Lab 1: RCT with 1 Endpoint

Grading: Turn in your first attempt at the tasks for a binary "fair attempt or not" grade on Canvas. That is, your first attempt need not be neat or correct.

Scientific context

Consider a randomized clinical trial designed to evaluate a new treatment for high cholesterol designed as follows:

- Recruit n subjects to be given placebo

- Recruit n subjects to be given the treatment

- For each subject, measure their serum total cholesterol.

The investigators want to determine if the mean serum total cholesterol in the treatment group is different from the mean serum total cholesterol in the control group. They test the null hypothesis of no difference in means using a two-tailed two-sample t-test with unequal variances, with a significance level/rejection threshold of 0.05. They report the point estimate for the difference in means, the p-value from the t-test, and a 95% t-confidence interval.

Serum total cholesterol is measured in mg/dL, and less than 200 mg/dL is generally considered normal for an adult in the USA. More than 240 mg/dL is considered high. According to the CDC, between 2017 and 2020, 10% of adults age 20 or older had serum total cholesterol levels above 240 mg/dL.

Task 1

Write a function that simulates the results of repeating the RCT once: that is, what happens if you recruit new subjects, measure their serum total cholesterol, and repeat the analysis.

The arguments to your function should include:

- Number of subjects in each group n

- Arguments that vary the distribution of the cholesterol measurements in the treatment group

- Arguments that vary the distribution of the cholesterol measurements in the control group

Important questions to ask yourself:

- How would you set the arguments in your function to have no difference in means?

- How would you set the arguments in your function to have a difference in means of 10?

- Are there multiple ways to set the arguments to have no difference in means?

Task 2

Use the function you made in Task 1 (or the code provided in class) to:

- Approximate the p-value distribution under the null hypothesis of no difference in means

- Numerically approximate the power function: that is, the probability of rejecting the null hypothesis of no difference in means at a given true difference in means.

- Numerically approximate the coverage function: that is, the probability of the 95% confidence interval containing the true difference in means, at a given true difference in means.

Comment on the results. Think about things like p-value validity, type I error rate, and power.