STAT 714 HW 5 SOLUTIONS

$\boxed{1}$ For matrices X, A, and B

(a) Show that $AX^TX = BX^TX \iff AX^T = BX^T$.

(b) Show that $X(X^TX)^-$ is a gen. inverse of X^T.

Solution:

(a) "$\Leftarrow$" Suppose $AX^T = BX^T$. Then post-multiplication by X gives $AX^TX = BX^TX$.

"$\Rightarrow$" Suppose $AX^TX = BX^TX$. This implies

$$X^TXA^T = X^TXB^T \quad \Rightarrow \quad X^TX(A^T - B^T) = 0,$$

so the columns of $A^T - B^T$ are in $\text{Nul}\, X^TX$.

Since $\text{Nul}\, X^TX = \text{Nul}\, X$ (we have shown this in class), we have

$$X(A^T - B^T) = 0 \quad \Rightarrow \quad XA^T = XB^T \quad \Rightarrow \quad AX^T = BX^T.$$

(b) Write $X^TX(X^TX)^-X^TX = X^TX$.

This implies $X^T\underbrace{X(X^TX)^-}X^T = X^T$ by the result proven in (a).

Therefore $X(X^TX)^-$ is a generalized inverse of X^T.

$\boxed{2}$ Let $Y_i = \mu + \sum_{j=1}^{d} g_{ij}\beta_j + \varepsilon_i$ for $i = 1, \dots, n$,

Let $(g_{ij})_{1 \le i \le n, 1 \le j \le d}$ have full-column rank and suppose $\sum_{j=1}^{d} g_{ij} = 1 \quad \forall\, i = 1, \dots, n$.

(a) Give representation $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$

(b) Check if μ is estimable

(c) Give matrix C such that $C\underset{\sim}{b} = \underset{\sim}{0}$ imposes $\sum_{j=1}^{d}\sum_{i=1}^{n} g_{ij}\beta_j = 0$.

(d) Find $\hat{\mu}$ under the constraint in (c).

Solution:

(a)
$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & g_{11} & \cdots & g_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & g_{n1} & \cdots & g_{nd} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \mu \\ \beta_1 \\ \vdots \\ \beta_d \end{bmatrix}}_{\underset{\sim}{b}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\underset{\sim}{e}}.$$

(b) $\mu = \underset{\sim}{c}^T \underset{\sim}{b}$ with $\underset{\sim}{c} = (1 \; 0 \cdots 0)^T$.

To check whether $\underset{\sim}{c} \in \mathrm{Col}\, X^T$, build augmented matrix for solving $X^T \underset{\sim}{x} = \underset{\sim}{c}$ and row reduce:

$$\left[\begin{array}{ccc|c} 1 & \cdots & 1 & 1 \\ g_{11} & \cdots & g_{n1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ g_{1d} & \cdots & g_{nd} & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \\ g_{11} & \cdots & g_{n1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ g_{1d} & \cdots & g_{nd} & 0 \end{array}\right] \leftarrow$$

Subtract from the first row the sum of all the other rows.

We see that there is no solution, so μ is <u>not</u> estimable.

(c) We may impose the constraint $\sum_{j=1}^{d} \sum_{i=1}^{n} g_{ij} \beta_j = 0$ as $C \underset{\sim}{b} = 0$ with

$$\underset{1\times(d+1)}{C} = \left[0 \quad \sum_{i=1}^{n} g_{i1} \quad \cdots \quad \sum_{i=1}^{n} g_{id} \right]$$

(d) The constrained estimator is the (unique) solution to

$$\begin{bmatrix} X^T X \\ C \end{bmatrix} \underset{\sim}{b} = \begin{bmatrix} X^T \underset{\sim}{y} \\ 0 \end{bmatrix}$$

