Chapter 7
Multiple Regression II

許湘伶

Applied Linear Regression Models
(Kutner, Nachtsheim, Neter, Li)
Some specialized topics

- extra sums of squares (額外平方和)
  - conducting tests about the regression coefficients
- standardized version of the multiple regression model
- multicollinearity: condition where the predictor variables are highly correlated
Illustration for the extra sum of squares

Example 1

A study of the relation of amount of body fat (體脂肪)

- a sample of 20 healthy females: 25-34 years old
- $Y$: body fat
- $X_1$: triceps skinfold thickness (三頭肌皮層厚度)
- $X_2$: thigh circumference (大腿圍)
- $X_3$: midarm circumference (中臂環圍)

It would be very helpful if a regression model with some or all these predictor variables could provide reliable estimates of amount of body fat.
Illustration for the extra sum of squares (cont.)

Table: Basic Data-Body Fat Example.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Triceps Skinfold Thickness $X_{i1}$</th>
<th>Thigh Circumference $X_{i2}$</th>
<th>Midarm Circumference $X_{i3}$</th>
<th>Body Fat $Y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.50</td>
<td>43.10</td>
<td>29.10</td>
<td>11.90</td>
</tr>
<tr>
<td>2</td>
<td>24.70</td>
<td>49.80</td>
<td>28.20</td>
<td>22.80</td>
</tr>
<tr>
<td>3</td>
<td>30.70</td>
<td>51.90</td>
<td>37.00</td>
<td>18.70</td>
</tr>
<tr>
<td>4</td>
<td>29.80</td>
<td>54.30</td>
<td>31.10</td>
<td>20.10</td>
</tr>
<tr>
<td>5</td>
<td>19.10</td>
<td>42.20</td>
<td>30.90</td>
<td>12.90</td>
</tr>
<tr>
<td>6</td>
<td>25.60</td>
<td>53.90</td>
<td>23.70</td>
<td>21.70</td>
</tr>
<tr>
<td>7</td>
<td>31.40</td>
<td>58.50</td>
<td>27.60</td>
<td>27.10</td>
</tr>
<tr>
<td>8</td>
<td>27.90</td>
<td>52.10</td>
<td>30.60</td>
<td>25.40</td>
</tr>
<tr>
<td>9</td>
<td>22.10</td>
<td>49.90</td>
<td>23.20</td>
<td>21.30</td>
</tr>
<tr>
<td>10</td>
<td>25.50</td>
<td>53.50</td>
<td>24.80</td>
<td>19.30</td>
</tr>
<tr>
<td>11</td>
<td>31.10</td>
<td>56.60</td>
<td>30.00</td>
<td>25.40</td>
</tr>
<tr>
<td>12</td>
<td>30.40</td>
<td>56.70</td>
<td>28.30</td>
<td>27.20</td>
</tr>
<tr>
<td>13</td>
<td>18.70</td>
<td>46.50</td>
<td>23.00</td>
<td>11.70</td>
</tr>
<tr>
<td>14</td>
<td>19.70</td>
<td>44.20</td>
<td>28.60</td>
<td>17.80</td>
</tr>
<tr>
<td>15</td>
<td>14.60</td>
<td>42.70</td>
<td>21.30</td>
<td>12.80</td>
</tr>
<tr>
<td>16</td>
<td>29.50</td>
<td>54.40</td>
<td>30.10</td>
<td>23.90</td>
</tr>
<tr>
<td>17</td>
<td>27.70</td>
<td>55.30</td>
<td>25.70</td>
<td>22.60</td>
</tr>
<tr>
<td>18</td>
<td>30.20</td>
<td>58.60</td>
<td>24.60</td>
<td>25.40</td>
</tr>
<tr>
<td>19</td>
<td>22.70</td>
<td>48.20</td>
<td>27.10</td>
<td>14.80</td>
</tr>
<tr>
<td>20</td>
<td>25.20</td>
<td>51.00</td>
<td>27.50</td>
<td>21.10</td>
</tr>
</tbody>
</table>
### Illustration for the extra sum of squares (cont.)

(a) Regression of $Y$ on $X_1$

(b) Regression of $Y$ on $X_2$

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>352.27</td>
<td>1</td>
<td>352.27</td>
</tr>
<tr>
<td>Error</td>
<td>143.12</td>
<td>18</td>
<td>7.95</td>
</tr>
<tr>
<td>Total</td>
<td>495.39</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

### Table 7.2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Regression Coefficient</th>
<th>Estimated Standard Deviation</th>
<th>$t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$b_1 = .8572$</td>
<td>$s(b_1) = .1288$</td>
<td>6.66</td>
</tr>
</tbody>
</table>

(b) Regression of $Y$ on $X_2$

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>381.97</td>
<td>1</td>
<td>381.97</td>
</tr>
<tr>
<td>Error</td>
<td>113.42</td>
<td>18</td>
<td>6.30</td>
</tr>
<tr>
<td>Total</td>
<td>495.39</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Regression Coefficient</th>
<th>Estimated Standard Deviation</th>
<th>$t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2$</td>
<td>$b_2 = .8565$</td>
<td>$s(b_2) = .1100$</td>
<td>7.79</td>
</tr>
</tbody>
</table>

(continued)

**Figure:** Regression Results for Several Fitted Models—Body Fat Example.
Illustration for the extra sum of squares (cont.)

