Session 2 : Exercises

We assume the fixed design model and the OLS estimator $\widehat{\boldsymbol{\theta}} = (X^\top X)^{-1} X^\top \mathbf{y}$.

Exercice 1. **(Moore-Penrose Inverse via Spectral Decomposition)** Show that $A^+ A A^+ = A^+$

Solution. For symmetric A :

$$A^+ = \sum_{j:\lambda_j \neq 0} \lambda_j^{-1} v_j v_j^\top$$

where v_j are eigenvectors of A. Provides the minimum-norm solution for rank-deficient systems.

$$A^+ = \sum_{j:\lambda_j \neq 0} \lambda_j^{-1} v_j v_j^\top$$

Then

$$A^+ A = \sum_{j:\lambda_j \neq 0} \lambda_j^{-1} v_j v_j^\top \left(\sum_k \lambda_k v_k v_k^\top \right) = \sum_{j:\lambda_j \neq 0} v_j v_j^\top$$

So

$$A^+ A A^+ = \left(\sum_{j:\lambda_j \neq 0} v_j v_j^\top \right) \left(\sum_{k:\lambda_k \neq 0} \lambda_k^{-1} v_k v_k^\top \right) = \sum_{j:\lambda_j \neq 0} \lambda_j^{-1} v_j v_j^\top = A^+$$

Exercice 2. **(Hat matrix)** Let X be full (column) rank, and $H = X(X^\top X)^{-1} X^\top$ the hat matrix. Show that H is an orthogonal projector.

Solution. Recall that a matrix H is an orthogonal projector if it satisfies :

1. $H^2 = H$ (idempotent)
2. $H^\top = H$ (symmetric)

We verify the properties :

$$\begin{aligned} H^2 &= X(X^\top X)^{-1} X^\top X (X^\top X)^{-1} X^\top = X(X^\top X)^{-1} \overbrace{X^\top X (X^\top X)^{-1}}^{I} X^\top \\ &= X(X^\top X)^{-1} X^\top = H, \end{aligned}$$

$$H^\top = \left(X(X^\top X)^{-1} X^\top \right)^\top = X(X^\top X)^{-1} X^\top = H.$$

Hence, H is symmetric and idempotent, so it is an orthogonal projector.

Exercice 3. **(P2 - Bias)** Show that whenever the matrix X has full rank, the least squares estimator is unbiased

Solution.

$$B = \mathbb{E}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\theta}^\star = \mathbb{E}\big((X^\top X)^{-1} X^\top \mathbf{y} \big) - \boldsymbol{\theta}^\star$$

$$B = \mathbb{E}\big((X^\top X)^{-1} X^\top (X\boldsymbol{\theta}^\star + \boldsymbol{\varepsilon}) \big) - \boldsymbol{\theta}^\star$$

$$B = (X^\top X)^{-1} X^\top X \boldsymbol{\theta}^\star + (X^\top X)^{-1} X^\top \mathbb{E}(\boldsymbol{\varepsilon}) - \boldsymbol{\theta}^\star = 0$$

Exercice 4. (Covariance) If $\mathrm{Cov}(\mathbf{y}) = \sigma^2 I$, the covariance matrix for $\widehat{\theta}$ is given by $\sigma^2(X^\top X)^{-1}$.

Solution.

$$\begin{aligned}
\mathrm{Cov}(\widehat{\theta}) &= \mathrm{Cov}\big[(X^\top X)^{-1}X^\top \mathbf{y}\big] \\
&= (X^\top X)^{-1}X^\top \,\mathrm{Cov}(\mathbf{y})\,\big[(X^\top X)^{-1}X^\top\big]^\top \\
&= (X^\top X)^{-1}X^\top(\sigma^2 I)X(X^\top X)^{-1} \\
&= \sigma^2(X^\top X)^{-1}X^\top X(X^\top X)^{-1} \\
&= \sigma^2(X^\top X)^{-1}.
\end{aligned}$$

Exercice 5. (P5 - BLUE, best linear unbiased estimator) If $\mathbb{E}(\mathbf{y}) = X\boldsymbol{\theta}^\star$ and $\mathrm{Cov}(\mathbf{y}) = \sigma^2 I$, then the least-squares estimators $\widehat{\theta}_j$, $j = 0, 1, \dots, k$, have minimum variance among all linear unbiased estimators.