Letting $\underset{\sim}{g}_j = (g_{1j}, \dots, g_{nj})^T$ for each $j = 1, \dots, d$, we may write

$$X^T X = \begin{bmatrix} n & \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & \cdots & \underset{\sim}{1}_n^T \underset{\sim}{g}_d \\ \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & & & \\ \vdots & & \left(\underset{\sim}{g}_i^T \underset{\sim}{g}_j \right)_{1 \le i,j \le d} & \\ \underset{\sim}{1}_n \underset{\sim}{g}_d & & & \end{bmatrix} \qquad X^T \underset{\sim}{y} = \begin{bmatrix} n\bar{Y} \\ \underset{\sim}{g}_1^T \underset{\sim}{y} \\ \vdots \\ \underset{\sim}{g}_d^T \underset{\sim}{y} \end{bmatrix}$$

From here we row-reduce the augmented matrix:

$$\left[\begin{array}{cccc|c} n & \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & \cdots & \underset{\sim}{1}_n^T \underset{\sim}{g}_d & n\bar{Y} \\ \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & & & & \underset{\sim}{g}_1^T \underset{\sim}{y} \\ \vdots & & \left(\underset{\sim}{g}_i^T \underset{\sim}{g}_j \right)_{1 \le i,j \le d} & & \vdots \\ \underset{\sim}{1}_n \underset{\sim}{g}_d & & & & \underset{\sim}{g}_d^T \underset{\sim}{y} \\ 0 & \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & \cdots & \underset{\sim}{1}_n^T \underset{\sim}{g}_d & 0 \end{array}\right]$$

$$\sim \left[\begin{array}{cccc|c} n & 0 & \cdots & 0 & n\bar{Y} \\ \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & & & & \underset{\sim}{g}_1^T \underset{\sim}{y} \\ \vdots & & \left(\underset{\sim}{g}_i^T \underset{\sim}{g}_j \right)_{1 \le i,j \le d} & & \vdots \\ \underset{\sim}{1}_n \underset{\sim}{g}_d & & & & \underset{\sim}{g}_d^T \underset{\sim}{y} \\ 0 & \underset{\sim}{1}_n^T \underset{\sim}{g}_1 & \cdots & \underset{\sim}{1}_n^T \underset{\sim}{g}_d & 0 \end{array}\right].$$

Subtract last row from first row

This shows that the first entry of the solution $\hat{\underset{\sim}{b}}$ to $\begin{bmatrix} X^TX \\ C \end{bmatrix} b = \begin{bmatrix} X^T\underset{\sim}{y} \\ 0 \end{bmatrix}$ is $\bar{y}$.

So $\hat{\mu} = \bar{Y}$.

$\boxed{3}$ Let $\underset{n \times p}{X}$ have full-column rank with first column a column of ones.
Let $\underset{n \times 1}{\underset{\sim}{Y}}$ a vector of responses not all equal to the same value.
Let P_1 be the orth. proj. onto $\operatorname{Span}\{\underset{\sim}{1}_n\}$.

(a) Interpret $\dfrac{\| (P_X - P_1) \underset{\sim}{Y} \|^2}{\| (I - P_1) \underset{\sim}{Y} \|^2}$.

(b) Give range of values for quantity in (a)

(c) Write quantity in terms of Y_i, $\hat{Y}_i$, $\bar{Y}_n$.

<u>Solution</u>:

(a) This is the ratio of the "model sum of squares" over the "total sum of squares".

The denominator measures the total variability in $\underset{\sim}{Y}$ around its mean.
The numerator measures the variability of $P_X \underset{\sim}{Y}$ around the mean of $\underset{\sim}{Y}$.
This is R^2, the coefficient of determination.

