HW2

Sam Olson

Outline

- Q1: Skeleton
- Q2: Skeleton
- Q3: Skeleton
- Q4: Skeleton
- Q5: Skeleton
- Q6: Skeleton
- Q7: Skeleton

Problem 1 (20 pt)

A city has a total of 100,000 dwelling units, of which 35,000 are houses, 45,000 are apartments, and 20,000 are condominiums. A stratified sample of size $n = 1000$ is selected using proportional allocation (and rounding the sample sizes to the nearest integer). The three strata are houses ($h = 1$), apartments ($h = 2$), and condominiums ($h = 3$). The table below gives the estimates of the mean energy consumption per dwelling unit for the three strata and the corresponding standard errors.

Stratum (h)	Estimated Mean Energy Consumption ($\bar{y}_h$) (kWh per dwelling unit)	Estimated Standard Error ($\hat{SE}(\bar{y}_h)$)
House ($h = 1$)	915	4.84
Apartments ($h = 2$)	641	2.98
Condominium ($h = 3$)	712	7.00

1.

Estimate the total energy consumption for the full population of 100,000 dwelling units.

Answer

$$\hat{T}_{str} = \sum_{h \in H} N_h \bar{y}_h = 915(35,000) + 641(45,000) + 712(20,000) = 75,110,000$$

2.

Estimate the standard error of the estimator used in (1).

Answer

$$SE(\hat{T}_{str}) = \sqrt{\text{Var}\left(\sum_{h \in H} N_h \bar{y}_h\right)} = \sqrt{\left(\sum_{h \in H} \text{Var}\left(N_h \bar{y}_h\right)\right)} = \sqrt{\left(\sum_{h \in H} N_h^2 \text{Var}\left(\bar{y}_h\right)\right)}$$

$$SE(\hat{T}_{str}) = \sqrt{(35,000^2)(4.84^2) + (45,000^2)(2.98^2) + (20,000^2)(7.00^2)} = 257,447$$

3.

What would be the sample size if the optimal allocation is to be used (under $n = 1000$) for this population? Assume that the survey costs are the same for each stratum.

Hint: Use the following steps:

a)

What is the sample size n_h for each stratum under proportional allocation?

Answer The overall sampling rate is

$$\frac{n}{N} = \frac{1,000}{100,000} = 0.01$$

Under proportional allocation, the sample sizes are

$$n_h = N_h \times 0.01$$

Thus, we have

$$n_1 = 350, \quad n_2 = 450, \quad n_3 = 200$$

b)

Note that:

$$\hat{SE}(\bar{y}_h) = \sqrt{\frac{1}{n_h}\left(1 - \frac{n_h}{N_h}\right) s_h^2}$$

Thus, you can obtain s_h^2.

Answer Now, using

$$SE(\bar{y}_h) = \sqrt{\left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2}$$

we can obtain S_h. That is, we may solve

$$\sqrt{\left(\frac{1}{350} - \frac{1}{35000}\right) S_1^2} = 4.84$$

$$\sqrt{\left(\frac{1}{450} - \frac{1}{45000}\right) S_2^2} = 2.98$$

$$\sqrt{\left(\frac{1}{200} - \frac{1}{20000}\right) S_3^2} = 7.00$$

to obtain

$$S_1 \approx 91.00, \quad S_2 \approx 63.53, \quad S_3 \approx 99.49$$

c)

Apply Neyman allocation (optimal allocation) using s_h in place of S_h.

Answer Finally, we can apply Neyman allocation

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} n$$

with $n = 1000$. Thus,

$$n_1 = \frac{35(91.00)}{35(91.00) + 45(63.53) + 20(99.49)} \cdot 1000 \approx 396$$

$$n_2 = \frac{45(63.53)}{35(91.00) + 45(63.53) + 20(99.49)} \cdot 1000 \approx 356$$

$$n_3 = \frac{10(99.49)}{35(91.00) + 45(63.53) + 20(99.49)} \cdot 1000 \approx 248$$

$$n_1 \approx 396 \quad n_2 \approx 356, \quad n_3 \approx 248$$

are the final sample sizes from Neyman allocation.

