STAT 510 Homework 2
Due Date: 11:00 A.M., Wednesday, January 25

1. Suppose $\boldsymbol{X}$ is an $n \times p$ matrix and $\boldsymbol{y}$ is an $n \times 1$ vector. Suppose $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X})$ and $\boldsymbol{z} \neq \boldsymbol{P_X y}$. Prove that $\|\boldsymbol{y}-\boldsymbol{z}\| > \|\boldsymbol{y}-\boldsymbol{P_X y}\|$. Hint: Note that for any vector $\boldsymbol{a}$ and any vector $\boldsymbol{b} \neq \boldsymbol{0}$ such that $\boldsymbol{a}'\boldsymbol{b} = 0$
$$\begin{aligned}\|\boldsymbol{a}+\boldsymbol{b}\|^2 &= (\boldsymbol{a}+\boldsymbol{b})'(\boldsymbol{a}+\boldsymbol{b}) = (\boldsymbol{a}'+\boldsymbol{b}')(\boldsymbol{a}+\boldsymbol{b}) \\ &= \boldsymbol{a}'\boldsymbol{a}+\boldsymbol{a}'\boldsymbol{b}+\boldsymbol{b}'\boldsymbol{a}+\boldsymbol{b}'\boldsymbol{b} = \boldsymbol{a}'\boldsymbol{a}+2\boldsymbol{a}'\boldsymbol{b}+\boldsymbol{b}'\boldsymbol{b} \\ &= \|\boldsymbol{a}\|^2+\|\boldsymbol{b}\|^2+2\boldsymbol{a}'\boldsymbol{b} \\ &= \|\boldsymbol{a}\|^2+\|\boldsymbol{b}\|^2 \text{ (because } \boldsymbol{a}'\boldsymbol{b}=0). \\ &> \|\boldsymbol{a}\|^2 \text{ (because } \boldsymbol{b}\neq\boldsymbol{0}).\end{aligned}$$
Now note that
$$\|\boldsymbol{y}-\boldsymbol{z}\|^2 = \|\boldsymbol{y}-\boldsymbol{P_X y}+\boldsymbol{P_X y}-\boldsymbol{z}\|^2 = \dots$$

2. Suppose $\boldsymbol{X}$ is an $n \times p$ matrix. Prove that $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{P_X})$.

3. Prove that $(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y}$ 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).

   (a) Suppose $\boldsymbol{C\beta}$ is estimable. Derive the distribution of $\boldsymbol{C}\hat{\boldsymbol{\beta}}$, the OLSE of $\boldsymbol{C\beta}$.
   (b) Now suppose $\boldsymbol{C\beta}$ is NOT estimable. Provide a fully simplified expression for $\mathrm{Var}\left(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y}\right)$.
   (c) Now suppose $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ is testable. Prove the result on slide 21 of slide set 2.

5. Consider a competition among 5 table tennis players labeled 1 through 5. For $1 \le i < j \le 5$, 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 $1 \le i < j \le 5$,
$$y_{ij} = \beta_i - \beta_j + \epsilon_{ij}, \tag{1}$$
where $\beta_1, \dots, \beta_5$ 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
$$\boldsymbol{y} = \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \quad \text{and } \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}.$$

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