MULTIPLE LINEAR REGRESSION MODEL

Q1) Prove Theorem 2-(b),(c)

$X \in \mathbb{R}^{n \times (p+1)}$, X is full rank

(a) $\text{rank}(X) = p+1$

(b) Number of observations must be greater than that of predictors

(c) The $(p+1) \times (p+1)$ real matrix X^TX is invertible

(b) $n \geq p+1$

$\text{row-rank}(A) = \text{column-rank}(A)$

if $n < p+1$, $\text{row-rank}(A) > \#\text{ of rows in } A \Rightarrow$ contradiction

$\because \text{row-rank}(A) \leq \#\text{ of rows in } A$

(c) $\text{rank}(X) = \text{rank}(X^TX)$

$\therefore \text{rank}(X^TX) = p+1 \Rightarrow X^TX$ is a square matrix and is full rank

$\Leftrightarrow X^TX$ is non-singular (= invertible)

Q2) Prove remark: the design matrix X must be full column rank in order for the model to be identifiable

$y = X\beta + \varepsilon$, $X = (\mathbb{1}\ x_1 \cdots x_p) \in \mathbb{R}^{n \times (p+1)}$, $\beta = (\beta_0 \cdots \beta_p)^T \in \mathbb{R}^{(p+1)\times 1}$

$\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$

(model 1): $y = X\beta + \varepsilon$ $\quad \varepsilon_i \overset{iid}{\sim} (0, \sigma^2)$

(model 2): $y' = X\beta' + \varepsilon$ $\quad \varepsilon_i \overset{iid}{\sim} (0, \sigma'^2)$

Suppose that y and y' are the same model.

Then, the distribution of y and y' are identical, meaning, $\sigma^2 = \sigma'^2$

$\therefore y = y' \Leftrightarrow X\beta = X\beta' \Rightarrow X^TX\beta = X^TX\beta' \Rightarrow (X^TX)^{-1}X^TX\beta = (X^TX)^{-1}(X^TX)\beta'$

$\Rightarrow \beta = \beta'$

$\therefore$ The design matrix must be full rank (= invertible) for the model to be identifiable.

Method of Estimation

Q3) Prove Theorem 3

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 = \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \|y - X\beta\|^2$$

Since we know that $C_X = \{X\beta : \beta \in \mathbb{R}^{p+1}\}$

it follows that $X\hat{\beta} = \hat{z}$ where $\hat{z} = \underset{z \in C_z}{\text{argmin}} \|y - z\|^2$

Now, Show that $X\hat{\beta} = \hat{z}$ where $\hat{z} = \text{argmin}_{z \in C_X} \|y - z\|^2$

* note 1 : there are two approaches to solve this question
* note 2 : Show that the Hessian matrix is positive definite.

(Approach #1)

$\hat{z} = \text{argmin}_{z \in C_z} \|y - z\|^2$,

$\hat{z}$ is the orthogonal projection of y onto the Column Space of X

$\therefore \hat{z} = \Pi_X y = X(X^TX)^{-1}X^Ty$ due to the "Closest vector property".

$X\hat{\beta} = \hat{z} \quad \therefore X\hat{\beta} = X(X^TX)^{-1}X^Ty \Rightarrow X^TX\hat{\beta} = X^TX(X^TX)^{-1}X^Ty$

$\Rightarrow (X^TX)^{-1}X^TX\hat{\beta} = (X^TX)^{-1}X^TX(X^TX)^{-1}X^Ty$

$\Rightarrow \hat{\beta} = (X^TX)^{-1}X^Ty$

(Approach #2)

$L(\beta) = \|y - X^T\beta\|^2 = (y - X^T\beta)^T(y - X^T\beta) = y^Ty - 2y^TX^T\beta + \beta^TX^TX\beta$

$\partial L(\beta)/\partial\beta = -2y^TX + 2X^TX\beta$

$[\partial L(\beta)/\partial\beta]_{\beta=\hat{\beta}} = -2y^TX + 2X^TX\hat{\beta} = 0$

$\Leftrightarrow X^TX\hat{\beta} = X^Ty$

$\Leftrightarrow \hat{\beta} = (X^TX)^{-1}X^Ty$

$\partial^2 L/\partial\beta^2 = 2X^TX \quad \forall c \in \mathbb{R}^{(p+1)}$

$c^TX^TXc = \|Xc\|^2 \geq 0 \quad \therefore X^TX$ is non-negative definite

+ X^TX is invertible : n.n.d & invertible $\Rightarrow X^TX$ is positive definite

$\Rightarrow \hat{\beta}$ is the minimum critical point

Q4) Prove Theorem 4

(a) $E[\hat{\beta}] = \beta$
(b) $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$

(a) $\hat{\beta} = (X^T X)^{-1} X^T y$, $\therefore E[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T E[y]$
$= (X^T X)^{-1} X^T X \beta = \beta$

(b) $Var((X^T X)^{-1} X^T y) = (X^T X)^{-1} X^T Var(y) X (X^T X)^{-1}$
$= (X^T X)^{-1} X^T (\sigma^2 I_n) X (X^T X)^{-1}$
$= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$
$= \sigma^2 (X^T X)^{-1}$

Q5) Prove Theorem 5

Decompose $X = (\mathbb{1} \; X_1)$ and define $X_{1,\perp} = X_{\mathbb{1}} - \Pi_{\mathbb{1}} X_1$. Prove the following:

(a) $C(X_{1,\perp})$ is orthogonal to $C(\mathbb{1})$
(b) $C(X) = C(\mathbb{1}) + C(X_{1,\perp})$

(a) $X_{1,\perp} = X_1 - \Pi_{\mathbb{1}} X_1$
$\langle X_1 - \Pi_{\mathbb{1}} X_1 , \mathbb{1} \rangle = X_1^T \mathbb{1} - (\Pi_{\mathbb{1}} X_1)^T \mathbb{1}$
$= X_1^T \mathbb{1} - X_1^T \Pi_{\mathbb{1}} \cdot \mathbb{1}$
$= X_1^T \mathbb{1} - X_1^T \mathbb{1} (\mathbb{1}^T \mathbb{1})^{-1} \mathbb{1}^T \mathbb{1}$
$= X_1^T \mathbb{1} - X_1^T \mathbb{1}$
$= 0$

$\therefore \mathbb{1}$ and $X_{1,\perp}$ are orthogonal $\Leftrightarrow C_1 = \mathbb{R}, C_2 \in \mathbb{R}^p$, $\mathbb{1} C_1$ and $X_{1,\perp} C_2$ are orthogonal
$\Leftrightarrow \mathbb{1} C_1 = C(\mathbb{1})$ and $X_{1,\perp} C_2 = C(X_{1,\perp})$
$\therefore C(\mathbb{1})$ and $C(X_{1,\perp})$ are orthogonal