4.

What would be the estimated standard error of the total estimator under the optimal allocation in (3)? Compare it with the answer in (2). Which one is smaller?

Answer

$$SE(\hat{T}_{str}) = \sqrt{\sum_{h=1}^{H} \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_h^2}$$

Given:

$$N_1 = 35,000, \quad N_2 = 45,000, \quad N_3 = 20,000$$

$$n_1 = 396, \quad n_2 = 356, \quad n_3 = 248$$

$$S_1 \approx 91.00, \quad S_2 \approx 63.53, \quad S_3 \approx 99.49$$

We compute:

$$\frac{N_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)S_1^2$$

$$\frac{(35,000)^2}{396}\left(1-\frac{396}{35,000}\right)(91.00)^2$$

$$\frac{N_2^2}{n_2}\left(1-\frac{n_2}{N_2}\right)S_2^2$$

$$\frac{(45,000)^2}{356}\left(1-\frac{356}{45,000}\right)(63.53)^2$$

$$\frac{N_3^2}{n_3}\left(1-\frac{n_3}{N_3}\right)S_3^2$$

$$\frac{(20,000)^2}{248}\left(1-\frac{248}{20,000}\right)(99.49)^2$$

Summing these three terms and taking the square root:

Calculating:

$$T_1=\frac{(35,000)^2}{396}\left(1-\frac{396}{35,000}\right)(91.00)^2=25,326,894,797.98$$

$$T_2=\frac{(45,000)^2}{356}\left(1-\frac{356}{45,000}\right)(63.53)^2=22,776,307,940.68$$

$$T_3=\frac{(20,000)^2}{248}\left(1-\frac{248}{20,000}\right)(99.49)^2=15,766,970,443.16$$

Summing these:

$$T_1+T_2+T_3=25,326,894,797.98+22,776,307,940.68+15,766,970,443.16=63,870,173,181.82$$

Taking the square root:

$$SE(\hat{T}_{str})=\sqrt{63,870,173,181.82}\approx 252,725$$

This is smaller than the SE under proportional allocation given previously, which is what we expected.

Problem 2 (10 pt)

Consider a simple random sample of size $n = 200$ from a finite population with size $N = 10,000$, measuring (X, Y), taking values on $\{(0,0), (0,1), (1,0), (1,1)\}$. The finite population has the following distribution.

	$X = 1$	$X = 0$	
$Y = 1$	N_{11}	N_{10}	N_{1+}
$Y = 0$	N_{01}	N_{00}	N_{0+}
	N_{+1}	N_{+0}	N

The population count N_{ij} are unknown.

Suppose that the realized sample has the following sample counts:

	$X = 1$	$X = 0$	
$Y = 1$	70	30	100
$Y = 0$	50	50	100
	120	80	200

1.

If it is known that $N_{+1} = N_{+0} = 5000$, how can you make use of this information to obtain a post-stratified estimator of $\theta = E(Y)$, using X as the post-stratification variable?

Answer

The poststratification estimator is

$$\hat{\theta} = W_1 \bar{y}_1 + W_2 \bar{y}_2 = 0.5 \times \left(\frac{30}{55} \right) + 0.5 \times \left(\frac{20}{45} \right).$$

2.

If we are interested in estimating $\theta = P(Y = 1 | X = 1)$, discuss how to estimate θ from the above sample and how to estimate its variance (Hint: Use Taylor expansion of ratio estimator to obtain the sampling variance).

Answer

Problem 2 (10 pt)

Consider a simple random sample of size $n = 200$ from a finite population with size $N = 10,000$, measuring (X, Y), taking values on $\{(0,0), (0,1), (1,0), (1,1)\}$. The finite population has the following distribution.

	$X = 1$	$X = 0$	
$Y = 1$	N_{11}	N_{10}	N_{1+}
$Y = 0$	N_{01}	N_{00}	N_{0+}
	N_{+1}	N_{+0}	N

The population count N_{ij} are unknown.