(b) We can write

$$\|(I-P_1)\underset{\sim}{y}\|^2 = \|(I-P_X)\underset{\sim}{y} + (P_X - P_1)\underset{\sim}{y}\|^2$$

$$= \|(I-P_X)\underset{\sim}{y}\|^2 + \|(P_X-P_1)\underset{\sim}{y}\|^2 + 2\left[(P_X-P_1)\underset{\sim}{y}\right]\cdot\left[(I-P_X)\underset{\sim}{y}\right],$$

where

$$\begin{aligned}
\left[(P_X-P_1)\underset{\sim}{y}\right]\cdot\left[(I-P_X)\underset{\sim}{y}\right] &= \underset{\sim}{y}^T(P_X-P_1)(I-P_X)\underset{\sim}{y} \\
&= -\underset{\sim}{y}^T P_1 (I-P_X)\underset{\sim}{y} \\
&= -\underset{\sim}{y}^T (P_1 - P_1 P_X)\underset{\sim}{y}, \\
&= 0,
\end{aligned}$$

since $P_1 P_X = P_1$.

To see why $P_1 P_X = P_1$, note that

since X has a column of ones

$$P_X \underbrace{P_1 \underset{\sim}{y}}_{\in \text{Col}\, X} = P_1 \underset{\sim}{y} \quad \forall\ \underset{\sim}{y} \quad \text{since} \quad \text{Span}\{\underset{\sim}{1}_n\} \subset \text{Col}\, X.$$

This gives $P_X P_1 = P_1 \implies (P_X P_1)^T = P_1^T = P_1 \implies P_1 P_X = P_1$.

From the above we have

$$\|(I-P_1)\underset{\sim}{y}\|^2 = \|(I-P_X)\underset{\sim}{y}\|^2 + \|(P_X-P_1)\underset{\sim}{y}\|^2 .$$

so the numerator $\|(P_X-P_1)\underset{\sim}{y}\|^2$ can never exceed the denominator $\|(I-P_1)\underset{\sim}{y}\|^2$.

The denominator is positive and the numerator is nonnegative, so

$$\frac{\|(P_X-P_1)\underset{\sim}{y}\|^2}{\|(I-P_1)\underset{\sim}{y}\|^2} \in [0, 1].$$

$\boxed{4}$ Let $X = [\underset{\sim}{x}_1 \; X_2]$, where X_2 has a column of ones and X has full-column rank. Let $\hat{\underset{\sim}{b}} = (X^T X)^{-1} X^T \underset{\sim}{y}$.

(a) Show that $\text{Var}\, \hat{b}_1 = \dfrac{\sigma^2}{1 - R_1^2} \dfrac{1}{\|(I - P_1)\underset{\sim}{x}_1\|^2}$, where $R_1^2 = \dfrac{\|(P_{X_2} - P_1)\underset{\sim}{x}_1\|^2}{\|(I - P_1)\underset{\sim}{x}_1\|^2}$.

(b) Interpret.

<u>Solution:</u>

(a) We have

$$\text{Cov}\, \hat{\underset{\sim}{b}} = \text{Cov}\left((X^T X)^{-1} X^T \underset{\sim}{y} \right)$$

$$= (X^T X)^{-1} X^T [\sigma^2 I_n] X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}.$$

The variance of $\hat{b}_1$ is the $(1,1)$ entry of $\sigma^2 (X^T X)^{-1}$. Using a block matrix inversion formula, we obtain

$$\text{Var}\, \hat{\underset{\sim}{b}}_1 = \sigma^2 \left[\underset{\sim}{x}_1^T \underset{\sim}{x}_1 - \underset{\sim}{x}_1^T X_2 (X_2^T X_2)^{-1} X_2^T \underset{\sim}{x}_1 \right]^{-1}$$

$$= \frac{\sigma^2}{\underset{\sim}{x}_1^T (I - P_{X_2}) \underset{\sim}{x}_1}$$

$$= \frac{\sigma^2}{\underset{\sim}{x}_1^T \left((I - P_1) - (P_{X_2} - P_1) \right) \underset{\sim}{x}_1}$$

$$= \frac{\sigma^2}{\dfrac{\underset{\sim}{x}_1^T (I - P_1) \underset{\sim}{x}_1 - \underset{\sim}{x}_1^T (P_{X_2} - P_1) \underset{\sim}{x}_1}{\underset{\sim}{x}_1^T (I - P_1) \underset{\sim}{x}_1}} \qquad \frac{1}{\underset{\sim}{x}_1^T (I - P_1) \underset{\sim}{x}_1}$$