(c) Regression of $Y$ on $X_1$ and $X_2$

(d) Regression of $Y$ on $X_1$, $X_2$ and $X_3$

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>385.43</td>
<td>2</td>
<td>192.72</td>
</tr>
<tr>
<td>Error</td>
<td>109.95</td>
<td>17</td>
<td>6.47</td>
</tr>
<tr>
<td>Total</td>
<td>495.39</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Regression Coefficient</th>
<th>Estimated Standard Deviation</th>
<th>$t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$b_1 = .2224$</td>
<td>$s(b_1) = .3034$</td>
<td>.73</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$b_2 = .6594$</td>
<td>$s(b_2) = .2912$</td>
<td>2.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>396.98</td>
<td>3</td>
<td>132.33</td>
</tr>
<tr>
<td>Error</td>
<td>98.41</td>
<td>16</td>
<td>6.15</td>
</tr>
<tr>
<td>Total</td>
<td>495.39</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Regression Coefficient</th>
<th>Estimated Standard Deviation</th>
<th>$t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$b_1 = -.334$</td>
<td>$s(b_1) = 3.016$</td>
<td>1.44</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$b_2 = -.2857$</td>
<td>$s(b_2) = 2.582$</td>
<td>-1.11</td>
</tr>
<tr>
<td>$X_3$</td>
<td>$b_3 = -.2186$</td>
<td>$s(b_3) = 1.596$</td>
<td>-1.37</td>
</tr>
</tbody>
</table>

Figure: Regression Results for Several Fitted Models-Body Fat Example.
Illustration for the extra sum of squares (cont.)

Notation:

- Assume $X_1$ is in the model
  - $SSR(X_1)$: The regression sum of squares
  - $SSE(X_1)$: The error sum of squares:
- measure the marginal effect of adding another variable to the regression model when $X_1$ is already in the model
  - $SSR(X_2|X_1)$: The extra sum of squares
Illustration for the extra sum of squares (cont.)

An extra sum of squares:

\[
SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) = 143.12 - 109.95 = 33.17
\]

\[
= SSR(X_1, X_2) - SSR(X_1) = 385.44 - 352.27 = 33.17
\]

**SSR(X_2|X_1)**

- reflects the **additional or extra** reduction in the error sum of squares (**SSE**) associated with \(X_2\), given that \(X_1\) is already included in the model
- the marginal **increase** in the regression sum of squares (**SSR**)
Illustration for the extra sum of squares (cont.)

Phenomenon
the marginal reduction in the $\text{SSE} =$ the marginal increase in $\text{SSR}$

- $\text{SSTO} = \text{SSR} + \text{SSE}$:
  - measure the variability of $Y_i$ and does not depend on the regression model fitted
  - Any reduction in $\text{SSE}$ implies an identical increase in $\text{SSR}$
7.1 Extra Sums of Squares

Illustration for the extra sum of squares (cont.)

**Figure**: Schematic Representation of Extra Sums of Squares-Body Fat Example.

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) = 143.12 - 109.95 = 33.17$$

$$= SSR(X_1, X_2) - SSR(X_1) = 385.44 - 352.27 = 33.17$$
Illustration for the extra sum of squares (cont.)

An extra sum of squares: adding $X_3$

$$SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = 109.95 - 98.41 = 11.54$$

$$= SSR(X_1, X_2, X_3) - SSR(X_1, X_2) = 396.98 - 385.44 = 11.54$$

An extra sum of squares: adding $X_2, X_3$

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) = 143.12 - 98.41 = 44.71$$

$$= SSR(X_1, X_2, X_3) - SSR(X_1) = 396.98 - 352.27 = 44.71$$
7.1 Extra Sums of Squares

Basic ideas

Extra Sums of Squares

- An extra sum of squares measures the marginal decrease in the error sum of squares when one or several predictor variables are added to the regression model, given that other variables are already in the model.

- Equivalently, one can view the extra sum of squares as measuring the marginal increase in the regression sum of squares.

Extra: $SSE \downarrow$; $SSR \uparrow$
Definitions

**Definition: Extra Sums of Squares for two variables**

If $X_1$ is the extra variable:

$$SSR(X_1|X_2) = SSE(X_2) - SSE(X_1, X_2)$$

$$= SSR(X_1, X_2) - SSR(X_2)$$

If $X_2$ is the extra variable:

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2)$$

$$= SSR(X_1, X_2) - SSR(X_1)$$
Definitions (cont.)

Definition: Extra Sums of Squares for three or more variables

If \( X_3 \) is the extra variable:

\[
SSR(X_3 | X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \\
= SSR(X_1, X_2, X_3) - SSR(X_1, X_2)
\]

If \( X_2, X_3 \) is the extra variable:

\[
SSR(X_2, X_3 | X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) \\
= SSR(X_1, X_2, X_3) - SSR(X_1)
\]
A variety (多樣化) of decompositions of SSR into extra sums of squares

Consider two $X$ variables:

\[ SSTO = SSR(X_1) + SSE(X_1) \]
\[ = SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2) \]
\[ SSTO = SSR(X_1, X_2) + SSE(X_1, X_2) \]
\[ \Rightarrow SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1) \]
Decomposition (分解) of $SSR$ into Extra Sums of Squares (cont.)

