7 Multiple Regression

7.1 The model

Multiple Linear Regression Model with k independent variables is defined as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \tag{7.1}$$

where $i = 1, ..., n$ (n is the sample size), or in matrix form

data $\cdots \{ y_i, \underset{\sim}{x_i}, i = 1, \cdots, n \}$

$$\mathbf{y}_{n\times 1} \;=\; \mathbf{X}_{n\times r}\boldsymbol{\beta}_{r\times 1} \;+\; \boldsymbol{\varepsilon}_{n\times 1}$$

where $r = k + 1$.

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} & x_{i2} \cdots & x_{ik} \\ 1 & x & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$\to (1 \;\; x_{i1}, \cdots, x_{ik}) = \underset{\sim}{x_i}^T$

What are $\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$?

Main assumptions of the model are

1. $E(\boldsymbol{\varepsilon}) \;=\; \mathbf{0}$
2. $Cov(\boldsymbol{\varepsilon}) \;=\; \sigma^2 \mathbf{I}$
3. $\mathbf{X}$ is full column rank

7.2 Least Squares Estimation of $\boldsymbol{\beta}$

Minimize the sum of squares of deviations of observed and predicted values $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Proof. — Remark 7.1

Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$\Rightarrow y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$

1°. when the model is given, how to do inference ?

2°. k ?   3°. model ?

Remark 7.1

$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \varepsilon_i, \quad i = 1, \cdots, n.$

LSE: To find $\hat{\underset{\sim}{\beta}}^T = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k)$ by minimizing

SSE: $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \underset{\sim}{x}_i^T \underset{\sim}{\beta})^2 = (\underset{\sim}{y} - X\underset{\sim}{\beta})^T (\underset{\sim}{y} - X\underset{\sim}{\beta})$ sum of square estimated errors.

$= (\underset{\sim}{y}^T - \underset{\sim}{\beta}^T X^T)(\underset{\sim}{y} - X\underset{\sim}{\beta}) = \underset{\sim}{y}^T\underset{\sim}{y} \underbrace{- \underset{\sim}{\beta}^T X^T \underset{\sim}{y} - \underset{\sim}{y}^T X \underset{\sim}{\beta}}_{\text{scalar}} + \underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}$

$\Rightarrow SSE(\underset{\sim}{\beta}) = \underset{\sim}{y}^T\underset{\sim}{y} - 2\underset{\sim}{y}^T X \underset{\sim}{\beta} + \underset{\sim}{\beta}^T (X^T X) \underset{\sim}{\beta}$

↓ 换成 β^T 在前的形式: $\underset{\sim}{\beta}' X' \underset{\sim}{y}$

$\Rightarrow \frac{\partial SSE(\underset{\sim}{\beta})}{\partial \underset{\sim}{\beta}} = 0 - 2X^T\underset{\sim}{y} + 2(X^T X)\underset{\sim}{\beta} = 0$

$\Rightarrow (X^T X)\underset{\sim}{\beta} = X^T \underset{\sim}{y} \Rightarrow \hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{y}$

Proof of Notes:

1° $X'(I - H)\underset{\sim}{y} = (X' - X'H)\underset{\sim}{y} = \underset{\sim}{0}$

2. $H = X(X'X)^{-1}X'$: $H^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = H$ idempotent.

$H' = (X(X'X)^{-1}X')' = X(X'X)^{-1}X' = H$ symmetric

3° $(\hat{\underset{\sim}{y}})'\hat{\underset{\sim}{e}} = (X\hat{\underset{\sim}{\beta}})'\hat{\underset{\sim}{e}} = \hat{\underset{\sim}{\beta}}' X' \hat{\underset{\sim}{e}} = 0$

4° $(I - H)(I - H) = I - H - H + H^2 = I - H$. idempotent.

$(I - H)' = I - H' = I - H$ symmetric

5°. $E(\hat{\underset{\sim}{\beta}}) = E((X'X)^{-1}X'\underset{\sim}{y}) = E((X'X)^{-1}X'[X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}]) = (X'X)^{-1}X'X\underset{\sim}{\beta} = \underset{\sim}{\beta}$

6°. $Cov(\hat{\underset{\sim}{\beta}}) = [(X'X)^{-1}X'](\sigma^2 I)X(X'X)^{-1} = \sigma^2 (X'X)^{-1}$

7°. $I - H$ is idempotent, $\Rightarrow tr(I - H) = r(I - H) \overset{?}{=} n - r.$

8°. $\hat{\underset{\sim}{e}}'\hat{\underset{\sim}{e}} = tr(\hat{\underset{\sim}{e}}'\hat{\underset{\sim}{e}}) = tr(\underset{\sim}{y}'[I - H]'[I - H]\underset{\sim}{y}) = tr(\underset{\sim}{y}'[I - H]\underset{\sim}{y}) = tr(\underset{\sim}{y}\underset{\sim}{y}'(I - H))$

9°. $E(\underset{\sim}{y}\underset{\sim}{y}') = E[(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon})(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon})'] = E[X\underset{\sim}{\beta}\underset{\sim}{\beta}'X' + \underset{\sim}{\varepsilon}\underset{\sim}{\beta}'X' + X\underset{\sim}{\beta}\underset{\sim}{\varepsilon}' + \underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}']$

$= X\underset{\sim}{\beta}\underset{\sim}{\beta}'X' + \sigma^2 I$

To construct an unbiased estimator of σ^2 based on $\hat{\boldsymbol{\beta}}$

Vector of ressduals $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$

Hence

$$\begin{aligned}\hat{\varepsilon} &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\ &= [\mathbf{I} - \mathbf{H}]\mathbf{y}\end{aligned}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ $\quad \hat{\underset{\sim}{\varepsilon}} = (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y}$

To estimate $\mathbf{y}$, we have

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$$

Notes:

1. $\mathbf{X}'\hat{\varepsilon} = \mathbf{0}$ $\quad (\mathbf{X}'\mathbf{H} = \mathbf{X}', \mathbf{H}\mathbf{X} = \mathbf{X}$ and $\mathbf{X}'(\mathbf{I} - \mathbf{H}) = \mathbf{0}, (\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0})$
2. $\mathbf{H}$ is symmetric idempotent
3. $\hat{\mathbf{y}}'\hat{\varepsilon} = 0$
4. $\mathbf{I} - \mathbf{H}$ is symmetric idempotent
5. $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ $\quad$ (unbiased estimator)
6. $Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$
7. $tr(\mathbf{I} - \mathbf{H}) = n - r$
8. $\hat{\varepsilon}'\hat{\varepsilon} = tr(\mathbf{y}\mathbf{y}'(\mathbf{I} - \mathbf{H}))$
9. $E(\mathbf{y}\mathbf{y}') = \sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'$
10. $E\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-r}\right) = \sigma^2$

Remark 7.2

Thus, the unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - r}.$$

Remark 7.2:

$\sigma^2 = E[(y_i - E(y_i))^2] = E[(y_i - \underset{\sim}{x}_i^T \underset{\sim}{\beta})^2]$

$(r = k+1)$

$S^{*2} = \dfrac{\sum_{i=1}^{n}(y_i - \underset{\sim}{x}_i'\hat{\underset{\sim}{\beta}})^2}{n} \approx S^2 = \dfrac{\sum_{i=1}^{n}(y_i - \underset{\sim}{x}_i^T\hat{\underset{\sim}{\beta}})^2}{n-r}$ note that $\sum_{i=1}^{n}(y_i - x_i^T\hat{\underset{\sim}{\beta}})^2 =$

$SSE = \underset{\sim}{y}^T(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y} \quad \underset{\sim}{y} \sim (\underset{\sim}{X}\underset{\sim}{\beta}, \sigma^2\underset{\sim}{I})$

$\boxed{E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y}) = tr(\underset{\sim}{A}\underset{\sim}{\Sigma}) + \underset{\sim}{\mu}'\underset{\sim}{A}\underset{\sim}{\mu}}$

$\Rightarrow E(\underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y}) = tr((\underset{\sim}{I} - \underset{\sim}{H})\sigma^2\underset{\sim}{I}) + (\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{I} - \underset{\sim}{H})(\underset{\sim}{X}\underset{\sim}{\beta})$

$= \sigma^2(n-r) = \sigma^2(n-k-1)$

$\Rightarrow E(S^2) = \sigma^2$

Example of weighted LSE:

$y_i = \underset{\sim}{x}_i'\underset{\sim}{\beta} + \varepsilon_i, \quad \varepsilon_i \sim (0, \sigma^2/m_i)$

$\Rightarrow \underset{\sim}{V}^{-1} = \underset{\sim}{W} = \begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_n \end{pmatrix}$

Weighted LS $= \sum_{i=1}^{n} m_i(y_i - \underset{\sim}{x}_i\underset{\sim}{\beta})^2$

7.2.1 Generalized least squares estimation

Assume that $Cov(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, $\mathbf{V}$ known

$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

$\underset{\sim}{\varepsilon} \sim (\underset{\sim}{0}, \sigma^2\underset{\sim}{V})$

- Estimation of $\boldsymbol{\beta}$
  - Ordinary Least Squares Estimator:
    minimize $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
  - Generalized Least Squares Estimator: minimize $(\mathbf{y} - \mathbf{X}\mathbf{a})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{a})$
    Let

$$\begin{aligned}
S &= (\mathbf{y} - \mathbf{X}\mathbf{a})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{a}) \\
&= \mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - 2\mathbf{y}'\mathbf{V}^{-1}\mathbf{X}\mathbf{a} + \mathbf{a}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{a} \\
\frac{\partial S}{\partial \mathbf{a}} &= -2\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{a} \;=\; 0 \\
\Rightarrow \quad \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{a} \\
\Rightarrow \quad \cancel{\underset{\sim}{\beta}} = \tilde{\mathbf{a}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}
\end{aligned}$$

Note: If $\mathbf{V} = \sigma^2\mathbf{I}$, OLS = GLS

  - Weighted Least Squares Estimator

subjective function: $= \sum_{i=1}^{n} w_i (y_i - \underset{\sim}{x}_i^T\underset{\sim}{\beta})^2 = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})^T \underset{\sim}{W} (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})$

$\Rightarrow \underset{\sim}{V}^{-1} = \underset{\sim}{W} = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{pmatrix}$

Remark 7.3

Remark 7.3

$E(\cancel{\underset{\sim}{\beta}}\,\tilde{a}) = E[(\underset{\sim}{X}^T\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{V}^{-1}\underset{\sim}{y}] = (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{V}^{-1}E(\underset{\sim}{y}) = (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{\beta}$ unbiased

$Var(\cancel{\underset{\sim}{\beta}}\,\tilde{a}) = (\underset{\sim}{X}^T\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{V}^{-1} \cdot \cancel{\sigma^2 I}\,\sigma^2\underset{\sim}{V}\; \underset{\sim}{V}'^{-1}\underset{\sim}{X}\,(\underset{\sim}{X}'\underset{\sim}{V}'^{-1}\underset{\sim}{X})^{-1}$

$\hat{\sigma}^2 = \frac{SSE}{n-k-1}\,(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})^T\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})$

Ex of weighted LSE:

$y_i = \underset{\sim}{x}_i^T\underset{\sim}{\beta} + \varepsilon_i, \; \varepsilon_i \sim (0, \frac{\sigma^2}{m_i}) \quad \Rightarrow \underset{\sim}{V}^{-1} = \underset{\sim}{W} = \begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_n \end{pmatrix}$ weighted LS: $\sum_{i=1}^{n} m_i (y_i - \underset{\sim}{x}_i\underset{\sim}{\beta})^2$

7.2.2 Properties of LSE

Gauss-Markov Theorem: The best linear unbiased estimator (b.l.u.e.)

