ESMA 6787: Homework 1

Alejandro M. Ouslan

Due Date: September 4, 2025

Problem 1: Syllabus Acknowledgment

I have read the syllabus, understand its contents,and have no questions.

Problem 2: Definitions in Your Own Words

- **Experiment:** Method to answer a research question, where the researcher controles an environment and only changes one or many treatment holding all other things constant
- **Experimental unit:** unit that will receive the treatment
- **Observational unit:** it is the person, thing or event where are studying.
- **Background variable:** Are all the factor that can affect the study but are not controlled for
- **Independent (predictor) variable:** It is what is applied as treatment
- **Dependent (response) variable:** It is the outcome or what we observe after we apply the treatment
- **Confounded factors:** Are un observe factors that affect both the treatment and the observed factor
- **Experimental error:** It is how far our experiment is from the reality.
- **Randomization:** Treatments are applied randomly
- **Replicate:** recreation of the experiments with the same settings and specification but with different data

Problem 3: Lady Tasting Tea

(a) Units in this experiment: The tea cups

(b) Treatments in this experiment: The order of what was added first to the tea

(c) Randomization method using physical devices: order cups from 1-8 then randomize the order with a program

(d) Adjustments if cups differ in material (porcelain vs china): Assigned cups type randomly to each group

Problem 4: Paper Airplane Experiment

(a) Experimental treatments: The experiments would be each of the 3 recipe, with the simple classic airplane as the controle group

(b) Experimental units and homogeneity: to ensure homogeneity we would use the same type paper, thrown from the same location.

(c) Randomization process: when we get the 20 papers we assign the folding recepte randomly to them

(d) Procedure for applying treatment to unit: when the recepte is chacen use the as much possible machines or tools ensure cons

(e) Measurement process: Mesure the distance the plane flew

Problem 5: Gasoline Mileage Study

(a) Comparison of strengths and weaknesses: Design experiment A is convenient but their are many variables that are not controlled like the brand of car, driving habits, etc. Design B has more strengths given that we use the same care for the same amount of time but a problem that could have the experiment is segregation of the vie vale over time. Design C addresses the concerns of the previouse experiemnt.

(b) Identification of true experiment(s) and justification: Experiment C would say it is the best experiment given that it controles for the most amount of factors buy using the same vehicle.

Problem 6: Baseball League as Experiment

- Treatments and units: The teams and the unites are the games
- Application of treatment to unit: The application of the treatment would be to that a particular team is playing at its expectations of victory is the outcome
- Randomization and replication: Randomizing the matching of the team and ensuring that each team plays each other for the replications
- Possibility and use of blocking: Could be used if there are confounding factors like home-field advantage or team skill differences, but it might not be necessary if teams are assumed to be comparable.

Problem 7: Tomato Fertilizer and Variety

- Experimental setup: We would devide the plot of land into quadrents and each will have a groups with all combinations of Fertilizer and seed type. In addition it will consist of a control group with no Fertilizer and a comon tomato seed.
- Use of replication and randomization: The randomization is applied to which section of the plot of land receives the combination of Fertilizer and seed
- Additional design principles in second season: rotate the treatments, long-term replication, seasonal variation consideration, increased precision in measurement and control of weather variability.

Problem 8: Hand Washing Experiment

(a) Experimental unit: The experimental unit is the individual subject (person) participating in the hand washing experiment

(b) Factors: Wash water temperature, detergent concentration

(c) Response:The response is the bacterial count on the palms of the subjects.

Problem 9: Real-life Application

I would say most economic problem would benefit form experiments designs given that it changes the results from corrolation to causality.

