QTM 385 - Experimental Methods

Assignment 03

# Instructions

This assignment covers the last two lectures of the course. As usual, it consists of 10 questions, each worth one point. You can answer the questions in any format you prefer, but I recommend using Jupyter Notebooks and converting the answers to PDF or html, as they are easier to read on Canvas. Please write at least one or two paragraphs for each written question.

If you have any questions about the assignment, feel free to email me at [danilo.freire@emory.edu](mailto:danilo.freire@emory.edu).

Good luck!

# Questions

1. Compare and contrast Type I and Type II errors. In causal inference experiments, why might a researcher be more concerned with one type of error over the other?
2. Explain the concept of randomisation inference and outline its advantages over traditional parametric tests, especially in the context of testing the sharp null hypothesis.
3. Compare Neyman’s hypothesis testing framework with Fisher’s sharp null hypothesis approach. What are the main advantages and disadvantages of each method in experimental settings?
4. Critically evaluate the use of p-values in hypothesis testing. What alternatives are suggested (or implied) in the lectures, and what are the potential benefits of these alternatives?
5. The code below simulates a dataset. Modify the code so that it adds a new variable called treat with 500 treated individuals and 500 control individuals (complete random assignment). Also include a binary covariate called gender (0 = male, 1 = female; with equal probability) and update the outcome (interviews) by adding 2 points if the individual is female.

## Set seed for reproducibility  
set.seed(385)  
  
# Load packages  
# install.packages("fabricatr")  
# install.packages("randomizr") # if you haven't installed them yet  
library(fabricatr)  
library(randomizr)  
  
## Simulate data  
data <- fabricate(  
 N = 1000,  
 interviews = round(rnorm(1000, mean = 10, sd = 2) + 5 \* treat, digits = 0)  
)  
head(data)

1. Using the dataset created in the previous question, estimate the average treatment effect on the outcome interviews using the lm\_robust() function from the estimatr package. Interpret the results.
2. Using the same dataset, estimate the average treatment effect of the treatment on the outcome interviews using randomisation inference. Interpret the results.
3. Explain how including covariates in an experimental regression model can increase the precision of the treatment effect estimate. Under what conditions might this adjustment lead to biased estimates?
4. Simulate a dataset with heterogeneous treatment effects (e.g., the treatment effect is larger for individuals with higher education). Estimate the treatment effect for different subgroups using an interaction term.
5. Why is the publication of null results important in experimental research? What are the main challenges in publishing null results, and how can the scientific community address these challenges?