Let $\mathbf{t}$ be a vector and we need to construct the b.l.u.e. of $\mathbf{t}'\boldsymbol{\beta}$.

Q: Find λ s.t. $\begin{cases} E(\lambda' y) = t'\beta \\ Var(\lambda' y) \text{ min} \end{cases}$

1. Let $\boldsymbol{\lambda}'\mathbf{y}$ be a linear function of the observations and an estimator of $\mathbf{t}'\boldsymbol{\beta}$.
2. If $\boldsymbol{\lambda}'\mathbf{y}$ is an unbiased estimator of $\mathbf{t}'\boldsymbol{\beta}$, $E(\boldsymbol{\lambda}'\mathbf{y}) = \mathbf{t}'\boldsymbol{\beta}$.

   But $E(\boldsymbol{\lambda}'\mathbf{y}) = \boldsymbol{\lambda}'E(\mathbf{y}) = \boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$

   Hence, $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta} = \mathbf{t}'\boldsymbol{\beta}$ which is true for all $\boldsymbol{\beta}$

   $\Rightarrow$ $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{t}'$ constraint $\Leftrightarrow$ $X'\lambda = t$
3. Find the linear unbiased estimator of $\mathbf{t}'\boldsymbol{\beta}$ which has minimum variance.

$$\text{Var}(\boldsymbol{\lambda}'\mathbf{y}) = (\boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda})\sigma^2$$

$$Var(\lambda' y) = E\left((\lambda' y - E(\lambda' y))(\lambda' y - E(\lambda' y))'\right)$$
$$= \lambda' E\left[(y - E(y))(y - E(y))'\right]\lambda$$
$$= \lambda' Cov(y)\lambda$$

Theorem: Remark 7.4

$W = \boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda}$ is minimum if

$$\boldsymbol{\lambda}' = \mathbf{t}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$$

subject to the constraint that
约束条件.

$$\mathbf{X}'\boldsymbol{\lambda} = \mathbf{t}$$

$y = X\beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2 V)$ V — known.

$\Rightarrow$ generalized LSE: $\hat{\alpha} = \underbrace{(X'VX)^{-1}X'V^{-1}}_{H} y = Hy$ — linear combination of y

$\begin{cases} E(t'\hat{\alpha}) = t'\beta & \text{— unbiased estimate of } t'\beta \\ Var(t'\hat{\alpha}) = Var(t'Hy) = \sigma^2 (t'H)V(t'H)' \end{cases}$

Gauss – Markov Thm: To prove $Var(t'\hat{\alpha}) = \min$ in $\{\lambda'y : E(\lambda'y) = t'\beta\} = \{\lambda'y, X'\lambda = t\}$ (BLUE)

$E(\lambda'y) = \lambda'X\beta = t'\beta$ for any $\beta \Leftrightarrow \lambda'X = t'$

$Var(\lambda'y) = \sigma^2(\lambda'V\lambda)$

$\lambda'V\lambda = (\lambda - H't + H't)'V(\lambda - H't + H't)$

$= (\lambda - H't)'V(\lambda - H't) + \underbrace{(H't)'V(H't)}_{Var(t'\hat{\alpha})/\sigma^2} + \underbrace{2(H't)'V(\lambda - H't)}_{=0\ ?}$

$\boxed{\begin{aligned} \text{note: } t'HV\lambda - t'HVH't &= t'(X'V^{-1}X)^{-1}X'V^{-1}V\lambda - t'(X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1}t \\ &= t'(X'V^{-1}X)^{-1}t - t'(X'V^{-1}X)^{-1}t = 0 \end{aligned}}$

Thus, $\lambda'V\lambda = (\lambda - H't)'V(\lambda - H't) + (H't)'V(H't) \geq (H't)'V(H't) = \dfrac{Var(t'\hat{\alpha})}{\sigma^2}$

$Var(\lambda'y) \geq Var(t'\hat{\alpha})$

$Var(\lambda'y) = Var(t'\hat{\alpha}) \Leftrightarrow \lambda = H't$

Remark 7.4.1

For the model $y = X\beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2 V)$

$E(\hat{\beta}) = E\left((X'X)^{-1}X'Y\right) = (X'X)^{-1}X'X\beta = \beta$ unbiased.

— but $\hat{\beta}$ is not the BLUE if $V \neq I$.

BLUE in $\hat{\alpha} = (X'V^{-1}X)^{-1}X'V^{-1}y$ need to consider the model

— $\hat{\beta}$ is the BLUE if $V = I$

$\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma^2)$

$\hat{\beta}_j \sim N(\beta_j, w_{jj}\sigma^2)$ w_{jj} – jth diagonal element of $(X'X)^{-1}$

$\Rightarrow \dfrac{\hat{\beta}_j}{(w_{jj}\sigma^2)^{\frac{1}{2}}} \sim N(0,1)$ $\Rightarrow \dfrac{\hat{\beta}_j}{(w_{jj}\hat{\sigma}^2)^{\frac{1}{2}}} \sim t_{n-k-1}$

?

7.3 Maximum likelihood estimation

Model: $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

A normal model is defined by (7.1) with an additional assumption

$$\varepsilon_{n\times 1} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \Rightarrow \underset{\sim}{y} \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \sigma^2 \underset{\sim}{I})$$

where σ^2 is unknown.

