# Homework 2: STA 721 Fall24

Your Name

September 5, 2024; Due in one week (see Gradescope)

1. Let $\mathbf{P}$ be an orthogonal projection matrix onto $\mathcal{M}$ (a $r$-dimesional subspace of $\mathbb{R}^n$). Use the spectral decomposition of $\mathbf{P}$, $\mathbf{P} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ with

$$\mathbf{U} = [u_1, \ldots, u_n]$$
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

with colunms of $\mathbf{U}$ forming an orthonormal basis for $\mathbb{R}^n$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ representing the eignevalues in the following. Prove that

   (a) the eigenvalues of $\mathbf{P}$, $\lambda_i$, are either zero or one
   (b) the trace of $\mathbf{P}$ is the rank of $\mathbf{P}$
   (c) the dimension of the subspace that $\mathbf{P}$ projects onto is the rank of $\mathbf{P}$
   (d) the columns of $\mathbf{U}_r = [u_1, u_2, \ldots u_r]$ form an ONB for the $C(\mathbf{P})$
   (e) the projection $\mathbf{P}$ has the representation $\mathbf{P} = \mathbf{U}_r\mathbf{U}_r^T = \sum_{i=1}^r u_i u_i^T$ (the sum of $r$ rank 1 projections)
   (f) the projection $\mathbf{I}_n - \mathbf{P} = \mathbf{I} - \mathbf{U}_r\mathbf{U}_r^T = \mathbf{U}_\perp\mathbf{U}_\perp^T$ where $\mathbf{U}_\perp = [u_{r+1}, \ldots u_n]$ is an orthogonal projection onto $\mathcal{N} = \mathcal{M}^\perp$

2. Show that $\mathbf{P}^- = \mathbf{P}$ for $\mathbf{P}$ a projection matrix is a generalized inverse using the definition of generalized inverse.

3. Use the singular value decomposition of $\mathbf{X} \in \mathbb{R}^{n \times p}$ with rank $r < p$ to construct a generalized inverse of $\mathbf{X}$ based on the definition in Christensen B.36. (Suggestion: look at the Moore-Penrose generalized inverse for $\mathbf{X}$ in the full rank case and the Moore-Penrose inverse of $\mathbf{X}^T\mathbf{X}$ in the rank $r < p$ case)

4. Consider the class of linear statistical models from class: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim P$,

$$\mathcal{P}_1 = \{P = \mathsf{N}(\mathbf{0}_n, \mathbf{I}_n)\} \tag{1}$$
$$\mathcal{P}_2 = \{P = \mathsf{N}(\mathbf{0}_n, \sigma^2\mathbf{I}_n), \sigma^2 > 0\} \tag{2}$$
$$\mathcal{P}_3 = \{P = \mathsf{N}(\mathbf{0}_n, \mathbf{\Sigma}), \mathbf{\Sigma} \in \mathcal{S}^+\} \tag{3}$$
$$\mathcal{P}_4 = \{\text{the set of distributions with } \mathsf{E}_P[\boldsymbol{\epsilon}] = \mathbf{0}_n \text{ and } \mathsf{E}_P[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] > 0\} \tag{4}$$
$$\mathcal{P}_5 = \{\text{the set of distributions with } \mathsf{E}_P[\boldsymbol{\epsilon}] = \mathbf{0}_n \text{ and } \mathsf{E}_P[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] \geq 0\} \tag{5}$$

Find an estimator that is unbiased for $\boldsymbol{\beta} \in \mathbb{R}^p$ and $P \in \mathcal{P}$ but is biased for $\boldsymbol{\beta} \in \mathbb{R}^p$ and $P \in \mathcal{P}$.

5. Centering in Regression: Consider the linear model $\mathbf{Y} \sim N(\mu\, \sigma^2\mathbf{I}_n)$ with $\boldsymbol{\mu} = \mathbf{1}_n\beta_0 + \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ a full rank matrix with (column) rank $p$, where $\mathbf{X}$ is linearly independent of the vector $\mathbf{1}_n$.

