Stat 5100 **Assignment 2**

Due: Wednesday, February 5th 11:59PM in gradescope.

Problem 1. *Suppose* $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *where* $\boldsymbol{\mu}^\top = \begin{bmatrix}1 & 2 & 3\end{bmatrix}$

$$\boldsymbol{\mu}^\top = \begin{bmatrix}1 & 2 & 3\end{bmatrix} \quad \textit{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix}.$$

Further, define a 3×3 *matrix* $\mathbf{A}$ *and a* 2×3 *matrix* $\mathbf{B}$ *as follows*

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad \textit{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}.$$

a) *Determine the distribution of* $u = \mathbf{1}_3^\top \mathbf{y}$.

b) *Determine the distribution of* $\mathbf{v} = \mathbf{A}\mathbf{y}$.

c) *Determine the distribution of* $\mathbf{w}$, *where* $\mathbf{w}^\top = \begin{bmatrix}\mathbf{Ay} & \mathbf{By}\end{bmatrix}$.

d) *Which of the distributions obtained in (a)–(c) are singular distributions? Recall that a distribution is singular if* $\boldsymbol{\Sigma}$ *is positive definite. Note that there are many algebraic properties of* $\boldsymbol{\Sigma}$ *that can be used to show that* $\boldsymbol{\Sigma}$ *is singular/nonsingular.*

Problem 2. *Suppose* $\boldsymbol{X}$ *and* $\boldsymbol{W}$ *are any two matrices with* n *rows for which* $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{W})$. *Show that* $\boldsymbol{P_X} = \boldsymbol{P_W}$.

Problem 3. *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* $1 \leq i < j \leq 5$,

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

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}. *Let*

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

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

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* $\boldsymbol{X}^\top \boldsymbol{X}$.

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

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

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

Problem 4. *Consider a linear model for which*

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \quad \textit{and} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix}.$$

a) *Obtain the normal equations for this model and solve them.*

b) *Are all functions* $\mathbf{c}^\top \boldsymbol{\beta}$ *estimable? Justify your answer.*

c) *Obtain the least squares estimator of* $\beta_1 + \beta_2 + \beta_3 + \beta_4$.

Problem 5. *Suppose the Gauss-Markov model with normal errors (GMMNE) holds.*

a) *Suppose* $\boldsymbol{C}\boldsymbol{\beta}$ *is estimable. Derive the distribution of* $\boldsymbol{C}\widehat{\boldsymbol{\beta}}$*, the OLSE of* $\boldsymbol{C}\boldsymbol{\beta}$.

b) *Now suppose* $\boldsymbol{C}\boldsymbol{\beta}$ *is NOT estimable. Provide a fully simplified expression for* $\mathrm{Var}\left(\boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{X}^\top \boldsymbol{y}\right)$.

c) *Now suppose* $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ *is testable and that* $\boldsymbol{C}$ *has only one row and* $\boldsymbol{d}$ *has only one element so that they may be written as* $\boldsymbol{c}^\top$ *and* $\mathbf{d}$*, respectively. Prove the result on slide 29 of slide set 2 of Key Linear Model Results.*

Problem 6. *Provide an example that shows that a generalized inverse of a symmetric matrix need not be symmetric. (*Comment: For this reason, we cannot assume that $(\boldsymbol{X}^\top \boldsymbol{X})^- = [(\boldsymbol{X}^\top \boldsymbol{X})^-]^\top$*.)*

Problem 7. [1] *The following "visualization" analogy is taken liberally from Christensen (2002).*

VISUALIZATION: *One can think about the geometry of least squares estimation in three dimensions (i.e., when* $n = 3$*). Consider your kitchen table and take one corner of the table to be the origin. Take* $\mathcal{C}(\mathbf{X})$ *as the two dimensional subspace determined by the surface of the table, and let* $\mathbf{y}$ *be any vector originating at the origin; i.e., any point in* $\mathbb{R}^3$*. The linear model says that* $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$*, which just says that* $\mathrm{E}(\mathbf{y})$ *is somewhere on the table. The least squares estimate* $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$ *is the perpendicular projection of* $\mathbf{y}$ *onto the surface of the table. The residual vector* $\widehat{\mathbf{e}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}$ *is the vector starting at the origin, perpendicular to the surface of the table, that reaches the same height as* $\mathbf{y}$*. Another way to think of the residual vector is to first connect* $\mathbf{y}$ *and* $\mathbf{P}_{\mathbf{X}}\mathbf{y}$ *with a line segment (that is perpendicular to the surface of the table). Then, shift the line segment along the surface (keeping it perpendicular) until the line segment has one end at the origin. The residual vector* $\widehat{\mathbf{e}}$ *is the perpendicular projection of* $\mathbf{y}$ *onto* $\mathcal{C}(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = \mathcal{N}(\mathbf{X}^\top)$*; that is, the projection onto the orthogonal complement of the table surface. The orthogonal complement* $\mathcal{C}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})$ *is the one-dimensional space in the vertical direction that goes through the origin. Once you have these vectors in place, sums of squares arise from Pythagorean's Theorem.*

[1] There is nothing to be turned in for this last "question" of the assignment.