Solution. We consider a linear estimator $A\mathbf{y}$ of $\boldsymbol{\theta}$ and seek the matrix A for which $A\mathbf{y}$ is a minimum variance unbiased estimator of $\boldsymbol{\theta}$. For $A\mathbf{y}$ to be unbiased, we must have

$$\mathbb{E}[A\mathbf{y}] = A\mathbb{E}[\mathbf{y}] = AX\boldsymbol{\theta}^\star = \boldsymbol{\theta}^\star.$$

This must hold for any possible $\boldsymbol{\theta}^\star$, which gives the condition

$$AX = I.$$

The covariance matrix of $A\mathbf{y}$ is

$$\mathrm{Cov}(A\mathbf{y}) = A(\sigma^2 I)A^\top = \sigma^2 AA^\top.$$

The variances of the estimators $\widehat{\theta}_j$ appear on the diagonal of $\sigma^2 AA^\top$, so we need to choose A (subject to $AX = I$) such that the diagonal elements of $AA^\top$ are minimized.

To relate $A\mathbf{y}$ to the least-squares estimator

$$\widehat{\boldsymbol{\theta}} = (X^\top X)^{-1}X^\top \mathbf{y},$$

we add and subtract $(X^\top X)^{-1}X^\top$:

$$AA^\top = \big[(A - (X^\top X)^{-1}X^\top) + ((X^\top X)^{-1}X^\top)\big]\big[(A - (X^\top X)^{-1}X^\top) + ((X^\top X)^{-1}X^\top)\big]^\top.$$

Expanding, we obtain four terms. Two vanish because of the restriction $AX = I$. The result is

$$AA^\top = \big[A - (X^\top X)^{-1}X^\top\big]\big[A - (X^\top X)^{-1}X^\top\big]^\top + (X^\top X)^{-1}.$$

The matrix $\big[A-(X^\top X)^{-1}X^\top\big]\big[A-(X^\top X)^{-1}X^\top\big]^\top$ is positive semidefinite, and its diagonal elements are therefore nonnegative. These diagonal elements can be made equal to zero by choosing

$$A = (X^\top X)^{-1}X^\top.$$

This value of A also satisfies $AX = I$. Hence, the minimum variance linear unbiased estimator is

$$A\mathbf{y} = (X^\top X)^{-1}X^\top \mathbf{y},$$

which coincides with the least-squares estimator $\widehat{\boldsymbol{\theta}}$.

Exercice 6. (P6 - Invariance) In this exercise we show that the predicted value $\hat{\mathbf{y}}$ is invariant under simple linear changes of scale on the design matrix X is the original desing matrix, Z is a transformed version in which each column i has been scaled by c_i, i.e., for each individual in the dataset $x = (1, x_1, \ldots, x_k)^\top$ and $z = (1, c_1 x_1, \ldots, c_k x_k)^\top$. Let $\widehat{\theta}$ and $\widehat{\theta}_z$ be the estimated coefficients for each design matrix. Show that the prediction is the same in both cases.

Solution. We are asked to show that $\widehat{\mathbf{y}} = \widehat{\theta}^\top x = \widehat{\theta}_z^\top z$, where $\widehat{\theta}_z$ is the least squares estimator from the regression of $\mathbf{y}$ on z. We can rewrite z as $z = Dx$, where $D = \operatorname{diag}(1, c_1, c_2, \ldots, c_k)$. Then, the X matrix is transformed to $Z = XD$. We substitute $Z = XD$ in the least-squares estimator

$$\widehat{\theta}_z = (Z^\top Z)^{-1} Z^\top \mathbf{y}$$

to obtain

$$\begin{aligned}
\widehat{\theta}_z &= (Z^\top Z)^{-1} Z^\top \mathbf{y} \\
&= \left[(XD)^\top (XD)\right]^{-1} (XD)^\top \mathbf{y} \\
&= D^{-1} (X^\top X)^{-1} X^\top \mathbf{y} \\
&= D^{-1} \widehat{\theta},
\end{aligned}$$

where $\widehat{\theta}$ is the usual estimator for $\mathbf{y}$ regressed on the x's. Then

$$\widehat{\theta}_z^\top z = (D^{-1}\widehat{\theta})^\top D x = \widehat{\theta}^\top x.$$

