ECON 7201
Applied Econometrics
Assignment 3

Due Date

Friday November 7, 2025 at 11:59 PM

Directions

Answer each question clearly and concisely. Write equations using LaTeX where appropriate and explain all assumptions in your own words. Submit both a PDF and Quarto file to the nexus assignment portal.

Git and GitHub

1. (a) Create a new R project in your **econ_3201** directory called **assignment_3**.
   (b) Download the assignment PDF and Quarto file the **assignment_3** folder.
   (c) Commit and push the changes to your **econ_3201** repository on GitHub.com.

Conceptual Foundations

1. (**5 pts**) Define the *potential outcomes framework*. Clearly explain what is meant by $Y_i(1)$, $Y_i(0)$, and D_i.

2. (**5 pts**) State and explain the **Stable Unit Treatment Value Assumption (SUTVA)**. Why is this assumption critical for causal inference?

3. (**5 pts**) What is the **Conditional Independence Assumption (CIA)**? Write it formally and explain its meaning in words.

4. (**5 pts**) Describe the **Overlap (Common Support)** condition and its practical importance.

5. (**5 pts**) Explain why the **Average Treatment Effect on the Treated (ATT)** may differ from the **Average Treatment Effect (ATE)**.

Identifying Causal Effects (25 points)

1. (**5 pts**) Suppose treatment $ D_i $ is assigned completely at random. Show that

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] = ATE$$

   Explain why randomization justifies this equality.

2. (**10 pts**) Suppose treatment depends on observed covariates $ X_i , *suchthat*$ (Y_i(1), Y_i(0)) D_i X_i $$ Derive an expression for ATE using the **law of iterated expectations**, and interpret it.

3. (**10 pts**) Explain how *selection bias* arises if the CIA fails. Use the decomposition:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = ATE + \text{Selection Bias.}$$

Applied Exercise in R (30 points)

Use R to simulate data, estimate treatment effects, and interpret your results. Include all R code and output.

```r
set.seed(123)

n <- 1000
X <- rnorm(n)
D <- ifelse(X + rnorm(n) > 0, 1, 0)
Y0 <- 2 + 0.5 * X + rnorm(n)
Y1 <- Y0 + 3 + 0.5 * X
Y <- ifelse(D == 1, Y1, Y0)
data <- data.frame(Y, D, X)
```

1. (**5 pts**) Compute the difference in average outcomes between treated and untreated. Interpret the result in terms of potential bias.

2. (**5 pts**)

   (a) Estimate the treatment effect controlling for (X):

$$Y_i = \alpha + \tau D_i + \beta X_i + \varepsilon_i$$

   (b) Compare the estimate of ( ) to the true treatment effect of 3.

3. (**5 pts**) Propensity Score Matching

   (a) Estimate treatment effects using matching. Search `?matchIt` in the console

(b) Compare the matched estimate to OLS and the naïve difference.

4. (**5 pts**) Simulate a violation of the Conditional Independence Assumption by introducing an unobserved confounder (U_i):

```r
U <- rnorm(n)
D <- ifelse(X + U + rnorm(n) > 0, 1, 0)
Y0 <- 2 + 0.5 * X + 0.5 * U + rnorm(n)
Y1 <- Y0 + 3 + 0.5 * X
Y <- ifelse(D == 1, Y1, Y0)
bias_model <- lm(Y ~ D + X, data = data.frame(Y, D, X))
summary(bias_model)
```

Visualize and discuss how this unobserved confounder biases the estimated treatment effect.

```r
plot(X, Y, col = ifelse(D==1, "red", "blue"), pch = 16,
     main = "Outcome vs Covariate with Confounding",
     xlab = "X", ylab = "Y")
legend("topleft", legend = c("Treated", "Untreated"),
       col = c("red", "blue"), pch = 16)
```