Suppose that the realized sample has the following sample counts:

	$X=1$	$X=0$	
$Y=1$	70	30	100
$Y=0$	50	50	100
	120	80	200

1.

If it is known that $N_{+1} = N_{+0} = 5000$, how can you make use of this information to obtain a post-stratified estimator of $\theta = E(Y)$, using X as the post-stratification variable?

Answer

For $\theta = P(Y = 1 \mid X = 1)$, we can use

$$\hat{\theta} = \frac{\hat{P}(X=1, Y=1)}{\hat{P}(X=1)} = \frac{n_{11}}{n_{1+}} = \frac{70}{120} = 0.7,$$

where n_{ij} is the number of sample elements with $(X = i, Y = j)$ and $n_{1+} = n_{11} + n_{10}$.

2.

If we are interested in estimating $\theta = P(Y = 1|X = 1)$, discuss how to estimate θ from the above sample and how to estimate its variance (Hint: Use Taylor expansion of ratio estimator to obtain the sampling variance).

Answer

Now, to obtain variance estimation, Taylor method can be used to get

$$\hat{\theta} \approx \theta + \frac{1}{E(n_{1+})}(n_{11} - \theta n_{1+})$$

$$= \theta + \frac{1}{E(n_{1+})/n}\left(\frac{n_{11}}{n} - \theta\frac{n_{1+}}{n}\right)$$

$$= \theta + \frac{1}{E(\hat{P}_{1+})}\left(\hat{P}_{11} - \theta\hat{P}_{1+}\right).$$

Thus,

$$V(\hat{\theta}) \approx \frac{1}{P_{1+}^2}\left\{V(\hat{P}_{11}) + \theta^2 V(\hat{P}_{1+}) - 2\theta \text{Cov}(\hat{P}_{11}, \hat{P}_{1+})\right\}. \tag{1}$$

Now, under simple random sampling, we have

$$V(\hat{P}_{11}) = \frac{1}{n}(1-f)P_{11}(1-P_{11})$$

$$V(\hat{P}_{1+}) = \frac{1}{n}(1-f)P_{1+}(1-P_{1+})$$

$$\text{Cov}(\hat{P}_{11}, \hat{P}_{1+}) = \frac{1}{n}(1-f)P_{11}(1-P_{1+}).$$

Also, using $\theta = P_{11}/P_{1+}$, we can simplify (1) to get

$$\begin{aligned}
V(\hat{\theta}) &= \frac{1}{n}(1-f)\frac{1}{P_{1+}^2}\left\{P_{11}(1-P_{11}) + \frac{P_{11}^2}{P_{1+}^2}P_{1+}(1-P_{1+}) - 2\frac{P_{11}}{P_{1+}}P_{11}(1-P_{1+})\right\} \\
&= \frac{1}{n}(1-f)\frac{1}{P_{1+}}\left\{P_{11} - \frac{P_{11}^2}{P_{1+}}\right\} \\
&= \frac{1}{n}(1-f)\frac{1}{P_{1+}}\theta(1-\theta).
\end{aligned}$$

Thus, we can estimate the variance of $\hat{\theta}$ by

$$\hat{V}(\hat{\theta}) = (1-f)\frac{1}{n_{1+}}\hat{\theta}(1-\hat{\theta}).$$

Using $f = 0.02$, $n_{1+} = 120$, and $\hat{\theta} = 70/120$, we obtain

$$\hat{V}(\hat{\theta}) = 0.001985.$$

Problem 3 (10 pt)

