QTM 385 - Experimental Methods

Assignment 04

# Instructions

As in previous assignments, here are 10 questions related to our recent classes. They cover randomisation inference, blocking, clustering, and statistical power.

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. Explain what the ICC represents in the context of a clustered randomised trial. Give a concrete example of a situation where you would expect a high ICC.
2. Imagine that you have designed a cluster-randomised trial with 10 schools (5 treatment, 5 control) and found an ICC of 0.15. Calculate the design effect (or variance inflation factor) for this study if the average cluster size is 2000 students.
3. You want to increase the statistical power of your experiment. What are the main factors that influence the power of a study, and how can you increase it?
4. A diet and exercise program advertises that it causes everyone who is currently starting a diet to lose at least seven pounds more than they otherwise would have during the first two weeks. Use randomisation inference to test the hypothesis that for all . The treatment group’s weight losses after two weeks are and the control group’s weight losses are . In order to test the hypothesis for all using the randomisation inference methods discussed in this chapter, subtract 7 from each outcome in the treatment group so that the exercise turns into the more familiar test of the sharp null hypothesis that for all . When describing your results, remember to state the null hypothesis clearly, and explain why you chose to use a one-sided or two-sided test. (Adapted from Gerber and Green, 2012, p. 88).
5. Below is an R function to calculate statistical power in R. Assuming that $\_t = 55, , , , and , calculate the statistical power of the test. How many individuals would be necessary for the statistical power to be 0.8?

power\_calculator <- function(mu\_t, mu\_c, sigma, alpha=0.05, N){   
 lowertail <- (abs(mu\_t - mu\_c)\*sqrt(N))/(2\*sigma)   
 uppertail <- -1\*lowertail   
 beta <- pnorm(lowertail- qnorm(1-alpha/2), lower.tail=TRUE) + 1- pnorm(uppertail- qnorm(1-alpha/2), lower.tail=FALSE)   
 return(beta)   
 }

1. Sometimes, rather than calculate a budget based on sample size, we have a maximum budget and need to decide whether it is worth doing the study (that is, whether we are sufficiently powered to detect a given effect size, conditional on budgetary limitations). You find out that you only have enough funds for a sample size of 2400 in total. Using the estimate of a roughly 5% increase in test scores and a standard deviation of test scores of 16.5, what is the power of your experiment? (An approximate answer is okay; it’s hard to get exact power on the calculator this way.) The mean test score in the control group is 37. Is it worth carrying out the study on just 2400 students? How would you determine this? (Adapted from J-PAL, 2019).

The text below, also adapted from Gerber and Green (2012, p. 88), discusses an article by Titiunik (2010) that analyses two natural experiments that can be combined. Using this text as a basis, answer questions 7 to 10. The dataset is available [in our GitHub repository](https://github.com/danilofreire/qtm385/blob/main/assignments/04-assignment/titiunik.csv). Use the packages we have been using in class to answer these questions (randomizr, estimatr, fabricatr, ri2, and others you may find useful). A table with the number of bills introduced by senators in both states is available below.

| Texas | # of bills introduced | Arkansas | # of bills introduced |
| --- | --- | --- | --- |
| Term Length: |  | Term Length: |  |
| 0 = four-year term; |  | 0 = four-year term; |  |
| 1 = two-year term |  | 1 = two-year term |  |
| ——————— | ——————— | ———————- | ——————— |
| 0 | 18 | 0 | 11 |
| 0 | 29 | 0 | 15 |
| 0 | 41 | 0 | 23 |
| 0 | 53 | 0 | 24 |
| 0 | 60 | 0 | 25 |
| 0 | 67 | 0 | 26 |
| 0 | 75 | 0 | 28 |
| 0 | 79 | 0 | 31 |
| 0 | 79 | 0 | 33 |
| 0 | 88 | 0 | 34 |
| 0 | 93 | 0 | 35 |
| 0 | 101 | 0 | 35 |
| 0 | 103 | 0 | 36 |
| 0 | 106 | 0 | 38 |
| 0 | 107 | 0 | 52 |
| 0 | 131 | 0 | 59 |
| 1 | 29 | 1 | 9 |
| 1 | 37 | 1 | 10 |
| 1 | 42 | 1 | 14 |
| 1 | 45 | 1 | 15 |
| 1 | 45 | 1 | 15 |
| 1 | 54 | 1 | 17 |
| 1 | 54 | 1 | 18 |
| 1 | 58 | 1 | 19 |
| 1 | 61 | 1 | 19 |
| 1 | 64 | 1 | 20 |
| 1 | 69 | 1 | 21 |
| 1 | 73 | 1 | 23 |
| 1 | 75 | 1 | 23 |
| 1 | 92 | 1 | 24 |
| 1 | 104 | 1 | 28 |
|  |  | 1 | 30 |
|  |  | 1 | 32 |
|  |  | 1 | 34 |
|  |  | 0 | 17 |

1. Naturally occurring experiments sometimes involve what is, in effect, block random assignment. For example, Titiunik studies the effect of lotteries that determine whether state senators in Texas and Arkansas serve two-year or four-year terms in the aftermath of decennial redistricting. These lotteries are conducted within each state, and so there are effectively two distinct experiments on the effects of term length. An interesting outcome variable is the number of bills (legislative proposals) that each senator introduces during a legislative session. The table below lists the number of bills introduced by senators in both states during 2003. For each state, estimate the effect of having a two-year term on the number of bills introduced.
2. Estimate the overall ATE for both states combined.
3. Explain why, in this study, simply pooling the data for the two states and comparing the average number of bills introduced by two-year senators to the average number of bills introduced by four-year senators leads to different estimates of the overall ATE.
4. Use randomisation inference to test the sharp null hypothesis that the treatment effect is zero for senators in both states.