$$= \frac{\sigma^2}{1 - R_1^2} \frac{1}{\|(I - P_1)\underset{\sim}{x}_1\|^2}.$$

[<u>Note:</u> $\underset{\sim}{x}_1^T (P_{X_2} - P_1) \underset{\sim}{x}_1 = \|(P_{X_2} - P_1)\underset{\sim}{x}_1\|^2$ uses the fact that $P_{X_2} - P_1$ is idempotent. This is idempotent because $\text{Span}\{\underset{\sim}{1}\} \subset \text{Col}\, X_2$]

(b) As R_1^2 approaches 1, $\text{Var}\, \hat{b}_1$ will "explode." The quantity R_1 expresses how correlated $\underset{\sim}{x}_1$ is with the other columns of X. High correlation / collinearity leads to high variance.

$\boxed{5}$ Let $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$, where $\mathbb{E}\,\underset{\sim}{e} = \underset{\sim}{0}$ and $\operatorname{Cov}\,\underset{\sim}{e} = \sigma^2 I_n$.

Assume eigenvalues of X^TX are all positive.

(a) Show that $X^TX\, b = X^Ty$ has a unique solution.

(b) Show $\mathbb{E}\,\|\hat{\underset{\sim}{b}} - \underset{\sim}{b}\|^2 = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}$, where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of X^TX.

Solution:

(a) Since X^TX has all positive eigenvalues its determinant is nonzero and therefore it is invertible. So

$$\hat{\underset{\sim}{b}} = (X^TX)^{-1}X^T\underset{\sim}{y}.$$

(b) We have

$$\mathbb{E}\,\|\hat{\underset{\sim}{b}} - \underset{\sim}{b}\|^2 = \mathbb{E}\left(\hat{\underset{\sim}{b}}^T\hat{\underset{\sim}{b}} - 2\,\hat{\underset{\sim}{b}}^T\underset{\sim}{b} + \underset{\sim}{b}^T\underset{\sim}{b}\right)$$

$$= \mathbb{E}\,\hat{\underset{\sim}{b}}^T\hat{\underset{\sim}{b}} - \underset{\sim}{b}^T\underset{\sim}{b}$$

$\mathbb{E}\,\underset{\sim}{z}^TA\underset{\sim}{z} = (\mathbb{E}\,\underset{\sim}{z})^TA\,\mathbb{E}\,\underset{\sim}{z} + \operatorname{tr}(A(\operatorname{Cov}\underset{\sim}{z}))$.

$$= \underset{\sim}{b}^T\underset{\sim}{b} + \operatorname{tr}(\operatorname{Cov}\hat{\underset{\sim}{b}}) - \underset{\sim}{b}^T\underset{\sim}{b}$$

$$= \operatorname{tr}(\operatorname{Cov}\hat{\underset{\sim}{b}})$$

$\operatorname{Cov}\hat{\underset{\sim}{b}} = \sigma^2(X^TX)^{-1}$.

$$= \sigma^2\operatorname{tr}\left((X^TX)^{-1}\right)$$

$$= \sigma^2\sum_{j=1}^{p}\frac{1}{\lambda_j},$$

Since λ_j^{-1}, $j = 1, \dots, p$ are the eigenvalues of $(X^TX)^{-1}$ and the trace of a matrix is the sum of its eigenvalues.

$\boxed{6}$ Show under $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e}$, $\mathbb{E}\,\underset{\sim}{e} = \underset{\sim}{0}$, a contrast $\underset{\sim}{c}^T \underset{\sim}{b}$ is estimable $\iff$ $\underset{\sim}{c} \in \text{Col}\, X^T$.

<u>Solution</u>: Firstly, note that $\text{Cov}\,\underset{\sim}{e}$ plays no role in determining estimability of a contrast (in class we established this result under $\mathbb{E}\,\underset{\sim}{e} = \underset{\sim}{0}$, so one could have any kind of covariance structure for $\underset{\sim}{e}$).