Suppose that we have a finite population of $(Y_{hi}(1), Y_{hi}(0))$ generated from the following superpopulation model:

$$\begin{pmatrix} Y_{hi}(0) \\ Y_{hi}(1) \end{pmatrix} \sim \left[\begin{pmatrix} \mu_{h0} \\ \mu_{h1} \end{pmatrix}, \begin{pmatrix} \sigma_{h0}^2 & \sigma_{h01} \\ \sigma_{h01} & \sigma_{h1}^2 \end{pmatrix} \right] \tag{1}$$

for $i = 1, \ldots, N_h$ and $h = 1, \ldots, H$. Instead of observing $(Y_{hi}(0), Y_{hi}(1))$, we observe $T_{hi} \in \{0, 1\}$ and

$$Y_{hi} = T_{hi} Y_{hi}(1) + (1 - T_{hi}) Y_{hi}(0)$$

The parameter of interest is the average treatment effect:

$$\tau = \sum_{h=1}^{H} W_h (\mu_{h1} - \mu_{h0}),$$

where $W_h = N_h / N$. The estimator is:

$$\hat{\tau}_{\text{sre}} = \sum_{h=1}^{H} W_h \hat{\tau}_h$$

where

$$\hat{\tau}_h = \frac{1}{N_{h1}} \sum_{i=1}^{N_h} T_{hi} Y_{hi} - \frac{1}{N_{h0}} \sum_{i=1}^{N_h} (1 - T_{hi}) Y_{hi}$$

1.

Compute the variance of $\hat{\tau}_{\text{sre}}$ using the model parameters in (1).

Answer

Recall that

$$E(\hat{\tau}_{sre} \mid \mathcal{F}_N) = \sum_{h=1}^{H} W_h \bar{\tau}_h$$

where

$$\bar{\tau}_h = N_h^{-1} \sum_{i=1}^{N_h} \{Y_{hi}(1) - Y_{hi}(0)\}.$$

Also, in the class, we have learned that

$$V(\hat{\tau}_{sre} \mid \mathcal{F}_N) = \sum_{h=1}^{H} W_h^2 \frac{1}{N_h} \left(\frac{N_{h0}}{N_{h1}} S_{h1}^2 + \frac{N_{h1}}{N_{h0}} S_{h0}^2 + 2 S_{h01} \right).$$

Hence, the total variance is

$$V(\hat{\tau}_{sre}) = V\{E(\hat{\tau}_{sre} \mid \mathcal{F}_N)\} + E\{V(\hat{\tau}_{sre} \mid \mathcal{F}_N)\}$$

$$= V\left\{\sum_{h=1}^{H} W_h \bar{\tau}_h\right\} + E\left\{\sum_{h=1}^{H} W_h^2 \frac{1}{N_h}\left(\frac{N_{h0}}{N_{h1}} S_{h1}^2 + \frac{N_{h1}}{N_{h0}} S_{h0}^2 + 2S_{h01}\right)\right\}.$$

Now, under model (2), we can obtain

$$V\left\{\sum_{h=1}^{H} W_h \bar{\tau}_h\right\} = \sum_{h=1}^{H} W_h^2 \frac{1}{N_h}(\sigma_{h1}^2 + \sigma_{h0}^2 - 2\sigma_{h01})$$

and

$$E\left\{\sum_{h=1}^{H} W_h^2 \frac{1}{N_h}\left(\frac{N_{h0}}{N_{h1}} S_{h1}^2 + \frac{N_{h1}}{N_{h0}} S_{h0}^2 + 2S_{h01}\right)\right\} = \sum_{h=1}^{H} W_h^2 \frac{1}{N_h}\left(\frac{N_{h0}}{N_{h1}} \sigma_{h1}^2 + \frac{N_{h1}}{N_{h0}} \sigma_{h0}^2 + 2\sigma_{h01}\right).$$

Therefore, combining the two, we obtain

$$V(\hat{\tau}_{sre}) = \sum_{h=1}^{H} W_h^2 \left(\frac{\sigma_{h1}^2}{N_{h1}} + \frac{\sigma_{h0}^2}{N_{h0}}\right).$$

2.

Assuming the model parameters are known, determine the optimal sample allocation to minimize $\mathrm{Var}(\hat{\tau}_{\mathrm{sre}})$.