Decomposition $SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$

1. $SSR(X_1)$: measuring the contribution by including $X_1$ alone in the model

2. $SSR(X_2|X_1)$: measuring the addition contribution when $X_2$ is included, given that $X_1$ is already in the model

The order of the $X$ variables is arbitrary

$$SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2)$$
7.1 Extra Sums of Squares

Decomposition (分解) of $SSR$ into Extra Sums of Squares (cont.)

**Figure**: Schematic Representation of Extra Sums of Squares-Body Fat Example.
Decomposition (分解) of $SSR$ into Extra Sums of Squares (cont.)

- When the regression model contains three $X$ variables $(X_1, X_2, X_3)$:

\[
SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \\
= SSR(X_2) + SSR(X_3|X_2) + SSR(X_1|X_2, X_3) \\
= SSR(X_3) + SSR(X_1|X_3) + SSR(X_2|X_1, X_3) \\
= SSR(X_1) + SSR(X_2, X_3|X_1)
\]

- The number of possible decompositions becomes vast (龐大的) as the number of $X$ variables in the regression model increases.
### ANOVA Table Containing Decomposition of $SSR$

**Figure**: Example of ANOVA Table with Decomposition of $SSR$ for Three $X$ Variables.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>$SSR(X_1, X_2, X_3)$</td>
<td>3</td>
<td>$MSR(X_1, X_2, X_3)$</td>
</tr>
<tr>
<td>$X_1$</td>
<td>$SSR(X_1)$</td>
<td>1</td>
<td>$MSR(X_1)$</td>
</tr>
<tr>
<td>$X_2</td>
<td>X_1$</td>
<td>$SSR(X_2</td>
<td>X_1)$</td>
</tr>
<tr>
<td>$X_3</td>
<td>X_1, X_2$</td>
<td>$SSR(X_3</td>
<td>X_1, X_2)$</td>
</tr>
<tr>
<td>Error</td>
<td>$SSE(X_1, X_2, X_3)$</td>
<td>$n-4$</td>
<td>$MSE(X_1, X_2, X_3)$</td>
</tr>
<tr>
<td>Total</td>
<td>$SSTO$</td>
<td>$n-1$</td>
<td></td>
</tr>
</tbody>
</table>
Each extra sum of squares involving
- a single extra $X$ variable has associated with it one degree of freedom
- two extra $X$ variables have two degrees of freedom

Mean squares:

$$MSR(X_2|X_1) = \frac{SSR(X_2|X_1)}{1}$$

$$MSR(X_2, X_3|X_1) = \frac{SSR(X_2, X_3|X_1)}{2}$$
Extra sums of squares are of interest because they occur in a variety of tests about regression coefficients where the question of concern is whether certain $X$ variables can be dropped from the regression model.
7.2 Use of Extra Sums of Squares in Tests for Regression Coefficients
Test whether a Single $\beta_k = 0$

- Test whether $\beta_k X_k$ can be dropped from a multiple regression model
- Interest:
  \[ H_0 : \beta_k = 0 \]
  \[ H_a : \beta_k \neq 0 \]

Test statistics in (6.51b): \[ t^* = \frac{b_k}{s\{b_k\}} \]

- The general linear test approach (Sec. 2.8): Full model vs. Reduced model

\[ F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \]
7.2 Use of Extra Sums of Squares in Tests for Regression Coefficients

Test whether a Single $\beta_k = 0$ (cont.)

- The general linear test approach (Sec. 2.8) involves an extra sum of squares
- Illustration:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{Full model}$$

Test the alternatives:

$$H_0 : \beta_3 = 0 \quad \text{vs.} \quad H_1 : \beta_3 \neq 0$$

When $H_0$ holds:

$$\Rightarrow Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad \text{Reduced model}$$

- The test whether or not $\beta_3 = 0$ is a marginal test, given $X_1, X_2$ are already in the model
Test whether a Single $\beta_k = 0$ (cont.)

- **Steps:**
  1. $SSE(F)$: $SSE(F) = SSE(X_1, X_2, X_3)$, $df_F = n - 4$
  2. $SSE(R)$: $SSE(R) = SSE(X_1, X_2)$, $df_F = n - 3$
  3. The general linear test statistic (2.70):

\[
F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \cdot \frac{SSE(F)}{df_F} = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{(n - 3) - (n - 4)} \cdot \frac{SSE(X_1, X_2, X_3)}{n - 4}
\]
\[
= \frac{SSR(X_3|X_1, X_2)}{1} \cdot \frac{SSE(X_1, X_2, X_3)}{n - 4}
\]
\[
= \frac{MSR(X_3|X_1, X_2)}{MSE(X_1, X_2, X_3)}
\]
Test whether a Single $\beta_k = 0$ (cont.)