(b) $C(X) = C(\mathbb{1}) + C(X_{1,\perp})$
WTS : ① $C(X) \subset C(\mathbb{1}) + C(X_{1,\perp})$
② $C(\mathbb{1}) + C(X_{1,\perp}) \subset C(X)$

② $\mathbb{1} \in C(X)$ is trivial

$X_{1,\perp} = (x_1 \cdots x_p) - \Pi_{\mathbb{1}} (x_1 \cdots x_p)$
$= (x_1 - \Pi_{\mathbb{1}} x_1) \cdots (x_p - \Pi_{\mathbb{1}} x_p)$
$= (x_1 - \bar{x}_1 \mathbb{1} \cdots x_p - \bar{x}_p \mathbb{1})$

$\Pi_{\mathbb{1}} x = (x_1 + \cdots + x_n) = (\bar{x} + \cdots + \bar{x}) = \bar{x} \mathbb{1}$

$\therefore C(X_{1,\perp})$: X 에서 $C(\mathbb{1})$을 뺀 것이기 때문에 $C(X_{1,\perp}) \in C(X)$

$\therefore C(\mathbb{1}) + C(X_{1,\perp}) \in C(X)$

① $C(X) \subset C(\mathbb{1}) + C(X_{1,\perp})$

We will show that the dim(LHS) = dim(RHS)

$\dim(C(\mathbb{1}) + C(X_{1,\perp})) = \dim(C(\mathbb{1})) + \dim(C(X_{1,\perp})) - \dim(C(\mathbb{1}) \cap C(X_{1,\perp}))$

X is full rank

$\therefore x_1 - \bar{x}_1 \mathbb{1}, x_2 - \bar{x}_2 \mathbb{1}, \cdots, x_p - \bar{x}_p \mathbb{1}$ are linearly independent.

(Reasoning) : $a_1 (x_1 - \bar{x}_1 \mathbb{1}) + \cdots + a_p (x_p - \bar{x}_p \mathbb{1}) = 0$
$\Rightarrow a_1 x_1 + \cdots + a_p x_p - (a_1 \bar{x}_1 + \cdots + a_p \bar{x}_p) \mathbb{1} = 0$
Since the column vectors of X are linearly independent,
$a_1 x_1 + \cdots + a_p x_p = 0 \Leftrightarrow a_1 = \cdots a_p = 0$

$\therefore a_1 x_1 + \cdots + a_p x_p = 0$ and $a_1 \bar{x}_1 + \cdots + a_p \bar{x}_p = 0$

$\therefore \dim(C(X_{1,\perp})) = p$

$\dim(C(\mathbb{1}) \cap C(X_{1,\perp})) = \dim(\{0\}) = 0$

$\therefore \dim(C(\mathbb{1}) + C(X_{1,\perp})) = 1 + p = \dim(C(X)) = \text{rank}(X)$

$\therefore C(\mathbb{1}) + C(X_{1,\perp}) = C(X)$

Q6) Prove Theorem 6

$$\pi_X = \pi_{\mathbb{1}} + \pi_{X_1,\perp}$$

$x = (\mathbb{1}\ X_1)$, $x^Tx = \begin{pmatrix} \mathbb{1}^T\mathbb{1} & \mathbb{1}^TX_1 \\ X_1^T\mathbb{1} & X_1^TX_1 \end{pmatrix}$

Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, $A^{-1} = \begin{bmatrix} A_{11\cdot 2}^{-1} & -A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} + A_{22}^{-1} \end{bmatrix}$

$$= \begin{bmatrix} A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} + A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1} \\ -A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} & A_{22\cdot 1}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

$A_{22\cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$
$A_{22\cdot 1} = X_1^TX_1 - X_1^T\mathbb{1}(\mathbb{1}^T\mathbb{1})^{-1}\mathbb{1}^TX_1 = X_1^T(I - \pi_{\mathbb{1}})X_1$

(··· too complicated)

Let A be an arbitrary vector / matrix

$(\pi_{\mathbb{1}} + \pi_{X_1,\perp})A = \pi_{\mathbb{1}}A + \pi_{X_1,\perp}A$

: projection of A onto $C(\mathbb{1})$ + projection of A onto $C(X_{1,\perp})$

∵ $C(\mathbb{1})$ and $C(X_{1,\perp})$ are orthogonal,
⇒ projection of A onto $(C(\mathbb{1}) + C(X_{1,\perp})) = C(X)$
⇒ $\pi_X A$

∴ $\pi_X = \pi_{\mathbb{1}} + \pi_{X_1,\perp}$

Q7) Prove Theorem 7.

Given that $X = (\mathbb{1}\ X_1)$, $X_1 = (x_1\ \cdots\ x_p)$
$\beta = (\beta_0\ \gamma_1^T)^T$, $(\beta_0, \cdots, \beta_p)^T$

Show the below:

$\hat{\beta}_0 = \bar{y} + \frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1$
$\hat{\gamma}_1 = (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty$

$X\hat{\beta} = \pi_X y = (\pi_{\mathbb{1}} + \pi_{X_1,\perp})y = \pi_{\mathbb{1}}y + \pi_{X_1,\perp}y$

$X\hat{\beta} = (\mathbb{1}\ X_1)\begin{pmatrix} \hat{\beta}_0 \\ \hat{\gamma}_1 \end{pmatrix} = \mathbb{1}\hat{\beta}_0 + X_1\hat{\gamma}_1 = \mathbb{1}\hat{\beta}_0 + \pi_{\mathbb{1}}X_1\hat{\gamma}_1 + X_{1,\perp}\hat{\gamma}_1$

∴ $\mathbb{1}\hat{\beta}_0 + \pi_{\mathbb{1}}X_1\hat{\gamma}_1 = \pi_{\mathbb{1}}y$ and $X_{1,\perp}\hat{\gamma}_1 = \pi_{X_{1,\perp}}y$

∴ $\hat{\gamma}_1 = (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty$ and $\hat{\beta}_0 = (\mathbb{1}^T\mathbb{1})^{-1}\mathbb{1}^Ty - (\mathbb{1}^T\mathbb{1})^{-1}\mathbb{1}^TX_1\hat{\gamma}_1$
$= \frac{1}{n}\mathbb{1}^Ty - \frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1$
$= \bar{y} - \frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1$

Q8) Prove Theorem 8

$E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\gamma}_1] = \gamma_1$

$\mathrm{Var}(\hat{\beta}_0) = \sigma^2\left[\frac{1}{n} + \left(\frac{\mathbb{1}^TX_1}{n}\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(\frac{X_1^T\mathbb{1}}{n}\right)\right]$
$\mathrm{Var}(\hat{\gamma}_1) = \sigma^2(X_{1,\perp}^TX_{1,\perp})^{-1}$