Answer

For each h, we wish to minimize

$$Q(N_{h1}, N_{h0}) = \frac{\sigma_{h1}^2}{N_{h1}} + \frac{\sigma_{h0}^2}{N_{h0}}$$

subject to $N_h = N_{h1} + N_{h0}$ being constant. Thus, by the Schwarz inequality, we can obtain

$$\left(\frac{\sigma_{h1}^2}{N_{h1}} + \frac{\sigma_{h0}^2}{N_{h0}}\right)(N_{h1} + N_{h0}) \geq (\sigma_{h1} + \sigma_{h0})^2.$$

which is equal to

$$\left(\frac{\sigma_{h1}^2}{N_{h1}} + \frac{\sigma_{h0}^2}{N_{h0}}\right)(N_{h1} + N_{h0}) \geq \frac{(\sigma_{h1} + \sigma_{h0})^2}{N_h}$$

with equality if and only if

$$\frac{\sigma_{ht}}{N_{ht}^{1/2}} \propto N_{ht}^{1/2}, \quad t = 0, 1.$$

That is, the minimum of $Q(N_{h1}, N_{h0})$ is achieved at

$$N_{h1}^* = N_h \frac{\sigma_{h1}}{\sigma_{h1} + \sigma_{h0}}$$

and

$$N_{h0}^* = N_h - N_{h1}^*.$$

Problem 4 (10 pt)

Assume that a simple random sample of size n is selected from a population of size N and (x_i, y_i) are observed in the sample. In addition, we assume that the population mean of x, denoted by $\bar{X}$, is known.

1.

Use a Taylor linearization method to find the variance of the product estimator $\frac{\bar{x}\bar{y}}{\bar{X}}$, where $(\bar{x}, \bar{y})$ is the sample mean of (x_i, y_i).

Answer

The product estimator is:

$$\hat{\theta} = \frac{\bar{x}\bar{y}}{\bar{X}}.$$

Using Taylor linearization, we approximate $\hat{\theta}$ using a first-order expansion around the true means:

$$\hat{\theta} \approx \frac{\bar{Y}\bar{X} + (\bar{x} - \bar{X})\bar{Y} + (\bar{y} - \bar{Y})\bar{X}}{\bar{X}}.$$

Simplifying,

$$\hat{\theta} \approx \bar{Y} + (\bar{x} - \bar{X})\frac{\bar{Y}}{\bar{X}} + (\bar{y} - \bar{Y}).$$

Taking variances,

$$V(\hat{\theta}) \approx V(\bar{y}) + \frac{\bar{Y}^2}{\bar{X}^2}V(\bar{x}) + 2\frac{\bar{Y}}{\bar{X}}\text{Cov}(\bar{x}, \bar{y}).$$

Since in simple random sampling:

$$V(\bar{x}) = \frac{S_x^2}{n}\left(1 - \frac{n}{N}\right), \quad V(\bar{y}) = \frac{S_y^2}{n}\left(1 - \frac{n}{N}\right),$$

$$\text{Cov}(\bar{x}, \bar{y}) = \frac{S_{xy}}{n}\left(1 - \frac{n}{N}\right).$$

Thus,

$$V(\hat{\theta}) \approx \left(\frac{S_y^2}{n} + \frac{\bar{Y}^2}{\bar{X}^2}\frac{S_x^2}{n} + 2\frac{\bar{Y}}{\bar{X}}\frac{S_{xy}}{n}\right)\left(1 - \frac{n}{N}\right).$$

2.

Find the condition that this product estimator has a smaller variance than the sample mean $\bar{y}$.

Answer

For $\hat{\theta}$ to be more efficient than $\bar{y}$, we require:

$$V(\hat{\theta}) < V(\bar{y}).$$

Substituting,

$$\frac{S_y^2}{n} + \frac{\bar{Y}^2}{\bar{X}^2}\frac{S_x^2}{n} + 2\frac{\bar{Y}}{\bar{X}}\frac{S_{xy}}{n} < \frac{S_y^2}{n}.$$

Canceling S_y^2/n,

$$\frac{\bar{Y}^2}{\bar{X}^2}S_x^2 + 2\frac{\bar{Y}}{\bar{X}}S_{xy} < 0.$$

Rearranging,

$$2\bar{Y}S_{xy} < -\frac{\bar{Y}^2}{\bar{X}^2}S_x^2.$$

Dividing by $\bar{Y}$,

$$2S_{xy} < -\frac{\bar{Y}}{\bar{X}^2}S_x^2.$$

Thus, the product estimator has lower variance when the covariance between x and y is sufficiently negative.