MLE for $\boldsymbol{\beta}$ and σ^2

The likelihood function is

where $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$$P(\underset{\sim}{y}) = \underline{L(\boldsymbol{\beta}, \sigma^2)} = (\frac{1}{2\pi\sigma^2})^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} = \prod_{i=1}^{n} P(y_i)$$

Take log, we have

$$\begin{aligned}
logL(\boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\
&= -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial logL}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\
\frac{\partial logL}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})
\end{aligned}$$

Put the above 2 equations to zero and we obtain the MLE of $\boldsymbol{\beta}$ and σ^2

$$\text{MLE}(\boldsymbol{\beta}) = \tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = LSE(\underset{\sim}{\beta}) = \hat{\underset{\sim}{\beta}}$$

$$\begin{aligned}
\text{MLE}(\sigma^2) &= \tilde{\sigma}^2 \\
&= \underline{\frac{1}{n}(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}})} \\
&= \frac{1}{n}(\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'(\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) \\
&= \frac{1}{n}\mathbf{y}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\
&= \frac{1}{n}\mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} \\
&= \frac{1}{n}[\mathbf{y}'\mathbf{y} - \tilde{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}]
\end{aligned}$$

unbiased estimation of σ^2.

$\hat{\sigma}^2 = \frac{1}{n-k-1}[\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}}]'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})$, $E(\hat{\sigma}^2) = \sigma^2$.

MLE: $E(\tilde{\sigma}^2) = E\left(\frac{n-k-1}{n}\hat{\sigma}^2\right) = \frac{n-k-1}{n}\sigma^2 \to \sigma^2$ as $n \to \infty$

asymptotic unbiased estimation of σ^2.

$$\underset{\sim}{y} \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \sigma^2\underset{\sim}{I})$$

Remarks:

1. Distribution of $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

$$\hat{\boldsymbol{\beta}} \sim N(E(\hat{\boldsymbol{\beta}}), Cov(\hat{\boldsymbol{\beta}}))$$

where

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \\ \text{Cov}\ \cancel{Var}(\hat{\boldsymbol{\beta}}) &= \text{Cov}\ \cancel{Var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \end{aligned}$$

$$\underset{\sim}{y} - \hat{\underset{\sim}{y}} = \underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta}$$
$$= \underset{\sim}{y} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$$
$$= (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y}$$

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2)$$

2. Let $SSE = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ —— $= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (\underset{\sim}{y} - \hat{\underset{\sim}{y}})'(\underset{\sim}{y} - \hat{\underset{\sim}{y}}) = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y}$
   $\tilde{\sigma}^2 = \frac{SSE}{n}$ is the MLE
   $\hat{\sigma}^2 = \frac{SSE}{n - r(\mathbf{X})}$ is an unbiased estimator of σ^2 with $r(\mathbf{X}) = \text{rank } of\ \mathbf{X}$

3. $\hat{\boldsymbol{\beta}}$ and SSE are independent

   $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

   $SSE = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ where $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$
   Since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\sigma}^2\mathbf{I})(\mathbf{I} - \mathbf{H}) = \boldsymbol{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{H}) = \mathbf{0}$
   Hence, $\hat{\boldsymbol{\beta}}$ and SSE are independent and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent

Theorem 6.2 $\quad \hat{\beta} \sim N(\underset{\sim}{\beta}, (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sigma^2)$

$\hat{\beta}_j \sim N(\beta_j, w_{jj}\sigma^2)$ $\qquad w_{jj}$ — jth diagonal element of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$

$$\frac{\hat{\beta}_j}{(w_{jj}\sigma^2)^{\frac{1}{2}}} \sim N(0,1)$$

$$\frac{\hat{\beta}_j}{(w_{jj}\hat{\sigma}^2)^{\frac{1}{2}}} \sim t_{n-k-1}$$

4. Distribution of $\hat{\sigma}^2$ Thm 6.1

Consider $\frac{SSE}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} \qquad (\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\boldsymbol{\sigma}^2))$

Since

$$\left(\frac{\mathbf{I}-\mathbf{H}}{\sigma^2}\right)(\mathbf{I}\sigma^2) = \mathbf{I}-\mathbf{H}$$

which is idempotent. Therefore,

$$\frac{SSE}{\sigma^2} \sim \chi^2_{\left(r\left(\frac{\mathbf{I}-\mathbf{H}}{\sigma^2}\right), \frac{1}{2}(\mathbf{X}\boldsymbol{\beta})'\left(\frac{\mathbf{I}-\mathbf{H}}{\sigma^2}\right)(\mathbf{X}\boldsymbol{\beta})\right)}$$

However,

$$(\mathbf{X}\boldsymbol{\beta})'\left(\frac{\mathbf{I}-\mathbf{H}}{\sigma^2}\right)(\mathbf{X}\boldsymbol{\beta}) = 0$$

So, noncentrality parameter $= 0$

$\mathrm{tr}\,\underset{\sim}{I} - \mathrm{tr}(\underset{\sim}{H}) = n - r(\underset{\sim}{X})$
$= n-k-1$
$\shortparallel$
$\mathrm{tr}(\underset{\sim}{I} - \underset{\sim}{H})$
$\shortparallel$

In addition,

$$r\left(\frac{\mathbf{I}-\mathbf{H}}{\sigma^2}\right) = r(\mathbf{I}-\mathbf{H}) = n - r(\mathbf{X})$$

Therefore,

$$\frac{SSE}{\sigma^2} \sim \chi^2_{(n-r(\mathbf{X}))}$$

$$\frac{[n-r(\mathbf{X})]\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-r(\mathbf{X}))}$$

Example 7.1:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n; \quad \varepsilon_i \overset{ind}{\sim} N(0,\ \sigma^2)$$

In matrix notation, $\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ $\quad \underset{\sim}{X} = \begin{pmatrix} 1 & x_{i1} & \cdots & x_{ik} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

