# 7 章

## 複迴歸之二

## Multiple Regression Ⅱ

- 本章在研究複迴歸所特有的主題，包含檢定迴歸係數時所需的額外平方和、複迴歸模型的標準化版本，以及用以描述預測變數間高度相關狀況的多重共線性等。

## 7.1 額外平方和

(Extra sums of squares)

### 基本想法

- 額外平方和(extra sums of squares)是指在原有的迴歸模型中，多加入一個或多個預測變數，造成誤差平方和 SSE 減少的量，也可以視為加入一個或多個預測變數後，造成迴歸平方和 SSR 增加的量。

# 額外平方和

**例題**

表7.1中的資料在研究人體脂肪量($Y$)與數個可能的預測變數間的關聯性，該研究樣本為二十位25至34歲的女性，可能的預測變數有三頭肌外表厚度($X_1$)、大腿周長($X_2$)與中臂周長($X_3$)。表7.1中的20個受測樣本其脂肪量是透過一些麻煩且昂貴之程序來取得的。由於預測變數很容易取得，因此如果可以透過部份或全部的預測變數來提供人體脂肪量的可靠估計，這樣的估計程序將十分實用。

**表 7.1**
脂肪量案例的基本資料。

| 樣本個體<br>$i$ | 三頭肌外表厚度<br>$X_{i1}$ | 大腿周長<br>$X_{i2}$ | 中臂周長<br>$X_{i3}$ | 人體脂肪量<br>$Y_i$ |
|---|---|---|---|---|
| 1 | 19.5 | 43.1 | 29.1 | 11.9 |
| 2 | 24.7 | 49.8 | 28.2 | 22.8 |
| 3 | 30.7 | 51.9 | 37.0 | 18.7 |
| ... | ... | ... | ... | ... |
| 18 | 30.2 | 58.6 | 24.6 | 25.4 |
| 19 | 22.7 | 48.2 | 27.1 | 14.8 |
| 20 | 25.2 | 51.0 | 27.5 | 21.1 |

表 7.2 為人體脂肪量 $Y$ 分別進行四種迴歸程序後之結果：(1)脂肪量 $Y$ 對三頭肌外表厚度 $X_1$，(2)脂肪量 $Y$ 對大腿周長 $X_2$，(3)脂肪量 $Y$ 同時對三頭肌外表厚度 $X_1$ 與大腿周長 $X_2$，(4)脂肪量 $Y$ 同時對三頭肌外表厚度 $X_1$、大腿周長 $X_2$ 與中臂周長 $X_3$。為了方便讀者參照四種不同之模型，我們將符號修改為當迴歸模型中僅存在預測變數 $X_1$ 時，其迴歸平方和為 $SSR(X_1) = 352.27$，誤差平方和為 $SSE(X_1) = 143.12$。

**表 7.2**

數個迴歸模型配適結果－脂肪量案例。

(a) $Y$ 對 $X_1$ 進行迴歸程序

$$\hat{Y} = -1.496 + .8572\, X_1$$

| 變異來源 | | 平方和 | 自由度 | 均方 |
|---|---|---|---|---|
| 迴歸 | $SSR(X_1) =$ | 352.27 | 1 | 352.27 |
| 誤差 | $SSE(X_1) =$ | 143.12 | 18 | 7.95 |
| 總合 | | 495.39 | 19 | |
| 變數 | | 迴歸係數估計 | 標準差估計 | $t^*$ |
| $X_1$ | | $b_1 = .8572$ | $s\{b_1\} = .1288$ | 6.66 |

(a) $Y$ 對$X_1$進行迴歸程序

$$\hat{Y} = -1.496 + .8572\ X_1$$

| 變異來源 | | 平方和 | 自由度 | 均方 |
|---|---|---|---|---|
| 迴歸 | $SSR(X_1) =$ | 352.27 | 1 | 352.27 |
| 誤差 | $SSE(X_1) =$ | 143.12 | 18 | 7.95 |
| 總合 | | 495.39 | 19 | |

| 變數 | 迴歸係數估計 | 標準差估計 | $t^*$ |
|---|---|---|---|
| $X_1$ | $b_1 = .8572$ | $s\{b_1\} = .1288$ | 6.66 |

(b) $Y$ 對$X_2$進行迴歸程序

$$\hat{Y} = -23.634 + .8565\ X_2$$

| 變異來源 | | 平方和 | 自由度 | 均方 |
|---|---|---|---|---|
| 迴歸 | $SSR(X_2) =$ | 381.97 | 1 | 381.97 |
| 誤差 | $SSE(X_2) =$ | 113.42 | 18 | 6.30 |
| 總合 | | 495.39 | 19 | |

| 變數 | 迴歸係數估計 | 標準差估計 | $t^*$ |
|---|---|---|---|
| $X_2$ | $b_2 = .8565$ | $s\{b_2\} = .1100$ | 7.79 |

類似的符號修改，當迴歸模型中預測變數有 $X_1$ 與 $X_2$ 時，其迴歸平方和為 $\boxed{SSR(X_1, X_2) = 385.44}$，誤差平方和為 $\boxed{SSE(X_1, X_2) = 109.95}$。

(c) $Y$ 對 $X_1$ 與 $X_2$ 進行迴歸程序

$$\hat{Y} = -19.174 + .2224\,X_1 + .6594\,X_2$$

| 變異來源 | | 平方和 | 自由度 | 均方 |
|---|---|---|---|---|
| 迴歸 | $SSR(X_1, X_2) =$ | $\boxed{385.44}$ | 2 | 192.72 |
| 誤差 | $SSE(X_1, X_2) =$ | $\boxed{109.95}$ | 17 | 6.47 |
| 總合 | | 495.39 | 19 | |

| 變數 | 迴歸係數估計 | 標準差估計 | $t^*$ |
|---|---|---|---|
| $X_1$ | $b_1 = .2224$ | $s\{b_1\} = .3034$ | .73 |
| $X_2$ | $b_2 = .6594$ | $s\{b_2\} = .2912$ | 2.26 |

(d) $Y$ 對 $X_1$、$X_2$ 與 $X_3$ 進行迴歸程序

$$\hat{Y} = 117.08 + 4.334\,X_1 - 2.857\,X_2 - 2.186\,X_3$$

| 變異來源 | | 平方和 | 自由度 | 均方 |
|---|---|---|---|---|
| 迴歸 | $SSR(X_1, X_2, X_3) =$ | $\boxed{396.98}$ | 3 | 132.33 |
| 誤差 | $SSE(X_1, X_2, X_3) =$ | $\boxed{98.41}$ | 16 | 6.15 |
| 總合 | | 495.39 | 19 | |

| 變數 | 迴歸係數估計 | 標準差估計 | $t^*$ |
|---|---|---|---|
| $X_1$ | $b_1 = 4.334$ | $s\{b_1\} = 3.016$ | 1.44 |
| $X_2$ | $b_2 = -2.857$ | $s\{b_2\} = 2.582$ | −1.11 |
| $X_3$ | $b_3 = -2.186$ | $s\{b_3\} = 1.596$ | −1.37 |

我們發現當預測變數 $X_1$ 與 $X_2$ 均在迴歸模型時，誤差平方和 $SSE(X_1, X_2) = 109.95$ 會小於模型中只有 $X_1$ 下的誤差平方和 $SSE(X_1) = 143.12$，其間之差即為額外平方和，用符號 $SSR(X_2|X_1)$ 來表示：

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

誤差平方和減少的量可以視為原本迴歸模型中已經存在 $X_1$ 時，如果再加入 $X_2$ 後的效果，因此額外平方和 $SSR(X_2|X_1)$ 是用來衡量當原本迴歸模型中已經存在 $X_1$ 時，增加 $X_2$ 後的邊際效果。

額外平方和 $SSR(X_2|X_1)$ 也可以視為迴歸平方和的邊際增量：

$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1)$$
$$= 385.44 - 352.27 = 33.17$$

根據(2.50)的變異數等式：

$$SSTO = SSR + SSE$$

由於 $SSTO$ 是用來衡量觀測值 $Y_i$ 的變異量，所以不受模型配適時預測變數不同之影響，故 $SSE$ 的縮減量將等於 $SSR$ 的增加量。

接下來我們考慮其他的額外平方和，例如當原本迴歸模型中已經存在$X_1$與$X_2$時，將預測變數$X_3$加入後的邊際效果：

$$SSR(X_3|X_1,X_2)=SSE(X_1,X_2)-SSE(X_1,X_2,X_3)$$
$$=\boxed{109.95}-\boxed{98.41}=11.54$$

亦即

$$SSR(X_3|X_1,X_2)=SSR(X_1,X_2,X_3)-SSR(X_1,X_2)$$
$$=\boxed{396.98}-\boxed{385.44}=11.54$$

另外，如果我們同時加入數個預測變數，例如當原本迴歸模型中僅存在$X_1$時，將預測變數$X_2$與$X_3$加入後的邊際效果：

$$SSR(X_2,X_3|X_1)=SSE(X_1)-SSE(X_1,X_2,X_3)$$
$$=\boxed{143.12}-\boxed{98.41}=44.71$$

亦即

$$SSR(X_2,X_3|X_1)=SSR(X_1,X_2,X_3)-SSR(X_1)$$
$$=\boxed{396.98}-\boxed{352.27}=44.71$$

