STAT 510 Homework 2 Solutions Spring 2017

1. Let $\boldsymbol{X}$ be an $n \times p$ matrix and $\boldsymbol{y}$ be an $n \times 1$ vector. Suppose that $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X})$ and $\boldsymbol{z} \neq \boldsymbol{P_X y}$, which implies $(\boldsymbol{P_X y} - \boldsymbol{z}) \neq \boldsymbol{0}_{n\times 1}$. Observe that $\boldsymbol{z} \in \mathcal{C}(\boldsymbol{X})$ implies that $\boldsymbol{P_X z} = \boldsymbol{z}$. Using this result and the fact that $\boldsymbol{P_X}$ is symmetric and idempotent, it follows that

$$\begin{aligned}
(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) &= (\boldsymbol{y}' - [\boldsymbol{P_X y}]')(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= (\boldsymbol{y}' - \boldsymbol{y}'\boldsymbol{P}'_{\boldsymbol{X}})(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= (\boldsymbol{y}' - \boldsymbol{y}'\boldsymbol{P_X})(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= \boldsymbol{y}'\boldsymbol{P_X y} - \boldsymbol{y}'\boldsymbol{z} - \boldsymbol{y}'\boldsymbol{P_X P_X y} + \boldsymbol{y}'\boldsymbol{P_X z} \\
&= \boldsymbol{y}'\boldsymbol{P_X y} - \boldsymbol{y}'\boldsymbol{z} - \boldsymbol{y}'\boldsymbol{P_X y} + \boldsymbol{y}'\boldsymbol{P_X z} \\
&= -\boldsymbol{y}'\boldsymbol{z} + \boldsymbol{y}'\boldsymbol{z} \\
&= 0.
\end{aligned}$$

Now that we have $(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) = 0$ and $(\boldsymbol{P_X y} - \boldsymbol{z}) \neq \boldsymbol{0}$, we can use the same argument provided in the homework with $\boldsymbol{a} = \boldsymbol{y} - \boldsymbol{P_X y}$ and $\boldsymbol{b} = \boldsymbol{P_X y} - \boldsymbol{z}$:

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{z}\|^2 &= \|\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z}\|^2 \\
&= (\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z})'(\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z}) \\
&= \big((\boldsymbol{y} - \boldsymbol{P_X y})' + (\boldsymbol{P_X y} - \boldsymbol{z})'\big)\big(\boldsymbol{y} - \boldsymbol{P_X y} + \boldsymbol{P_X y} - \boldsymbol{z}\big) \\
&= (\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{y} - \boldsymbol{P_X y}) + 2(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) + (\boldsymbol{P_X y} - \boldsymbol{z})'(\boldsymbol{P_X y} - \boldsymbol{z}) \\
&= \|\boldsymbol{y} - \boldsymbol{P_X y}\|^2 + \|\boldsymbol{P_X y} - \boldsymbol{z}\|^2 \\
&> \|\boldsymbol{y} - \boldsymbol{P_X y}\|^2 .
\end{aligned}$$

Hence, $\|\boldsymbol{y} - \boldsymbol{z}\| > \|\boldsymbol{y} - \boldsymbol{P_X y}\|$, which says that $\boldsymbol{P_X y}$ is the unique point in $\mathcal{C}(\boldsymbol{X})$ that is closest to $\boldsymbol{y}$ in Euclidean distance.

Comments: You can instead show that $(\boldsymbol{y} - \boldsymbol{P_X y})'(\boldsymbol{P_X y} - \boldsymbol{z}) = 0$ by orthogonality, but as this is a proof, you need to provide sufficient reasoning or work to establish this.

2. Key:
   1. $\underset{n\times 1}{\boldsymbol{a}} \in \mathcal{C}(\boldsymbol{X}) \iff \boldsymbol{a} = \boldsymbol{X} \underset{p\times 1}{\boldsymbol{b}}$ for some $\boldsymbol{b}$
   2. $\boldsymbol{P_X X} = \boldsymbol{X}$ by property of projection matrix

