Problem Set 1: Randomization Inference

MGMT 737 — Spring 2025

This problem set uses the Dehejia-Wahba sample from the Lalonde NSW experiment (`lalonde_nsw.csv`). The outcome is `re78` (real earnings in 1978), treatment is `treat`, and remaining variables are covariates.

You may use AI coding assistants, but you must submit a brief collaboration log (see end of assignment).

Part A: Baseline Implementation (20 points)

1. Calculate the ATE using a difference in means. Store as `tau_ate`.
2. Calculate the ATT. Store as `tau_att`. In 2–3 sentences, explain why these are numerically identical under complete randomization.
3. Implement a randomization test for the sharp null ($\tau_i = 0$ for all i). Use 1,000 permutations, holding fixed the number treated. Report the two-sided p-value as `p_ri`.
4. Compare to the p-value from a regression with HC2 robust standard errors. Report as `p_robust`.

Part B: Extend the Estimator (35 points)

Now we'll stress-test and extend your randomization inference implementation.

5. **Stratified randomization.** Suppose treatment was actually randomized *within* strata defined by `married` $\times$ `nodegree` (four cells). Modify your randomization test to permute treatment labels only within strata. Report the new p-value as `p_ri_strat`.

   In 3–4 sentences: Why might this matter? When would ignoring stratification lead to incorrect inference?

6. **Covariate adjustment.** Modify your randomization test to use a regression-adjusted test statistic instead of the simple difference in means. Specifically:

   - Regress `re78` on `age`, `education`, `black`, `hispanic`, `married`, `nodegree`, and `re74`
   - Use the coefficient on `treat` as your test statistic
   - Permute treatment and re-run the regression 1,000 times

   Report the p-value as `p_ri_adjusted`.

   In 3–4 sentences: What is the theoretical justification for this procedure? Does covariate adjustment "buy" you anything under the sharp null?

7. **Studentized test statistic.** Your current implementation uses $\hat{\tau}$ as the test statistic. Modify it to use $\hat{\tau}/\text{SE}(\hat{\tau})$ instead, recomputing the standard error for each permutation. Report as `p_ri_studentized`.

   In 2–3 sentences: When would you expect the studentized version to have better properties? (Hint: think about heteroskedasticity.)

8. **Confidence interval inversion.** Using your randomization test machinery, construct a 95% confidence interval for the ATE by inverting a family of hypothesis tests. Test $H_0 : \tau_i = c$ for $c \in \{-5000, -4500, \ldots, 9500, 10000\}$. Report the interval as `ci_ri`.

   Compare this interval to the one from robust regression. Which is wider? In 2–3 sentences, explain why.

Part C: Debugging (25 points)

9. **Find the bug.** Below is a randomization inference implementation that produces incorrect p-values. It consistently *under-rejects* the null (p-values are too large). Identify the error and explain why it causes this specific problem.

```
run_ri_test <- function(Y, W, n_perms = 1000) {
  obs_diff <- mean(Y[W == 1]) - mean(Y[W == 0])

  perm_diffs <- numeric(n_perms)
  for (i in 1:n_perms) {
    W_perm <- sample(W)
    perm_diffs[i] <- mean(Y[W_perm == 1]) - mean(Y[W_perm == 0])
  }

  p_value <- mean(perm_diffs >= obs_diff)
  return(p_value)
}
```

10. **Diagnose the simulation.** A researcher runs the following simulation to verify that randomization inference controls size at 5%:

```
set.seed(42)
reject <- numeric(1000)

for (sim in 1:1000) {
  # Generate data under sharp null
  n <- 200
  Y0 <- rnorm(n, mean = 10, sd = 5)
  Y1 <- Y0  # Sharp null: no effect
  W <- rbinom(n, 1, 0.5)
  Y <- ifelse(W == 1, Y1, Y0)

  # Run RI test
  p <- run_ri_test_correct(Y, W, n_perms = 500)
  reject[sim] <- (p < 0.05)
}

mean(reject)  # Returns 0.038
```

The researcher expected a rejection rate of 0.05 but gets 0.038. The `run_ri_test_correct` function is correctly implemented. What explains the under-rejection?

(Hint: the issue is not the code per se.)

Part D: Conceptual Questions (20 points)

AI Free Zone

11. A colleague argues: "Randomization inference is pointless. If you have a randomized experiment, just run a regression with robust standard errors—it's easier and gives the same answer asymptotically."

    Write a paragraph (5–8 sentences) evaluating this claim. Under what conditions is it correct? When might you prefer randomization inference despite its computational cost?

12. Another colleague is analyzing an RCT with $n = 30$ (15 treated, 15 control). They find a large positive effect but the randomization inference p-value is 0.08. They say: "The effect is clearly real—randomization inference just lacks power with small samples. I'll report the t-test p-value of 0.03 instead."

    In a paragraph, explain what's wrong with this reasoning.

Submission Requirements

Submit to GitHub Classroom:

- `homework1-code.R` — your code
- `homework1-writeup.pdf` — answers to written questions
- `homework1-collab.txt` — collaboration log (see below)

Collaboration log: List any AI tools used (e.g., ChatGPT, Claude, Copilot). For each, briefly note what you used it for and whether you had to correct its output. Example:

> "Used Claude to debug my studentized test statistic code. Initial output had an error in how it recomputed SEs—I fixed by [description]."

If you didn't use AI tools, just write "No AI tools used."