## 額外平方和 SSR（$X_1$ | $X_2$）定義

- 由於額外平方和是在衡量新舊兩種迴歸模型間的誤差平方和 SSE 之差，若$X_1$是後來才加入的新預測變數，則因此可以定義：

$$\boxed{SSR\left(X_1 \middle| X_2\right) = SSE\left(X_2\right) - SSE\left(X_1, X_2\right)} \qquad (7.1a)$$

或

$$\boxed{SSR\left(X_1 \middle| X_2\right) = SSR\left(X_1, X_2\right) - SSR\left(X_2\right)} \qquad (7.1b)$$

若$X_2$是後來才加入的新預測變數，則

$$SSR\left(X_2 \middle| X_1\right) = SSE\left(X_1\right) - SSE\left(X_1, X_2\right) \qquad (7.2a)$$

或

$$SSR\left(X_2 \middle| X_1\right) = SSR\left(X_1, X_2\right) - SSR\left(X_1\right) \qquad (7.2b)$$

將上述定義擴增至三個或三個以上時，例如：

$$\boxed{SSR\left(X_3 \middle| X_1, X_2\right) = SSE\left(X_1, X_2\right) - SSE\left(X_1, X_2, X_3\right)} \qquad (7.3a)$$

$$\boxed{SSR(X_3|X_1,X_2)=SSE(X_1,X_2)-SSE(X_1,X_2,X_3)} \quad (7.3a)$$

或 $$SSR(X_3|X_1,X_2)=SSR(X_1,X_2,X_3)-SSR(X_1,X_2) \quad (7.3b)$$

而且 $$\boxed{SSR(X_2,X_3|X_1)=SSE(X_1)-SSE(X_1,X_2,X_3)} \quad (7.4a)$$

等價於 $$SSR(X_2,X_3|X_1)=SSR(X_1,X_2,X_3)-SSR(X_1) \quad (7.4b)$$

## 分解SSR成為額外平方和

複迴歸的迴歸平方和 *SSR* 可以依照不同的額外平方和形式，而有多種不同的總和分解方式，首先考慮(2.50)的等式

$$SSTO=SSR(X_1)+SSE(X_1) \quad (7.5)$$

上式指出模型的預測變數為$X_1$，現在利用(7.2a)的等式代入

$$SSR(X_2|X_1)=SSE(X_1)-SSE(X_1,X_2) \quad (7.2a)$$

$$\boxed{SSTO=SSR(X_1)+SSR(X_2|X_1)+SSE(X_1,X_2)} \quad (7.6)$$

$$SSR(X_2|X_1)=SSR(X_1,X_2)-SSR(X_1) \quad (7.2b)$$

$$\boxed{SSTO = SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2)}$$

- 於是形成兩個預測變數的複迴歸平方和等式，亦即：

$$\boxed{SSTO = SSR(X_1, X_2) + SSE(X_1, X_2)} \tag{7.7}$$

根據(7.7)與(7.6)的結果，可以得到：

$$\boxed{SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)} \tag{7.8}$$

平方和 $SSR(X_1, X_2)$ 分解成兩部份：(1) SSR（$X_1$）表示模型中預測變數$X_1$的貢獻，（2）SSR（$X_2$ | $X_1$）表示當模型中已經存在有預測變數$X_1$時，再引進預測變數$X_2$後所增加的貢獻。

- 事實上，引進預測變數的順序可能會改變，所以平方和也可以分解為：

$$\boxed{SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2)} \tag{7.9}$$

- 圖7.1表示人體脂肪量案例中兩種不同的 $SSR(X_1, X_2)$ 分解方式 $\boxed{SSR(X_1,X_2)=SSR(X_1)+SSR(X_2|X_1)}$，左圖長條為 SSTO 以及(7.9)的分解方式，該圖中的空白部份為 $SSR(X_2)$，接著為額外平方和 $SSR(X_1 \mid X_2)$， 類似的做法，右圖長條(7.8) 的分解方式 $\boxed{SSR(X_1,X_2)=SSR(X_2)+SSR(X_1|X_2)}$ 。

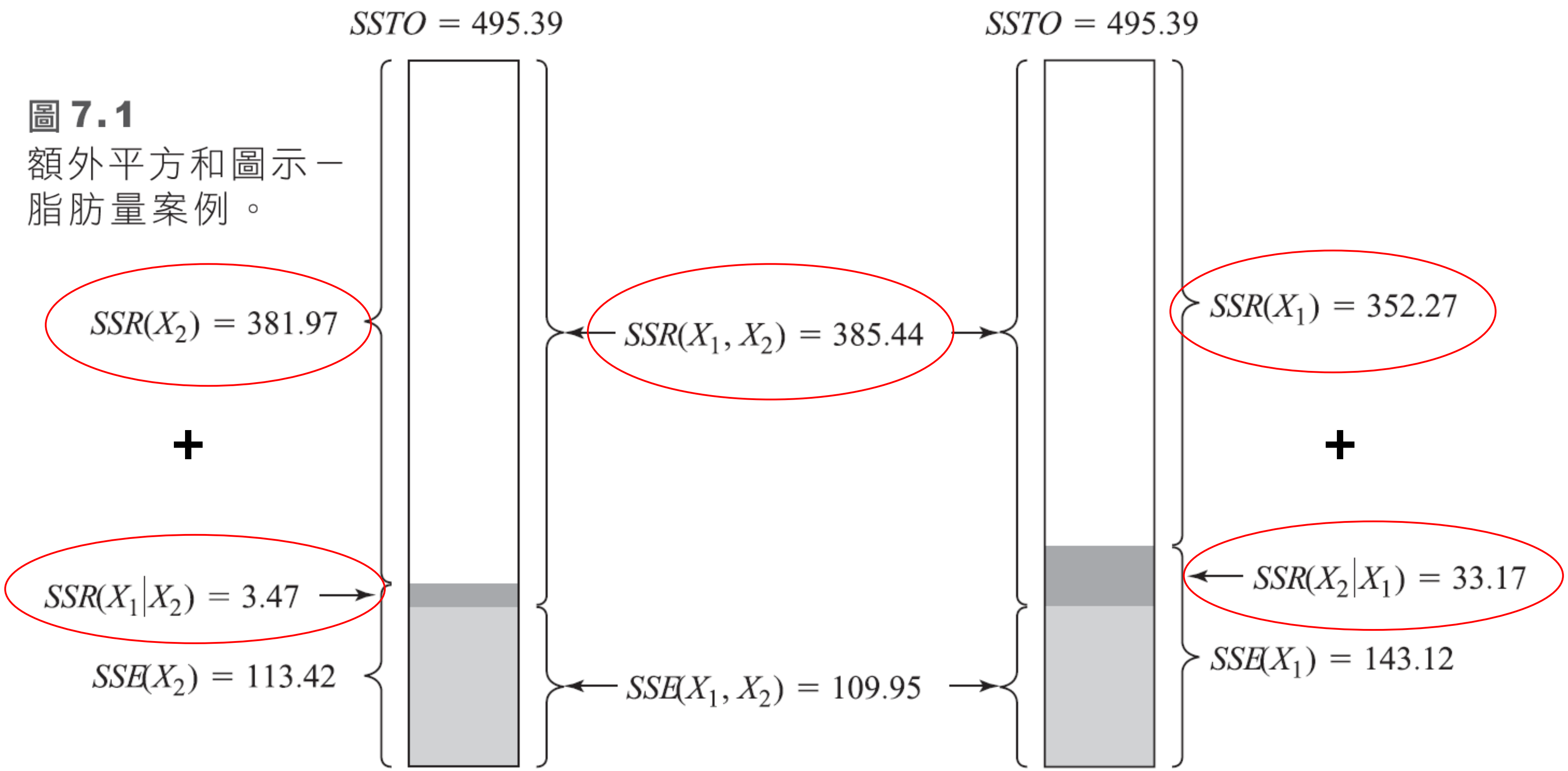
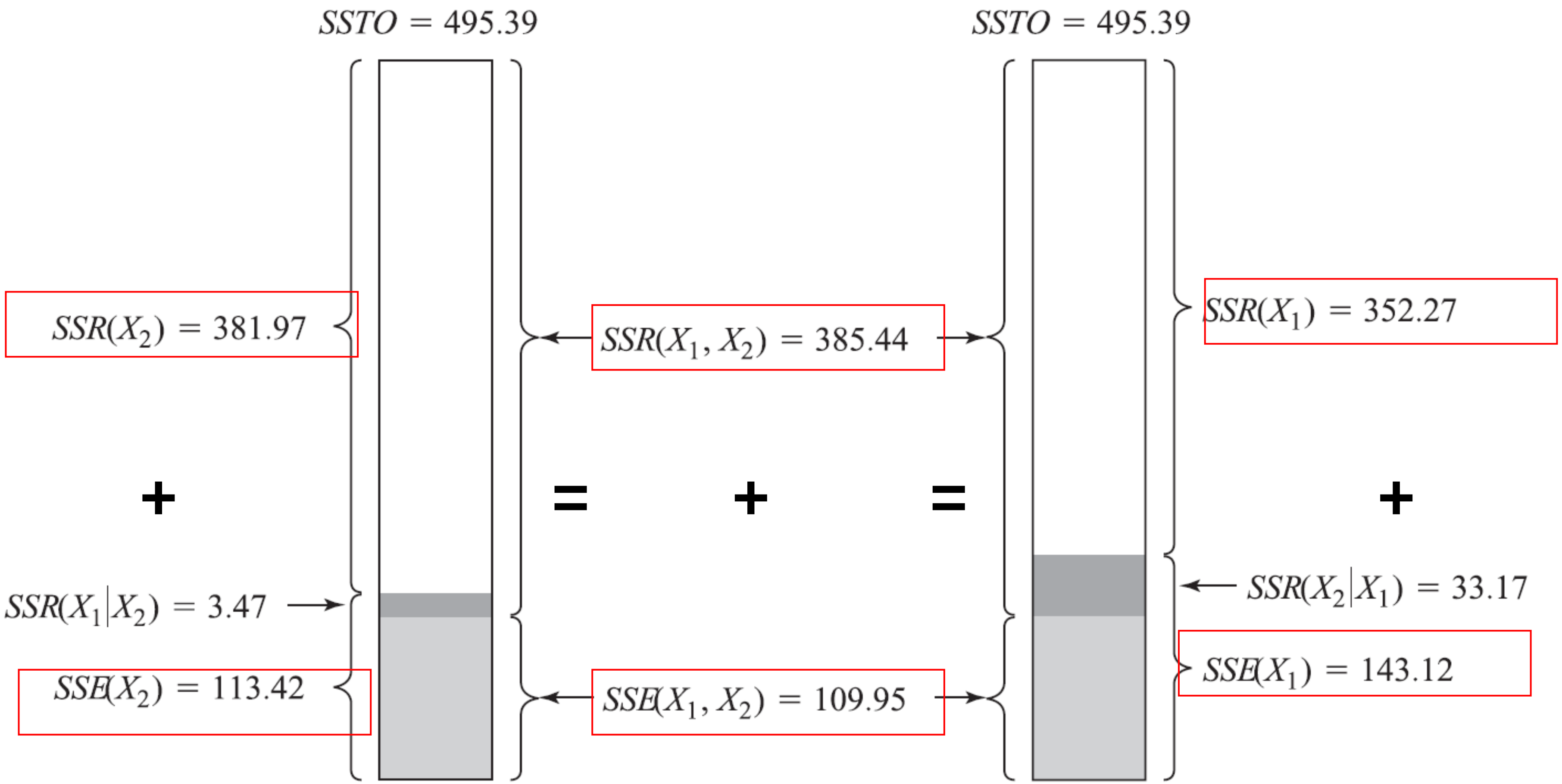