3.

Prove that if the population covariance of x and y is zero, then the product estimator is less efficient than $\bar{y}$.

Answer

If $S_{xy} = 0$, then the variance formula simplifies to:

$$V(\hat{\theta}) = V(\bar{y}) + \frac{\bar{Y}^2}{\bar{X}^2}V(\bar{x}).$$

Since $\frac{\bar{Y}^2}{\bar{X}^2}V(\bar{x}) > 0$, it follows that:

$$V(\hat{\theta}) > V(\bar{y}).$$

Thus, when x and y are uncorrelated, the product estimator is less efficient than $\bar{y}$.

Problem 5 (10 pt)

In a population of 10,000 businesses, we want to estimate the average sales $\bar{Y}$. For that, we sample $n = 100$ businesses using simple random sampling. Furthermore, we have at our disposal the auxiliary information "number of employees", denoted by x, for each business. It is known that $\bar{X} = 50$ in the population. From the sample, we computed the following statistics:

- $\bar{y}_n = 5.2 \times 10^6$ (average sales in the sample)
- $\bar{x}_n = 45$ employees (sample mean)
- $s_y^2 = 25 \times 10^{10}$ (sample variance of y_k)
- $s_x^2 = 15$ (sample variance of x_k)
- $r = 0.8$ (sample correlation coefficient between x and y)

Answer the following questions:

1.

Compute a 95% confidence interval for $\bar{Y}$ using the ratio estimator.

Answer

The **ratio estimator** for the population mean sales is:

$$\hat{\bar{Y}}_R = \bar{y}_n \frac{\bar{X}}{\bar{x}_n}.$$

Substituting the given values:

$$\hat{\bar{Y}}_R = (5.2 \times 10^6) \times \frac{50}{45} = 5.778 \times 10^6.$$

The variance of the ratio estimator is approximated by:

$$V(\hat{\bar{Y}}_R) \approx \bar{Y}^2 \left(\frac{1}{n}\right) \left(\frac{s_y^2}{\bar{y}_n^2} + \frac{s_x^2}{\bar{x}_n^2} - 2r\frac{s_y}{\bar{y}_n}\frac{s_x}{\bar{x}_n}\right).$$

Substituting the values:

$$V(\hat{\bar{Y}}_R) = (5.778 \times 10^6)^2 \times \frac{1}{100} \left(\frac{25 \times 10^{10}}{(5.2 \times 10^6)^2} + \frac{15}{45^2} - 2(0.8)\frac{5 \times 10^5}{5.2 \times 10^6}\frac{3.873}{45}\right).$$

Computing the standard error, the **95% confidence interval** is given by:

$$\hat{\bar{Y}}_R \pm 1.96 \times \sqrt{V(\hat{\bar{Y}}_R)}.$$

2.

Compute a 95% confidence interval for $\bar{Y}$ using the regression estimator based on the simple linear regression of y on x (with intercept).

Answer

The **regression estimator** for the population mean is:

$$\hat{\bar{Y}}_{reg} = \bar{y}_n + b(\bar{X} - \bar{x}_n),$$

where the estimated slope b is given by:

$$b = r\frac{s_y}{s_x}.$$

Substituting the values:

$$b = 0.8 \times \frac{5 \times 10^5}{3.873} = 1.033 \times 10^5.$$

Thus, the regression estimator is:

$$\hat{\bar{Y}}_{reg} = (5.2 \times 10^6) + (1.033 \times 10^5)(50 - 45) = 5.7165 \times 10^6.$$