   Prove that $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{P_X})$:

$$\begin{aligned}
\boldsymbol{a} \in \mathcal{C}(\boldsymbol{X}) &\iff \boldsymbol{a} = \boldsymbol{X b} && \text{for some } \boldsymbol{b} \text{ by key 1} \\
&\iff \boldsymbol{a} = \underbrace{\boldsymbol{P_X X}}_{\boldsymbol{X}} \boldsymbol{b} && \text{for some } \boldsymbol{b} \text{ by key 2} \\
&\iff \boldsymbol{a} = \boldsymbol{P_X} \underbrace{\boldsymbol{X b}}_{n\times 1} && \text{treat as } \boldsymbol{P_X} \text{ product a } n \times 1 \text{ vector} \\
&\implies \boldsymbol{a} \in \mathcal{C}(\boldsymbol{P_X}) && \text{by key 1}
\end{aligned}$$

   So $\mathcal{C}(\boldsymbol{X}) \subseteq \mathcal{C}(\boldsymbol{P_X})$.

then similarly,

$$
\begin{aligned}
\boldsymbol{g} \in \mathcal{C}(\boldsymbol{P_X}) &\iff \boldsymbol{g} = \boldsymbol{P_X h} && \text{for some } n \times 1 \text{ vector } \boldsymbol{h} \text{ by key 1} \\
&\iff \boldsymbol{g} = \underbrace{\boldsymbol{X}(\boldsymbol{X'X})^{-}\boldsymbol{X'}}_{\boldsymbol{P_X}}\boldsymbol{h} && \text{for some } \boldsymbol{h} \\
&\iff \boldsymbol{g} = \boldsymbol{X}\underbrace{(\boldsymbol{X'X})^{-}\boldsymbol{X'h}}_{p\times 1} && \text{treat as } \boldsymbol{X} \text{ product a } p \times 1 \text{ vector} \\
&\implies \boldsymbol{g} \in \mathcal{C}(\boldsymbol{P_X}) && \text{by key 1}
\end{aligned}
$$

So $\mathcal{C}(\boldsymbol{P_X}) \subseteq \mathcal{C}(\boldsymbol{X})$. According to the results above, $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{P_X})$.

3. Prove $(\boldsymbol{X'X})^{-}\boldsymbol{X'y}$ is a solution to the normal equations $\boldsymbol{X'Xb} = \boldsymbol{X'y}$ (by slide 8 of set 2).

   Let $\boldsymbol{b} = (\boldsymbol{X'X})^{-}\boldsymbol{X'y}$:

$$
\begin{aligned}
\boldsymbol{X'Xb} &= \boldsymbol{X'}\underbrace{\boldsymbol{X}(\boldsymbol{X'X})^{-}\boldsymbol{X'}}_{\boldsymbol{P_X}}\boldsymbol{y} \\
&= \boldsymbol{X'P_X y} \\
&= \boldsymbol{X'y} && \boldsymbol{X'P_X} = \boldsymbol{X'} \text{ by property of projection matrix in slice 5 of set 2}
\end{aligned}
$$

   Therefore $(\boldsymbol{X'X})^{-}\boldsymbol{X'y}$ is a solution to 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 $\boldsymbol{C\hat{\beta}}$, the OLSE of $\boldsymbol{C\beta}$.

   $\boldsymbol{C\beta}$ is estimable $\implies$ there exists $\boldsymbol{A}$ that $\boldsymbol{C} = \boldsymbol{AX}$

$$
\begin{aligned}
\boldsymbol{C\hat{\beta}} &= \boldsymbol{C}(\boldsymbol{X'X})^{-}\boldsymbol{X'y} \\
&= \boldsymbol{AX}(\boldsymbol{X'X})^{-}\boldsymbol{X'y} && \boldsymbol{C} = \boldsymbol{AX} \\
&= \boldsymbol{AP_X y} && \boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X'X})^{-}\boldsymbol{X'}
\end{aligned}
$$

   Base on the model assumption, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$. Then $\boldsymbol{C\hat{\beta}} = \boldsymbol{AP_X y}$ is also multivariate normal by silce 24 of set 1, $\boldsymbol{AP_X y} \sim \mathcal{N}(\boldsymbol{AP_X X\beta}, \boldsymbol{AP_X}\sigma^2\boldsymbol{I}(\boldsymbol{AP_X})')$

$$
\boldsymbol{AP_X X\beta} = \boldsymbol{AX\beta} = \boldsymbol{C\beta}
$$

$$
\begin{aligned}
\boldsymbol{AP_X}\sigma^2\boldsymbol{I}(\boldsymbol{AP_X})' &= \sigma^2\boldsymbol{AP_X P_X' A'} \\
&= \sigma^2\boldsymbol{AP_X A'} && \boldsymbol{P_X} \text{ is symmetric and idempotent} \\
&= \sigma^2\boldsymbol{AX}(\boldsymbol{X'X})^{-}\boldsymbol{X'A'} \\
&= \sigma^2\boldsymbol{C}(\boldsymbol{X'X})^{-}\boldsymbol{C'}
\end{aligned}
$$

   Therefore $\boldsymbol{C\hat{\beta}} \sim \mathcal{N}(\boldsymbol{C\beta}, \sigma^2\boldsymbol{C}(\boldsymbol{X'X})^{-}\boldsymbol{C'})$.