**圖 7.1**
額外平方和圖示－脂肪量案例。




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

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

當迴歸模型所引進的預測變數有三個時，則可能有多種 SSR($X_1$, $X_2$, $X_3$)的分解形式，例如：

$$\boxed{SSR(X_1,X_2,X_3)=SSR(X_1)+SSR(X_2|X_1)+SSR(X_3|X_1,X_2)} \quad (7.10a)$$

$$\boxed{SSR(X_1,X_2,X_3)=SSR(X_2)+SSR(X_3|X_2)+SSR(X_1|X_2,X_3)} \quad (7.10b)$$

$$\boxed{SSR(X_1,X_2,X_3)=SSR(X_1)+SSR(X_2,X_3|X_1)} \quad (7.10c)$$

顯然當迴歸模型所引進的預測變數越多時，可能的分解形式將大幅增加。

- **包含 *SSR* 之分解的 ANOVA 表**

此處我們透過ANOVA表來表示迴歸平方和SSR的分解，

• 首先表7.3是當迴歸模型引進三個預測變數($P$-1=3)時，迴歸分析套裝軟體所常用的分解方式，而表7.4則是將此一分解方式應用在人體脂肪量案例中，$n = 20, p = 4$（程式例題練習：SS1，SS2與之對照）。

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | $SSR(X_1, X_2, X_3)$ | 3 (P-1=3) | $MSR(X_1, X_2, X_3)$ |
| $X_1$ | $SSR(X_1)$ | 1 | $MSR(X_1)$ |
| $X_2 \| X_1$ | $SSR(X_2 \| X_1)$ | 1 | $MSR(X_2 \| X_1)$ |
| $X_3 \| X_1, X_2$ | $SSR(X_3 \| X_1, X_2)$ | 1 | $MSR(X_3 \| X_1, X_2)$ |
| 誤差 | $SSE(X_1, X_2, X_3)$ | $n-4$ (n-p) | $MSE(X_1, X_2, X_3)$ |
| 總合 | $SSTO$ | $n-1$ | |

**表 7.3**
引進三個預測變數下 ANOVA 的表例。

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | 396.98 | 3 | 132.33 |
| $X_1$ | 352.27 | 1 | 352.27 |
| $X_2 \| X_1$ | 33.17 | 1 | 33.17 |
| $X_3 \| X_1, X_2$ | 11.54 | 1 | 11.54 |
| 誤差 | 98.41 | 16 | 6.15 |
| 總合 | 495.39 | 19 | |

**表 7.4**
引進三個預測變數下的 ANOVA 表－脂肪量案例。

- 當引進一個單獨的預測變數X時，額外平方和SSR($X_2$ | $X_1$)都將具有一個自由度，所以有關均方的計算方式與一般的平方和一樣，例如表7.3的 MSR($X_2$ | $X_1$) 之計算方式為：

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

- 如果同時引進兩個預測變數時，則額外平方和SSR($X_2$ ,$X_3$|$X_1$)將具有兩個自由度，因為它可以表示成兩個各具有一個自由度的額外平方和之總和，例如：

$$\boxed{SSR(X_2, X_3 | X_1) = SSR(X_2 | X_1) + SSR(X_3 | X_1, X_2)} \qquad (7.11)$$

所以均方可以計算如下：

$$MSR(X_2, X_3 | X_1) = \frac{SSR(X_2, X_3 | X_1)}{2}$$

- 許多的電腦套裝軟體關於迴歸分析的程序，均提供分解SSR為單一自由度下額外平方和的功能。一般而言，通常是依照預測變數進入模型中之順序來進行分解，因此如果預測變數進入的順序為$X_1$、$X_2$與$X_3$，則輸出的額外平方和為：

| | | | |
|---|---|---|---|
| | $SSR(X_1)$ | | $SSR(X_1 \mid X_2, X_3)$ |
| 連續 SSR | $SSR(X_2 \mid X_1)$ | 部分 SSR | $SSR(X_2 \mid X_1, X_3)$ |
| | $SSR(X_3 \mid X_1, X_2)$ | | $SSR(X_3 \mid X_1, X_2)$ |

- 假設有必要一次引進多個預測變數時，則額外平方和可藉由一個適當的單一自由度額外平方和的加總獲得，如前例中的$SSR(X_2, X_3 \mid X_1)$，可以利用（7.11）將$SSR(X_2 \mid X_1)$與$SSR(X_3 \mid X_1, X_2)$加總起來即可。

- 如果需要計算SSR（$X_1, X_3 | X_2$），則可以改變預測變數進入模型中之順序，指定為$X_2$、$X_1$與$X_3$，或是$X_2$、$X_3$與$X_1$，前者所得到的結果為：

$$SSR（X_2）$$

$$SSR（X_1 | X_2）$$

$$SSR（X_3 | X_1, X_2）$$

而後面兩項之加總便是SSR（$X_1, X_3 | X_2$）。

$$SSR（X_1, X_3 | X_2）= SSR（X_1 | X_2）+ SSR（X_3 | X_1, X_2）$$

- 會對於額外平方和產生興趣，是因為它們牽涉到迴歸係數的檢定，也就某些預測變數是否可以從模型中將它們踢除。

# 7.2 利用額外平方和檢定迴歸係數

Use of Extra Sums of Squares in Tests for Regression Coefficients

- **檢定單一 $\beta_k = 0$**

A.此處想檢定，複迴歸模型中的某一項（如 $\beta_k k_k$）是否可以剔除，首先考慮：

$$H_o：\beta_3 = 0$$

$$H_\alpha：\beta_3 \neq 0$$

所適用的檢定統計量為（6.51b）：

$$\boxed{t^* = \frac{b_k}{s\{b_k\}}}$$

B.此外，上述檢定也可以根據2.8節所介紹的一般線性檢定方法來進行。我們先考慮三個預測變數的一階迴歸模型：

$$\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{全模型}} \qquad (7.12)$$

假設檢定為：

$$\boxed{\begin{aligned} H_0 &: \beta_3 = 0 \\ H_a &: \beta_3 \neq 0 \end{aligned}} \qquad (7.13)$$

- 配適全模型後可以得到誤差平方和 SSE(F)，亦即：

$$\mathrm{SSE(F)} = \mathrm{SSE}(X_1, X_2, X_3)$$

由於(7.12)之全模型中有四個參數$P$，所以上式之自由度為
$df_f = n - 4$。 $\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{全模型}}$

- (7.13)的$H_0$成立時可以得到縮減模型為：

$$\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad \text{縮減模型}} \qquad (7.14)$$

配適此縮減模型後($P$=3， $\beta_3$=0)，可得到誤差平方和SSE(R)
$= \mathrm{SSE}(X_1, X_2)$，自由度 $df_R = n - 3$。

一般線性檢定之統計量為(2.70)：

亦即：

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

亦即：
$$F^* = \frac{SSE(X_1,X_2)-SSE(X_1,X_2,X_3)}{(n-3)-(n-4)} \div \frac{SSE(X_1,X_2,X_3)}{n-4}$$

