MA206, Lesson 7 - Generalization

What is **generalization**?

Define:

   Population:

   Sample:

   Convenience Sample:

   Biased Sampling Method:

   Simple Random Sample:

Does a larger sample size fix sampling bias?

What are some Nonsampling concerns which might also introduce bias into our data?

1) Suppose you have a massive dessert bowl containing 40% red skittles and 60% green skittles. You take many, many random samples of 25 skittles and each time note the proportion that are red. From this, you create a distribution of all your sample proportions of red skittles.

a) What is the expected mean of your distribution of sample proportions?

b) What is the expected standard deviation of your distribution?

2) Suppose the leadership at Arvin Gym wants to get a sense of how many cadets actually want to reopen the weight room on the third floor. They know that sending a survey out to the entire Corps is destined to fail, and so have come up with four courses of action.

Which course of action below should be used? Justify your answer.

a) Send a survey to the Football Team to gather their opinion.

b) Have the front desk ask everybody who comes into Arvin.

c) Compile a list of names of all the Cadets on Corps Squad teams, randomly select a sample of those names using a random number generator, and survey those cadets.

d) Compile a list of names of all Cadets, Staff, and Faculty at USMA, randomly select a sample of those names using a random number generator, and survey those individuals.

e) Compile a list of names of all Cadets, randomly select a sample of those names using a random number generator, and survey those individuals.

3) As part of the General Social Survey (GSS) in 2018, a random sample of U.S. adults were asked whether they have ever been told by a health professional that they have depression. In the sample of 1,414 people that received this question, 271 of them said that they have been told that they did have depression.

a) Suppose in the population of all U.S. adults, 20% have been told by a health professional that they had depression. What should be the mean and standard deviation be if we were to sample from this population many times?

b) How many standard deviations below the mean of the distribution described in part (a) is the sample proportion from the GSS?

c) Based on your answer from part (b), is it very unlikely that a random sample of 1,414 U.S. adults would only find 271 of them that would say that they have been told they had depression? Explain.

d) Using your work, calculate a p-value for an alternative hypothesis that the true proportion of adults that have been told by a health professional that they had depression is not 20% .

4) A survey was conducted on 56 West Point cadets in MAJ McD's AY23-2 MA206 class about their preferences for original Starburst flavors between Pink, Orange, and Yellow. The results are compiled in the Starburst.csv file on Teams. You may use the course guide as a reference to read in this file. We want to validate the claim that, if given a choice between Pink, Orange, and Yellow from the original starburst colors, cadets think that Yellow starburst are the worst.

   a) Write the null and alternate hypotheses using symbols and describe the parameter of interest in words.

   b) List your observed statistic, sample size, standardized statistic, and p-value (using appropriate methods).

   c) Interpret the results of your analysis. Ensure you include your calculated p-value.

   d) Do you feel comfortable generalizing these results to the entire Corps of Cadets?

   e) To what population would you feel comfortable generalizing these results to?

5) According to the National Coffee Drinking Study from the National Coffee Association, 40% of 18-to 24-year-olds in the United States regularly drink coffee every day. Suppose this number is accurate. Researchers are interested in testing if the daily coffee-drinking habits of cadets at West Point differ from the national average. To test this, they sampled 140 cadets in Grant Hall after lunch. Of those, 73 indicated they regularly drank coffee every day.

a) Write the null and alternate hypotheses, in words and symbols.

b) Do we meet validity conditions to use theoretical methods in this analysis?

c) Using simulation, report the standardized statistic and p-value.

d) Using theoretical methods, report the standardized statistic and p-value.

e) In words, summarize your findings. Use your findings from **c)** or **d)** to justify your conclusion, using the appropriate based on validity conditions.

f) Comment on the generalize-ability of these results.