Exercice 7. Using the matrix $(X^\top X)^{-1}$ for simple linear regression give the variance and covariance of the estimators.

Solution.

$$\operatorname{Cov}(\widehat{\boldsymbol{\theta}}) = \operatorname{Cov}\begin{pmatrix} \widehat{\theta}_0 \\ \widehat{\theta}_1 \end{pmatrix} = \widehat{\sigma}^2 (X^\top X)^{-1}.$$

That is,

$$\operatorname{Cov}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \operatorname{Var}(\widehat{\theta}_0) & \operatorname{Cov}(\widehat{\theta}_0, \widehat{\theta}_1) \\ \operatorname{Cov}(\widehat{\theta}_0, \widehat{\theta}_1) & \operatorname{Var}(\widehat{\theta}_1) \end{pmatrix} = \widehat{\sigma}^2 (X^\top X)^{-1}.$$

We can express the inverse of $(X^\top X)$ as

$$(X^\top X)^{-1} = \frac{1}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix}.$$

Thus,

$$\operatorname{Var}(\widehat{\theta}_0) = \widehat{\sigma}^2 \frac{\sum_i x_i^2}{n\sum_i (x_i - \bar{x})^2}, \quad \operatorname{Var}(\widehat{\theta}_1) = \frac{\widehat{\sigma}^2}{\sum_i (x_i - \bar{x})^2}, \quad \operatorname{Cov}(\widehat{\theta}_0, \widehat{\theta}_1) = -\widehat{\sigma}^2 \frac{\bar{x}}{\sum_i (x_i - \bar{x})^2}.$$

Exercice 8. Show that the predicted value $\widehat{\mathbf{y}}$ is invariant to a full-rank linear transformation on the x's.

Solution. We can express a full-rank linear transformation of the x's as

$$Z = XK = \begin{pmatrix} j & X_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & K_1 \end{pmatrix} = \begin{pmatrix} j + X_1 0 & j0 + X_1 K_1 \end{pmatrix} = \begin{pmatrix} j & X_1 K_1 \end{pmatrix},$$

where K_1 is nonsingular and

$$X_1 = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}.$$

We partition X and K in this way to transform only the x's in X_1, leaving the first column of X unaffected. Now the estimator becomes

$$\widehat{\theta}_z = (Z^\top Z)^{-1} Z^\top \mathbf{y} = K^{-1}\widehat{\theta},$$

and we have

$$\widehat{\mathbf{y}} = \widehat{\theta}_z^\top z = \widehat{\theta}^\top x,$$

where $z = K^\top x$.

Exercice 9. Hat Matrix with Generalized Inverse is a Projection

Solution. Let $X \in \mathbb{R}^{n\times p}$ (possibly rank-deficient). The hat matrix using the Moore–Penrose inverse is

$$H_+ = X(X^\top X)^+ X^\top.$$

P is an orthogonal projector iff

$$P^2 = P \quad \text{and} \quad P^\top = P.$$

We verify for H_+.

$$H_+^\top = \big(X(X^\top X)^+ X^\top\big)^\top = X(X^\top X)^+ X^\top = H_+.$$

$$\begin{aligned} H_+^2 &= X(X^\top X)^+ X^\top X(X^\top X)^+ X^\top \\ &= X(X^\top X)^+ (X^\top X)(X^\top X)^+ X^\top \\ &= X(X^\top X)^+ X^\top \\ &= H_+. \end{aligned}$$

where we used (Moore–Penrose property : $(X^\top X)^+(X^\top X)(X^\top X)^+ = (X^\top X)^+$.