上式分子中的兩個誤差平方和之差正是（7.3a）式

$\boxed{SSR(X_3|X_1,X_2)=SSE(X_1,X_2)-SSE(X_1,X_2,X_3)}$ 的額外平方和：

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

所以一般線性檢定之統計量為：

$$\boxed{F^* = \frac{SSR(X_3|X_1,X_2)}{1} \div \frac{SSE(X_1,X_2,X_3)}{n-4} = \frac{MSR(X_3|X_1,X_2)}{MSE(X_1,X_2,X_3)}} \quad (7.15)$$

又 $F(1-\alpha,1,n-4)$，根據上述結果可以得知，在給定 $X_1$ 與 $X_2$ 後之模型，對虛無假設 $\beta_3=0$ 之檢定可視為一種邊際檢定，同時，額外平方和SSR（$X_3|X_1$, $X_2$）具有一個自由度。

在人體脂肪量的案例中，我們想要檢定預測變數中臂周長$X_3$是否可以從模型中剔除，檢定的假說如(7.13)：$\beta_3 = 0$。表 7.4 是迴歸模型(7.12)配適後的 ANOVA 結果，包含了依序為$X_1$、$X_2$與$X_3$的額外平方和，所以得到(7.15)的檢定統計量值為：

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

在$\alpha = .01$下，$F(.99;1,16) = 8.53 \geq 1.88 = F^*$，故結論無法拒絕$H_0$，亦即在模型引進$X_1$、$X_2$後，預測變數$X_3$可以被忽略掉。

表 7.2d 計算出本案例之$t^*$檢定統計量值為：

$$t^* = \frac{b_3}{s\{b_3\}} = \frac{-2.186}{1.596} = -1.37$$

由於$(t^*)^2 = (-1.37)^2 = 1.88 = F^*$，顯示此二檢定統計量為等價，所以結果與簡單線性迴歸相同。

說明

透過(7.15)的檢定統計量$F^*$來檢定$\beta_3 = 0$，稱為**部份 *F* 檢定**(*partial F test*)統計量，主要在區分(6.39b)中有關於檢定所有的$\beta_k = 0$，後者稱為**整體 *F* 檢定**(*overall F test*)。■

## 檢定多個 $\beta_k=0$

在複迴歸中經常關心的是迴歸模型中的多個項是否可以剔除掉。例如是否可以從全模型（7.12）中剔除掉 $\beta_2 X_2$ 與 $\beta_3 X_3$? 此時檢定虛無假設為：

$$\boxed{\begin{aligned} &H_0 : \beta_2 = \beta_3 = 0 \\ &H_a : \beta_2 \text{ 與 } \beta_3 \text{ 非全為 } 0 \end{aligned}} \tag{7.16}$$

採用一般的線性檢定，在成立下之縮減模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \qquad \text{縮減模型} \tag{7.17}$$

且縮減模型的誤差平方和為

$$SSE（R）= SSE（X_1）$$

誤差平方和之自由度 $df_R = n - 2$。

一般線性檢定之統計量為（2.70）：

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

上式分子中的兩個誤差平方和相減，是另外一個額外平方和，亦即：

$$\boxed{SSE(X_1) - SSE(X_1, X_2, X_3) = SSR(X_2, X_3 | X_1)}$$

檢定之統計量為：

$$\boxed{F^* = \frac{SSR(X_2, X_3 | X_1)}{2} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{MSR(X_2, X_3 | X_1)}{MSE(X_1, X_2, X_3)}} \quad (7.18)$$

又額外平方和SSR（ $X_2, X_3 | X_1$）具有兩個自由度。此 $F^*$ 與 $\boxed{F(1-\alpha, 2, n-4)}$ 比較。

**例題**

在人體脂肪量的案例中，我們想要檢定預測變數大腿周長$X_2$與中臂周長$X_3$是否可以從全模型(7.12)中剔除掉，檢定的虛無假設如(7.16)，額外平方和在表 7.4 中，根據(7.11)我們可以得到：

$$SSR(X_2,X_3|X_1)=SSR(X_2|X_1)+SSR(X_3|X_1,X_2)$$
$$=33.17+11.54=44.71$$

因此檢定之統計量(7.18)可以計算出：

$$F^*=\frac{SSR(X_2,X_3|X_1)}{2}\div MSE(X_1,X_2,X_3)$$
$$=\frac{44.71}{2}\div 6.15=3.63$$

在$\alpha$ = .05 下，$F(.95;2,16)$ = 3.63 = $F^*$，恰好落在決策邊界（此時 $P$- 值 = .05），故需要更進一步之分析，才可以做出是否可以從全模型(7.12)中剔除掉 $X_2$ 與 $X_3$ 之結論。

# 7.3 檢定迴歸係數之總結
Summary of Tests Concerning Regression Coefficients

- **檢定所有的 $\beta_k$ =0**

式子(6.39)的整體$F$檢定是針對反應變數$Y$對於所有的預測變數$X$，是否存在迴歸關係之檢定，檢定的虛無假設為：

$$\boxed{\begin{aligned} &H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \\ &H_a: \beta_k \ (k = 1, \cdots, p-1) \text{ 非全為 } 0 \end{aligned}} \tag{7.21}$$

而檢定的統計量為：

$$\boxed{F^* = \frac{SSR(X_1, \cdots, X_{p-1})}{p-1} \div \frac{SSE(X_1, \cdots, X_{p-1})}{n-p}} \tag{7.22}$$

$$\boxed{= \frac{MSR}{MSE}}$$

- **檢定單一 $\beta_k = 0$**

針對特定迴歸係數$\beta_k = 0$ 的部份*F*檢定，檢定的虛無假設為：

$$\boxed{\begin{aligned} H_0 &: \beta_k = 0 \\ H_a &: \beta_k \neq 0 \end{aligned}} \tag{7.23}$$

而檢定的統計量為：

$$\boxed{F^* = \frac{SSR\left(X_k \middle| X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_{p-1}\right)}{1} \div \frac{SSE\left(X_1, \cdots, X_{p-1}\right)}{n-p}}$$

$$\boxed{= \frac{MSR\left(X_k \middle| X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_{p-1}\right)}{MSE}} \tag{7.24}$$

當$H_o$成立時$F^* \sim F(1, n-p)$，$F^*$越大越不容易拒絕虛無假設$H_o$

- **檢定多個 $\beta_k$ =0**

這是另一種部份$F$檢定，檢定的虛無假設為：

$$\boxed{\begin{aligned} &H_0 : \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0 \\ &H_a : H_0 \text{ 的 } \beta_k \text{ 非全為 } 0 \end{aligned}} \tag{7.26}$$

為了方便表示，把要進行檢定的$(p - q)$個變數移至最後，檢定的統計量為：

$$\boxed{F^* = \frac{SSR\left(X_q,\cdots,X_{p-1} \middle| X_1,\cdots,X_{q-1}\right)}{p-q} \div \frac{SSE\left(X_1,\cdots,X_{p-1}\right)}{n-p}}$$

$$\boxed{= \frac{MSR\left(X_q,\cdots,X_{p-1} \middle| X_1,\cdots,X_{q-1}\right)}{MSE}} \tag{7.27}$$

當$H_o$成立時，F*~F（p-q,n-p），F*越大越容易拒絕虛無假設$H_o$。

- 檢定統計量（7.27）包含了前述的兩種情形。當 q = 1時，即為檢定是否全部的迴歸係數均為零，而當 q = p-1時便成為特定迴歸係數是否為零之檢定。如果統計套裝軟體有提供所需的額外平方和，則

$$
\begin{aligned}
& SSR\left(X_q,\cdots,X_{p-1} \middle| X_1,\cdots,X_{q-1}\right) \\
& \quad = SSR\left(X_q \middle| X_1,\cdots,X_{q-1}\right)+\cdots+SSR\left(X_{p-1} \middle| X_1,\cdots,X_{p-2}\right)
\end{aligned}
\tag{7.28}
$$

## 其他檢定

有時迴歸係數的檢定並不是針對單一或多個 $\beta_k=0$，則無法使用額外平方和，所以一定要分別配適全模型與縮減模型，然後進行一般線性檢定，例如全模型中有三個預測變數$X$：

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\beta_3X_{i3}+\varepsilon_i \quad \text{全模型} \tag{7.30}$$

而想檢定：

$$\boxed{\begin{aligned}H_0&:\beta_1=\beta_2\\ H_a&:\beta_1\neq\beta_2\end{aligned}} \tag{7.31}$$

於是可以先配適全模型，然後配適縮減模型：

$$Y_i=\beta_0+\beta_c\left(X_{i1}+X_{i2}\right)+\beta_3X_{i3}+\varepsilon_i \quad \text{縮減模型} \tag{7.32}$$

其中$\beta_c$表示當$H_0$成立時，$\beta_1=\beta_2=\beta_c$，而$X_{i1}+X_{i2}$對應的新預測變數X，然後可以應用（2.70）的一般F*統計量，自由度為1與（n-4）。

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

- 另外一個額外平方和無法適用的例子為檢定迴歸模型(7.30)中：

$$\boxed{\begin{aligned}&H_0: \beta_1 = 3, \beta_3 = 5\\ &H_a: H_0 \text{ 的等式不全成立}\end{aligned}} \qquad (7.33)$$

縮減模型為：$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$　全模型

$$Y_i - 3X_{i1} - 5X_{i3} = \beta_0 + \beta_2 X_{i2} + \varepsilon_i \quad \text{縮減模型} \qquad (7.34)$$

