Ben Goodwin

MSDS6371-HW1

1. To begin, I think its important to note that cause-and-effect relationships can be inferred from randomized experiments, but not from observational studies. Confounding variables may interfere with observational studies, sometimes being the reason for observed differences. In randomized experiments, we do not have the same issue, they mitigate this by ensuring that differences between groups are due to chance alone. When we have randomization, we have a firm basis for drawing inferences. A random sample of size *n* is a subset of a population obtained by a procedure giving all set of n distinct items in the population an equal chance of being chosen. A random sample deals with the methodologies behind selecting a representative population, and a random experiment uses a random sample to draw inference.
2. This is an interesting article, because I believe the authors had been lucky with presidential elections up until the Landon one. The magazine claimed they had access to 1 in 5 American voters, and this is an impressive feat. However, I think that the magazine fails to realize that their subscriber base may not have been a representative sample of the greater population. Assuming they sampled a homogenous population, they had at best 20% of the opinion of the votes. This is a huge problem with opt in surveys, they sometimes only capture what you want to hear!
3. A) The population of interest in part A are the customers who have purchased the old fertilizer, and some of their crop yields, and for those who completed the survey we can comment on yields

B) The population of interest in part B, we are concerned with crop yields, since we have either tied them to the old or new fertilizer.

C) The population of interest in part C, we are again concerned with crop yields, but this time instead of relying on self-reporting of the farmers, we are relying on a team to measure the farmers’ crop yields.

D) The population of interest in part D, we are concerned with the farmers who have previously purchased the old fertilizer and the difference between the crop yields of the old fertilizer and the new one.

1. A screenshot of a social media post

   Description automatically generatedA screenshot of a social media post

   Description automatically generated

From the SAS histograms, the initial EDA seems to indicate that there is a difference in means. I think that the visual variance in Seattle U and SMU indicates that there will be a difference in means. Additionally, I believe the student carrying $1200 will have an impact on this outcome.

4b. A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

5C) Since the subjects cannot be randomly assigned to be SMU students or Seattle U students, this is an observational study, and thus the difference in mean scores is only associated with the SMU/Seattle U status. We cannot tell if school caused the difference or not.

The sample was not a random sample, and therefore, these findings cannot be generalized to all SMU and Seattle U students. In conclusion, it cannot be inferred that the mean amount of money in each students pocket is any different between the schools.

Our null hypothesis: H\_0: The mean amount of money in each students pocket is the same.

Our alternative hypothesis: H\_A: The mean amount of money in each students pocket is different.

P-Value: 0.1732, We will fail to reject the null hypothesis at the 0.05 significance level.

Permutation Test code results below:

A picture containing bird

Description automatically generated