

Problem Set 2

STAT II (Spring 2025)

```
pacman::p_load(tidyverse, haven)
```

Disclaimer: Please read the guidelines below carefully. Good luck!

Guidelines

- Please upload your answers to Brightspace by **Thursday, March 6th, at 11:59 PM**.
- Direct all questions about the problem set to Brightspace under Contents > Discussions > Class Feed > Problem Set 2.
- For precise and expedited responses, we will address questions about this problem set on Brightspace until 6:00 PM on March 5rd.
- Collaboration is allowed, but I encourage you to attempt the problems on your own before seeking help from others. Regardless of collaboration, you must individually write up and submit your answers.
- Late submissions will not be accepted unless prior approval is obtained from the instructor at least one day before the due date.
- **The total points for the problem set are 100.** Individual bonus questions may appear. If your total points exceed 100, the excess points will carry over to subsequent problem sets (e.g., if you earn 120 points, 20 points will be added to future problem sets).
- **Grading:** Show every step of your derivations. We grade the steps of derivations as well as the final answers. If you are unable to solve the problem completely, partial credit can be given for your derivations. Conversely, a correct final answer without complete derivations may not receive full credit.
- **Stylistic Requirements:** Adhere to the guidelines for the format of your submitted answers. Ensure you follow these rules:
 1. Submit your answers as a PDF file compiled from L^AT_EX, R Markdown, or (ideally) a Quarto Markdown document.
 2. If using raw L^AT_EX(.tex) for your answers, submit an accompanying .R file for any computational tasks, with referenced line numbers corresponding to each specific task.
 3. If using R Markdown (.Rmd) or Quarto Markdown (.qmd), include your code as code chunks in the source file. Additionally, submit the source .Rmd or .qmd file along with the compiled PDF to allow us to run your code easily.
 4. To ensure reproducibility of your simulation results, use `set.seed()` at the beginning of your document.

Problem 1. Important concepts (10 points)

- (a) What is a standard error? What is the difference between a standard error and a standard deviation?
- (b) How is randomization (Fisher) inference used to test the sharp null hypothesis of no effect for any subject?
- (c) What is a 95% confidence interval?
- (d) How does complete random assignment differ from block random assignment and cluster random assignment?
- (e) Experiments that assign the same number of subjects to the treatment group and control group are said to have a “balanced design.” What are some desirable statistical properties of balanced designs?

Problem 2. Potential outcomes (10 points)

Consider the schedule of outcomes in the table below. If treatment A is administered, the potential outcome is $Y_i(A)$, and if treatment B is administered, the potential outcome is $Y_i(B)$. If no treatment is administered, the potential outcome is $Y_i(0)$. The treatment effects are defined as $Y_i(A) - Y_i(0)$ or $Y_i(B) - Y_i(0)$.

Subject	$Y_i(0)$	$Y_i(A)$	$Y_i(B)$
Miriam	1	2	3
Benjamin	2	3	3
Helen	3	4	3
Eva	4	5	3
Billie	5	6	3

Suppose a researcher plans to assign two observations to the control group and the remaining three observations to just one of the two treatment conditions. The researcher is unsure which treatment to use.

- (a) Applying equation for $\mathbb{V}[\hat{\tau}_{DiM} \mid \mathcal{O}_N]$ from slide 18 on Randomized Experiments, determine which treatment, A or B, will generate a sampling distribution with a smaller standard error.
- (b) What does the result in part (a) imply about the feasibility of studying interventions that attempt to close an existing “achievement gap”?

Problem 3. Randomization inference (20 points)

The file `Clingingsmith_subset.dta` contains data from a study by Clingingsmith, Khwaja, and Kremer (2009) that focuses on Pakistani Muslims who participated in a lottery to obtain visa for the pilgrimage to Mecca. The study is described as follows in chapter 3.5 of the Gerber and Green textbook:

“By comparing lottery winners to lottery losers, the authors are able to estimate the effects of the pilgrimage on the social, religious, and political views of the participants. Here, we consider the effect of winning the visa lottery on attitudes toward people from other countries. Winners and losers were asked to rate the Saudi, Indonesian, Turkish, African, European, and Chinese people on a five-point scale ranging from very negative (−2) to very positive (+2). Adding the responses to all six items creates an index ranging from −12 to +12.”