**TABLE 7.4**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>396.98</td>
<td>3</td>
<td>132.33</td>
</tr>
<tr>
<td>$X_1$</td>
<td>352.27</td>
<td>1</td>
<td>352.27</td>
</tr>
<tr>
<td>$X_2</td>
<td>X_1$</td>
<td>33.17</td>
<td>1</td>
</tr>
<tr>
<td>$X_3</td>
<td>X_1$, $X_2$</td>
<td>11.54</td>
<td>1</td>
</tr>
<tr>
<td>Error</td>
<td>98.41</td>
<td>16</td>
<td>6.15</td>
</tr>
<tr>
<td>Total</td>
<td>495.39</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

Body Fat Example

- **Test statistic:**
  \[
  F^* = \frac{SSR(X_3|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n - 4} = \frac{11.54}{1} \div \frac{98.41}{16} = 1.88
  \]

  \[
  F^* = 1.88 \leq 8.53 = F(0.99; 1, 16) \Rightarrow \text{conclude } H_0 (\alpha = 0.01)
  \]

- $X_3$ can be dropped from the regression model that already contains $X_1$, $X_2$
R codes for Extra Sum of Squares with the body fat example

```r
ex <- read.table("CH07TA01.txt",header=F)
n<-length(ex$V1)
frm1 <- lm(V4~V1+V2+V3,data=ex)
frm2 <- lm(V4~V1+V2,data=ex)
SSE1 <- deviance(frm1)
SSE2 <- deviance(frm2)
F<-(SSE2-SSE1)/1/(SSE1/(n-4))
```
Test whether Several $\beta_k = 0$

- Interest in whether several terms in the regression model can be dropped
- Illustration: whether both $\beta_2X_2$ and $\beta_3X_3$ can be dropped from the full model

$$Y_i = \beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \beta_3X_{i3} + \varepsilon_i \quad \text{Full model}$$

- Alternative:

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_a : \text{not both } \beta_2 \text{ and } \beta_3 \text{ equal zero}$$
Test whether Several $\beta_k = 0$ (cont.)

- The reduced model under $H_0$:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$  \hspace{1cm} \text{Reduced model}$$

$$\Rightarrow F^* = \frac{SSE(X_1) - SSE(X_1, X_2, X_3)}{(n - 2) - (n - 4)} \div \frac{SSE(X_1, X_2, X_3)}{n - 4}$$

$$= \frac{SSR(X_2, X_3|X_1)}{2} \div \frac{SSE(X_1, X_2, X_3)}{n - 4}$$

$$= \frac{MSR(X_2, X_3|X_1)}{MSE(X_1, X_2, X_3)}$$

where $SSR(X_2, X_3|X_1)$ is an extra sum of squares
Test whether Several $\beta_k = 0$ (cont.)

**Body Fat Example**

- Can both $X_2$ and $X_3$ be dropped from the full model?
- Test statistic:
  \[
  F^* = \frac{SSR(X_2, X_3|X_1)}{1} \div MSE(X_1, X_2, X_3) = \frac{44.71}{2} \div 6.15 = 3.63
  \]
  \[
  F^* = 3.63 \approx 3.63 = F(0.95; 2, 16)(\alpha = 0.05)
  \]
  \[\Rightarrow \text{at the boundary of the decision rule with } P\text{-value} \approx 0.05\]
- We may wish to **make further analyses** before deciding whether $X_2$ and $X_3$ should be dropped from the regression model that already contains $X_1$. 
Comments

- Testing whether a single $\beta_k$ equals zero:
  1. the $t^*$ test statistic
  2. the $F^*$ general linear test statistic

- Testing whether several $\beta_k$ equal zero:
  1. the $F^*$ general linear test statistic

General linear test statistic can be expressed in terms of the coefficients of multiple determination $R^2$

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \cdot \frac{SSE(F)}{df_F}$$

$$= \frac{R_F^2 - R_R^2}{df_R - df_F} \cdot \frac{1 - R_F^2}{df_F}$$

Body Fat Example

Can both $X_2$ and $X_3$ be dropped from the full model?

$$F^* = 3.63 = \frac{0.80135 - 0.71110}{(20 - 2) - (20 - 4)} \div \frac{1 - 0.80135}{16}$$

$$= \frac{R^2_{Y|123} - R^2_{R|1}}{(n - 2) - (n - 4)} \div \frac{1 - R^2_{Y|123}}{n - 4} = 3.63$$

Test statistic:

$$F^* = \frac{R^2_F - R^2_R}{df_R - df_F} \div \frac{1 - R^2_F}{df_F}$$

is not appropriate when the full and reduced regression models do not contain $\beta_0$
7.3 Summary of Tests Concerning Regression Coefficients
Summary

- Test whether all $\beta_k = 0$

  overall $F$ test: $F^* = \frac{MSR}{MSE} \sim F(p - 1, n - p)$

- Test whether a single $\beta_k = 0$

  partial $F$ test: $F^* = \frac{MSR(X_k|X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{p-1})}{MSE} \sim F(1, n - p)$

  $\Leftrightarrow t^* = \frac{b_k}{s\{b_k\}}$
Test whether some $\beta_k = 0$

$$H_0 : \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$$

Partial $F$ test: $F^* = \frac{MSR(X_q, \ldots, X_{p-1}|X_1, \ldots, X_{q-1})}{MSE} \sim F(p - q, n - p)$
Summary (cont.)

