P-Set I

POLI 784 - Spring 2026

2026-01-17

Complete all of the questions below together with your group members. Please submit both (1) a compiled PDF or HTML file with your responses and (2) the `R Markdown`, `Quarto`, etc. file containing both your responses and code as a group. You are encouraged to use `R Studio` projects and virtual environments to complete this p-set, although it is not required. Similarly, you are encouraged to use consistent style and linters when writing code.

Question 1

a. *(0.5 points)* Simulate the following data generating process in `R` and save the results in dataframe `df`:

   1. set seed to `42` to make your results reproducible
   2. sample size is $N = 100$
   3. X_i is drawn from the uniform distribution on $[0, 1]$
   4. $Y_i(0)$ is a quadratic function of X_i: $Y_i(0) = 3 + 2 * X_i + X_i^2 + \varepsilon_i$, where $\varepsilon_i \sim N(0, 1)$
   5. treatment effect τ_i is drawn from a normal distribution with mean $5 * \sin(X_i)$ and standard deviation 1 for each unit
   6. the outcome under treatment is $Y_i(1) = Y_i(0) + \tau_i$ for each unit
   7. treatment status $D_i = 1$ is assigned randomly with probability 0.4 for each unit

b. *(0.5 points)* Reveal the observed outcome for each unit by saving it in a new column `Y`. Print the individualistic treatment effect (ITE), $\tau_i = Y_i(1) - Y_i(0)$, for unit `i = 1` (i.e., the 1st row in your dataframe) to console. How can we estimate the ITE for this (or any other) unit following the potential outcomes framework? Briefly explain.

c. *(1 point)* Calculate the *true* ATE in your sample, an estimate of the ATE using $\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$, and the bias of this (sample) ATE estimate. Print all three quantities to console. Under which condition(s) can we identify the ATE? Is this condition / are these conditions met here? Briefly explain. *(Hint: Think about how treatment was assigned and what that buys you!)*

d. *(0.5 points)* Create a new treatment variable `df$D2` where treatment status D_i is assigned randomly via a Bernoulli trial for each unit (just like before), but the probability of treatment now varies between units: $p_i = X_i$. Create another column `df$D3` where treatment status is assigned with probability $p_i = (1 - X_i)$. Compute the ATE estimates and the bias of these (sample) ATE estimates (like you did in *1c*) using `df$D2` and `df$D3`, respectively. Create a plot showing the value of the true ATE and the point estimates of the three ATE estimates you computed (i.e., like a coefficient plot without confidence intervals).

e. *(0.5 points)* Create the following plot three times, once each for `df$D`, `df$D2`, and `df$D3`: Plot the values of $Y_i(0)$ and $Y_i(1)$ on the Y-axis of a scatterplot against X_i on the X-axis for all i, give the dots different colors for $Y_i(0)$ and $Y_i(1)$, and draw separate regression lines through the points for the two sets of potential outcomes. Give those units assigned to treatment and control different shapes (make sure that the shapes are very different so you can distinguish them easily!). *(Note: Only the last part will differ between the three plots.)*

f. *(1 point)* Based on the plots from *1d* and *1e*: What do you observe about the ATE estimates / their biases? Which condition for identifying the ATE is violated and for which of the above ATE estimates? How do the relationship between X_i and D_i and the relationship between X_i and Y_i factor into this?

Question 2

a. *(1 point)* Consider the following assumptions. First explain what each of them states. Then prove that this condition is sufficient for identifying the ATT. Briefly explain. *(Hint: For the proof, begin by stating the* ***theoretical*** *definition of the ATT and restate it until you arrive at a definition that is empirically assessable.)*

$$D_i \perp \{Y_i(0)\}_{i=1}^N$$
$$0 < P(D_i = 1) < 1, \forall i$$

b. *(1 point)* Consider the assumption Imbens (2004) calls "weak ignorability".

$$\mathbf{1}\{D_i = d\} \perp \{Y_i(d)\}_{i=1}^N, d \in \{0, 1\}$$
$$0 < P(D_i = 1) < 1, \forall i$$

How does the assumption differ from the one below?

$$\mathbf{1}\{D_i = d\} \perp \{Y_i(0), Y_i(1)\}_{i=1}^N, d \in \{0, 1\}$$
$$0 < P(D_i = 1) < 1, \forall i$$

Is the first assumption (i.e., weak ignorability) sufficient for the identification of the ATE? Why or why not?