The variance of the regression estimator is:

$$V(\hat{\bar{Y}}_{reg}) = \frac{s_y^2}{n}(1 - r^2).$$

Substituting the values:

$$V(\hat{\bar{Y}}_{reg}) = \frac{25 \times 10^{10}}{100}(1 - 0.64) = 9 \times 10^9.$$

Computing the standard error, the **95% confidence interval** is given by:

$$\hat{\bar{Y}}_{reg} \pm 1.96 \times \sqrt{V(\hat{\bar{Y}}_{reg})}.$$

Problem 6 (10 pt)

Under the setup of Chapter 6, Part 1 lecture, prove the last two equalities on page 23:

$$\text{Cov}\left(\frac{1}{N_1}\sum_{i=1}^{N} T_i e_i(1), \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right) = 0$$

$$\text{Cov}\left(\frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)e_i(0), \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right) = 0$$

Answer

Setup

- T_i is the treatment indicator (1 for treated, 0 for control).
- $N_1 = \sum_{i=1}^{N} T_i$ is the number of treated units.
- $N_0 = \sum_{i=1}^{N}(1-T_i)$ is the number of control units.
- $e_i(1)$ and $e_i(0)$ are error terms under treatment and control.
- $\mathbf{x}_i$ is the covariate vector.
- $\mathbf{B}_0$ is a fixed coefficient vector.

Proof

Since treatment assignments T_i are independent of errors and covariates, we have:

$$E[T_i e_i(1)|\mathcal{F}_N] = \pi_i e_i(1), \quad E[(1-T_i)e_i(0)|\mathcal{F}_N] = (1-\pi_i)e_i(0)$$

$$E[T_i \mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N] = \pi_i \mathbf{x}_i'\mathbf{B}_0, \quad E[(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N] = (1-\pi_i)\mathbf{x}_i'\mathbf{B}_0$$

where $\pi_i = P(T_i = 1)$.

Expanding the first covariance:

$$\text{Cov}\left(\frac{1}{N_1}\sum_{i=1}^{N} T_i e_i(1), \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right)$$

Expanding linearly:

$$\frac{1}{N_1 N_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\text{Cov}\left(T_i e_i(1), (1-T_j)\mathbf{x}_j'\mathbf{B}_0|\mathcal{F}_N\right)$$

For $i \neq j$, T_i and $(1-T_j)$ are independent, making cross terms vanish. For $i = j$:

$$\text{Cov}(T_i e_i(1), (1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N) = E[T_i(1-T_i)|\mathcal{F}_N]E[e_i(1)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N]$$

Since $T_i(1-T_i) = 0$, the covariance is zero.

Similarly, for the second covariance:

$$\mathrm{Cov}\left(\frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)e_i(0), \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right)$$

Expanding:

$$\frac{1}{N_0^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\left((1-T_i)e_i(0), (1-T_j)\mathbf{x}_j'\mathbf{B}_0|\mathcal{F}_N\right)$$

For $i \neq j$, the terms are independent and vanish. For $i = j$:

$$\mathrm{Cov}\left((1-T_i)e_i(0), (1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right) = 0.$$

Thus, we have proven:

$$\mathrm{Cov}\left(\frac{1}{N_1}\sum_{i=1}^{N}T_ie_i(1), \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right) = 0$$

$$\mathrm{Cov}\left(\frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)e_i(0), \frac{1}{N_0}\sum_{i=1}^{N}(1-T_i)\mathbf{x}_i'\mathbf{B}_0|\mathcal{F}_N\right) = 0$$

These results follow from the independence of treatment assignments and errors, ensuring that their covariances vanish.

Problem 7 (10 pt)

Under the setup of Chapter 6, Part 2 lecture:

1.

Prove Lemma 3.

Lemma 3:

Let X be a $n \times p$ matrix such that

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$$

and $\omega = (\omega_1, \ldots, \omega_n)'$ be an n-dimensional weight vector ($n = N_1$). Given

$$\bar{x} = N^{-1} \sum_{i=1}^{N} x_i',$$

and D ($p \times p$ symmetric, invertible matrix), the minimizer of

$$Q(\omega) = \gamma(\omega' X - \bar{x})' D(\omega' X - \bar{x}) + \omega' \omega$$

$$= \gamma(X'\omega - \bar{x})' D(X'\omega - \bar{x}) + \omega'\omega$$

is given by

$$\hat{\omega} = (\gamma X D X' + I_n)^{-1} \gamma X D \bar{x} \tag{10}$$

$$= X(X'X + \gamma^{-1} D^{-1})^{-1} \bar{x} \tag{11}$$

Answer

We seek to minimize the quadratic objective function:

$$Q(\omega) = \gamma(X'\omega - \bar{x})' D(X'\omega - \bar{x}) + \omega'\omega.$$

Taking the gradient with respect to ω:

$$\frac{dQ}{d\omega} = 2\gamma X D(X'\omega - \bar{x}) + 2\omega.$$

Setting this to zero:

$$\gamma X D X'\omega - \gamma X D \bar{x} + \omega = 0.$$

Rearranging:

$$(\gamma XDX' + I_n)\omega = \gamma XD\bar{x}.$$

Multiplying by $(\gamma XDX' + I_n)^{-1}$:

$$\hat{\omega} = (\gamma XDX' + I_n)^{-1}\gamma XD\bar{x}.$$

which proves equation (10), giving us Lemma 3.

2.

Show that the final weight in (13) satisfies a hard calibration for $\mathbf{x}_1$:

$$\sum_{i \in A} \hat{\omega}_i \mathbf{x}_{1i} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{1i}.$$

(9):

The implicit model is that

$$Y(1) = x_1' \beta + x_2' u + e(1) \tag{9}$$

where $u \sim (0, D_q \sigma_u^2)$ with known D_q and $e(1) \sim (0, \sigma_e^2)$.

(10-11): Given in Lemma 3

(12):

Using (11), the solution can be written as

$$\hat{\omega} = X \left(X'X + \Omega^{-1} \right)^{-1} \bar{x} \tag{12}$$

where $\Omega^{-1} = \text{Diag}\{\gamma_1^{-1} D_p^{-1}, \gamma_2^{-1} D_q^{-1}\}$ and $\gamma_1 \to \infty$.

(13):

Under the mixed model setup in (9), the solution (12) can be written as

$$\hat{\omega}_i = \left(N^{-1} \sum_{i=1}^{N} x_i \right)' \left\{ \sum_{i=1}^{N} T_i x_i x_i' + \Omega^{-1} \right\}^{-1} x_i, \tag{13}$$

where $\Omega^{-1} = \text{Diag}\{0_p, \gamma_2^{-1} D_q^{-1}\}$ and $\gamma_2 = \sigma_u^2 / \sigma_e^2$.

Answer

We express the solution using the Woodbury identity:

$$(\gamma X D X' + I_n)^{-1} \gamma X D = X (X'X + \gamma^{-1} D^{-1})^{-1}.$$

Thus, we obtain equation (11):

$$\hat{\omega} = X (X'X + \gamma^{-1} D^{-1})^{-1} \bar{x}.$$

We need to show:

$$\sum_{i \in A} \hat{\omega}_i \mathbf{x}_{1i} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{1i}.$$

From equation (13):

$$\hat{\omega}_i = \left(N^{-1} \sum_{i=1}^{N} x_i \right)' \left\{ \sum_{i=1}^{N} T_i x_i x_i' + \Omega^{-1} \right\}^{-1} x_i.$$

Summing over $i \in A$:

$$\sum_{i \in A} \hat{\omega}_i \mathbf{x}_{1i} = \sum_{i \in A} \left(N^{-1} \sum_{i=1}^{N} x_i \right)' \left\{ \sum_{i=1}^{N} T_i x_i x_i' + \Omega^{-1} \right\}^{-1} x_i \mathbf{x}_{1i}.$$

Since $\Omega^{-1} = \mathrm{Diag}\{0_p, \gamma_2^{-1} D_q^{-1}\}$, for large γ_1, the term simplifies to:

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{1i}.$$

Thus, the final weight satisfies hard calibration.