"Estimable" means there exists a_0 and $\underset{\sim}{a}$ such that $\mathbb{E}[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}] = \underset{\sim}{c}^T \underset{\sim}{b}$ $\forall\, \underset{\sim}{b}$.

"$\Leftarrow$" Let $\underset{\sim}{c} \in \text{Col}\, X^T$. Then $\exists\, \underset{\sim}{a}$ such that $\underset{\sim}{c} = X^T \underset{\sim}{a}$.

Then with $a_0 = 0$ we have $\mathbb{E}[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}] = \underset{\sim}{a}^T X \underset{\sim}{b} = \underset{\sim}{c}^T \underset{\sim}{b}$ $\forall\, \underset{\sim}{b}$.

"$\Rightarrow$" Suppose $\mathbb{E}[a_0 + \underset{\sim}{a}^T \underset{\sim}{y}] = \underset{\sim}{c}^T \underset{\sim}{b}$ for all $\underset{\sim}{b}$.

Then $a_0 + \underset{\sim}{a}^T X \underset{\sim}{b} = \underset{\sim}{c}^T \underset{\sim}{b}$ $\forall\, \underset{\sim}{b}$,

which implies $a_0 = 0$ and $\underset{\sim}{a}^T X = \underset{\sim}{c}$, so $\underset{\sim}{c} \in \text{Col}\, X^T$.

$\boxed{7}$ Let $X = [\underset{\sim}{x}_1 \cdots \underset{\sim}{x}_p]$ be full-rank with unit columns.

Find the unit vector $\underset{\sim}{a}$ and the scalar A such that $\underset{\sim}{a}^T \underset{\sim}{x}_j = A$ for $j = 1, \ldots, p$.

Solution:

The condition $\underset{\sim}{a}^T \underset{\sim}{x}_j = A$ for $j = 1, \ldots, p$ may be expressed as $X^T \underset{\sim}{a} = A \underset{\sim}{1}_p$.

A solution is given by

$$\underset{\sim}{a}^* = (X^T)^- A \underset{\sim}{1}_p,$$

where $(X^T)^-$ is a generalized inverse of X^T.

A generalized inverse of X^T is $X(X^TX)^{-1}$, since $X^TX(X^TX)^{-1}X^T = X^T$.

So we can write $\underset{\sim}{a}^* = A\, X(X^TX)^{-1} \underset{\sim}{1}_p$.

But we must scale the solution so that it has unit norm. This gives

$$\underset{\sim}{a} = \frac{1}{\|\underset{\sim}{a}^*\|} \underset{\sim}{a}^* = \frac{1}{\sqrt{A^2 \underset{\sim}{1}_p^T (X^TX)^{-1} X^T X (X^TX)^{-1} \underset{\sim}{1}_p}}\, A\, X(X^TX)^{-1} \underset{\sim}{1}_p = \frac{1}{\|X(X^TX)^{-1}\underset{\sim}{1}_p\|} X(X^TX)^{-1} \underset{\sim}{1}_p.$$

Then

$$X^T \underset{\sim}{a} = X^TX(X^TX)^{-1} \underset{\sim}{1}_p \frac{1}{\|X(X^TX)^{-1}\underset{\sim}{1}_p\|} = \frac{1}{\|X(X^TX)^{-1}\underset{\sim}{1}_p\|} \underset{\sim}{1}_p,$$

so $A = \dfrac{1}{\|X(X^TX)^{-1}\underset{\sim}{1}_p\|}$. We can also write

$$A = \left[\underset{\sim}{1}_p^T (X^TX)^{-1} \underset{\sim}{1}_p\right]^{-1/2}$$

$$\underset{\sim}{a} = \left[\underset{\sim}{1}_p^T (X^TX)^{-1} \underset{\sim}{1}_p\right]^{-1/2} X(X^TX)^{-1} \underset{\sim}{1}_p.$$