- **Other test:**

  \[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{Full model} \]

  \[
  H_0 : \beta_1 = \beta_2 \\
  H_a : \beta_1 \neq \beta_2 \\
  \Rightarrow Y_i = \beta_0 + \beta_c (X_{i1} + X_{i2}) + \beta_3 X_{i3} + \varepsilon_i \quad \text{Reduced model} \\
  \Rightarrow \text{the general } F^* \text{ test statistic } \sim F(1, n - 4)
  
  \[
  H_0 : \beta_1 = 3, \beta_3 = 5 \\
  H_a : \text{not both equalities in } H_0 \text{ hold} \\
  \Rightarrow Y_i - 3X_{i1} - 5X_{i3} = \beta_0 + \beta_2 X_{i2} + \varepsilon_i \quad \text{Reduced model} \\
  \Rightarrow \text{the general } F^* \text{ test statistic } \sim F(2, n - 4) \]
7.4 Coefficients of Partial Determination
7.4 Coefficients of Partial Determination

**partial determination**

- $R^2$: measures the proportionate reduction in the variation of $Y$ achieved by the introduction of the entire set of $X$ considered in the model.

- Coefficient of partial determination: measures the marginal contribution on one $X$ variable when all others are already included in the model.
Illustration: two predictor variables

Model \( Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \)

- \( \text{SSE}(X_2) \): measures the variation in \( Y \) when \( X_2 \) is included in the model
- \( \text{SSE}(X_1, X_2) \): measures the variation in \( Y \) when \( X_1, X_2 \) are included in the model
- \( R^2_{Y_{1|2}} \): the coefficient of partial determination between \( Y \) and \( X_{i1} \), given that \( X_2 \) is in the model

\[
R^2_{Y_{1|2}} = \frac{\text{SSE}(X_2) - \text{SSE}(X_1, X_2)}{\text{SSE}(X_2)} = \frac{\text{SSR}(X_1|X_2)}{\text{SSE}(X_2)}
\]
General Case: coefficients of partial determination to three or more $X$ variables in the model

\[
R_{Y1|23}^2 = \frac{SSR(X_1|X_2, X_3)}{SSE(X_2, X_3)}
\]

\[
R_{Y2|13}^2 = \frac{SSR(X_2|X_1, X_3)}{SSE(X_1, X_3)}
\]

\[
R_{Y3|12}^2 = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)}
\]

\[
R_{Y4|123}^2 = \frac{SSR(X_4|X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)}
\]
The coefficients of partial determination: $0 \sim 1$

Other interpretation: with a coefficient of simple determination

- Residuals-regress $Y$ on $X_2$

$$e_i(Y|X_2) = Y_i - \hat{Y}_i(X_2)$$

- Residuals- regress $X_1$ on $X_2$

$$e_i(X_1|X_2) = X_{i1} - \hat{X}_{i1}(X_2)$$

$R^2$ between $e_i(Y|X_2)$ and $e_i(X_1|X_2)$ will be the same as $R^2_{Y1|2}$

added variable plots or partial regression plots (Chap. 10.1): the strength of the relationship between $Y$ and $X_1$ adjusted for $X_2$

$$e_i(Y|X_2) \text{ vs. } e_i(X_1|X_2)$$
Body Fat example

\[ R^2_{Y1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = 0.031 \]

```r
ex<-read.table("CH07TA01.txt",header=F)
attach(ex)
fit1<-lm(V4~V2)
SSE1<-deviance(fit1)
SSE2<-SSE1-deviance(lm(V4~V1+V2))
RY12<-SSE2/SSE1
[1] 0.03061875
```
7.4 Coefficients of Partial Determination

Comments (cont.)

```r
ex<-read.table("CH07TA01.txt",header=F)
res1<-lm(V4~V2,data=ex)$residuals
res2<-lm(V1~V2,data=ex)$residuals
fitres<-summary(lm(res1~res2))
fitres$ r.squared
[1] 0.03061875
```
Coefficients of Partial Correlation

- Coefficient of partial correlation: (Chap 9.)
  
  \[ r_{Y2|1} = \sqrt{R_{Y2|1}^2} \]

- The same sign with the regression coefficient

- Expressed in terms of simple or other partial correlation coefficients:

  \[ R_{Y2|1}^2 = [r_{Y2|1}]^2 = \frac{(r_{Y2} - r_{12}r_{Y1})^2}{(1 - r_{12}^2)(1 - r_{Y1}^2)} \]

  \[ R_{Y2|13}^2 = [r_{Y2|13}]^2 = \frac{(r_{Y2|3} - r_{12|3}r_{Y1|3})^2}{(1 - r_{12|3}^2)(1 - r_{Y1|3}^2)} \]

- \( r_{Y1} \): correlation of \( Y \) and \( X_1 \)
- \( r_{12} \): correlation of \( X_1 \) and \( X_2 \)
7.5 Standardized Multiple Regression Model
Roundoff errors (捨入誤差):

- $X'X$
  - determinant that is close to zero: some variables are highly intercorrelated
  - the element of $X'X$ substantially different: the entries in $X'X$ cover a wide range magnitudes
  ⇒ standardized regression model
- transformation: correlation transformation
  - Transformed variables fall between -1 and 1
  - becomes much less subject to roundoff errors
Lack of Comparability in Regression Coefficients

- differences in the units

Illustration:

\[ \hat{Y} = 200 + 20000X_1 + 0.2X_2 \]

- \( Y \): dollars; \( X_1 \): thousand dollars; \( X_2 \): cents
- Is \( X_1 \) the only important predictor variable?
Correlation Transformation

- help with controlling roundoff errors
- expressing the regression coefficients in the same units

### Standard Normal Random Variable

\[ Y : \text{a normal r.v.} \]

\[ \Rightarrow \text{The standard normal r.v. } z = \frac{Y - \mu}{\sigma} \]

- **Standardizing**: involving **centering** and **scaling** the variable
Correlation Transformation (cont.)