Question 3

A software company is planning on providing a new AI coding assistant to its software developers to increase productivity. Since the provision of the assistant is expensive and empirical results on the efficiency gains of AI-assisted developers are mixed, you have been tasked with evaluating the causal effect of this new tool before it is approved for company-wide rollout. Given limited resources and multiple potential approaches, you have decided to draw on the `DeclareDesign` package in `R` to evaluate which research design is most appropriate for this task *before* conducting your experiment. The outcome of interest is the average number of features shipped by a developer per week and the treatment is receiving access to the AI assistant (*and being forced to use it* – we'll deal with non-compliance in a later p-set). You are interested in estimating the ATE. Make sure to set seed `42` to ensure reproducibility. *(Note: You should draw heavily on the DeclareDesign documentation and the book linked at the website to solve this problem. All steps required below are covered in those resources. Use the week 3 lab for an applied introduction.)*

a. *(1 point)* Begin by declaring a classical two-arm (i.e., one control and one treatment group) experiment with simple random assignment. You have been told to aim for a sample size of $n = 150$, which will be drawn via complete random sample from the total population of developers $N = 3,000$. Include in your declared design a variable $U_i \sim U(0,1)$ *(note: this is a standard uniform distribution)* capturing unknown heterogeneity and a covariate $X_i \sim unif\{0,1,2\}$ *(note: this is a discrete uniform distribution)* representing developer experience level. Simulate potential outcomes in your design using $Y_i = \tau * Z_i + X_i + U_i$, where developers are assigned to treatment Z_i with $p = 0.5$ and the average treatment effect is assumed to be $\tau = 0.25$ based on early research you have reviewed. Us the difference-in-means estimator for this design.

b. *(1 point)* Diagnose the design with `diagnose_design()` (using default arguments for all parameters of the function is fine) and print the diagnostic summary to console. Briefly define the concepts of bias, statistical power, and coverage. Interpret the bias, power, and coverage diagnosand values provided for the declared design.

c. *(0.5 points)* Given the above diagnosand values, redesign the baseline declaration using `redesign()` by varying the sample size (i.e., $n = \{150, 200, 250, 300, 350, 400\}$). Diagnose the resulting designs using `diagnose_design()`, print the results to console, and create a single plot with three lines showing how bias, power, and coverage change across sample size. What effect does sample size have on the three? *(Note: Since these diagnosand values are estimates, we would normally plot confidence intervals around them, but we'll skip those for now.)*

d. *(0.5 points)* Fixing the ATE at a single value makes for a strong assumption, so declare a new design that is identical to the one from *3a* except that it varies the ATE across a range of plausible values: $\tau \sim U(0, 0.5)$. Diagnose the design using `diagnose_design()`, print the results to console, and plot the true effect size on the X-axis and a LOESS curve representing power on the Y-axis. Briefly interpret the plot. *(Note: See Figure 13.3 and associated code in section 13.6.1. of the DeclareDesign book far an example of how to do this.)*

e. *(1 point)* Given that both sample size and true effect size appear to matter, use `redesign()` on the design object you declared in *3a* to vary both the sample size (i.e., $n = \{150, 250, 350\}$) and the true effect size (i.e., $\tau = \{0.2, 0.3, 0.4\}$) simultaneously. Diagnose the resulting designs using `diagnose_design()` and print the results to console. Finally, consider the trade-offs between the different designs tested: Which sample size would you recommend for the research design and on what grounds? Consider bias, power, coverage, cost related to sample size, and the expected size of the true effect in your response. *(Note: There is no single correct response here, so a recommendation based on a sensible interpretation of the simulation results will suffice.)*