Exercise 2.4.1

By solving the questions of this exercise, you provide a proof of the Gauss-Markov theorem. We use the following notation.

- The OLS estimator is $b = A_0 y$ where $A_0 = (X^T X)^{-1} X^T$, with A_0 $(k \times n)$ matrix
- Let $\hat{\beta} = Ay$ be linear unbiased, with A $(k \times n)$ matrix
- Define the difference matrix $D = A - A_0$

(a) Prove the following three results:

1. $Var(\hat{\beta}) = \sigma^2 A A^T$
2. $\hat{\beta}$ unbiased implies $AX = I$ and $DX = 0$.
3. Part (2) implies $AA^T = DD^T + (X^T X)^{-1}$

Part (1):

$$\begin{aligned}
&y = X\beta + \epsilon \\
&\hat{\beta} = Ay = AX\beta + A\epsilon \\
&E(\hat{\beta}) = E(AX\beta + A\epsilon) = AX\beta + AE(\epsilon) = AX\beta
\end{aligned}$$

$$\begin{aligned}
Var(\hat{\beta}) &= E((\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^T) \\
&= E((AX\beta + A\epsilon - AX\beta)(AX\beta + A\epsilon - AX\beta)^T) \\
&= E(A\epsilon(A\epsilon)^T) \\
&= E(A\epsilon\epsilon^T A^T) \\
&= AE(\epsilon\epsilon^T)A^T \\
&= \sigma^2 AA^T
\end{aligned}$$

Part (2):

$$\begin{aligned}
&\text{Unbiased } E(\hat{\beta}) = AX\beta = \beta \\
&\therefore AX = I
\end{aligned}$$

$$DX = (A - A_0)X = AX - A_0 X = I - (X^T X)^{-1} X^T X = 0$$

Part (3):

$$A_0^T = [(X^TX)^{-1}X^T]^T = X(X^TX)^{-1}$$

$$\begin{aligned}
AA^T &= (D + A_0)(D + A_0)^T \\
&= (D + A_0)(D^T + A_0^T) \\
&= DD^T + DA_0^T + A_0D^T + A_0A_0^T \\
&= DD^T + \underbrace{DX(X^TX)^{-1}}_{=0} + \underbrace{(X^TX)^{-1}X^TD^T}_{=0} + (X^TX)^{-1}X^TX(X^TX)^{-1} \\
&= DD^T + (X^TX)^{-1}
\end{aligned}$$

(b) Prove that part (a)(3) implies $Var(\hat{\beta}) = Var(b) + \sigma^2 DD^T$

$$\begin{aligned}
Var(\hat{\beta}) &= \sigma^2(DD^T + (X^TX)^{-1}) \\
&= \sigma^2 DD^T + \sigma^2(X^TX)^{-1} \\
&= \sigma^2 DD^T + Var(b)
\end{aligned}$$

(c) Prove that part (b) implies $Var(\hat{\beta}) - Var(b)$ is positive semidefinite (Gauss-Markov)

$$Var(\hat{\beta}) - Var(b) = \sigma^2 DD^T$$

Define:

z $(k \times 1)$ vector:

$d = D^T z$ $(n \times 1)$ vector with components d_i

$$z^T DD^T z = (D^T z)^T D^T z = d^T d = \sum_{i=1}^{n} d_i^2 \geq 0$$

$\therefore DD^T$ is positive semi-definite (PSD)

(d) Prove that $Var(\hat{\beta}) \geq Var(b)$ for every $j = 1, ..., k$

Let c_j be a $(k \times 1)$ unit vector such that the j-th element equals to 1:

$$c_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{gathered} c_j^T(Var(\hat{\beta}) - Var(b))c_j = c_j^T j(\sigma^2 DD^T)c_j \geq 0 \\ c_j^T Var(\hat{\beta})c_j \geq c_j^T Var(b)c_j \\ Var(c_j^T \hat{\beta}) \geq Var(c_j^T b) \\ Var(\hat{\beta}_j) \geq Var(b_j) \end{gathered}$$

Therefore, $Var(\hat{\beta}_j) \geq Var(b_j)$ for every $j = 1, ..., k$