- The usual standardizations of the variables:

\[
\frac{Y_i - \bar{Y}}{s_Y} ; \quad s_Y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n - 1}}
\]

\[
\frac{X_{ik} - \bar{X}_k}{s_k} ; \quad s_k = \sqrt{\frac{\sum (X_{ik} - \bar{X}_k)^2}{n - 1}} (k = 1, \ldots, p - 1)
\]

- The correlation transformation:

\[
Y^*_i = \frac{1}{\sqrt{n - 1}} \left( \frac{Y_i - \bar{Y}}{s_Y} \right)
\]

\[
X^*_{ik} = \frac{1}{\sqrt{n - 1}} \left( \frac{X_{ik} - \bar{X}_k}{s_k} \right) (k = 1, \ldots, p - 1)
\]
A standardized regression model:

\[ Y_i^* = \beta_1^{*} X_{i1}^{*} + \cdots + \beta_{p-1}^{*} X_{i,p-1}^{*} + \varepsilon_i^{*} \]

- no need for intercept

\[ \beta_k = \left( \frac{S_{Y_k}}{S_k} \right) \beta_k^{*} \quad (k = 1, \ldots, p - 1) \]

\[ \beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \cdots - \beta_{p-1} \bar{X}_{p-1} \]
**X'X Matrix for Transformed Variables**

- \( r_{XX} \): correlation matrix of the \( X \) variables

\[
\begin{align*}
\mathbf{r}_{XX} &= \begin{bmatrix}
1 & r_{12} & \cdots & r_{1,p-1} \\
r_{21} & 1 & \cdots & r_{2,p-1} \\
& \vdots & \ddots & \vdots \\
r_{p-1,1} & r_{p-1,2} & \cdots & 1
\end{bmatrix}
\end{align*}
\]

- \( r_{YX} \): correlation between \( Y \) and each of \( X \) variables:

\[
\begin{align*}
\mathbf{r}_{YX} &= \begin{bmatrix}
r_Y1 \\
r_Y2 \\
& \vdots \\
r_{Y,p-1}
\end{bmatrix}
\end{align*}
\]
X′X Matrix for Transformed Variables (cont.)

- The transformed variables: (no column of 1 in X)

\[
X = \begin{bmatrix}
X_{11}^* & \cdots & X_{1,p-1}^* \\
X_{21}^* & \cdots & X_{2,p-1}^* \\
\vdots & & \vdots \\
X_{n1}^* & \cdots & X_{n,p-1}^*
\end{bmatrix}
\]

\[X_n \times (p-1) = X_1 \cdots X_{p-1}, \quad X_{p-1} \cdots X_{n,p-1}\]

⇒ \[X'X = r_{XX}\]

- All of the elements of \(X'X\) are between -1 and 1

\[\sum (X_{i1}^*)^2 = 1\]

\[\sum X_{i1}^* X_{i2}^* = \frac{\sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\left[\sum (X_{i1} - \bar{X}_1)^2 \sum (X_{i2} - \bar{X}_2)^2\right]^2}\]
Estimated Standard Regression Coefficients

- the least squares estimator:

\[ b = (X'X)^{-1}XY \]

- The least squares normal equations and estimators of the regression coefficients of the standardized regression model:

\[ r_{XX}b = r_{YX} \Rightarrow b = r_{XX}^{-1}r_{YX} \]

\[
\begin{bmatrix}
  b_1^* \\
  b_2^* \\
  \vdots \\
  b_{p-1}^*
\end{bmatrix}
\]

- \( b_1^*, \ldots, b_{p-1}^* \): standardized regression coefficients
Estimated Standard Regression Coefficients (cont.)

the standardized parameters vs. the original parameters

\[ \beta_k = \left( \frac{S_Y}{s_k} \right) \beta_k^* \quad (k = 1, \ldots, p - 1) \]
\[ \beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \cdots - \beta_{p-1} \bar{X}_{p-1} \]

\[ b_k = \left( \frac{S_Y}{s_k} \right) b_k^* \quad (k = 1, \ldots, p - 1) \]
\[ b_0 = \bar{Y} - b_1 \bar{X}_1 - \cdots - b_{p-1} \bar{X}_{p-1} \]
Estimated Standard Regression Coefficients (cont.)