(b) Now suppose $\boldsymbol{C\beta}$ is NOT estimable.

$$\begin{aligned} Var(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}) &= (\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}')\sigma^2\boldsymbol{I}(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}')' \\ &= \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-'}\boldsymbol{C}' \end{aligned}$$

We can not simply this further when $\boldsymbol{C\beta}$ is NOT estimable.

(c) Now suppose $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ is testable. Prove the result on slide 21 of set 2.

Given the hypothesis is testable (see slide 16 of set 2), $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is estimable and from the resluts in part (a), we have $\boldsymbol{c}'\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{c}'\boldsymbol{\beta}, \sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c})$, by linear transformation,

$$\frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}} \sim \mathcal{N}(\frac{\boldsymbol{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}, 1)$$

let $u = \frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}$ and $\delta = \frac{\boldsymbol{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}$, $u \sim \mathcal{N}(\delta, 1)$.

Then by slide 17 of set 2,

$$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{n-r}}{n-r} \implies w = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$$

$\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent, so their function u and w are also independent (see Theorem 4.3.5 in Casella and Berger,2002). By slide 29 of set 1,

$$\frac{u}{\sqrt{w/(n-r)}} = \frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - d}{\sqrt{\hat{\sigma}^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}} \sim t_{n-r}(\delta)$$

Therefore it follows a t distribution with non-central parameter $\delta = \frac{\boldsymbol{c}'\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{c}}}$ and degrees of freedom $n - r$.

Note: The independence between u and w are necessary, we can first check $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$. $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ is estimable so we can write it as $\boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{a}'\boldsymbol{P_X}\boldsymbol{y}$ for some $\boldsymbol{a}'$, which is a function of $\boldsymbol{P_X}\boldsymbol{y}$, and $\hat{\sigma}^2 = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/(n-r) = ||(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}||^2/(n-r)$ which is a function of $(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}$.

By **the independence results on slide 34 in set 1**, when $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ in GMMNE (slide 16 of set 2), let $\boldsymbol{A}_1 = \boldsymbol{P_X}$, and $\boldsymbol{A}_2 = \boldsymbol{I} - \boldsymbol{P_X}$.

$$\begin{aligned} \boldsymbol{A}_1\sigma^2\boldsymbol{I}\boldsymbol{A}_2' &= \boldsymbol{P_X}\sigma^2\boldsymbol{I}(\boldsymbol{I} - \boldsymbol{P_X})' \\ &= \sigma^2\boldsymbol{P_X}(\boldsymbol{I} - \boldsymbol{P_X})' \\ &= \sigma^2\boldsymbol{P_X}(\boldsymbol{I} - \boldsymbol{P_X}) \\ &= \sigma^2(\boldsymbol{P_X} - \boldsymbol{P_X}\boldsymbol{P_X}) \\ &= 0 \qquad\qquad \text{because } \boldsymbol{P_X} \text{ is idempotent and symmetric} \end{aligned}$$

then by Theorem 4.3.5 in Casella and Berger(2002), we have
$\boldsymbol{P_X}\boldsymbol{y} \perp (\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \implies \boldsymbol{c}'\hat{\boldsymbol{\beta}} \perp \hat{\sigma}^2 \implies u \perp w$.

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 $\beta_1, \ldots, \beta_5$ 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}.

(a) Define a design matrix $\boldsymbol{X}$ so that model (1) may be written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

$$\boldsymbol{y} = \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_4 - \beta_4 \\ \beta_2 - \beta_5 \\ \beta_1 - \beta_5 \end{bmatrix} + \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}}_{\boldsymbol{X}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}$$

(b) Is $\beta_1 - \beta_2$ estimable?
Let $\boldsymbol{c}' = (1, -1, 0, 0, 0)$, then $\beta_1 - \beta_2$ can be written as $\boldsymbol{c}'\boldsymbol{\beta}$.
If $\boldsymbol{c}'\boldsymbol{\beta}$ is estimable, there exists a $\boldsymbol{a}'$ so that $\boldsymbol{c}' = \boldsymbol{a}'\boldsymbol{X}$ by slide 7 of set 2. We can find such $\boldsymbol{a}' = (1, 0, 0, 0)'$, so $\beta_1 - \beta_2$ is estimable.