$(\underset{\sim}{H}'\underset{\sim}{t})'\underset{\sim}{y}\,(\underset{\sim}{H}'$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1} &= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \\
&= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i \\ \\ n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \end{bmatrix} \\
&= \begin{bmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}) \\
&= \frac{1}{n-2}[\sum_{i=1}^n y_i^2 - \hat{\beta}_0 \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i y_i] \\
&= \frac{1}{n-2}[\sum_{i=1}^n y_i^2 - \bar{y}\sum_{i=1}^n y_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i y_i] \\
&= \frac{1}{n-2}[\sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\beta}_1 (\sum_{i=1}^n x_i y_i - n\sum_{i=1}^n x_i \sum_{i=1}^n y_i)] \\
&= \frac{1}{n-2}[\sum_{i=1}^n (y_i - \bar{y})^2 - \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]
\end{aligned}$$

Review:

CH 7. model. $y, x_1, x_2, \cdots, x_k$

$$y = f(x_1, \cdots, x_k) + \varepsilon = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

$$D = \{y_i, x_{i1}, x_{i2}, \cdots, x_{ik}, \ i = 1, \cdots, n\} \rightarrow \text{Inference}.$$

For the data: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \ i = 1, \cdots, n$

or in matrix form: $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ where $\underset{\sim}{\varepsilon} \sim (0, \sigma^2 \underset{\sim}{I}) \Rightarrow$ LSE: $\hat{\beta} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$

Gauss-Markov theorem.

if $\varepsilon \sim (\underset{\sim}{0}, \sigma^2 \underset{\sim}{V})$ $\underset{\sim}{V}$ - given

LSE: $\hat{\beta} = (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y} \Rightarrow$ BLUE

MLE: $\underset{\sim}{y} \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \sigma^2 \underset{\sim}{I}) \Rightarrow \hat{\beta} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$. $= \frac{SSE}{n} \approx \frac{SSE}{n-k-1}$

$\tilde{\sigma}^2$ asymptotic unbiased estimator

$\hat{\sigma}^2 = \frac{SSE}{n-k-1}$ unbiased estimate of σ^2.

7.3.1 The model in centered form Remark 7.5

Recall that the Multiple Linear Regression Model with k independent variables in (7.1) is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i$$

and the model in matrix form is

$$\mathbf{y}_{n\times 1} = \mathbf{X}\boldsymbol{\beta}_{(k+1)\times 1} + \boldsymbol{\varepsilon}_{n\times 1}.$$

The estimation is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\quad var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$

Let $\mathbf{X} = (\mathbf{1} \quad \mathbf{X}_1)$ $\qquad \boldsymbol{\beta}' = (\beta_0 \quad \underset{\sim}{\beta}_1')$

where $\mathbf{X}_1 = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$, $\qquad \underset{\sim}{\beta}_1' = (\beta_1 \quad \beta_2 \quad \cdots \quad \beta_k)$

Rewrite $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= \left[\begin{pmatrix} \mathbf{1}' \\ \mathbf{X}_1' \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{X}_1 \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{1}' \\ \mathbf{X}_1' \end{pmatrix} \mathbf{y} \\
&= \begin{bmatrix} n & n\bar{\mathbf{X}}' \\ n\bar{\mathbf{X}} & \mathbf{X}_1'\mathbf{X}_1 \end{bmatrix}^{-1} \begin{bmatrix} n\bar{y} \\ \mathbf{X}_1'\mathbf{y} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n} + \bar{\mathbf{X}}'\mathbf{B}^{-1}\bar{\mathbf{X}} & -\bar{\mathbf{X}}'\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\bar{\mathbf{X}} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \mathbf{X}_1'\mathbf{y} \end{bmatrix}
\end{aligned}$$

where $\mathbf{B} = \mathbf{X}_1'\mathbf{X}_1 - n\bar{\mathbf{X}}\bar{\mathbf{X}}' = \mathbf{Z}'\mathbf{Z}$.

$$\Rightarrow \qquad \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \bar{y} - \bar{\mathbf{X}}'\mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{X}}) \\ \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{X}}) \end{bmatrix}$$

Remark 7.5

the model in centered form $\quad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad \cdots (*)$

$\Rightarrow \hat\beta_0, \hat\beta_1, \cdots, \hat\beta_k$

Centered form: $\quad y_i = \alpha + b_1(x_{i1} - \bar{x}_1) + b_2(x_{i2} - \bar{x}_2) + \cdots + b_k(x_{ik} - \bar{x}_k) + \varepsilon_i \quad \cdots (*)(*)$

or $y_i = \alpha + b_1 z_{i1} + \cdots + b_k z_{ik} + \varepsilon_i^*$ $\qquad \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \; j = 1, \cdots, k$

$\Rightarrow \hat\alpha, \hat b_1, \hat b_2, \cdots, \hat b_k$ $\qquad \bar{z}_j = \frac{1}{n}\sum_{i=1}^{n} z_{ij} = 0$

Comparing $(*)$ and $(*)(*)$

$$y_i = \alpha - b_1\bar{x}_1 - \cdots - b_k\bar{x}_k + b_1 x_{i1} + \cdots + b_k x_{ik} + \varepsilon_i^*$$

$$\boxed{\text{We want to know whether } \begin{cases} \hat b_k = \hat\beta_k \\ \hat\alpha = \hat\beta_0 + \hat b_1 \bar{x}_1 + \cdots + \hat b_k \bar{x}_k \end{cases} \quad ?}$$