因為當$H_o$成立時，$\beta_1 X_1$與$\beta_3 X_3$均為已知，所以此縮減模型採用了新的反應變數$Y - 3X_1 - 5X_3$，然後應用（2.70）的一般F*統計量，自由度為 2 與（n-4）。

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

# 7.4 偏判定係數

## Coefficients of Partial Determination

- 額外平方和SSR不僅可以使用複迴歸模型中關於迴歸係數檢定，也可以利用來作為變數間關係程度的指標，稱之為偏判定係數（coefficients of partial determination） 。
- 之前所提的複判定係數$R^2$，是用來衡量當模型引進了整組的X變數，可以降低多少關於Y變數的變異比率，而偏判定係數則是用以衡量當模型引進了整組的X變數後，如果再增加一個特定的X變數，可以降低Y變數變異的邊際貢獻比率。

## 雙預測變數

- 首先考慮（6.1）中的第一階模型：

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

誤差平方和 $SSE(X_2)$ 是可以用來衡量變數 $X_2$ 引進模型中時，對於 $Y$ 變數的變異影響，$SSE(X_1, X_2)$則是用來衡量變數 $X_1$ 與 $X_2$ 引進模型中時，對於 $Y$ 變數的變異影響，因此當變數 $X_2$ 已經在模型中時，變數 $X_1$ 對於 $Y$ 變數的變異之邊際貢獻比率為：

$$\frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1 | X_2)}{SSE(X_2)}$$

此一數值可以視為給定 $X_2$ 已經在模型中時，$Y$ 變數與變數 $X_1$ 之間的偏判定係數，用符號 $R^2_{Y1|2}$ 表示：

$$\boxed{R^2_{Y1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1 | X_2)}{SSE(X_2)}} \qquad (7.35)$$

• 類似的作法可以用來定義給定 $X_1$ 已經在模型中時，$Y$ 變數與變數 $X_2$ 之間的偏判定係數 $R^2_{Y2|1}$ 表示：

$$\boxed{R^2_{Y2|1} = \frac{SSR(X_2|X_1)}{SSE(X_1)}} \tag{7.36}$$

## 一般情形

上述關於偏判定係數的定義可以立即推廣至三個或三個以上$X$變數的模型如：

$$R^2_{Y1|23} = \frac{SSR(X_1|X_2,X_3)}{SSE(X_2,X_3)} \tag{7.37}$$

$$R^2_{Y2|13} = \frac{SSR(X_2|X_1,X_3)}{SSE(X_1,X_3)} \tag{7.38}$$

$$R^2_{Y3|12} = \frac{SSR(X_3|X_1,X_2)}{SSE(X_1,X_2)} \tag{7.39}$$

$$R^2_{Y4|123} = \frac{SSR(X_4|X_1,X_2,X_3)}{SSE(X_1,X_2,X_3)} \tag{7.40}$$

例題

在人體脂肪量的案例中，我們可以計算不同的偏判定係數，例如下面三個例子（參考表 7.2 與表 7.4）：

$$R^2_{Y2|1} = \frac{SSR(X_2|X_1)}{SSE(X_1)} = \frac{33.17}{143.12} = .232$$

$$R^2_{Y3|12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.54}{109.95} = .105$$

$$R^2_{Y1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = \frac{3.47}{113.42} = .031$$

上述結果顯示了當模型中已經存在有變數$X_1$時，繼續引進變數$X_2$，原來的誤差平方和 $SSE(X_1)$可以再度縮減 23.2%，而模型中已經存在變數 $X_1$ 與變數 $X_2$ 時，繼續引進變數 $X_3$，則誤差平方和 $SSE(X_1, X_2)$僅能縮減 10.5%，從另一方面來看，若是模型中先引進變數$X_2$，然後再引進變數$X_1$，則僅能縮減誤差平方和 $SSE(X_2)$的 3.1%。

## 說明

1. 根據定義可以明顯看出偏判定係數的值必然介於 0 與 1 之間。

## 偏相關係數

偏判定係數的平方根稱為偏相關係數（coefficients of partial correlation），其符號的正負與迴歸函數中對應的迴歸係數相同，雖然偏相關係數的意義不如偏判定係數那麼明確，不過在實務中經常使用，特別是在電腦程式中，對於如何找出關於引進迴歸模型預測變數的最佳順序。

**例題** 在人體脂肪量的案例中，我們可以計算：

$$r_{Y2|1} = \sqrt{.232} = .482$$

$$r_{Y3|12} = -\sqrt{.105} = -.324$$

$$r_{Y1|2} = \sqrt{.031} = .176$$

在上面的計算過程中，由於偏相關係數其符號的正負與迴歸函數中對應的迴歸係數相同，根據表 7.2c 我們知道 $b_2 = .6594$ 、 $b_1 = .2224$ ，所以 $r_{Y2|1}$ 與 $r_{Y1|2}$ 的符號均為正，而表 7.2d 中顯示 $b_3 = -2.186$ ，所以 $r_{Y3|12}$ 的符號為負。

# 7.5 標準化複迴歸模型
## Standardized Multiple Regression Model

- 對於（6.7）第一般複迴歸模型其標準化的形式，可以透過控制標準方程式在計算上的捨入誤差，使得估計迴歸係數具有相同單位後，以資比較。 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$

**標準方程式計算捨入誤差**

- 透過標準方程式在計算結果，可能因中間過程的四捨五入而相當地敏感，是有可能發生嚴重的捨入誤差。
- 下面兩種情形是造成對X`X取反矩陣時，具有嚴重捨入誤差的高危險性：（1）X`X的行列式值接近零；（2）X`X內部元素數值差距甚大。
- 標準化迴歸模型的轉換稱為相關轉換（correlation transformation），轉換後的X`X內部元素值均介於±1之間，於是可以使得反矩陣的計算相對於原來的矩陣，比較不會因數字差異過大而發生嚴重的捨入誤差。

## 迴歸係數的不可比較性

對於非標準化迴歸模型(6.7)的第二個缺點是由於單位不同，迴歸係數通常無法直接進行比較，例如：

1. 考慮配適的反應函數如下：　$\hat{Y}=200+20{,}000X_1+0.2X_2$

   可能會以為$X_1$是為一重要的預測變數，$X_2$對於反應變數Y作用很小，事實上進行這種結論必須更加小心，因為並不知道兩種變數的單位是什麼。假設反應變數Y的單位是元，而$X_1$的單位是千元，$X_2$的單位是分，此時若固定$X_2$，則$X_1$每增加一元（一單位），對平均反應的影響效果是兩萬元，其影響效果跟固定$X_1$時，$X_2$每增加一千千元(一百萬單位)之效果相同。

2. 圖 6.5 Dwaine工作室的例子中，因$X_1$是指十六歲以下的人口數(單位是千人)，而$X_2$是平均每人可支配所得(單位是千元)，所以$b_1$與$b_2$完全不能比較。

# 相關轉換

- 應用相關轉換有助對於控制捨入誤差，同時由於各迴歸變數之同單位化後，其迴歸係數也因此可以進行比較。
- 相關轉換是對於變數標準化做法再進行一個簡單的修改，對於變數的標準化做法是藉由置中(centering)與定尺度(scaling)兩種程序來完成的。
- 置中是取每一個觀測值與該變數的平均值之差，而定尺度則是將已經置中過後的觀測值，轉化為以該變數之標準差為單位，因此一般反應變數 $Y$ 與預測變數 $X_1, \ldots, X_{p-1}$ 其標準化後如下：

$$\boxed{\frac{Y_i - \bar{Y}}{s_Y}} \qquad (7.43a)$$

$$\boxed{\frac{X_{ik} - \bar{X}_k}{s_k} \quad (k = 1, \cdots, p-1)} \qquad (7.43b)$$

其中$\bar{Y}$與$\bar{X}_k$分別為$Y$與$X_k$觀測值的平均，而$S_Y$與$S_k$則分別是其標準差，定義如下：

$$\frac{Y_i-\bar{Y}}{s_Y} \quad \frac{X_{ik}-\bar{X}_k}{s_k} \quad (k=1,\cdots,p-1)$$

$$\boxed{s_Y=\sqrt{\frac{\sum_i (Y_i-\bar{Y})^2}{n-1}}} \quad (7.43c) \qquad \boxed{s_k=\sqrt{\frac{\sum_i (X_{ik}-\bar{X}_k)^2}{n-1}} \quad (k=1,\cdots,p-1)} \quad (7.43d)$$

於是相關轉換可以藉由(7.43a,b)的標準化，簡單修改為：

$$\boxed{Y_i^*=\frac{1}{\sqrt{n-1}}\left(\frac{Y_i-\bar{Y}}{s_Y}\right)} \qquad (7.44a)$$

$$\boxed{X_{ik}^*=\frac{1}{\sqrt{n-1}}\left(\frac{X_{ik}-\bar{X}_k}{s_k}\right) \quad (k=1,\cdots,p-1)} \qquad (7.44b)$$

- 標準化迴歸模型

透過(7.44)的相關轉換後，進行配適的迴歸模型稱為標準化迴歸模型：

$$\boxed{Y_i^*=\beta_1^* X_{i1}^*+\cdots+\beta_{p-1}^* X_{i,p-1}^*+\varepsilon_i^*} \qquad (7.45)$$

