***If truly randomized the sample could probably be much smaller than ¼ of the population and provide more accurate predictions than the actual poll.***

1. 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.
   * 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.
   * ***I would not infer any outcome from this experiment. Confounding variables are introduced by offering a discount. Because the farmers are aware that this is a new fertilizer they may be influenced or biased resulting in changes in their farming behavior; for example, monitoring the crop more closely.***
   * 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.
   * ***Because the participation is voluntary the result may be skewed. (Confounding variables)***
   * 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.
   * ***If the selection is randomized and both the new and old fertilizer orders are evenly represented we may draw inference from this experiment.***
   * ***I would prefer to see this experiment conducted over several seasons. Weather patterns may influence the behavior of both fertilizers; for example, one fertilizer may be more reactive to moisture or temperature than the other.***
   * 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.
   * ***Confounding variables are introduced by offering a discount.***
2. 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.

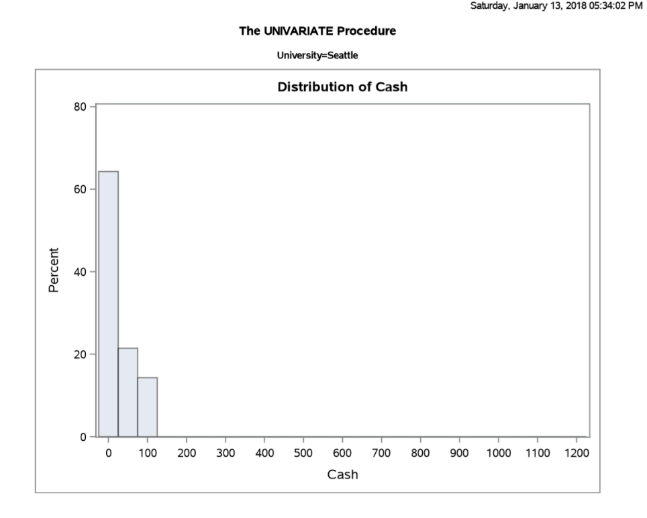
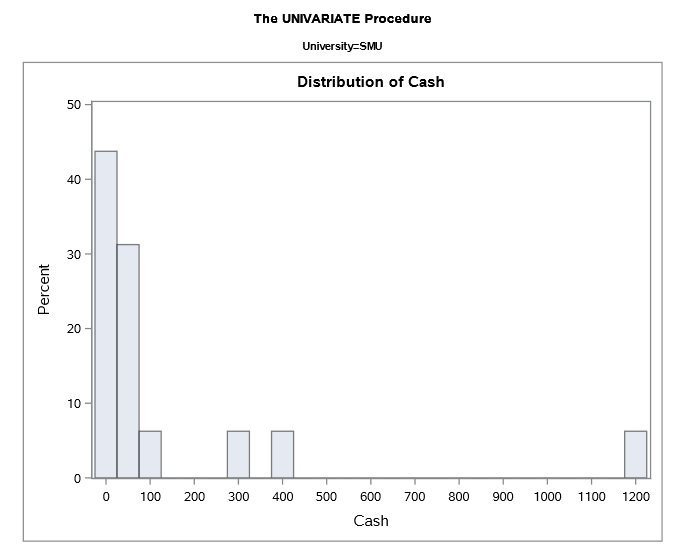
**SMU**

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.
* ***We cannot determine if there is a difference in population means because this data is only from stats students. Also, the data from Seattle is stale; changes may have occurred since last year; SMU and Seattle may have different demographics (confounding variables). Students are volunteering the information; it doesn’t seem that we are asking them to show the cash but tell us how much they have.***
* ***We can usually say that inference can be drawn only about the samples in this case because the students are volunteering the information. I wouldn’t draw inference even about these 2 samples since there are significant extreme values.***
* ***Our sample data does have a difference in the means. The SMU mean is higher than the Seattle mean. There are some extreme values in the SMU sample (right skewed) which is suspicious.***



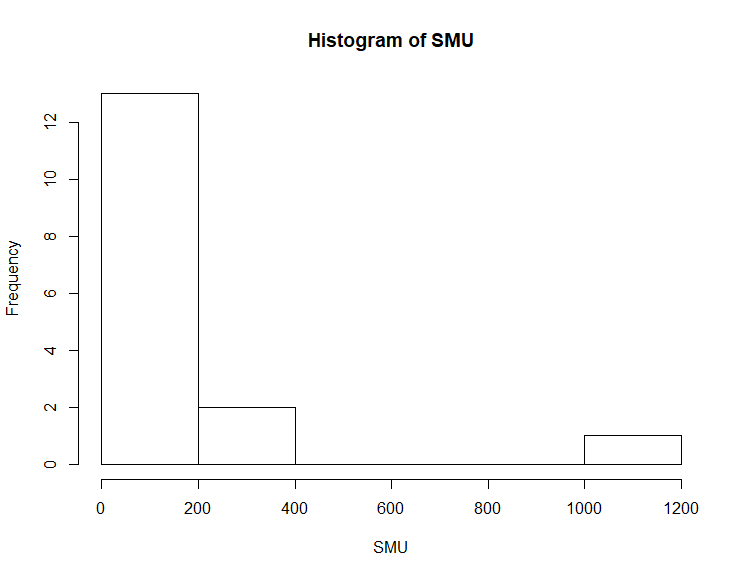
* + - * 1. Use the following R code to reproduce your histograms. Simply cut and paste the histograms into your HW.

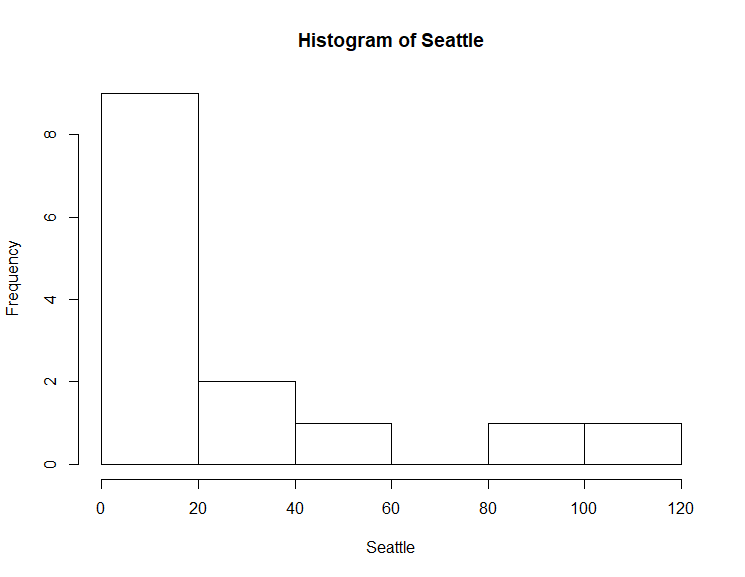
***SMU = c(34, 1200, 23, 50, 60, 50, 0, 0, 30, 89, 0, 300, 400, 20, 10, 0)***

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

***hist(SMU)***

***hist(Seattle)***





* + - * 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.)
    - ***Ho Mean SMU = Seattle, Ha Mean SMU <> Seattle***

******

* + - ***There is not enough evidence to conclude that there is a difference between the cash carried SMU students is different than the Seattle students with 95% confidence (p .149 from permutation test)***
    - ***Inference cannot be drawn because we fail to reject the null hypothesis and confounding variables.***