Proof: From $(*)$, $\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} = [\,\mathbf{1} \;\; \mathbf{X}_1\,]_{n \times (k+1)}$, where $\mathbf{X}_1$ is the block $\begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}$

$$\hat{\boldsymbol\beta} = \begin{pmatrix} \hat\beta_0 \\ \hat{\boldsymbol\beta}_1 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left[\begin{pmatrix} \mathbf{1}' \\ \mathbf{X}_1' \end{pmatrix}(\mathbf{1}\;\mathbf{X}_1)\right]^{-1}\begin{pmatrix} \mathbf{1}' \\ \mathbf{X}_1' \end{pmatrix}\mathbf{y} = \begin{bmatrix} n & n\bar{\mathbf{x}}' \\ n\bar{\mathbf{x}} & \mathbf{X}_1'\mathbf{X} \end{bmatrix}^{-1}\begin{pmatrix} n\bar{y} \\ \mathbf{X}_1'\mathbf{y} \end{pmatrix}$$

$$\hat{\boldsymbol\beta}_1 = \begin{pmatrix} \hat\beta_1 \\ \vdots \\ \hat\beta_k \end{pmatrix} \qquad \bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_k \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \qquad \mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}^{-1} \end{pmatrix}$$

where $\mathbf{B} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{X}_1'\mathbf{X}_1 - n\bar{\mathbf{x}}\bar{\mathbf{x}}'$

$$\Rightarrow \begin{pmatrix} n & n\bar{\mathbf{x}}' \\ n\bar{\mathbf{x}} & \mathbf{X}_1'\mathbf{X}_1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{n} + \bar{\mathbf{x}}'\mathbf{B}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}'\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\bar{\mathbf{x}} & \mathbf{B}^{-1} \end{pmatrix}$$

$$\begin{pmatrix} \hat\beta_0 \\ \hat{\boldsymbol\beta}_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} + \bar{\mathbf{x}}'\mathbf{B}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}'\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\bar{\mathbf{x}} & \mathbf{B}^{-1} \end{pmatrix}\begin{pmatrix} n\bar{y} \\ \mathbf{X}_1'\mathbf{y} \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{\mathbf{x}}'\mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) \\ \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol\beta}_1 \\ \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) \end{pmatrix} \quad \cdots (***)$$

Proof: From $(*)(*)$ $\quad \mathbf{y} = (\mathbf{1}'\,\mathbf{Z})\begin{pmatrix} \alpha \\ \mathbf{b} \end{pmatrix} + \boldsymbol\varepsilon \qquad \mathbf{Z} = (\mathbf{X}_1 - \mathbf{1}\bar{\mathbf{x}}'),\; \bar{\mathbf{z}} = \begin{pmatrix} \bar{z}_1 \\ \vdots \\ \bar{z}_k \end{pmatrix} = \mathbf{0}$

Similar to $(***)$, we have

$$\begin{pmatrix} \hat\alpha \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \mathbf{B}^{-1}(\mathbf{Z}\mathbf{y}) \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \mathbf{B}^{-1}(\mathbf{X}_1 - \mathbf{1}\bar{\mathbf{x}})\mathbf{y} \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) \end{pmatrix}$$

$\Rightarrow$ thus we have $\begin{cases} \hat\beta_0 = \bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol\beta}_1 \\ \hat\alpha = \bar{y}. \\ \hat{\mathbf{b}} = \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) \\ \hat{\boldsymbol\beta}_1 = \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) \end{cases}$ $\quad \Rightarrow \begin{cases} \hat\beta_0 = \hat\alpha - \bar{\mathbf{x}}'\hat{\boldsymbol\beta}_1 = \hat\alpha - (\bar{x}_1\hat\beta_1 + \bar{x}_2\hat\beta_2 + \cdots + \bar{x}_n\hat\beta_n) \\ \hat{\mathbf{b}} = \mathbf{B}^{-1}(\mathbf{X}_1'\mathbf{y} - n\bar{y}\bar{\mathbf{x}}) = \hat{\boldsymbol\beta}_1 \end{cases}$ $\square$

The centered form of the multiple linear regression model is

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \\ &= \alpha + b_1(x_{i1} - \bar{x}_1) + b_2(x_{i2} - \bar{x}_2) + \cdots + b_k(x_{ik} - \bar{x}_k) + \epsilon_i \end{aligned} \quad (7.2)$$

where $\alpha = \beta_0 + \beta_1\bar{x}_1 + \beta_2\bar{x}_2 + \ldots + \beta_k\bar{x}_k$, and $\bar{x}_j$ is the average of $\{x_{ij}, i = 1, \ldots, n\}$ for $j = 1, 2, \ldots, k$. The matrix form of (7.2) can be expressed as

$$\mathbf{y}_{n\times 1} = \begin{pmatrix} \mathbf{1} & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \alpha \\ \mathbf{b} \end{pmatrix} + \boldsymbol{\varepsilon}_{n\times 1},$$

where $(\mathbf{Z} = \mathbf{X}_1 - \mathbf{1}\bar{\mathbf{X}}')$, $\bar{\mathbf{X}}' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)$ and $\mathbf{b}' = (b_1, \ldots, b_k)$. We can prove that

$$\hat{\alpha} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \ldots + \hat{\beta}_k\bar{x}_k$$

and $\hat{\mathbf{b}} = \hat{\boldsymbol{\beta}}_1$.

Similarly

$$\begin{aligned} var(\hat{\boldsymbol{\beta}}) &= var\begin{pmatrix} \hat{\beta}_0 \\ \hat{\mathbf{b}} \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \\ &= \begin{bmatrix} \frac{1}{n} + \bar{\mathbf{X}}'(\mathbf{Z}'\mathbf{Z})^{-1}\bar{\mathbf{X}} & -\bar{\mathbf{X}}'(\mathbf{Z}'\mathbf{Z})^{-1} \\ -(\mathbf{Z}'\mathbf{Z})^{-1}\bar{\mathbf{X}} & (\mathbf{Z}'\mathbf{Z})^{-1} \end{bmatrix}\sigma^2 \end{aligned}$$

$$\boxed{\begin{aligned} \Rightarrow \quad & var(\hat{\mathbf{b}}) \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2 \end{aligned}}$$

$$\begin{aligned} & var(\hat{\beta}_0) \\ &= \frac{\sigma^2}{n} + \bar{\mathbf{X}}'(\mathbf{Z}'\mathbf{Z})^{-1}\bar{\mathbf{X}}\sigma^2 \\ &= \frac{\sigma^2}{n} + \bar{\mathbf{X}}'\ var(\hat{\mathbf{b}})\ \bar{\mathbf{X}} \end{aligned}$$

$$
\begin{aligned}
& cov(\hat{\beta}_0,\ \hat{\mathbf{b}}') \\
= \quad & -\bar{\mathbf{X}}'(\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2 \\
= \quad & -\bar{\mathbf{X}}'\ var(\hat{\mathbf{b}})
\end{aligned}
$$