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

- 因為在（7.45）的標準化迴歸模型下，透過最小平方估計量或是最大概似估計量，所得到的截距項參數均為零，所以（7.45）的標準化迴歸模型並沒有截距項。
- 標準化迴歸模型的參數$\beta_1^*,\ldots,\beta_{p-1}^*$與一般複迴歸模型(6.7)之參數$\beta_0, \beta_1,\ldots,\beta_{p-1}$有下列的關係：

$$\boxed{\beta_k = \left(\frac{s_Y}{s_k}\right)\beta_k^* \quad (k = 1,\cdots,p-1)} \qquad (7.46a)$$

$$\boxed{\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \cdots - \beta_{p-1}\bar{X}_{p-1}} \qquad (7.46b)$$

所以標準化迴歸模型的參數$\beta_k^*$，與原迴歸係數$\beta_k$，k＝1,…p-1，兩者間的關係只差一個標準差比例 (7.46a) 的尺度因子

# 轉換變數後的 X'X 矩陣

- 為了研究經過相關轉換後的**X'X** 與最小平方標準方程式的特殊性質，可以將(6.67)中有關反應變數與預測變數兩兩相關係數之相關矩陣分解為二：

1. 第一個矩陣以$\mathbf{r}_{\mathbf{xx}}$表示，**稱為*X*變數的相關矩陣**（correlation matrix），組成元素為所有*X*變數兩兩間的簡單相關係數，其定義如下：

$$\underset{(p-1)\times(p-1)}{\mathbf{r}_{XX}} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{21} & 1 & \cdots & r_{2,p-1} \\ \vdots & \vdots & & \vdots \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix} \tag{7.47}$$

這裡的$r_{12}$仍然代表變數$X_1$與$X_2$簡單相關係數，其餘以此類推，因為各變數與本身自己的相關係數都是1，所以上述矩陣之主要對角線元素都是1，又因為$r_{kk`} = r_{k`k}$，所以矩陣$r_{xx}$為一個對稱性。

2.第二個矩陣以$\mathbf{r}_{\mathbf{YX}}$表示，是由反應變數$Y$與每一個$X$變數間的簡單相關係數(一樣用符號 $\mathrm{r}_{\mathrm{Y1}}$，$\mathrm{r}_{\mathrm{Y2}}$等等表示)所組成的向量：

$$\boxed{\underset{(p-1)\times 1}{\mathbf{r}_{YX}} = \begin{bmatrix} r_{Y1} \\ r_{Y2} \\ \vdots \\ r_{Y,p-1} \end{bmatrix}} \tag{7.48}$$

現在考慮 (7.45) $Y_i^* = \beta_1^* X_{i1}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^*$ 的標準化迴歸模型中經過轉換變數的 **X'X** 矩陣，此時的 **X** 矩陣為：

$$\underset{n\times(p-1)}{\mathbf{X}} = \begin{bmatrix} X_{11}^* & \cdots & X_{1,p-1}^* \\ X_{21}^* & \cdots & X_{2,p-1}^* \\ \vdots & & \vdots \\ X_{n1}^* & \cdots & X_{n,p-1}^* \end{bmatrix} \tag{7.49}$$

由於(7.45)標準化迴歸模型並沒有截距項，因此在 **X** 矩陣中並沒有全為 1 的行，

- 同時可以經證明而得到轉換變數後的X`X 就是(7.47)中的相關矩陣，亦即

$$\boxed{\underset{(p-1)\times(p-1)}{\mathbf{X}'\mathbf{X}} = \mathbf{r}_{XX}} \qquad (7.50)$$

因為轉換變數後的X`X矩陣是由X變數間的簡單相關係數所組成，所有元素均介於 +1~-1 之間，故具有相同大小之等級，如前所述，這對於計算X`X反矩陣時，捨入誤差的結果很有幫助。

## 標準化迴歸係數的估計

一般複迴歸模型的最小平方標準方程式(6.24)：$\mathbf{X'Xb=X'Y}$

與最小平方估計量(6.25)：$\mathbf{b}=\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y}$

對於經轉換過後之變數，可以證明為：

$$\boxed{\underset{(p-1)\times(p-1)}{\mathbf{X'X}} = \mathbf{r}_{XX}} \quad (7.50) \qquad \boxed{\underset{(p-1)\times 1}{\mathbf{X'Y}} = \mathbf{r}_{YX}} \quad (7.51)$$

其中$\mathbf{r}_{YX}$之定義如(7.48)，是由反應變數Y與每一X變數的簡單相關係數所組成的向量，根據（7.50）與（7.51）可以得到標準化迴歸模型（7.45）的最小平方標準方程式，以及迴歸係數之估計量如下：

$$\boxed{Y_i^* = \beta_1^* X_{i1}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^*} \quad (7.45)$$

$$\mathbf{r}_{XX}\mathbf{b} = \mathbf{r}_{YX} \qquad (7.52a)$$

$$\boxed{\mathbf{b} = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX}} \qquad (7.52b)$$

$$\boxed{\mathbf{b} = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX}}$$

其中，

$$\underset{(p-1)\times 1}{\mathbf{b}} = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_{p-1}^* \end{bmatrix} \tag{7.52c}$$

而 $b_1^*, \ldots, b_{p-1}^*$ 迴歸係數習慣上稱之為 標準化迴歸係數（standardized regression coefficients）。

- 藉由下面的關係式可以轉換回原變數下的迴歸模型(6.7)中估計之迴歸係數：

$$\boxed{b_k = \left(\frac{s_Y}{s_k}\right) b_k^* \quad (k = 1, \cdots, p-1)} \tag{7.53a}$$

$$\boxed{b_0 = \bar{Y} - b_1\bar{X}_1 - \cdots - b_{p-1}\bar{X}_{p-1}} \tag{7.53b}$$

表 7.5a 重複了圖 6.5b 中有關 Dwaine 工作室例子的部份資料，而表 7.5b 是根據(7.44)之相關轉換後的部份資料，我們以第一筆資料來說明轉換的計算程序。使用表 7.5a 的平均數與標準差：

例題

**表 7.5**

相關轉換以及配適的標準化迴歸模型—Dwaine 工作室案例。

(a) 原始資料

| 個案 $i$ | 營業額 $Y_i$ | 目標人口 $X_{i1}$ | 平均每人可支配所得 $X_{i2}$ |
|---|---|---|---|
| 1 | 174.4 | 68.5 | 16.7 |
| 2 | 164.4 | 45.2 | 16.8 |
| ... | ... | ... | ... |
| 20 | 224.1 | 82.7 | 19.1 |
| 21 | 166.5 | 52.3 | 16.0 |
| | $\bar{Y} = 181.90$ | $\bar{X}_1 = 62.019$ | $\bar{X}_2 = 17.143$ |
| | $s_Y = 36.191$ | $s_1 = 18.620$ | $s_2 = .97035$ |

對 $i=1$

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i-\bar{Y}}{s_Y}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{174.4-181.90}{36.191}\right) = -.04634$$

$$X_{11}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{11}-\bar{X}_1}{s_1}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{68.5-62.019}{18.620}\right) = .07783$$

$$X_{12}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{12}-\bar{X}_2}{s_2}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{16.7-17.143}{.97035}\right) = -.10208$$

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{s_Y}\right) \qquad X_{11}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{11} - \bar{X}_1}{s_1}\right)$$

$$= \frac{1}{\sqrt{21-1}}\left(\frac{174.4-181.90}{36.191}\right) \qquad = \frac{1}{\sqrt{21-1}}\left(\frac{68.5-62.019}{18.620}\right)$$

$$= -.04634 \qquad = .07783$$

$$X_{12}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{12} - \bar{X}_2}{s_2}\right) = \frac{1}{\sqrt{21-1}}\left(\frac{16.7-17.143}{.97035}\right) = -.10208$$

**表 7.5**
相關轉換以及配適的標準化迴歸模型—Dwaine 工作室案例。

(b) 轉換後資料

| $i$ | $Y_i^*$ | $X_{i1}^*$ | $X_{i2}^*$ |
|---|---|---|---|
| 1 | −.04637 | .07783 | −.10205 |
| 2 | −.10815 | −.20198 | −.07901 |
| ... | ... | ... | ... |
| 20 | .26070 | .24835 | .45100 |
| 21 | −.09518 | −.11671 | −.26336 |

(c) 配適的標準化模型

$$\boxed{\hat{Y}^* = .7484\, X_1^* + .2511\, X_2^*}$$

對於資料經轉換後配適標準化迴歸模型(7.45)之結果在表 7.5c 中：

$$\hat{Y}^* = .7484X_1^* + .2511X_2^*$$

標準化迴歸係數 $b_1^* = .7484$、$b_2^* = .2511$，統計套裝軟體 SYSTAT 的輸出在圖 6.5a 中標示地方於 STD COEF 的位置，根據標準化迴歸係數可以知道當 $X_2$（平均每人可支配所得）固定時，每增加一個標準差的 $X_1$（目標人口），會造成期望營業額（以 $Y$ 的標準差為單位）的增加量，將遠大於固定 $X_1$，而增加一個標準差的 $X_2$ 之量。

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T | P(2 TAIL) |
|---|---|---|---|---|---|---|
| CONSTANT | -68.8571 | 60.0170 | 0.0000 | . | -1.1473 | 0.2663 |
| TARGTPOP | 1.4546 | 0.2118 | 0.7484 | 0.3896 | 6.8682 | 0.0000 |
| DISPOINC | 9.3655 | 4.0640 | 0.2511 | 0.3896 | 2.3045 | 0.0333 |