Illustration: \( p - 1 = 2 \)

\[
\begin{align*}
\mathbf{b} &= \frac{1}{1 - r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \begin{bmatrix} r_{Y_1} \\ r_{Y_2} \end{bmatrix} \\
\mathbf{b}^*_1 &= \frac{r_{Y_1} - r_{12}r_{Y_2}}{1 - r_{12}^2} \\
\mathbf{b}^*_2 &= \frac{r_{Y_2} - r_{12}r_{Y_1}}{1 - r_{12}^2}
\end{align*}
\]
Estimated Standard Regression Coefficients (cont.)

Dwane Studios example

Figure: Correlation Transformation and Fitted Standardized Regression Model-Dwaine Studios Example.

<table>
<thead>
<tr>
<th>Case</th>
<th>Sales</th>
<th>Target Population</th>
<th>Per Capita Disposable Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$Y_i$</td>
<td>$X_{i1}$</td>
<td>$X_{i2}$</td>
</tr>
<tr>
<td>1</td>
<td>174.4</td>
<td>68.5</td>
<td>16.7</td>
</tr>
<tr>
<td>2</td>
<td>164.4</td>
<td>45.2</td>
<td>16.8</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>20</td>
<td>224.1</td>
<td>82.7</td>
<td>19.1</td>
</tr>
<tr>
<td>21</td>
<td>166.5</td>
<td>52.3</td>
<td>16.0</td>
</tr>
</tbody>
</table>

\[ \bar{Y} = 181.90 \quad \bar{X}_1 = 62.019 \quad \bar{X}_2 = 17.143 \]
\[ s_y = 36.191 \quad s_1 = 18.620 \quad s_2 = 9.7035 \]
Estimated Standard Regression Coefficients (cont.)

(b) Transformed Data

<table>
<thead>
<tr>
<th>(i)</th>
<th>(y_i^*)</th>
<th>(x_{i1}^*)</th>
<th>(x_{i2}^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.04637</td>
<td>.07783</td>
<td>-.10205</td>
</tr>
<tr>
<td>2</td>
<td>-.10815</td>
<td>-.20198</td>
<td>-.07901</td>
</tr>
<tr>
<td>20</td>
<td>.26070</td>
<td>.24835</td>
<td>.45100</td>
</tr>
<tr>
<td>21</td>
<td>-.09518</td>
<td>-.11671</td>
<td>-.26336</td>
</tr>
</tbody>
</table>

(c) Fitted Standardized Model

\[
\hat{y}^* = 0.7484 x_1^* + 0.2511 x_2^*
\]

\[
\hat{y} = -68.860 + 1.455 x_1 + 9.365 x_2
\]
## Estimated Standard Regression Coefficients (cont.)

```r
## Ex p277
library(QuantPsyc)
ex7.5<-read.table("CH07TA05.txt")
fit<-lm(V1~V2+V3,data=ex7.5)
fit
  Call:
  lm(formula = V1 ~ V2 + V3, data = ex7.5)

  Coefficients:
  (Intercept) V2 V3
  -68.857 1.455 9.366

lm.beta(fit)
  V2 V3
  0.7483670 0.2511039
```

- Estimated coefficients for the standardized multiple regression model.
- The model is specified as `V1 ~ V2 + V3`.
- The intercept is -68.857, coefficient for `V2` is 1.455, and for `V3` is 9.366.
- The standardized coefficients (`lm.beta(fit)`) are 0.7483670 for `V2` and 0.2511039 for `V3`.
Estimated Standard Regression Coefficients (cont.)

\[ \hat{Y}^* = 0.7484X_1^* + 0.2511X_2^* \]

- Does \( X_1 \) have a much greater impact on sales than \( X_2 \)? (\( \because b_1^* > b_2^* \))
- One must be cautious about interpreting any regression coefficient whether standardized or not.
- correlated among the predictor variables
- \( r_{12} = 0.781 \) in the Dwaine Studios data
7.6 Multicollinearity and Its Effects
Multicollinearity and Its Effects

Questions:

1. the relative importance of the effects of $X$
2. the magnitude of the effect of a given $X$ on $Y$
3. Can any $X$ be dropped from the model?
4. Should any $X$ not yet included in the model?

*intercorrelation* or *multicollinearity*:
the predictor variables are correlated among themselves
Uncorrelated Predicted Variables

- Table 7.6: $Y$ - crew productivity; $X_1$ - the effect of work crew size; $X_2$ - level of bonus pay
- $r_{12}^2 = 0$: the predictor variables are uncorrelated
- $SSR(X_1|X_2) = 231.125 = SSR(X_1)$
- $SSR(X_2|X_1) = 171.125 = SSR(X_2)$
Multicollinearity and Its Effects

Figure: Regression Results when Predictor Variables Are Uncorrelated—Work Crew Productivity Example.