$E[\hat{\gamma}_1] = E[(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty] = (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^T(\mathbb{1}\beta_0 + X_1\gamma_1)$
$= (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^T(\mathbb{1}\beta_0 + \pi_{\mathbb{1}}X_1\gamma_1 + X_{1,\perp}\gamma_1)$
$= (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^TX_{1,\perp}\gamma_1$
$= \gamma_1$

$E[\hat{\beta}_0] = E[\frac{1}{n}\mathbb{1}^Ty - \frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1] = \frac{1}{n}\mathbb{1}^T[\mathbb{1}\beta_0 + X_1\gamma_1] - \frac{1}{n}\mathbb{1}^TX_1\gamma_1$
$= \frac{1}{n}\cdot n\beta_0 + \frac{1}{n}\mathbb{1}^TX_1\gamma_1 - \frac{1}{n}\mathbb{1}^TX_1\gamma_1$
$= \beta_0$

$\mathrm{Var}(\hat{\gamma}_1) = \mathrm{Var}[(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty)$
$= (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^T\mathrm{Var}(y)X_{1,\perp}(X_{1,\perp}^TX_{1,\perp})^{-1}$
$= (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^T\sigma^2I_nX_{1,\perp}(X_{1,\perp}^TX_{1,\perp})^{-1}$
$= \sigma^2(X_{1,\perp}^TX_{1,\perp})^{-1}$

$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\frac{1}{n}\mathbb{1}^Ty - \frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1)$
$= \mathrm{Var}(\frac{1}{n}\mathbb{1}^Ty) + \mathrm{Var}(\frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1) - \mathrm{cov}(\frac{1}{n}\mathbb{1}^Ty, \frac{1}{n}\mathbb{1}^TX_1\hat{\gamma}_1)$

* $\mathrm{cov}(\frac{1}{n}\mathbb{1}^Ty, \frac{1}{n}\mathbb{1}^TX_1(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty) = \frac{1}{n^2}\mathbb{1}^T\mathrm{cov}(y,y)X_{1,\perp}(X_{1,\perp}^TX_{1,\perp})^{-1}X_1^T\mathbb{1}$
$= 0$ (∵ $\mathbb{1}^TX_{1,\perp} = 0$)

$\mathrm{Var}(\frac{1}{n}\mathbb{1}^Ty) = \frac{1}{n^2}\mathbb{1}^T\mathrm{var}(y)\mathbb{1} = \sigma^2\frac{1}{n^2}\mathbb{1}^T\mathbb{1} = \sigma^2/n$

∴ $\mathrm{Var}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \left(\frac{\mathbb{1}^TX_1}{n}\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(\frac{X_1^T\mathbb{1}}{n}\right)\right)$ □

Q9) what is the matrix representation of the following estimates:

(a) $E[\hat{\beta}_0]$
(b) $E[\hat{\gamma}_1]$
(c) $Var(\hat{\beta}_0)$
(d) $Var(\hat{\gamma}_1)$

(a) $E[\hat{\beta}] = \beta = (1\ 0^T)\beta$
(b) $E[\hat{\gamma}] = (0\ I_p)\beta$

(c) $Var(\hat{\beta}_0) = \sigma^2 (1\ 0^T)(X^TX)^{-1}\begin{pmatrix}1\\0\end{pmatrix}$
(d) $Var(\hat{\gamma}_1) = \sigma^2 \cdot (0\ I_p)(X^TX)^{-1}\begin{pmatrix}0\\I_p\end{pmatrix}$

$Var(\hat{\beta}_0) = (1\ 0^T)\, Var(\hat{\beta})\begin{pmatrix}1\\0\end{pmatrix}$
$= \sigma^2 (1\ 0^T)(X^TX)^{-1}\begin{pmatrix}1\\0\end{pmatrix}$

$Var(\hat{\gamma}) = \sigma^2 (0\ I_p)(X^TX)^{-1}\begin{pmatrix}0\\I_p\end{pmatrix}$

Q10) Prove Corollary 10

(a) $E[\hat{\mu}_i] = \mu_i$
(b) $E[\hat{y}_Z] = E[y_Z]$
(c) $E[\hat{\varepsilon}_i] = E[\varepsilon_i]$
(d) $E[\hat{\sigma}^2] = \sigma^2$

(e) $Var(\hat{\mu}_i) = \sigma^2\left[\frac{1}{n} + \left(z_i^T - \frac{1^TX_1}{n}\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(z_i - \frac{X_1^T1}{n}\right)\right]$
(f) $Var(\hat{y}_Z) = \sigma^2\left[\frac{1}{n} + \left(z^T - \frac{1^TX_1}{n}\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(z - \frac{X_1^T1}{n}\right)\right]$
(g) $Var(\hat{\varepsilon}_i) = \sigma^2\left[1 - \frac{1}{n} - \left(z_i^T - \frac{1^TX_1}{n}\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(z_i - \frac{X_1^T1}{n}\right)\right]$

(a) $\hat{\mu}_i = \hat{\beta}_0 + z_i^T\hat{\gamma}_1$, $\therefore E[\hat{\mu}_i] = E[\hat{\beta}_0 + z_i^T\hat{\gamma}_1] = \beta_0 + z_i^T\gamma_1 = \mu_i$
(b) $\hat{y}_Z = \hat{\beta}_0 + z^T\hat{\gamma}_1$ $\therefore E[\hat{y}_Z] = E[\hat{\beta}_0 + z^T\hat{\gamma}_1] = \beta_0 + z^T\gamma_1 = E[y_Z]$
(c) $\hat{\varepsilon}_i = y_i - \hat{y}_i$, $E[\hat{\varepsilon}_i] = E[y_i] - E[\hat{y}_i] = \mu_i - E[\hat{\mu}_i] = 0 = E[\varepsilon_i]$

(d) $\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n-p-1}\|\hat{\varepsilon}\|^2 = \frac{1}{n-p-1}\|(I_n - \pi_X)y\|^2 = \frac{1}{n-p-1}\, y^T(I_n - \pi_X)y$

$y^T\left(\frac{I_n - \pi_X}{\sigma^2}\right)y \sim \chi^2(r;\delta)$, $r = \text{rank}\left(\frac{I_n - \pi_X}{\sigma^2}\right)$, $\delta = \mu^T\left(\frac{I_n - \pi_X}{\sigma^2}\right)\mu$

$\mu = E[y] = X\beta$, $\mu^T(I_n - \pi_X)\mu = \beta^TX^T(I_n - \pi_X)X\beta = \beta^TX^T(X\beta - X(X^TX)^{-1}X^TX\beta) = 0$
$\therefore \delta = 0$

$r = \text{rank}\left(\frac{I_n - \pi_X}{\sigma^2}\right) = \text{rank}(I_n - \pi_X) = tr(I_n - \pi_X) = tr(I_n) - tr(X(X^TX)^{-1}X^T)$
$= tr(I_n) - tr((X^TX)^{-1}X^TX) = n - (p+1) = n - p - 1$

$A\Sigma = \left(\frac{I_n - \pi_X}{\sigma^2}\right)\sigma^2 I_n = I_n - \pi_X$: idempotent

$\therefore\ y^T\left(\frac{I_n - \pi_X}{\sigma^2}\right)y \sim \chi^2(n-p-1)$

$\Rightarrow E[y^T(I_n - \pi_X)y] = \sigma^2 \cdot (n-p-1)$
$\Rightarrow E[\hat{\sigma}^2] = E\left[\frac{1}{n-p-1}\, y^T(I_n - \pi_X)y\right] = \sigma^2$

(e) Approach #1

$Var(\hat{\mu}_i) = Var(\hat{y}_{z_i}) = Var(\hat{\beta}_0 + z_i^T\hat{\gamma}_1)$
$= Var(\hat{\beta}_0) + Var(z_i^T\hat{\gamma}_1) + 2Cov(\hat{\beta}_0, z_i^T\hat{\gamma}_1)$
$= \underbrace{Var\left(\frac{1}{n}1^Ty - \frac{1}{n}1^TX_1\hat{\gamma}_1\right)}_{①} + \underbrace{z_i^T Var(\hat{\gamma}_1) z_i}_{②} + \underbrace{2Cov(\hat{\beta}_0, z_i^T\hat{\gamma}_1)}_{③}$

① $Var\left(\frac{1}{n}1^Ty - \frac{1}{n}1^TX_1\hat{\gamma}_1\right)$

$= \frac{1}{n^2}1^T Var(y)1 + \left(\frac{1^TX_1}{n}\right)Var(\hat{\gamma}_1)\left(\frac{X_1^T1}{n}\right) - 2Cov\left(\frac{1}{n}1^Ty, \frac{1}{n}1^TX_1\hat{\gamma}_1\right)$

$= \frac{1}{n^2}1^T Var(y)1 + \left(\frac{1^TX_1}{n}\right)Var(\hat{\gamma}_1)\left(\frac{X_1^T1}{n}\right)$ $\because Cov\left(\frac{1}{n}1^Ty, \frac{1}{n}1^TX_1\hat{\gamma}_1\right)$
$1^Ty \in C(1)$, $\hat{\gamma}_1 \in C(X_{1,\perp})$

② $z_i^T Var(\hat{\gamma}_1) z_i = z_i^T\left(\sigma^2(X_{1,\perp}^TX_{1,\perp})^{-1}\right)z_i = \sigma^2 \cdot z_i^T(X_{1,\perp}^TX_{1,\perp})^{-1}z_i$

③ $2Cov(\hat{\beta}_0, z_i^T\hat{\gamma}_1) = 2\,Cov(\hat{\beta}_0\hat{\gamma}_1)z_i = 0$ $\because \hat{\beta}_0 \in C(1)$, $\hat{\gamma}_1 \in C(X_{1,\perp})$

$\therefore Var(\hat{\mu}_i) = \sigma^2\left(\frac{1}{n} + \left(\frac{1^TX_1}{n}\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(\frac{X_1^T1}{n}\right) + z_i^T(X_{1,\perp}^TX_{1,\perp})^{-1}z_i\right)$

Approach #2

$Var(\hat{\mu}_i) = Var(\hat{\beta}_0 + z_i^T\hat{\gamma}_1) = Var\left(\frac{1}{n}1^Ty - \frac{1}{n}1^TX_1\hat{\gamma}_1 + z_i^T\hat{\gamma}_1\right)$
$= Var\left(\frac{1}{n}1^Ty\right) + Var\left(\left(z_i^T - \frac{1}{n}1^TX_1\right)\hat{\gamma}_1\right)$
$= \frac{1}{n^2}1^T1\, Var(y) + \left(z_i^T - \frac{1}{n}1^TX_1\right)Var(\hat{\gamma}_1)\left(z_i - \frac{1}{n}X_1^T1\right)$
$= \sigma^2\left(\frac{1}{n} + \left(z_i^T - \frac{1}{n}1^TX_1\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(z_i - \frac{1}{n}X_1^T1\right)\right)$

(f) $Var(\hat{y}_Z) = Var(\hat{\beta}_0 + z^T\hat{\gamma}_1)$
$= Var\left(\frac{1}{n}1^Ty - \frac{1}{n}1^TX_1\hat{\gamma}_1 - z^T\gamma\right)$
$= \sigma^2\left(\frac{1}{n} + \left(z^T - \frac{1}{n}1^TX_1\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(z - \frac{1}{n}X_1^T1\right)\right)$

(g) $Var(\hat{\varepsilon}_i) = Var(y_i - \hat{y}_i) = Var(y_i) + Var(\hat{y}_i) - 2Cov(y_i, \hat{y}_i)$
$= Var(y_i) + Var(\hat{y}_i) - 2Cov(\hat{y}_i + \hat{\varepsilon}_i, \hat{y}_i)$
$= Var(y_i) + Var(\hat{y}_i) - 2\,Var(\hat{y}_i) - 2Cov(\hat{\varepsilon}_i, \hat{y}_i)$
$= Var(y_i) - Var(\hat{y}_i)$ $(\because Cov(\hat{\varepsilon}_i, \hat{y}_i) = 0)$
$= \sigma^2\left(1 - \frac{1}{n} - \left(z_i^T - \frac{1}{n}1^TX_1\right)(X_{1,\perp}^TX_{1,\perp})^{-1}\left(z_i - \frac{1}{n}X_1^T1\right)\right)$

* $Cov(\hat{\varepsilon}, \hat{y}) = Cov((I_n - \pi_X)y, \pi_X y)$
$= (I_n - \pi_X)Var(y)\pi_X$
$= \sigma^2(I_n - \pi_X)\pi_X$
$= 0$

Q11) Derive the following:

(a) $\mathrm{Var}(\hat{\mu}_i) = \sigma^2 (1 \; z_i^T)(X^TX)^{-1}\begin{pmatrix}1\\ z_i\end{pmatrix}$

(b) $\mathrm{Var}(\hat{y}_z) = \sigma^2 (1 \; z^T)(X^TX)^{-1}\begin{pmatrix}1\\ z\end{pmatrix}$

[Solution]:

(a) $\mathrm{Var}(\hat{\mu}_i) = \mathrm{Var}(\hat{\beta}_0 + z_i^T\hat{\gamma}_1) = \mathrm{Var}\left([1 \; z_i^T]\begin{bmatrix}\hat{\beta}_0\\ \hat{\gamma}_1\end{bmatrix}\right)$

$= (1 \; z_i^T)\,\mathrm{Var}(\hat{\beta}_0)\begin{pmatrix}1\\ z_i\end{pmatrix}$

$= (1 \; z_i^T)\,\mathrm{Var}\left((X_1^TX_1)^{-1}X_1^Ty\right)\begin{pmatrix}1\\ z_i\end{pmatrix}$

$= (1 \; z_i^T)(X_1^TX_1)^{-1}X_1^T\,\mathrm{Var}(y)\,X_1(X_1^TX_1)^{-1}\begin{pmatrix}1\\ z_i\end{pmatrix}$

$= \sigma^2(1 \; z_i^T)(X_1^TX_1)^{-1}\begin{pmatrix}1\\ z_i\end{pmatrix}$

(b) $\mathrm{Var}(\hat{y}_z)$: $z_i \Rightarrow z$ in $\mathrm{Var}(\hat{\mu}_i)$

$= \sigma^2(1 \; z^T)(X_1^TX_1)^{-1}\begin{pmatrix}1\\ z\end{pmatrix}$

ANOVA [Analysis of Variance]

Q12) Prove Theorem 11

$SST = \|y - \mathbb{1}\bar{y}\|^2 = \sum_{i=1}^n (y_i - \bar{y})^2$

$SSE = \|y - \hat{y}\|^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

$SSR = \|\hat{y} - \bar{y}\|^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

(Solution)

$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$

$= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (y_i - \bar{y})^2$

$\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = [(I_n - \Pi_X)y]^T[(\Pi_X - \Pi_{\mathbb{1}})y]$

$= y^T(I_n - \Pi_X)(\Pi_X - \Pi_{\mathbb{1}})y$

$I_n - \Pi_X \in C(X)^\perp$ and $\Pi_X - \Pi_{\mathbb{1}} \in C(X)$

$\therefore (I_n - \Pi_X)(\Pi_X - \Pi_{\mathbb{1}}) = 0$

$\therefore SST = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SSE + SSR$

Q13) Prove Theorem 12

(a) $SST = SSR + SSE$

$SST = y^T(I_n - \Pi_{\mathbb{1}})y$

$SSE = y^T(I_n - \Pi_X)y$

$SSR = y^T(\Pi_X - \Pi_{\mathbb{1}})y$

(b) For the regression sum of square,

$SSR = y^T\Pi_{X_1,\perp}y = \hat{\gamma}_1^T X_{1,\perp}^T X_{1,\perp}\hat{\gamma}_1$

(a)

$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = [(I_n - \Pi_{\mathbb{1}})y]^T[(I_n - \Pi_{\mathbb{1}})y]$

$= y^T(I_n - \Pi_{\mathbb{1}})y$

$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = [(I_n - \Pi_X)y]^T[(I_n - \Pi_X)y]$

$= y^T(I_n - \Pi_X)y$

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = [(\Pi_X - \Pi_{\mathbb{1}})y]^T[(\Pi_X - \Pi_{\mathbb{1}})y]$

$= y^T(\Pi_X - \Pi_{\mathbb{1}})y$

(b) $SSR = y^T(\Pi_X - \Pi_{\mathbb{1}})y = y^T(\Pi_{\mathbb{1}} + \Pi_{X_1,\perp} - \Pi_{\mathbb{1}})y$

$= y^T\Pi_{X_1,\perp}y = (\Pi_{X_1,\perp}y)^T(\Pi_{X_1,\perp}y)$

$= (X_{1,\perp}(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty)^T(X_{1,\perp}(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty)$

$= (X_{1,\perp}\hat{\gamma}_1)^T(X_{1,\perp}\hat{\gamma}_1)$

$= \hat{\gamma}_1^TX_{1,\perp}^TX_{1,\perp}\hat{\gamma}_1$

Q14) Define the coefficient of Determination R^2 in terms of SSR and SST

$$R^2 = \frac{SSR}{SST}$$

Q15) Prove Theorem 14

(a) $R^2 = 0 \Leftrightarrow \hat{\beta}_1 = \cdots = \hat{\beta}_p = 0$

(b) $R^2 = 0 \Leftrightarrow y = X\hat{\beta}$

(c) R^2 = square of the sample correlation coefficient between y and $\hat{y}$.

$$R^2 = \frac{\left[\sum_{i=1}^n (y_i - \bar{y})(\hat{y} - \tilde{y})\right]^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \tilde{y})^2}, \quad \tilde{y} = \frac{1}{n}\sum_{i=1}^n \hat{y}_i$$

(a) $R^2 = 0 \Leftrightarrow SSR = 0 \quad \therefore \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 0 \quad \therefore \hat{y}_i = \bar{y} \;\; (i = 1, 2, \cdots, n)$

$y^T(\Pi_X - \Pi_{\mathbb{1}})y = 0$

$\|(\Pi_X - \Pi_{\mathbb{1}})y\|^2 = 0 \quad \therefore \Pi_X y = \Pi_{\mathbb{1}} y$

$\Pi_X = \Pi_{\mathbb{1}} + \Pi_{X_1,\perp}$, $\therefore \Pi_{\mathbb{1}} y + \Pi_{X_1,\perp}y = \Pi_{\mathbb{1}} y \Rightarrow \Pi_{X_1,\perp}y = 0$

$\Rightarrow X_{1,\perp}(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty = 0 \Leftrightarrow (X_{1,\perp}^TX_{1,\perp})(X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty = 0$

$\Leftrightarrow (X_{1,\perp}^TX_{1,\perp})^{-1}X_{1,\perp}^Ty = 0$

$\Leftrightarrow \hat{\gamma}_1 = 0$

$\therefore \hat{\beta}_1 = \cdots = \hat{\beta}_p = 0$

(b) $R^2=1 \Leftrightarrow SSR = SST \Leftrightarrow \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2$

$\Leftrightarrow y^T(\Pi_X-\Pi_{\mathbb{1}})y = y^T(I_n-\Pi_{\mathbb{1}})y$

$\Leftrightarrow \|(\Pi_X-\Pi_{\mathbb{1}})y\|^2 = \|(I_n-\Pi_{\mathbb{1}})y\|$

$\therefore \Pi_X y = y \Leftrightarrow X\hat{\beta} = y$

(c) $$\frac{\left[\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}-\bar{\hat{y}})\right]^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}$$

nominator: $[(I_n-\Pi_{\mathbb{1}})y]^T[(\Pi_X-\Pi_{\mathbb{1}}\Pi_X)y]$

$= y^T(I_n-\Pi_{\mathbb{1}})(\Pi_X-\Pi_{\mathbb{1}})y$

$= y^T(\Pi_X-\Pi_{\mathbb{1}}-\Pi_{\mathbb{1}}+\Pi_{\mathbb{1}})y$

$= y^T(\Pi_X-\Pi_{\mathbb{1}})y = SSR$

denumerator: $\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2 = [(\Pi_X-\Pi_{\mathbb{1}}\Pi_X)y]^T[(\Pi_X-\Pi_{\mathbb{1}}\Pi_X)y]$

$= y^T(\Pi_X-\Pi_{\mathbb{1}}\Pi_X)y$

$= y^T(\Pi_X-\Pi_{\mathbb{1}})y = SSR$

$$\therefore \frac{\left[\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}-\bar{\hat{y}})\right]^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2} = \frac{SSR^2}{SST\cdot SSR} = \frac{SSR}{SST} = R^2$$

Distribution of Estimators:

Q16) Prove Theorem 16

(a) $\hat{\gamma}_1 \sim N_p(\gamma_1, \sigma^2(X_{1,\perp}^T X_{1,\perp})^{-1})$

(b) $(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-p-1)$

(c) $\hat{\gamma}_1$ and $\hat{\sigma}^2$ are independent

(a) $E[\hat{\gamma}_1] = \gamma_1$ and $Var(\hat{\gamma}_1) = \sigma^2(X_{1,\perp}^T X_{1,\perp})^{-1}$

$\therefore \hat{\gamma}_1 \sim N(\gamma_1, \sigma^2(X_{1,\perp}^T X_{1,\perp})^{-1})$

(b) $$(n-p-1)\hat{\sigma}^2/\sigma^2 = \frac{(n-p-1)}{\sigma^2}\cdot\frac{1}{n-p-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

$$= \frac{1}{\sigma^2}y^T(I_n-\Pi_X)y = y^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)y \sim \chi^2(r;\delta), \quad r = \operatorname{rank}\left(\frac{I_n-\Pi_X}{\sigma^2}\right),\ \delta = \mu^T A\mu,\ \mu = E[y],\ A = \left(\frac{I_n-\Pi_X}{\sigma^2}\right)$$

$\Leftrightarrow A\Sigma$ is idempotent

$$A\Sigma = \left(\frac{I_n-\Pi_X}{\sigma^2}\right)\cdot\sigma^2 I_n = I_n-\Pi_X : \text{idempotent}.$$

$$\therefore y^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)y \sim \chi^2(n-p-1) \quad \because \operatorname{rank}\left(\frac{I_n-\Pi_X}{\sigma^2}\right) = \operatorname{trace}(I_n-\Pi_X)$$

$= tr(I_n) - tr(X(X^TX)^{-1}X^T)$

$= tr(I_n) - tr(X^TX(X^TX)^{-1})$

$= n-(p+1) = n-p-1$

$$\delta = \mu^T A\mu = (X\beta)^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)(X\beta)$$

$(I_n-\Pi_X)X = X - X(X^TX)^{-1}X^TX = 0 \quad \therefore \delta = 0$

(c) $\hat{\gamma}_1 = (X_{1,\perp}^T X_{1,\perp})^{-1}X_{1,\perp}^T y$ and $\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \frac{1}{n-p-1}y^T(I_n-\Pi_X)y$

$\hat{\gamma}_1 \in C(X_{1,\perp}) \subset C(X)$ and $\hat{\sigma}^2 \in C(X)^{\perp}$

$\therefore \hat{\gamma}_1$ and $\hat{\sigma}^2$ are orthogonal

$$A\Sigma B = \left[(X_{1,\perp}^T X_{1,\perp})^{-1}X_{1,\perp}^T\right]\left[\sigma^2 I_n\right]\left[\frac{I_n-\Pi_X}{\sigma^2}\right] = 0$$

$\therefore Ay = \hat{\gamma}_1$ and $y^TBy = (n-p-1)\hat{\sigma}^2$ are independent

$\Rightarrow \hat{\gamma}_1$ and $\hat{\sigma}^2$ are independent.

Q17) Prove Remark in Chapter 5.3

(a) $SSE/\sigma^2 \sim \chi^2(n-p-1)$

(b) $SSR/\sigma^2 \sim \chi^2(p;\delta)$

(a) $$SSE/\sigma^2 = y^T(I_n-\Pi_X)y/\sigma^2 = y^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)y = y^TBy$$

$$B\Sigma = \left(\frac{I_n-\Pi_X}{\sigma^2}\right)\cdot\sigma^2 I_n = I_n-\Pi_X : \text{idempotent}$$

$$\therefore y^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)y \sim \chi^2(r;\delta)$$

$$r = \operatorname{rank}\left(\frac{I_n-\Pi_X}{\sigma^2}\right) = \operatorname{rank}(I_n-\Pi_X) = n-p-1$$

$$\delta = \mu^T A\mu = E[y]^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)E[y] = \beta^T X^T\left(\frac{I_n-\Pi_X}{\sigma^2}\right)X\beta = 0 \quad \because \langle I_n-\Pi_X, X\rangle = 0$$

(b) $$SSR/\sigma^2 = y^T\left(\frac{\Pi_X-\Pi_{\mathbb{1}}}{\sigma^2}\right)y = y^TAy$$

$$A\Sigma = \left(\frac{\Pi_X-\Pi_{\mathbb{1}}}{\sigma^2}\right)\cdot\sigma^2 I_n = \Pi_X-\Pi_{\mathbb{1}} : \text{idempotent}$$

$$\therefore y^T\left(\frac{\Pi_X-\Pi_{\mathbb{1}}}{\sigma^2}\right)y \sim \chi^2(r;\delta)$$

$$r = \operatorname{rank}(\Pi_X-\Pi_{\mathbb{1}}) = \operatorname{trace}(\Pi_X-\Pi_{\mathbb{1}}) = \operatorname{trace}(\Pi_X) - \operatorname{trace}(\Pi_{\mathbb{1}})$$

$$= (p+1) - 1 = p$$

$$\delta = \mu^T A\mu = E[y]^T\left(\frac{\Pi_X-\Pi_{\mathbb{1}}}{\sigma^2}\right)E[y] = (X\beta)^T\left(\frac{\Pi_X-\Pi_{\mathbb{1}}}{\sigma^2}\right)(X\beta)$$

$$= (X\beta)^T\left(\frac{\Pi_{X_{1,\perp}}+\Pi_{\mathbb{1}}-\Pi_{\mathbb{1}}}{\sigma^2}\right)(X\beta)$$

$$= \sigma^{-2}(X_{1,\perp}\gamma_1)^T\Pi_{X_{1,\perp}}(X_{1,\perp}\gamma_1)$$

$$= \sigma^{-2}\gamma_1^T(X_{1,\perp}^TX_{1,\perp})\gamma_1$$

Q18) Prove Theorem 17

$H_0: \gamma_1 = \gamma_1^*$ vs $H_0: \gamma_1 \neq \gamma_1^*$

$$F = \frac{(\hat{\gamma}_1 - \gamma_1^*)^T X_{1,\perp}^T X_{1,\perp} (\hat{\gamma}_1 - \gamma_1^*)/p}{SSE/(n-p-1)}$$

Under H_0: $F \sim F(p, n-p-1)$

Under H_1: $F \sim F(p, n-p-1; \delta)$, $\delta = \frac{(\gamma_1 - \gamma_1^*)^T X_{1,\perp}^T X_{1,\perp} (\gamma_1 - \gamma_1^*)}{\sigma^2}$

$\hat{\gamma}_1 \sim N(\gamma_1, \sigma^2 (X_{1,\perp}^T X_{1,\perp})^{-1})$

$(\hat{\gamma}_1 - \gamma_1)/\sigma^2 (X_{1,\perp}^T X_{1,\perp})^{-1/2} \sim N(0, I_p)$

$\therefore (\hat{\gamma}_1 - \gamma_1)^T (X_{1,\perp}^T X_{1,\perp})(\hat{\gamma}_1 - \gamma_1)/\sigma^2 \sim \chi^2(p)$

$SSE = y^T (I - \Pi_X) y$, $SSE/\sigma^2 = y^T \left(\frac{I_n - \Pi_X}{\sigma^2}\right) y \sim \chi^2(n-p-1)$

$$\therefore \frac{(\hat{\gamma}_1 - \gamma_1)^T (X_{1,\perp}^T X_{1,\perp})(\hat{\gamma}_1 - \gamma_1)/p}{y^T (I_n - \Pi_X) y/(n-p-1)} \sim \frac{V_1/n_1}{V_2/n_2}, \quad V_1 \sim \chi^2(n_1), V_2 \sim \chi^2(n_2)$$

$\therefore$ under H_0, $F \sim F(p, n-p-1)$

under H_1, $\hat{\gamma}_1 \sim \gamma_1^* \sim N(\gamma_1 - \gamma_1^*, \sigma^2 (X_{1,\perp}^T X_{1,\perp})^{-1})$

$\therefore (\hat{\gamma}_1 - \gamma_1^*)^T (X_{1,\perp}^T X_{1,\perp})(\hat{\gamma}_1 - \gamma_1^*)/\sigma^2 \sim \chi^2(p; \delta)$, $\delta = (\gamma_1 - \gamma_1^*)^T (X_{1,\perp}^T X_{1,\perp})(\gamma_1 - \gamma_1^*)/\sigma^2$

$\therefore F \sim F(p, n-p-1; \delta)$, $\delta = \frac{(\gamma_1 - \gamma_1^*)^T (X_{1,\perp}^T X_{1,\perp})(\gamma_1 - \gamma_1^*)}{\sigma^2}$ □

Q19) Prove Theorem 18

$H_0: \gamma_1 = 0$ vs $H_1: \gamma_1 \neq 0$

$$F = \frac{\hat{\gamma}_1^T X_{1,\perp}^T X_{1,\perp} \hat{\gamma}_1/p}{SSE/(n-p-1)}$$

Under H_0: $F \sim F(p, n-p-1)$

Under H_1: $F \sim F(p, n-p-1; \delta)$, $\delta = \frac{\gamma_1 X_{1,\perp}^T X_{1,\perp} \gamma_1}{\sigma^2}$

insert γ_1^* into Theorem 18 (previous question)

Q20) Prove Theorem 19

$H_0: \beta_j = \beta_j^*$ vs $H_1: \beta_j \neq \beta_j^*$

for some $\beta_j^* \in \mathbb{R}$. Consider a test statistic T defined as below

$$T = \frac{\beta_j - \beta_j^*}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

Let $e_j = [0 \cdots 1 \cdots 0]^T$ 1 is in the jth row

$e_j^T \hat{\gamma}_1 \sim N(e_j^T \gamma_1, e_j^T \sigma^2 (X_{1,\perp}^T X_{1,\perp})^{-1} e_j)$

$\Rightarrow \hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{jj})$, C_{jj}: j^{th} diagonal entry of $(X_{1,\perp}^T X_{1,\perp})^{-1}$

$(\beta_j - \hat{\beta}_j)/\sqrt{\sigma^2 C_{jj}} \sim N(0, I_n)$

$$\frac{\beta_j - \hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\beta_j - \hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}} \Big/ \sqrt{\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2}}/(n-p-1) = \frac{Z}{\sqrt{V/(n-p-1)}}, \quad Z \sim N(0,1), \; V \sim \chi^2(n-p-1)$$

$$\therefore \frac{\beta_j - \hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t(n-p-1)$$

Q21) Prove Theorem 20

$H_0: \beta_0 = \beta_0^*$ vs. $H_1: \beta_0 \neq \beta_0^*$

for some constant $\beta_0^* \in \mathbb{R}$. Consider the test statistic T below:

$$T = \frac{\beta_0 - \beta_0^*}{\sqrt{\hat{\sigma}^2 D_{11}}}$$

under H_0, $T \sim t(n-p-1)$

D_{11}: 1st diagonal entry of $(X^T X)^{-1}$

D_{11}: $\left[\frac{1}{n} + \left(\frac{1^T X_1}{n}\right)(X_{1,\perp}^T X_{1,\perp})^{-1}\left(\frac{X_1^T 1}{n}\right)\right]$ * ↓ 아래 피테이지

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left[\frac{1}{n} + \left(\frac{1^T X_1}{n}\right)(X_{1,\perp}^T X_{1,\perp})^{-1}\left(\frac{X_1^T 1}{n}\right)\right]\right)$$
$$\overset{d}{=} N(\beta_0, \sigma^2 D_{11})$$

$$\therefore \text{under } H_0, \; \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\sigma^2 D_{11}}} \sim N(0,1)$$

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 D_{11}}} = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\sigma^2 D_{11}}} \Big/ \sqrt{\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2}}/(n-p-1) = \frac{Z}{\sqrt{V/(n-p-1)}}, \quad Z \sim N(0,1), \; V \sim \chi^2(n-p-1)$$

$\therefore (\hat{\beta}_0 - \beta_0)/\sqrt{\hat{\sigma}^2 D_{11}} \sim t(n-p-1)$ □

CONFIDENCE INTERVAL

Q22) Prove Theorem 22

The $(1-\alpha)100\%$ CI for β_j and β_0 are:

β_j: $\hat{\beta}_j \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 C_{jj}}$

β_0: $\hat{\beta}_0 \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 D_{11}}$

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t(n-p-1)$$

$\therefore P(|(\hat{\beta}_j - \beta_j)/\sqrt{\hat{\sigma}^2 C_{jj}}| > t_{\alpha/2}) = \alpha$

$\Rightarrow P(\hat{\beta}_j - \beta_j/\sqrt{\hat{\sigma}^2 C_{jj}} > t_{\alpha/2}, \; \hat{\beta}_j - \beta_j/\sqrt{\hat{\sigma}^2 C_{jj}} < -t_{\alpha/2})$

$$\therefore CI: -t_{\alpha/2} < (\hat{\beta}_j - \beta_j)/\sqrt{\hat{\sigma}^2 C_{jj}} < t_{\alpha/2}$$
$$= \hat{\beta}_j - t_{\alpha/2}\sqrt{\hat{\sigma}^2 C_{jj}} < \beta_j < \hat{\beta}_j + t_{\alpha/2}\sqrt{\hat{\sigma}^2 C_{jj}}$$

$\therefore \beta_j$ CI: $\hat{\beta}_j \pm t_{\alpha/2}\sqrt{\hat{\sigma}^2 C_{jj}}$, $t_{\alpha/2} = t_{\alpha/2}(n-p-1)$

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 D_{11}}} \sim t(n-p-1)$$

$\therefore$ Similar to above $-t_{\alpha/2}(n-p-1) < \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 D_{11}}} < t_{\alpha/2}(n-p-1)$

$\therefore$ CI of β_0: $\hat{\beta}_0 \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 D_{11}}$

Q23) Prove Theorem 2.2

Let $C_z = \left[\frac{1}{n} + \left(z^T - \frac{\mathbb{1}^T X_1}{n}\right)\left(X_{1,\perp}^T X_{1,\perp}\right)^{-1}\left(z^T - \frac{X_1^T \mathbb{1}}{n}\right)\right]$

(a) CI of μ_i at $z_i\,(i=1,\cdots,n)$: $\hat{\mu}_i \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 C_{jj}}$

(b) CI of y_z at $z \in \mathbb{R}^p$ (y_z : out-of-sample response)
: $\hat{y}_z \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2(1+C_z)}$

(proof)

(a) $\hat{\mu}_i = \hat{\beta}_0 + z_i^T\hat{\gamma}_i \sim N(\mu_i, \mathrm{var}(\hat{\mu}_i)) \overset{d}{=} N(\mu_i, \sigma^2 C_z)$

$\therefore \dfrac{\hat{\mu}_i - \mu_i}{\sqrt{\sigma^2 C_z}} \sim N(0,1)$

$\Rightarrow \dfrac{\hat{\mu}_i - \mu_i}{\sqrt{\hat{\sigma}^2 C_z}} = \dfrac{\hat{\mu}_i - \mu_i}{\sqrt{\sigma^2 C_z}} \Big/ \sqrt{\dfrac{(n-p-1)\hat{\sigma}^2}{\sigma^2}/(n-p-1)} = \dfrac{Z}{\sqrt{V/(n-p-1)}}$, $\begin{matrix} Z \sim N(0,1) \\ V \sim \chi^2(n-p-1) \end{matrix}$

$\therefore \dfrac{\hat{\mu}_i - \mu_i}{\sqrt{\hat{\sigma}^2 C_z}} \sim t(n-p-1)$, μ_i CI : $\hat{\mu}_i \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 C_z}$

(b) $y_z \sim N(\beta_0 + z^T\gamma, \sigma^2)$

$\hat{y}_z - y_z \sim N(0, \sigma^2(1+C_z))$ $\because$ $\hat{y}_z$ and y_z are independent

$\dfrac{\hat{y}_z - y_z}{\sqrt{\sigma^2(1+C_z)}} \sim N(0,1)$

$\therefore \dfrac{\hat{y}_z - y_z}{\sqrt{\hat{\sigma}^2(1+C_z)}} = \dfrac{\hat{y}_z - y_z}{\sqrt{\sigma^2(1+C_z)}} \Big/ \sqrt{\dfrac{(n-p-1)\hat{\sigma}^2}{\sigma^2}/(n-p-1)} = \dfrac{Z}{\sqrt{V/(n-p-1)}}$, $\begin{matrix} Z \sim N(0,1) \\ V \sim \chi^2(n-p-1) \end{matrix}$

$\therefore \dfrac{\hat{y}_z - y_z}{\sqrt{\hat{\sigma}^2(1+C_z)}} \sim t(n-p-1) \Rightarrow$ CI y_z : $\hat{y}_z \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2(1+C_z)}$

$* \; D_{11} = \left[\frac{1}{n} + \left(\frac{\mathbb{1}^T X_1}{n}\right)^T \left(X_{1,\perp}^T X_{1,\perp}\right)^{-1}\left(\frac{X_1^T \mathbb{1}}{n}\right)\right]$

$D_{11} = e_1^T (X^T X)^{-1} e_1$

$X = [\mathbb{1}\ X_1]$, $X^T X = \begin{bmatrix} \mathbb{1}^T \\ X_1^T \end{bmatrix}[\mathbb{1}\ X_1] = \begin{bmatrix} \mathbb{1}^T\mathbb{1} & \mathbb{1}^T X_1 \\ X_1^T\mathbb{1} & X_1^T X_1 \end{bmatrix}$

let $P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, $P^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & \cdots \\ \cdots & \cdots \end{bmatrix}$

$= \begin{bmatrix} (A - BD^{-1}C)^{-1} & \cdots \\ \cdots & \cdots \end{bmatrix}$

we will use $A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}$
$= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$

$D_{11} = (\mathbb{1}^T\mathbb{1})^{-1} + (\mathbb{1}^T\mathbb{1})^{-1}\mathbb{1}^T X_1 \left(X_1^T X_1 - X_1^T\mathbb{1}(\mathbb{1}^T\mathbb{1})^{-1}\mathbb{1}^T X_1\right)^{-1} X_1^T\mathbb{1}(\mathbb{1}^T\mathbb{1})^{-1}$

$= \frac{1}{n} + \left(\frac{\mathbb{1}^T X_1}{n}\right)\left(X_1^T(I - \Pi_{\mathbb{1}})X_1\right)^{-1}\left(\frac{X_1^T\mathbb{1}}{n}\right)$

$= \frac{1}{n} + \left(\frac{\mathbb{1}^T X_1}{n}\right)\left(X_{1,\perp}^T X_{1,\perp}\right)^{-1}\left(\frac{X_1^T\mathbb{1}}{n}\right)$