Emily Johnson

Senior Project: How Correlated Data Affects Power and Sample Size Calculations

The objective of this senior project was to learn about power and sample size calculations with correlated data. The motivation for this project was inspired by a researcher in the kinesiology department at Cal Poly, San Luis Obispo. The researcher, Dr. Ventura, wanted to examine the effectiveness of bottle feeding an infant using a clear bottle, which represents the control group, versus using an opaque bottle, which is the treatment group. The infant’s intake, measured in mL, was used as a way to assess if weight gain is lower or higher in the treatment group. The idea behind this was that if amount of milk left in the bottle was not visible, then possibly the infant would not finish the bottle and drink as much milk. Each infant’s intake was assessed at the baseline and at the four weeks follow up. Some of the infants were bottle fed with the clear bottle on the first visit and the opaque bottle on the second visit, while other infants were bottle fed with the opaque bottle on the first visit and the clear bottle on the second visit. The primary goal of this project was to use simulation to examine how the correlation in infant intake between the baseline and four weeks follow up affects the power and sample size calculations.

To begin my investigation, I manipulated the equation for the difference of the standard deviations of two groups in order to incorporate correlation. Then, the new functions that I created in R use this formula for the standard deviation of the differences to calculate the power of a paired t-test and a 2-sample t-test. The overall results of my project concluded that as correlation between data increases, the standard deviation of the differences is going to decrease due to the nature of the formula. Thus, as the standard deviation of the differences decreases, the power of the test will increase. This information is helpful when designing a study because if you know that the data you will be collecting will have some sort of strong correlation, the less conservative you need to be in your design.