Problem 10: Variance as Quadratic Form

Suppose $y_1, \dots, y_n \sim N(\mu, \sigma^2)$. Let $y = (y_1, \dots, y_n)'$ and let $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Show that $\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2$ can be written as $y'Ay$ for sum matrix A. Identify the A matrix.

$$\frac{1}{n-1}(y - \bar{y}\mathbf{1})'(y - \bar{y}\mathbf{1})$$

$$y - \bar{y}\mathbf{1} = y - \left(\frac{1}{n}\mathbf{1}'y\right)\mathbf{1} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)y$$

$$\frac{1}{n-1}(y - \bar{y}\mathbf{1})'(y - \bar{y}\mathbf{1}) = \frac{1}{n-1}y'C'Cy$$

$$\frac{1}{n-1}y'Cy$$

$$A = \frac{1}{n-1}\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)$$

Problem 11: Cell Means Model with Unequal Group Sizes

consider a cells means model with $T = 4$ treatments and $n_1 = n_2 = 5, n_3 = 3, n_4 = 7$

(a) Proposed cell means model:

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, 3, 4$$

(b) Design matrix X and its rank:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \text{rank}(X) = 4$$

(c) Computation of $X'X$:

$$X'X = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

(d) OLS estimates as a function of y, i.e, $\hat{\beta} = (X'X)_{-}X'y$:

$$\hat{\beta} = \begin{bmatrix} \frac{1}{5}\sum_{j=1}^5 y_{1j} \\ \frac{1}{5}\sum_{j=1}^5 y_{2j} \\ \frac{1}{3}\sum_{j=1}^3 y_{3j} \\ \frac{1}{7}\sum_{j=1}^7 y_{4j} \end{bmatrix}$$

(e) Analysis of Table 1 dataset:

i. OLS estimates:

$$\hat{\mu}_1 = \frac{4.25 + 3.92 + 8.48 + 6.66 + 1.15}{5} = 4.89,$$
$$\hat{\mu}_2 = \frac{9.55 + 6.08 + 11.53 + 4.78 + 6.18}{5} = 7.224$$
$$\hat{\mu}_3 = \frac{15.74 + 11.70 + 10.76}{3} = 12.7333,$$
$$\hat{\mu}_4 = \frac{14.85 + 14.34 + 13.89 + 14.40 + 15.53 + 12.74 + 12.97}{7} = 14.103$$

ii. Projection matrix P_X:

$$P_X = X(X'X)^{-1}X', \quad \text{where } X \text{ is the design matrix from part (b)}$$

Since $X'X$ is diagonal, its inverse is:

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{5} & 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{7} \end{bmatrix}$$

So, P_X projects each observation to its group mean.

iii. Compute $y'(I - P_X)y$:

$$y'(I - P_X)y = \sum_{i=1}^{4}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i)^2$$
$$= \sum_{j=1}^{5}(y_{1j} - 4.89)^2 + \sum_{j=1}^{5}(y_{2j} - 7.224)^2 + \sum_{j=1}^{3}(y_{3j} - 12.7333)^2 + \sum_{j=1}^{7}(y_{4j} - 14.103)^2 = 111.0322$$

iv. Compute $\bar{y}_{..}$ and $\bar{y}_{i.}$ for all i:

$$\bar{y}_{..} = \frac{1}{20}\sum y = \frac{261.77}{20} = 13.0885$$
$$\bar{y}_{1.} = 4.89, \quad \bar{y}_{2.} = 7.224, \quad \bar{y}_{3.} = 12.7333, \quad \bar{y}_{4.} = 14.103$$

v. Estimability of μ_1:

$$\mu_1 \text{ is estimable since it corresponds to a column in } X$$

vi. Estimability of $\mu_2 - \mu_3$:

$$\mu_2 - \mu_3 \text{ is estimable since both } \mu_2 \text{ and } \mu_3 \text{ are estimable}$$

vii. Estimability of $\mu_1 - \frac{\mu_2+\mu_3}{2}$:

$$\mu_1 - \frac{\mu_2 + \mu_3}{2} \text{ is a linear combination of estimable functions} \Rightarrow \text{estimable}$$

Problem 12: Fixed-Effect Model with Unequal Group Sizes

(a) Proposed fixed-effect model:

(b) Design matrix X and its rank:

(c) Computation of $X'X$:

(d) OLS estimates as a function of y:

(e) Using Table 1 dataset:

i. OLS estimates:

ii. Projection matrix P_X:

iii. Compute $y'(I - P_X)y$:

iv. Compute $\bar{y}_{\cdot\cdot}$ and $\bar{y}_{i\cdot}$ for all i:

v. Estimability of α_1:

vi. Estimability of $\alpha_2 - \alpha_3$:

vii. Estimability of $\alpha_1 - \frac{\alpha_3+\alpha_4}{2}$: