Causal Inference

Homework 3

2026

Problem 1 (Regression approach)

Consider the dataset in Problem 1 of Homework 2 again.

A completely randomized experiment was carried out in 12 first-grade classrooms spread across one school district to investigate the effectiveness of a new reading curriculum. Six of the classrooms were randomly assigned to use the new curriculum while the other six used the traditional curriculum. Average scores for each class on a standardized test were measured at the end of the year. The data:

new curriculum: 57.6, 43.8, 66.5, 70.9, 72.7, 80.8;

old curriculum: 29.3, 39.3, 51.1, 59.9, 62.2, 63.5.

Students in the district were all given a pretest at the start of the year. For the new curriculum group the average pretest scores (in the same order as the classes listed above) were 46.6, 63.2, 65.4, 66.0, 73.4, 84.8. For the old curriculum group the average pretest scores were: 50.0, 55.9, 55.8, 69.8, 81.1, 91.5.

Suppose these 12 classrooms can be viewed as i.i.d. samples from all the classrooms in this district. Our interested estimand is the average treatment effect for all the classrooms in this district (that is τ_{SP}).

a. Assume that given X_i, the potential outcomes $Y_i(1)$ and $Y_i(0)$ have equal variances, and they are all the same for all $X_i, \forall i$. Please apply linear regression **with adjusting for pretest scores** to estimate τ_{SP}, and provide an approximate 95% confidence interval for τ_{SP}.

b. (optional) We have no assumptions about the conditional variances of the potential outcomes $Y_i(1)$ and $Y_i(0)$ given X_i. Please apply linear regression **with adjusting for pretest scores** to estimate τ_{SP}, and provide an approximate 95% confidence interval for τ_{SP}.

c. Compare the results from (a) (or (b)) with your results from Problem 1 (d), (f) in Homework 2. What do you find? Comment on your findings.

Problem 2 (Model-based Analysis)

A completely randomized experiment with $N_t = 2$ and $N_c = 2$ is implemented on a finite population of 4 units. The data received are illustrated in the table below.

Unit	W_i	$Y_i(1)$	$Y_i(0)$
1	0	?	2
2	0	?	6
3	1	7	?
4	1	5	?

Assume that

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \mid \theta \sim N \left(\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

For the unknown parameter $\theta = (\mu_c, \mu_t)'$, we specify its prior distribution as $P(\theta) \propto 1$, implying that μ_c and μ_t are independent in prior.

a. Please **list the key steps** that you need to derive the posterior mean of the causal estimand $\tau_{sp} = E(Y(1) - Y(0))$, and $\tau_{fs} = \bar{Y}(1) - \bar{Y}(0)$, respectively. (Just formulas, without detailed calculations.)

b. (Optional) Please **derive** the posterior mean of the causal estimand $\tau_{sp} = E(Y(1) - Y(0))$, and $\tau_{fs} = \bar{Y}(1) - \bar{Y}(0)$, respectively. (That is, find the results based on the data.)

Some hints:

$$L(\theta \mid Y^{\text{obs}}, W) \propto \prod_{i:W_i=0} \exp\left\{ -\frac{1}{2} \left(Y_i^{\text{obs}} - \mu_c \right)^2 \right\} \cdot \prod_{i:W_i=1} \exp\left\{ -\frac{1}{2} \left(Y_i^{\text{obs}} - \mu_t \right)^2 \right\},$$

$$P(\theta \mid Y^{\text{obs}}, W) \propto L(\theta \mid Y^{\text{obs}}, W) \cdot P(\theta).$$

Problem 3 (Treatment Effect Heterogeneity)

In completely randomized experiments:

1. Show that $S^2(\tau) = 0$ implies that $S^2(1) = S^2(0)$.

2. Give a counterexample with $S^2(1) = S^2(0)$ but $S^2(\tau) \neq 0$. (Therefore, in the real application, $S^2(1) = S^2(0)$ is a weaker assumption than $S^2(\tau) \neq 0$, and thus is suitable in a wider range.)

3. (Optional) Show that $S^2(1) < S^2(0)$ implies that

$$S(Y(0), \tau) = \frac{1}{n-1} \sum_{i=1}^{n} \{Y_i(0) - \bar{Y}(0)\}(\tau_i - \tau) < 0.$$

4. (Optional) Give a counterexample with $S^2(1) > S^2(0)$ but $S(Y(0), \tau) < 0$.