(a) Show that the mean function can be re-written as

$$\begin{aligned}\boldsymbol{\mu} =& \mathbf{1}_n\beta_0 + \mathbf{X}\boldsymbol{\beta} \\ =& \mathbf{1}_n\alpha_0 + \mathbf{X}_c\boldsymbol{\beta}_c\end{aligned}$$

where $\mathbf{X}_c = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X} = \mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}^T$ (which centers each column of $\mathbf{X}$), $\mathbf{P}_1$ is the orthogonal projection onto the column of ones, and $\alpha = \beta_0 + \bar{\mathbf{x}}^T\boldsymbol{\beta}$ is the intercept in the model with the centered $\mathbf{X}$. (Hint: add and subtract $\mathbf{P}_1\mathbf{X}\boldsymbol{\beta}$ )

(b) Show that $\mathbf{X}_c$ and $\mathbf{1}_n$ are orthogonal.

(c) Show that $\mathbf{P}_1 + \mathbf{P}_{\mathbf{X}_c}$ is an orthogonal projection, where $\mathbf{P}_{\mathbf{X}_c}$ is the orthogonal projection onto $\mathbf{X}_c$. Does $\mathbf{P}_1 + \mathbf{P}_{\mathbf{X}_c}$ project onto $C(\mathbf{1}_n, \mathbf{X})$?

(d) Find the OLS/MLE estimates of $\alpha$ and $\boldsymbol{\beta}_c$. What are the MLEs of $\beta_0, \boldsymbol{\beta}$ in the original model in terms of $\alpha$ and $\boldsymbol{\beta}_c$?

(e) Show that diagonal elements of $\mathbf{P}_1 + \mathbf{P}_{\mathbf{X}_c}$ are

$$h_{ii} = \frac{1}{n} + (\mathbf{x}_i - \bar{\mathbf{x}})^T \left((\mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}^T)\right)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$$

(recall all vectors are column vectors). The $h_{ii}$ are known as the leverage values.

(f) Suppose unknown to you the response $\mathbf{Y}$ has been centered. What will be the OLS/MLEs of $\alpha$ and $\boldsymbol{\beta}_c$? Can you recover $\beta_0$? $\boldsymbol{\beta}$?

6. Gauss-Markov: Let $\mathsf{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ and $\mathsf{Cov}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ for some known $\mathbf{X} \in \mathbb{R}^{n\times p}$ of rank $r$, unknown $\boldsymbol{\beta} \in \mathbb{R}^p$ and unknown $\sigma^2 > 0$.

(a) Let $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in C(\mathbf{X})$ be vectors such that $\mathsf{E}[\mathbf{a}^T\mathbf{Y}] = \mathsf{E}[\mathbf{b}^T\mathbf{Y}]$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. How does the space in which $\delta = \mathbf{a} - \boldsymbol{\beta}$ lives relate to $C(\mathbf{X})$? Specifically, what can you say about $\delta^T\mathbf{m}$ for $\mathbf{m} \in C(\mathbf{X})$?

(b) Derive an expression for the variance of $\mathbf{a}^T\mathbf{Y}$ in terms of the variance of $\boldsymbol{\beta}^T\mathbf{Y}$, and determine conditions under which the latter variance is smaller. (see class slides or text)

(c) Does your result depend on whether $\mathbf{X}$ is full rank?

(d) Let $\boldsymbol{\lambda} = \mathbf{X}^T\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \mathfrak{M}$. Show that $\mathsf{E}[\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}] = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$ even if $r < p$ and $\hat{\boldsymbol{\beta}}$ is not unique.

(e) Suppose that $\mathbf{a}^T\mathbf{Y}$ is another unbiased estimator of $\lambda^T\boldsymbol{\beta}$. Under what conditions will $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^T\mathbf{P}\mathbf{Y}$ have the smallest variance among linear unbiased estimators? (repeat proof outline from class writing $\mathbf{a}^T\mathbf{Y} = \boldsymbol{\beta}^T\mathbf{P}\mathbf{Y} + \delta^T\mathbf{Y}$)