**表 7.5** 相關轉換以及配適的標準化迴歸模型—Dwaine 工作室案例。

(a) 原始資料

| 個案 $i$ | 營業額 $Y_i$ | 目標人口 $X_{i1}$ | 平均每人可支配所得 $X_{i2}$ |
|---|---|---|---|
| 1 | 174.4 | 68.5 | 16.7 |
| 2 | 164.4 | 45.2 | 16.8 |
| ⋯ | ⋯ | ⋯ | ⋯ |
| 20 | 224.1 | 82.7 | 19.1 |
| 21 | 166.5 | 52.3 | 16.0 |
| | $\bar{Y} = 181.90$ | $\bar{X}_1 = 62.019$ | $\bar{X}_2 = 17.143$ |
| | $s_Y = 36.191$ | $s_1 = 18.620$ | $s_2 = .97035$ |

(b) 轉換後資料

| $i$ | $Y_i^*$ | $X_{i1}^*$ | $X_{i2}^*$ |
|---|---|---|---|
| 1 | −.04637 | .07783 | −.10205 |
| 2 | −.10815 | −.20198 | −.07901 |
| ⋯ | ⋯ | ⋯ | ⋯ |
| 20 | .26070 | .24835 | .45100 |
| 21 | −.09518 | −.11671 | −.26336 |

(c) 配適的標準化模型

$$\hat{Y}^* = .7484\ X_1^* + .2511\ X_2^*$$

如果要將標準化迴歸係數$b_1^*$與$b_2^*$轉換回原變數下模型之迴歸係數，可以利用(7.53)。使用表 7.5 的資料，我們有以下之結果：

$$b_1 = \left(\frac{s_Y}{s_1}\right)b_1^* = \frac{36.191}{18.620}(.7484) = 1.4546$$

$$b_2 = \left(\frac{s_Y}{s_2}\right)b_2^* = \frac{36.191}{.97035}(.2511) = 9.3652$$

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 = 181.90 - 1.4546(62.019) - 9.3652(17.143) = -68.860$$

故原變數下的複迴歸模型之估計迴歸函數為：

$$\boxed{\hat{Y} = -68.860 + 1.4546X_1 + 9.3652X_2}$$

此時 $b_1$ 與 $b_2$ 不能比較，因為 $X_1$ 單位是千人，而 $X_2$ 是以千元為單位。

| 個案 $i$ | 營業額 $Y_i$ | 目標人口 $X_{i1}$ | 平均每人可支配所得 $X_{i2}$ |
|---|---|---|---|
| 1 | 174.4 | 68.5 | 16.7 |
| 2 | 164.4 | 45.2 | 16.8 |
| ... | ... | ... | ... |
| 20 | 224.1 | 82.7 | 19.1 |
| 21 | 166.5 | 52.3 | 16.0 |
| | $\bar{Y}$ = 181.90 | $\bar{X}_1$ = 62.019 | $\bar{X}_2$ = 17.143 |
| | $s_Y$ = 36.191 | $s_1$ = 18.620 | $s_2$ = .97035 |

# 7.6 多重共線性及其效果

Multicollnearity and Its Effects

- 在複迴歸分析中，有關於預測變數與反應變數間的關係以及顯著性是最受關心的部份，而常見的問題：

1. 不同預測變數其效應的相對重要性如何？
2. 某一預測變數對反應變數之效應大小如何？
3. 是否存在有預測變數對反應變數之效應太小而可以從模型中將它剔除？
4. 是否有某一預測變數目前不在模型內，而有必要考慮納入模型中？

- 然而，在許多商業、經濟、社會與生物科學方面的非實驗情形，預測變數彼此間都有相關性，同時某些目前不在模型中，卻跟反應變數具有相關的預測變數，也同時跟原先的預測變數彼此間都有相關性。
- 當預測變數彼此間都有相關性時，稱為其中存在有**交互相關（intercorrelation）或多重共線性（multicollinearity）**
- 多重共線性經常是特別針對預測變數彼此間相關程度很高的情形，在此將討論多種由預測變數間的多重共線性而引發的相關問題

# 預測變數無相關

- 表 7.6 為工作團隊的規模（$X_1$）、獎金水準（$X_2$）對於（Y）效應的小型實驗資料。預測變數 $X_1$ 與 $X_2$ 無關，亦即 $r_{12}^2=0$，其中 $r_{12}^2$ 表示$X_1$與$X_2$間的簡單判定係數。

**表 7.6**
無相關預測變數－工作團隊之案例。

| 個案 $i$ | 團隊規模 $X_{i1}$ | 獎金（元）$X_{i2}$ | 團隊生產力 $Y_i$ |
|---|---|---|---|
| 1 | 4 | 2 | 42 |
| 2 | 4 | 2 | 39 |
| 3 | 4 | 3 | 48 |
| 4 | 4 | 3 | 51 |
| 5 | 6 | 2 | 49 |
| 6 | 6 | 2 | 53 |
| 7 | 6 | 3 | 61 |
| 8 | 6 | 3 | 60 |

- 表 7.7a則配適了迴歸函數，以及$X_1$與$X_2$同時在模型中的變異數分析表。
- 表 7.7b只有預測變數X1在模型中。
- 表 7.7c只有預測變數X2在模型中。

**表 7.7**

無相關預測變數之 迴歸結果－工作團 隊之案例。

(a) $Y$ 對$X_1$與$X_2$之迴歸

$$\hat{Y} = .375 + 5.375 X_1 + 9.250 X_2$$

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | 402.250 | 2 | 201.125 |
| 誤差 | 17.625 | 5 | 3.525 |
| 總合 | 419.875 | 7 | |

(b) $Y$ 對$X_1$之迴歸

$$\hat{Y} = 23.500 + 5.375 X_1$$

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | 231.125 | 1 | 231.125 |
| 誤差 | 188.750 | 6 | 31.458 |
| 總合 | 419.875 | 7 | |

(c) $Y$ 對$X_2$之迴歸

$$\hat{Y} = 27.250 + 9.250 X_2$$

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | 171.125 | 1 | 171.125 |
| 誤差 | 248.750 | 6 | 41.458 |
| 總合 | 419.875 | 7 | |

- 值得注意的是表 7.7中$X_1$迴歸係數$b_1$=5.375，此一結果不論$X_2$是否在模型中都一樣，而同樣的現象是$b_2$=9.250，會發生此一現象的原因為兩個預測變數$X_1$與$X_2$無相關性所致。

(a) $Y$對$X_1$與$X_2$之迴歸

$$\hat{Y} = .375 \boxed{+\ 5.375\ X_1} \boxed{+\ 9.250\ X_2}$$

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | 402.250 | 2 | 201.125 |
| 誤差 | 17.625 | 5 | 3.525 |
| 總合 | 419.875 | 7 | |

(b) $Y$對$X_1$之迴歸

$$\hat{Y} = 23.500 \boxed{+\ 5.375\ X_1}$$

| 變異來源 | 平方和 | 自由度 | 均方 |
|---|---|---|---|
| 迴歸 | 231.125 | 1 | 231.125 |
| 誤差 | 188.750 | 6 | 31.458 |
| 總合 | 419.875 | 7 | |

(c) $Y$對$X_2$之迴歸

$$\hat{Y} = 27.250 \boxed{+\ 9.250\ X_2}$$

另外關於誤差平方和部份，表 7.7中額外平方和SSR（$X_1|X_2$）恰等於只有$X_1$在模型中時的迴歸平方和SSR（$X_1$）：

$$SSR(X_1|X_2)=SSE(X_2)-SSE(X_1,X_2)$$
$$=248.750-17.625=231.125$$

$$SSR(X_1)=231.125$$

額外平方和SSR（$X_2|X_1$）恰等於只有$X_2$在模型中時的迴歸平方和SSR（$X_2$）：

$$SSR(X_2|X_1)=SSE(X_1)-SSE(X_1,X_2)$$
$$=188.750-17.625=171.125$$

$$SSR(X_2)=171.125$$

當兩個或兩個以上的預測變數無相關性時，某一預測變數在其他預測變數已經納入模型下，所能對降低誤差平方和之邊際貢獻，會等同於該預測變數單獨在模型中時的貢獻。

# 預測變數完全相關下的問題本質

在此假設當兩個預測變數為完全相關的情形，表7.8為一個反應變數與兩個預測變數下，收集四個樣本觀測值的例子，A先生配適了一個第一階複迴歸函數：$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ (7.57)

A先生配適後之結果如下：$\hat{Y} = -87 + X_1 + 18X_2$ (7.58)

在表7.8中列出了配適值。

表 **7.8**
完全相關預測變數之例。

| 個案 $i$ | $X_{i1}$ | $X_{i2}$ | $Y_i$ | 迴歸函數配適值 (7.58) | 迴歸函數配適值 (7.59) |
|---|---|---|---|---|---|
| 1 | 2 | 6 | 23 | 23 | 23 |
| 2 | 8 | 9 | 83 | 83 | 83 |
| 3 | 6 | 8 | 63 | 63 | 63 |
| 4 | 10 | 10 | 103 | 103 | 103 |

反應函數：

$$\hat{Y} = -87 + X_1 + 18X_2 \quad (7.58)$$
$$\hat{Y} = -7 + 9X_1 + 2X_2 \quad (7.59)$$