### TABLE 7.7
Regression Results when Predictor Variables Are Uncorrelated—Work Crew Productivity Example.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>402.250</td>
<td>2</td>
<td>201.125</td>
</tr>
<tr>
<td>Error</td>
<td>17.625</td>
<td>5</td>
<td>3.525</td>
</tr>
<tr>
<td>Total</td>
<td>419.875</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

(a) Regression of \( \hat{Y} \) on \( X_1 \) and \( X_2 \)
\[ \hat{Y} = .375 + 5.375X_1 + 9.250X_2 \]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>231.125</td>
<td>1</td>
<td>231.125</td>
</tr>
<tr>
<td>Error</td>
<td>188.750</td>
<td>6</td>
<td>31.458</td>
</tr>
<tr>
<td>Total</td>
<td>419.875</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

(b) Regression of \( \hat{Y} \) on \( X_1 \)
\[ \hat{Y} = 23.500 + 5.375X_1 \]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>171.125</td>
<td>1</td>
<td>171.125</td>
</tr>
<tr>
<td>Error</td>
<td>248.750</td>
<td>6</td>
<td>41.458</td>
</tr>
<tr>
<td>Total</td>
<td>419.875</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

(c) Regression of \( \hat{Y} \) on \( X_2 \)
\[ \hat{Y} = 27.250 + 9.250X_2 \]
Uncorrelated Predicted Variables

When two or more predictor variables are uncorrelated, the marginal contribution of one predictor variable in reducing the error sum of squares when the other predictor variables are in the model is exactly the same as when this predictor variable is in the model alone.

\[ b_1 = \frac{r_{Y1} - r_{12}r_{Y2}}{1 - r_{12}^2} \]

\[ \Rightarrow b_1 = \frac{\sum (X_{i1} - \bar{X}_1)(Y_i - \bar{Y})}{\sum (X_{i1} - \bar{X}_1)^2} \quad \text{when} \quad r_{12} = 0 \]
7.6 Multicollinearity and Its Effects

Multicollinearity and Its Effects (cont.)

Predictor variables are perfectly correlated

\[ E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \]

Figure: Example of Perfectly Correlated Predictor Variables.

<table>
<thead>
<tr>
<th>Case</th>
<th>( X_{i1} )</th>
<th>( X_{i2} )</th>
<th>( Y_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>8</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>10</td>
<td>103</td>
</tr>
</tbody>
</table>

Fitted Values for Regression Function

\[ \hat{Y} = -87 + 18X_1 + 18X_2 \quad (7.58) \]
\[ \hat{Y} = -7 + 9X_1 + 2X_2 \quad (7.59) \]
Multicollinearity and Its Effects (cont.)

Figure: Two Response Planes That Intersect when $X_2 = 5 + 0.5X_1$.

FIGURE 7.2
Two Response Planes That Intersect when $X_2 = 5 + .5X_1$. 
When $X_1$ and $X_2$ are perfectly correlated, many different response functions will lead to the same perfectly fitted values for the observations.

The perfect relation between $X_1$ and $X_2$ did not inhibit (約束) the ability to obtain a good fit to the data.

Since many different response functions provide the same good fit, we cannot interpret any one set of regression coefficients as reflecting the effects of the different predictor variables.
Figure: Scatter Plot Matrix and Correlation Matrix of the Predictor Variables—Body Fat Example.

**FIGURE 7.3**
Scatter Plot Matrix and Correlation Matrix of the Predictor Variables—Body Fat Example.

(a) Scatter Plot Matrix of $X$ Variables

(b) Correlation Matrix of $X$ Variables

$$
\mathbf{r}_{xx} = \begin{bmatrix}
1.0 & .924 & .458 \\
.924 & 1.0 & .085 \\
.458 & .085 & 1.0
\end{bmatrix}
$$
### Body Fat Example - Effects on Regression Coefficients

<table>
<thead>
<tr>
<th>Variables in Model</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0.8572</td>
<td>—</td>
</tr>
<tr>
<td>$X_2$</td>
<td>—</td>
<td>0.8565</td>
</tr>
<tr>
<td>$X_1, X_2$</td>
<td>0.2224</td>
<td>0.6594</td>
</tr>
<tr>
<td>$X_1, X_2, X_3$</td>
<td>4.334</td>
<td>-2.857</td>
</tr>
</tbody>
</table>
Multicollinearity and Its Effects (cont.)

The high degree of multicollinearity among the predictor variables is responsible for the inflated variability of the estimated regression coefficients.

<table>
<thead>
<tr>
<th>Variables in Model</th>
<th>$s{b_1}$</th>
<th>$s{b_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>.1288</td>
<td>—</td>
</tr>
<tr>
<td>$X_2$</td>
<td>—</td>
<td>.1100</td>
</tr>
<tr>
<td>$X_1, X_2$</td>
<td>.3034</td>
<td>.2912</td>
</tr>
<tr>
<td>$X_1, X_2, X_3$</td>
<td>3.016</td>
<td>2.582</td>
</tr>
</tbody>
</table>
7.6 Multicollinearity and Its Effects (cont.)

Effects on fitted values and predictions

<table>
<thead>
<tr>
<th>Variables in Model</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>7.95</td>
</tr>
<tr>
<td>$X_1, X_2$</td>
<td>6.47</td>
</tr>
<tr>
<td>$X_1, X_2, X_3$</td>
<td>6.15</td>
</tr>
</tbody>
</table>

- Estimated means and Predicted values are not affected
Effects on simultaneous tests of $\beta_k$

- It is possible that when individual $t$ tests are performed, neither $\beta_1$ or $\beta_2$ is significant.
- However, when the $F$ test is performed for both $\beta_1$ and $\beta_2$, the results may still be significant.

- Need for more powerful diagnostics for multicollinearity