# 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: $\{ y_i, \underset{\sim}{x}_i', i=1 \cdots n \}$

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

where $r = k + 1$.

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

$$\underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \underset{\sim}{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \qquad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$(1, x_{i1}, \cdots, x_{ik}) = \underset{\sim}{x}_i'$$

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

## <u>To construct an unbiased estimator of $\sigma^2$ based on $\hat{\boldsymbol{\beta}}$</u>

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

Hence

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

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

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

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

<u>Notes</u>:

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

$E(\hat\beta) = E[(X'X)^{-1}X'Y]$
$= (X'X)^{-1}X'X\beta = \beta$

$cov(\hat\beta) = [(X'X)^{-1}X'][\sigma^2 I]X(X'X)^{-1}$
$= \sigma^2(X'X)^{-1}$

$(\hat\varepsilon'\hat\varepsilon) = tr(\hat\varepsilon'\hat\varepsilon) = tr(Y'(I-H)Y)$
$= tr(YY'(I-H))$

Thus, the unbiased estimator of $\sigma^2$ is

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

<u>Remark 7.2</u>

### 7.2.1 Generalized least squares estimation

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

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

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

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

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

$\tilde{\beta}$

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

  - Weighted Least Squares Estimator

subjective function $= \sum_{i=1}^{n} w_i (y_i - \underset{\sim}{x}_i' \underset{\sim}{\beta})^2$

$V^{-1} = \underset{\sim}{W} = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{pmatrix} \quad = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})'\underset{\sim}{W}(\underset{\sim}{y} - \underset{\sim}{X}\beta)$

Remark 7.3

### 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 $\mathbf{t}'\boldsymbol{\beta}$.

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 $\boldsymbol{\beta}$

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

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

Remark 7.4

Theorem:

$W = \boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda}$ 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

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

## 7.3 Maximum likelihood estimation

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

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

$$\boldsymbol{\varepsilon}_{n\times 1} \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$$

where $\sigma^2$ is unknown.

MLE for $\boldsymbol{\beta}$ and $\sigma^2$

$p(\underset{\sim}{y})$ $\quad \underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

The likelihood function is

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

Take log, we have

$$\begin{aligned} logL(\boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{n}{2}log2\pi - \frac{n}{2}log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial logL}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}) = 0 \\ \frac{\partial logL}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = 0 \end{aligned}$$

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

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

unbiased estimate of $\sigma^2$

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

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

$E(\hat{\sigma}^2) = \sigma^2$

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

$\longrightarrow \sigma^2$ as $n \to \infty$

asymptotic unbiased estimate of $\sigma^2$

Remarks:

$Y \sim N(X\beta, \sigma^2 I)$

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

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

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

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

2. Let $SSE = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ $= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (Y - \hat{Y})'(Y - \hat{Y})$ $= Y'(I - H)Y$

   $\tilde{\sigma}^2 = \frac{SSE}{n}$ is the MLE

   $\hat{\sigma}^2 = \frac{SSE}{n - r(\mathbf{X})}$ is an unbiased estimator of $\sigma^2$ with $r(\mathbf{X}) =$ 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

Theorem 6.1

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

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,

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

However,

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

So, noncentrality parameter $= 0$

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, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\underset{\sim}{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

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

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

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

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

### 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 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\qquad var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$

Let $\quad \mathbf{X} = (\mathbf{1} \quad \mathbf{X}_1) \qquad \boldsymbol{\beta}' = (\beta_0 \quad \beta_1')$

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

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

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

The centered form of the multiple linear regression model is

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

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

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

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

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

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

### 7.4 Partitioning Total Sum of Squares

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

$\bar{y} = \frac{1}{n}\mathbf{1}'y$

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

$$\text{SST} = \mathbf{y}'\mathbf{y} - \tfrac{1}{n}\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y}$$
$$= \mathbf{y}'\underbrace{(\mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}')}_{A}\mathbf{y}$$
$$\frac{SST}{\sigma^2} \sim \chi^2_{(n-1,\ \frac{\boldsymbol{\beta}'\mathbf{x}'\mathbf{x}\boldsymbol{\beta} - \frac{1}{n}(\mathbf{1}'\mathbf{x}\boldsymbol{\beta})^2}{2\sigma^2})}$$

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

W.T. prove

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

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

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

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

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

Since $\hat{\mathbf{b}} \sim N[\ \mathbf{b},\ (\mathbf{Z}'\mathbf{Z})^{-1}\sigma^2]$

$$\Rightarrow \quad \frac{SSR}{\sigma^2} \sim \chi^2_{(k,\ \frac{\mathbf{b}'(\mathbf{Z}'\mathbf{Z})\mathbf{b}}{2\sigma^2})}$$

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

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

$= \hat{b}'(Z'Z)\hat{b}$

The model in centered form.

$\hat{y}_i = \hat{\alpha} + \hat{b}_1 z_{i1} + \cdots + \hat{b}_k z_{ik}$ $\quad \hat{\alpha} = \bar{y}$

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

$= Z\hat{b}$

$k$ $\qquad n-k-1$

Anova Table

| Source | df | SS | MS | F-statistics |
|---|---|---|---|---|
| Regression | $r(\mathbf{X})-1$ | $\hat{\mathbf{b}}'\mathbf{Z}'\mathbf{y}$ | $\frac{SSR}{r(\mathbf{X})-1}$ | $F = \frac{MSR}{MSE}$ |
| Error | $n-r(\mathbf{X})$ | $\mathbf{y}'\mathbf{y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ | $\frac{SSE}{n-r(\mathbf{X})}$ | |
| 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 $H_0:\ \mathbf{b} = \mathbf{0}$

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

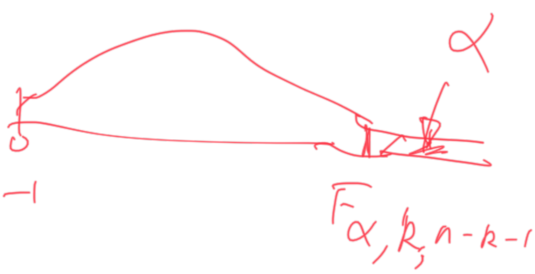
Hypothesis test:

$H_0$: $\underset{\sim}{b} = 0$ or $\underset{\sim}{\beta_1} = 0$ or $\beta_1 = \beta_2 = \cdots \beta_k = 0$

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

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

Reject $H_0$ if $F \geq F_{\alpha, k, n-k-1}$




OR P-value $= P(F_{k,n-k-1} \geq F)$

## 7.5 Model Misspecification

Remark 7.6

### 7.5.1 Misspecification of the error structure

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).