- (a) Use the data in `Clingingsmith_subset.dta` to test the sharp null hypothesis that winning the visa lottery (variable `success` in dataset) for the pilgrimage to Mecca had no effect on the views of Pakistani Muslims toward people from other countries (variable `views` in data). Assume that the Pakistani authorities assigned visas using complete random assignment. Conduct 10,000 simulated random assignments under the sharp null hypothesis.
- How many of the simulated random assignments generate an estimated ATE that is at least as large as the actual estimate of the ATE ?
 - What is the implied one-tailed (upper) p -value?
 - How many of the simulated random assignments generate an estimated ATE that is at least as large in absolute value as the actual estimate of the ATE ?
 - What is the implied two-tailed p -value?
- (b) Now, test the sharp null hypothesis that the effect of winning the visa lottery for the pilgrimage to Mecca on the views of Pakistani Muslims toward people from other countries equals your estimate of the ATE . What is the implied two-tailed p -value?

Problem 4. Block random assignment (30 points)

Naturally occurring experiments sometimes involve what is, in effect, block random assignment. For example, Titiunik (2016) 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 data set `Titiunik.dta` contains `term2year`—an indicator for whether a senator was assigned to treatment (a two-year rather than a four-year term); `bills_introduced`—the number of bills introduced by the senator (the outcome); `texas0_arkansas1`—an indicator for whether the senator is from Texas or Arkansas (the blocks).

- For each state, estimate the effect of having a two-year term on the number of bills introduced.
- For each state, estimate the standard error of the estimated ATE (using the conservative estimator from slide 19 on Randomized Experiments).
- Estimate the overall ATE for both states combined through a weighted average of the state-level ATE estimates (as shown on slide 72 on Randomized Experiments).
- 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 biased estimates of the overall ATE .
- Show that the weighted average of the state-level ATE estimates gives the same overall ATE estimate as a weighted regression that weights each observation by the inverse of its probability of being assigned to the condition that the observation was assigned to. For this task you can rely on the function `estimatr::lm_robust()` to run the regression and pass the inverse probability weights to the `weights` argument of this function.
- Estimate the standard error for the overall ATE by combining estimates of the block-level standard errors as shown on the slide 73 on Randomized Experiments.

- (g) Use randomization inference to test the sharp null hypothesis that the treatment effect is zero for all senators.

Problem 5. Cluster random assignment (30 points)

Consider the 2003 Kansas City voter mobilization experiment studied by Arceneaux (2005). In the experiment a group called ACORN targeted 28 low-income precincts in an effort to mobilize voters on behalf of a ballot measure designed to fund municipal bus service. ACORN wanted to work within selected precincts in order to make it easier to train and supervise its canvassers. Of the 28 precincts in the sample, 14 were randomly allocated to the treatment group, and ACORN made repeated attempts to canvass and call voters on its target list in those precincts. The 28 precincts contain a total of 9,712 voters, and the number of targeted voters per precinct ranges from 31 to 655; the marked difference in cluster size leaves open the possibility for bias if precincts with large numbers of potential ACORN sympathizers have different potential outcomes from precincts with relatively few.

The study data in `Arcaneaux_subset.dta` contains a wealth of covariates: the registrar recorded whether each voter participated in elections dating back to 1996.

- (a) Assess the balance of the treatment and control groups by looking at whether past turnout predicts treatment assignment. To do so first regress treatment assignment on the entire set of past votes, and retrieve the observed F -statistic for the test of goodness of fit (e.g. using `estimatr::lm_robust()$fstatistic`). Use randomization inference to test the null hypothesis that none of the past turnout variables predict treatment assignment.¹ Judging from the p -value of this test, what does the F -statistic seem to suggest about whether subjects in the treatment and control groups have comparable background characteristics?
- (b) Explain what intraclass correlation is and why it is important in the context of this experiment. Calculate the intraclass correlation for the 2003 Kansas City voter mobilization experiment using the turnout data from 2003. Interpret your results and calculate Moulton factor (design effect) using formula from slide 63 on Randomized Experiments.
- (c) Regress turnout in 2003 (after the treatment was administered) on the experimental assignment and the full set of covariates. Interpret the estimated ATE . Use randomization inference to test the sharp null hypothesis that experimental assignment had no effect on any subject's decision to vote.
- (d) Using the turnout data from 2003, compare the parametric variance estimates obtained with and without accounting for clustering (both can be done using `estimatr::lm_robust()` and specifying appropriate `se_type` argument). What do your results suggest about the importance of considering clustering in this experiment?

¹**Hint:** To simulate the distribution of the F -statistic, you must generate a large number of random cluster assignments and calculate the F -statistic for from regressing each simulated treatment assignment on the entire set of past votes.

References

- Arceneaux, Kevin. 2005. "Using Cluster Randomized Field Experiments to Study Voting Behavior." *The Annals of the American Academy of Political and Social Science* 601 (1): 169–79.
- Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer. 2009. "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering." *The Quarterly Journal of Economics* 124 (3): 1133–70.
- Titunik, Rocio. 2016. "Drawing Your Senator from a Jar: Term Length and Legislative Behavior." *Political Science Research and Methods* 4 (2): 293–316.