(c) Is $\beta_1 - \beta_3$ estimable?
Let $\boldsymbol{c}_2' = (1, 0, -1, 0, 0)$, then $\beta_1 - \beta_3$ can be written as $\boldsymbol{c}_2'\boldsymbol{\beta}$.
Assume $\boldsymbol{c}_2'\boldsymbol{\beta}$ is estimable, then there must be an $\boldsymbol{a}_2' = (a_1, a_2, a_3, a_4)$ so that

$$\boldsymbol{c}_2' = \boldsymbol{a}_2'\boldsymbol{X}$$

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

which implies

$$1 = a_1 + a_4; \quad 0 = -a_1 + a_3; \quad \underbrace{-1 = a_2; \quad 0 = -a_2;}_{contradiction!} \quad 0 = -a_3 - a_4 \tag{2}$$

So $\beta_1 - \beta_3$ is not estimable.

Comments: Some students claimed that they can not find an $\boldsymbol{a}' = \boldsymbol{c}'\boldsymbol{X}$ but without any further proof, you need show work to support your statement.

(d) Find a generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$.
Use the `R` function `ginv` in the `MASS` package.

```
> X=matrix(c(1,0,0,1,-1,0,1,0,0,1,0,0,0,-1,0,0,0,0,-1,-1),nrow=4)
> MASS::ginv( t(X) %*% X)
```

a generalized inverse matrix $(\boldsymbol{X}'\boldsymbol{X})^-$ (not unique) is

$$(\boldsymbol{X}'\boldsymbol{X})^- = \begin{bmatrix} 0.2222 & -0.1111 & 0 & 0 & -0.1111 \\ -0.1111 & 0.2222 & 0 & 0 & -0.1111 \\ 0 & 0 & 0.25 & -0.25 & 0 \\ 0 & 0 & -0.25 & 0.25 & 0 \\ -0.1111 & -0.1111 & 0 & 0 & 0.2222 \end{bmatrix}$$

(e) Write down a general expression for the normal equations.

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$$

(f) Find a solution to the normal equations in this particular problem involving table tennis players.

$$\begin{aligned} \boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y} \\ &= \begin{bmatrix} 1/3 & 0 & 0 & 1/3 \\ -1/3 & 0 & 1/3 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & -1/2 & 0 & 0 \\ 0 & 0 & -1/3 & 1/3 \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} y_{12}/3 + y_{15}/3 \\ -y_{12}/3 + y_{25}/3 \\ y_{34}/2 \\ -y_{34}/2 \\ -y_{25}/3 - y_{15}/3 \end{bmatrix} \end{aligned}$$

The solution is not unique since $(\boldsymbol{X}'\boldsymbol{X})^-$ is not unique.

Comments: In this problem, you need to write out $\boldsymbol{b}$ as a 5×1 matrix involving y_{ij} instead of dot product of two matrices.

(g) Find the Ordinary Least Squares (OLS) estimator of $\beta_1 - \beta_5$.

Let $\boldsymbol{c}' = (1, 0, 0, 0, -1)$ then $\beta_1 - \beta_5 = \boldsymbol{c}'\boldsymbol{\beta}$.
The OLSE of $\boldsymbol{c}'\boldsymbol{\beta}$ is $\boldsymbol{c}'\boldsymbol{b}$ by slide 7 of set 2, based on the resluts in part (f),

$$\boldsymbol{c}'\boldsymbol{b} = y_{12}/3 + 2y_{15}/3 + y_{25}/3$$

Note: $\boldsymbol{c}'\boldsymbol{b}$ is unique since it is estimable and can be written as $\boldsymbol{a}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{a}'\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$, where $\boldsymbol{P}_{\boldsymbol{X}}$ is unique by slide 5 of set 2.

(h) Give a linear unbiased estimator of $\beta_1 - \beta_5$ that is not the OLS estimator.

We need to find $\boldsymbol{a}'$ that $E(\boldsymbol{a}'\boldsymbol{y}) = \beta_1 - \beta_5$, the simpliest one is y_{15} when $\boldsymbol{a}' = (0, 0, 0, 1)$, y_{15} is linear unbiased but not the OLSE of $\beta_1 - \beta_5$.