

STAT 510 Homework 2

Due Date: 11:00 A.M., Wednesday, January 25

1. Suppose \mathbf{X} is an $n \times p$ matrix and \mathbf{y} is an $n \times 1$ vector. Suppose $\mathbf{z} \in \mathcal{C}(\mathbf{X})$ and $\mathbf{z} \neq \mathbf{P}_\mathbf{X}\mathbf{y}$. Prove that $\|\mathbf{y} - \mathbf{z}\| > \|\mathbf{y} - \mathbf{P}_\mathbf{X}\mathbf{y}\|$. Hint: Note that for any vector \mathbf{a} and any vector $\mathbf{b} \neq \mathbf{0}$ such that $\mathbf{a}'\mathbf{b} = 0$

$$\begin{aligned}\|\mathbf{a} + \mathbf{b}\|^2 &= (\mathbf{a} + \mathbf{b})'(\mathbf{a} + \mathbf{b}) = (\mathbf{a}' + \mathbf{b}')(\mathbf{a} + \mathbf{b}) \\ &= \mathbf{a}'\mathbf{a} + \mathbf{a}'\mathbf{b} + \mathbf{b}'\mathbf{a} + \mathbf{b}'\mathbf{b} = \mathbf{a}'\mathbf{a} + 2\mathbf{a}'\mathbf{b} + \mathbf{b}'\mathbf{b} \\ &= \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 + 2\mathbf{a}'\mathbf{b} \\ &= \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 \text{ (because } \mathbf{a}'\mathbf{b} = 0\text{).} \\ &> \|\mathbf{a}\|^2 \text{ (because } \mathbf{b} \neq \mathbf{0}\text{).}\end{aligned}$$

Now note that

$$\|\mathbf{y} - \mathbf{z}\|^2 = \|\mathbf{y} - \mathbf{P}_\mathbf{X}\mathbf{y} + \mathbf{P}_\mathbf{X}\mathbf{y} - \mathbf{z}\|^2 = \dots$$

2. Suppose \mathbf{X} is an $n \times p$ matrix. Prove that $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{P}_\mathbf{X})$.
3. Prove that $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ is a solution to the normal equations (see slide 8 of slide set 2 for the definition of the normal equations).
4. Suppose the Gauss-Markov model with normal errors holds (see slide 14 of slide set 2 for a precise statement of the model).
- Suppose $\mathbf{C}\beta$ is estimable. Derive the distribution of $\mathbf{C}\hat{\beta}$, the OLSE of $\mathbf{C}\beta$.
 - Now suppose $\mathbf{C}\beta$ is NOT estimable. Provide a fully simplified expression for $\text{Var}(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y})$.
 - Now suppose $H_0: \mathbf{C}\beta = \mathbf{d}$ is testable and that \mathbf{C} has only one row and \mathbf{d} has only one element so that they may be written as c' and d , respectively. Prove the result on slide 21 of slide set 2.
5. Consider a competition among 5 table tennis players labeled 1 through 5. For $1 \leq i < j \leq 5$, define y_{ij} to be the score for player i minus the score for player j when player i plays a game against player j . Suppose for $1 \leq i < j \leq 5$,

$$y_{ij} = \beta_i - \beta_j + \epsilon_{ij}, \tag{1}$$

where β_1, \dots, β_5 are unknown parameters and the ϵ_{ij} terms are random errors with mean 0. Suppose four games will be played that will allow us to observe y_{12} , y_{34} , y_{25} , and y_{15} . Let

$$\mathbf{y} = \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \quad \text{and } \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}.$$

- Define a design matrix \mathbf{X} so that model (1) may be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
- Is $\beta_1 - \beta_2$ estimable? Prove that your answer is correct.
- Is $\beta_1 - \beta_3$ estimable? Prove that your answer is correct.
- Find a generalized inverse of $\mathbf{X}'\mathbf{X}$.
- Write down a general expression for the normal equations.
- Find a solution to the normal equations in this particular problem involving table tennis players.
- Find the Ordinary Least Squares (OLS) estimator of $\beta_1 - \beta_5$.
- Give a linear unbiased estimator of $\beta_1 - \beta_5$ that is not the OLS estimator.