$\sum_{i=1}^{n}(y_i-\bar{y})^2 \quad \bar{y}=\frac{1}{n}\mathbf{1}'\mathbf{y}$

$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \underbrace{\sigma^2\mathbf{I}}_{V})$

Need to prove $(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}')(\sigma^2\mathbf{I})$

$= (\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}')$ is idempotent

7.4 Partitioning Total Sum of Squares

- SST (Total sum of squares corrected for the mean)

$$\text{SST} = \underline{\mathbf{y}'\mathbf{y} - \tfrac{1}{n}\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y}}$$
$$= \mathbf{y}'\underbrace{(\mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}')}_{A}\mathbf{y}$$

$$\frac{SST}{\sigma^2} \sim \chi^2_{(n-1,\ \frac{\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}-\frac{1}{n}(\mathbf{1}'\mathbf{X}\boldsymbol{\beta})^2}{2\sigma^2})}$$

- SSR (Sum of squares of regression) = SST - SSE

$$\text{SSR} = \mathbf{y}'(\mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{y} - \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$$
$$= \mathbf{y}'(\mathbf{H} - \tfrac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{y}$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 = (\hat{\mathbf{y}}-\bar{\mathbf{y}})'(\hat{\mathbf{y}}-\bar{\mathbf{y}}) = \hat{\mathbf{b}}'(\mathbf{Z}'\mathbf{Z})\hat{\mathbf{b}}$$

$\hat{a}=\bar{y}$

$$SSR = \hat{\mathbf{b}}'\mathbf{Z}'\mathbf{y} = \hat{\mathbf{b}}'(\mathbf{Z}'\mathbf{Z})\hat{\mathbf{b}} \quad (\text{Since } \hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}, (\mathbf{Z}'\mathbf{Z})\hat{\mathbf{b}} = \mathbf{Z}'\mathbf{y})$$

$$\text{Since } \hat{\mathbf{b}} \sim N[\,\mathbf{b},\ (\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2]$$
$$\Rightarrow \frac{SSR}{\sigma^2} \sim \chi^2_{(k,\ \frac{\mathbf{b}'(\mathbf{Z}'\mathbf{Z})\mathbf{b}}{2\sigma^2})}$$

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

The model in centered form: $\hat{y}_i = \hat{a} + \hat{b}_1 z_{i1} + \cdots + \hat{b}_k z_{ik}$

$\hat{y}_i - \bar{y} = \hat{b}_1 z_{i1} + \cdots + \hat{b}_k z_{ik} = \mathbf{z}\hat{\mathbf{b}}$

Anova Table

Source	df	SS	MS	F-statistics
Regression	K $\underline{r(\mathbf{X})-1}$	$\hat{\mathbf{b}}'\mathbf{Z}'\mathbf{y}$	$\frac{SSR}{r(\mathbf{X})-1}$	$F=\frac{MSR}{MSE}$
Error	n-K-1 $\underline{n-r(\mathbf{X})}$	$\mathbf{y}'\mathbf{y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	$\frac{SSE}{n-r(\mathbf{X})}$	
Total	$n-1$	$\mathbf{y}'(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{y}$		

Note: Since $\quad \text{SSR} = \hat{\mathbf{b}}'\mathbf{Z}'\mathbf{y} = \mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$

and $\quad \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I}\sigma^2)(\mathbf{I}-\mathbf{H}) = 0$

$\Rightarrow$ SSR is independent of SSE

$\Rightarrow \quad F \sim F_{\left(r(\mathbf{X})-1,\ n-r(\mathbf{X}),\ \frac{\mathbf{b}'(\mathbf{Z}'\mathbf{Z})\mathbf{b}}{2\sigma^2}\right)}$

Under H_0 : $\mathbf{b} = \mathbf{0}$

$$F \sim F_{(r(\mathbf{X})-1,\ n-r(\mathbf{X}),\ 0)}$$

Hypothesis test:

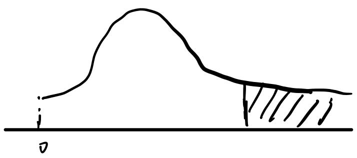
H_0: $\underset{\sim}{b} = 0$ or $\underset{\sim}{\beta_1} = 0$

or $\beta_1 = \cdots = \beta_k = 0$

H_1: at least one of $\beta_j \neq 0$. $j = 1, \cdots, k$.

$F = \frac{MSR}{MSE} \overset{H_0}{\sim} F_{k, n-k-1}$

We reject H_0 if $F \geq F_{\alpha, k, n-k-1}$.



or consider p-value $= \mathbb{P}(F_{k,n-k-1} \geq F)$

7.5 Model Misspecification

7.5.1 Misspecification of the error structure — Remark 7.6

Suppose the true model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{V},$$

but the working model is

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}.$$

This will still have an unbiased estimate of $\boldsymbol{\beta}$, but it is not the BLUE.

7.5.2 Misspecification of the mean — Remark 7.7

We consider the following two models

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \quad \text{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}, \tag{7.3}$$

and

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \quad \text{Cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}. \tag{7.4}$$

- Under-fitting: if the true model is (7.3), but use the model (7.4);
- Over-fitting: if the true model is (7.4), but use the model (7.3).

remark 7.8

§ 7.5

Data $D = \{y_i, \underset{\sim}{x}_i, i=1,\cdots,n\}$

true model

working model.

