Jake Caldwell and Riley Coburn Project Proposal for Math 420

For our semester project in Statistical modelling, we would like to choose a theory project. The topic we would like to examine is a deeper dive into Randomization-Based Inference. Specifically, how it is related to Monte Carlo Simulations and why this method of inference even works. This project would allow us to expand our knowledge of randomization-based inference using new tools for randomization and discovering integral theories. Some new tools that will be introduced are Monte Carlo Integration, Smoothed Bootstrap, and the Parametric Bootstrap, to name a few. Furthermore, we will gain a better understanding of the foundation of randomization-based inference, why it works, and a deeper dive into examples and case studies using the idea. We can expand on our coverage of randomization-based inference in class by exploring the aforementioned new topics, performing exercises from the supplemental textbook, and applying these examples in R.

High Level Overview

A Randomization-Based Inference is a process that introduces random resampling to generate a normal distribution, representative of the population, that allows the user to estimate certain statistics, generate p-values, or create a normal sample. Starting from a birds eye view, a Randomization-Based Inference is a resampling method. It is heavily based upon the law of large numbers in that a large enough number of random samples, coming from the original sample of the population, can generate the required normal distribution. Using this distribution, made up of each sample, we have essentially estimated the sampling distribution. Taking this to the next level, what we hope to do in this project, is find new ways to use this idea of resampling to account for error, skewness, but also understand the underlying structure that even allows this method to work. This allows us to answer questions where the statistics that we are currently

using fail to provide any information, as well as have ample theory and support to back our findings. Furthermore, we look to connect this to Monte Carlo Simulations, which form part of the basis for why resampling works. These powerful yet simple versions of resampling can be expanded as far as solving extremely complex problems based on probability or as simple as generating an estimated normal distribution. These models and the theories behind them will help us understand the way randomization-based inference works.

Outline

The rough outline for our research paper will contain four main sections: an introduction, a review of the literature, applied exercises, and a conclusion. Our introduction will begin to explain the topic of our research. This will likely contain a high level overview of randomization-based inference and introduction to the Monte Carlo simulation method. It will be here where we may include some cases or examples which exemplify the importance of our research. From here, we will transition into current literature in the field. Some interesting sources that Dr. White has provided so far include *What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum* by Tim Hesterberg and *Mathematical Statistics with Resampling and R* by Laura Chihara and Tim Hesterberg. The first of these sources will be useful for gaining a better understanding of bootstrapping methods, permutation tests, and some foundational theory behind both of these. This will cover most if not all chapters in the document. The latter of these sources will give us more pertinent information on these subjects including more theory, but most importantly, examples and exercises that we will be able to include and work through. Finally, our conclusion will synthesize information from the resources we will have reviewed. As mentioned in the project introduction, we will be

formatting this paper much like the textbook giving the reader a mix of theory, applications, and real-world examples.

Exercises:

As aforementioned, we will be including examples from *Mathematical Statistics with Resampling and R.* We have some hope of what kind of problems we'd like to incorporate. These include one question from Chapter 4: Sampling Distributions, two questions from Chapter 5: The Bootstrap, one question from Chapter 6: Estimation, the case study from Chapter 9: Regression, and three to four questions from Chapter 11: Additional Topics. One of these three to four questions will include number eleven which allows the reader to see the implementation of Monte Carlo simulations.

Resources:

Book: *Mathematical Statistics with Resampling and R* -- Laura Chihara and Tim Hesterberg

- This has theory, exercises, and applications of those exercises in R. This also contains full case studies and how to carry them out in R.

Supplemental Paper: *What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum* -- Tim Hesterberg

- This paper provides more supplemental material, including theory, at a level digestible by a less advanced statistics student.