Remark 7.6

7.5.1 true model: $Y = X\beta + \varepsilon$ $Cov(\varepsilon) = \sigma^2 V$ V is given. (*)

The BLUE is (Gauss-Markov theorem)

$$\hat\beta = (X'V^{-1}X)^{-1}X'V^{-1}y$$

$$E(\hat\beta) = \beta, \quad Var(\hat\beta) = \sigma^2(X'V^{-1}X)$$

Working model: $y = X\beta + \varepsilon$, $Cov(\varepsilon) = \sigma^2 I$ (**)

LSE: $\hat\beta^* = (X'X)^{-1}X'y$. $y_i \sim$ (*)

$$E(\hat\beta^*) = (X'X)^{-1}X'E(y) = \beta$$

$$Var(\hat\beta^*) = (X'X)^{-1}X' \cdot Var(y) X (X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1}X'VX(X'X)^{-1} \geq \sigma^2(X'V^{-1}X)$$ G-M Thm.

使用$\sigma^2 I$仍然为一个无偏估计. $\Rightarrow$ uncertainity增加.

Remark 7.7

7.5.2 Under-fit:

example: height ~ father, mother. X_1 first consider grandfather, grandmother X_2

True model: $y = (X_1\ X_2)\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} + \varepsilon = X\beta + \varepsilon$ (7.3)

$X = (X_1\ X_2)$ $\beta = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$. $Cov(\varepsilon) = \sigma^2 I$.

Working model: $y = X_1\beta_1^* + \varepsilon$ (7.4)

Note that $\hat\beta_1^* = (X_1'X_1)^{-1}X_1'y$. $\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} = \hat\beta = (X'X)^{-1}X'y$

$dim(\beta_1) = P+1$
$dim(\beta) = K+1$

Conclusion 1,2,3,4

$\Rightarrow$ (1) $E(\hat\beta_1^*) = \beta_1 + A\beta_2$. $A = (X_1'X_1)^{-1}X_1'X_2$

Note that $E(\hat\beta_1^*) = \beta_1$, only when $X_1'X_2 = 0$.

$\Rightarrow$ (2). $Cov(\hat\beta_1^*) = \sigma^2(X_1'X_1)^{-1}$

$\Rightarrow$ (3). $S^2 = \dfrac{(y - X\hat\beta)'(y - X\hat\beta)}{n-k-1}$, $S_1^2 = \dfrac{(y - X_1\hat\beta_1^*)'(y - X_1\hat\beta_1^*)}{n-p-1}$

$$E(S_1^2) = \sigma^2 + \frac{\beta_2' X_2'(I - H_1)X_2\beta_2}{n-p-1}, \quad H_1 = X_1(X_1'X_1)^{-1}X_1'$$

$\Rightarrow$ (4). Prediction: y^* at x^*. $\hat y_1^* = x_1^{*}\hat\beta_1^*$ $x^* = (x_1^*\ x_2^*)$

true model: $y^* = x_1^*\beta_1 + x_2^*\beta_2 + \varepsilon^*$

$$E(\hat y_1^*) = x_1^*(\beta_1 + A\beta_2) = \underbrace{x_1^*\beta_1 + x_2^*\beta_2}_{\text{true } E(y^*)} - \underbrace{(x_2^* - x_1^*A)\beta_2}_{\text{bias}}$$

$\neq E(y^*)$ if $\beta_2 \neq 0$. $\hat y_1^*$ is biased prediction of y^*

Proof: (1) $E(\hat{\beta}_1^*) = E[(X_1'X_1)^{-1}X_1'y]$

$= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2)$

$= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2$

Note: true model

$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

$E(y) = X_1\beta_1 + X_2\beta_2$

(2) $SSE_1 = (y - X_1\hat{\beta}_1^*)'(y - X_1\hat{\beta}_1^*) = y'(I-H_1)y$

$\begin{cases} Ey = X\beta \\ Cov(y) = \sigma^2 I \end{cases}$

$E(SSE_1) = tr[(I-H_1)\sigma^2 I] + (X_1\beta_1 + X_2\beta_2)'[I-H_1](X_1\beta_1 + X_2\beta_2)$

$= \sigma^2(n-p-1) + \beta_2'X_2'(I-H_1)X_2\beta_2$

7.5.2 Remark 7.8 Over-fitting

Working model: $y = (X_1\ X_2)\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} + \varepsilon = X\beta + \varepsilon$

True model: $y = X_1\beta_1^* + \varepsilon$

Note: $Ey = X_1\beta_1 = X_1\beta_1 + X_2 0 = X\begin{pmatrix}\beta_1 \\ 0\end{pmatrix}$

$\hat{\beta} = \begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix} = (X'X)^{-1}X'y$

$(X'X) = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}$ ($C = X_1'X_2$, $C' = X_2'X_1$) $\Rightarrow (X'X)^{-1} = \begin{pmatrix} (X_1'X_1)^{-1} + (X_1'X_1)^{-1}CBC'(X_1'X_1)^{-1} & * \\ -BC'(X_1'X_1)^{-1} & B \end{pmatrix}$

where $B = (X_2'X_2 - C'(X_1'X_1)^{-1}C)^{-1}$

$E\hat{\beta} = \begin{pmatrix}E\hat{\beta}_1 \\ E\hat{\beta}_2\end{pmatrix} = (X'X)^{-1}X'(Ey) = (X'X)^{-1}X'X\begin{pmatrix}\beta_1 \\ 0\end{pmatrix} = \begin{pmatrix}\beta_1 \\ 0\end{pmatrix}$ 本身那一部分，仍为 unbiased.

$\Rightarrow E\hat{\beta}_1 = \beta_1,\ E\hat{\beta}_2 = 0.$

Note: $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$

$\Rightarrow Cov(\hat{\beta}_1) = \sigma^2[(X_1'X_1)^{-1} + (X_1'X_1)^{-1}CBC'(X_1'X_1)^{-1}]$ B - positive definite.

$\geq \sigma^2(X_1'X_1)^{-1} = Cov(\hat{\beta}_1^*)$ Estimate from (7.4)