- 而同一時間B小姐也針對同樣的資料配適了(5.57)的第一階複迴歸函數，卻得到結果如下： $\hat{Y} = -7 + 9X_1 + 2X_2$ (7.59)
- 事實上有無窮多個可能的反應函數可以完全配適，其原因是兩個預測變數具有下列的完全相關性： $X_2 = 5 + .5X_1$ (7.60)
- 配適的反應函數（7.58）與（7.59）是全然不同的反應曲面，如圖7.2所示而兩個反應函數只在其相交處具有相同的配適值，其條件正是滿足(7.60)。

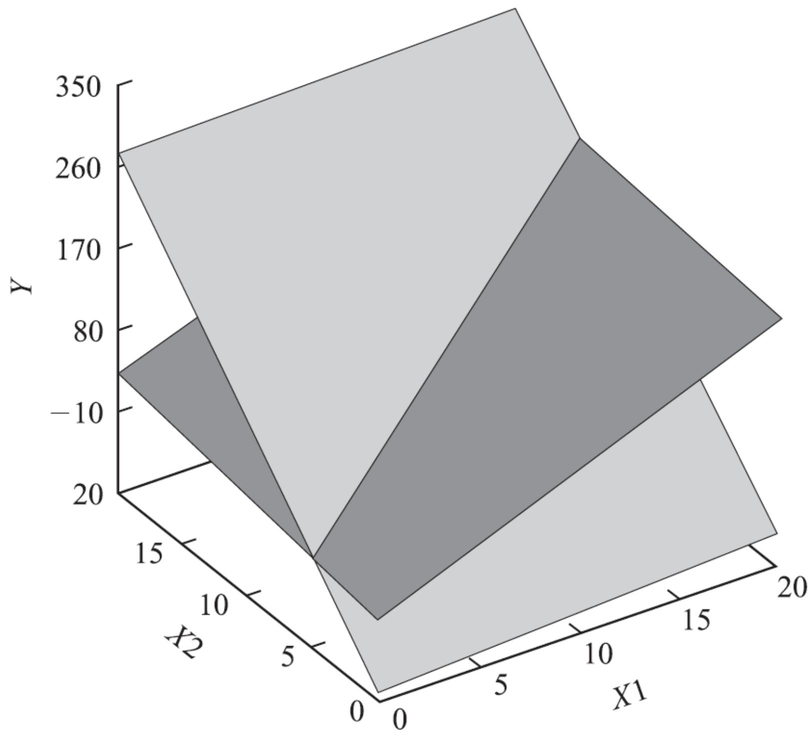



**圖 7.2**
兩個反應平面在 $X_2 = 5 + .5X_1$ 處相交。

- 因此若是像上述的例子一樣，兩個預測變數為完全相關的情形，而資料中無任何誤差成分存在，則有無窮多個可能的反應函數可以完全配適資料，且當（$X_1, X_2$）之組合恰巧滿足$X_1$與$X_2$之關係時，將會得到相同的配適值，不過卻是在不同的反應曲面上，而且當（$X_1, X_2$）之組合不滿足$X_1$與$X_2$之關係時，得到不同的配適值。

上例中隱含了兩個重點：

1. 當$X_1, X_2$為完全相關時，對於資料做出良好配適之能力並不會受到影響。

2. 由於許多不同的反應函數可能都具有良好的配適，而無法將任何一組迴歸係數視為反應不同預測變數的效應，所以在（7.58）中$b_1 = 1$而$b_2 = 18$並不表示$X_2$就是主要的預測變數，而$X_1$是次要角色，因反應函數（7.59）提供了同樣好的配適，但是其迴歸係數的大小卻恰好相反。

## 多重共線性效果

1. 一般而言，當預測變數中的某一部份，或是全部的預測變數存在相關性時，只要推論或預測範圍是保持在觀測區域內，則上述問題並不會妨礙到配適出一個適當的反應函數，也不會影響對於平均反應的推論或是新觀測值的預測
2. 前述案例中關於同樣有良好的配適，卻存在有不同形式的迴歸函數，在實務上的意義是，如果預測變數間有高度相關性，則所估計的迴歸函數將有很大的變異，也就是每一組樣本所估計出來的迴歸係數將會差異很大，也因此而可能對於真實的迴歸係數，較有可能得到不精確的訊息。
3. 通常將迴歸係數解釋為在其他預測變數固定的情形下，對應的預測變數每增加一各單位，會造成反應變數期望值的改變量。不過在多重共線性存在時，上述解釋就不完全適合了。

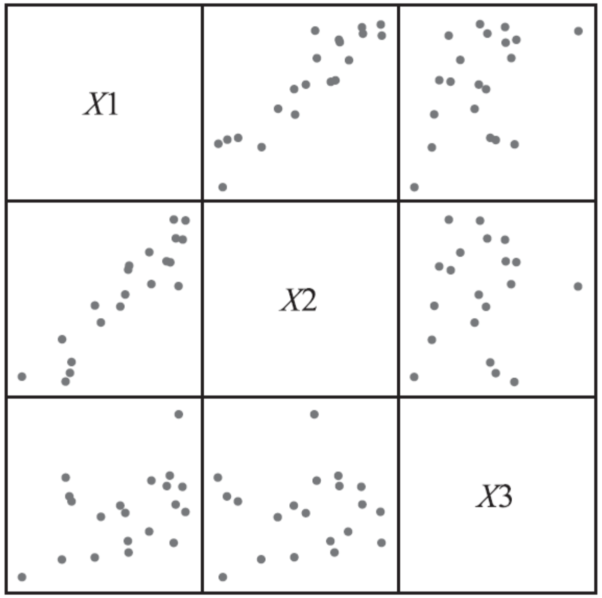
3a.雖然在觀念上是指維持其他預測變數的固定，而僅僅只變動某一個預測變數，然而對於高相關性的預測變數而言，實務上是很難做到的。

例如，當我們考慮雨量與日照時數對於農作物產量的影響時，由於迴歸模型中的兩個預測變數間之高相關性，因而很難控制成為固定某一個預測變數，而僅僅改變特定的預測變數，因此，對於高相關性的預測變數而言，不能簡單地將迴歸係數解釋為邊際效應的測量。

- 利用人體脂肪量的例子來說明多重共線性效果，表7.1是部份基礎資料，表7.2是迴歸程序的結果，圖7.3則是關於預測變數的散佈圖矩陣以及相關矩陣，

**圖 7.3**
預測變數的散佈圖矩陣及相關矩陣－脂肪量案例。

(a) *X* 變數的散佈圖矩陣




(b) *X* 變數的相關矩陣

$$r_{XX} = \begin{bmatrix} 1.0 & .924 & .458 \\ .924 & 1.0 & .085 \\ .458 & .085 & 1.0 \end{bmatrix}$$

從散佈圖矩陣中可以看出預測變數$X_1$與$X_2$的高相關性，同時相關矩陣也顯示出 $r_{12}=0.924$，從另一方面來看，$X_3$與$X_1$或$X_3$與$X_2$的相關性則沒有很高，而相關也顯示出$r_{13}=0.458$、$r_{23}=0.085$。（不過如果將$X_1$與$X_2$聯合起來，與$X_3$的複判定係數則高達.998。）

# 對迴歸係數的效果

- 對迴歸係數的效果 從表7.2知道三頭肌外表厚度 $X_1$ 的迴歸係數，明顯受到模型中包含了哪些預測變數而不同：

| 模型中的變數 | $b_1$ | $b_2$ |
|---|---|---|
| $X_1$ | 0.8572 | — |
| $X_2$ | — | 0.8565 |
| $X_1, X_2$ | 0.2224 | 0.6594 |
| $X_1, X_2, X_3$ | 4.334 | -2.857 |

- 上述問題同樣也發生在 $X_2$的迴歸係數，事實上當$X_3$加入了同時含有$X_1$與$X_2$的模型時，$b_2$甚至會改變正負符號。

## 對額外平方和的效果

- 當預測變數間存在有相關性時，任何一個預測變數被引進模型中時，對於降低誤差平方和的邊際貢獻，將會因模型中有哪些預測變數而不同，例如表7.2所提供$X_1$的額外平方和：

  $$SSR(X_1)=352.27$$

  $$SSR(X_1|X_2)=3.47$$

- 由於$X_1$與$X_2$間以及分別對反應變數均存在有高度相關性，以至於 $SSR(X_1|X_2)$ 相對於 $SSR(X_1)$ 顯得很小，因此當$X_2$已經存在模型中時，則加入預測變數$X_1$，對於降低誤差平方和的邊際貢獻將會相對地小很多，這是由於$X_2$已經提供了許多$X_1$所能提供的訊息。相同的情形出現在表7.2中的 $X_2$，此時 $SSR(X_2| X_1)＝33.17$，相對於$SSR(X_2)=381.97$顯得很小。

## 對 $s\{b_k\}$ 的效果

- 表7.2中人體脂肪量的例子，看到迴歸係數 $b_1$ 與 $b_2$ 如何隨著預測變數的增加而更不精確(標準差數大)：

| 模型中的變數 | $s\{b_1\}$ | $s\{b_2\}$ |
|---|---|---|
| $X_1$ | 0.1288 | — |
| $X_2$ | — | 0.1100 |
| $X_1, X_2$ | 0.3034 | 0.2912 |
| $X_1, X_2, X_3$ | 3.016 | 2.582 |

- 這是由於預測變數間有著高度的多重共線性，因而擴大了迴歸係數估計的變異性。

## 對配適值及預測值的效果

表7.2人體脂肪量的例子，注意到隨著模型中加入了新的變數後，預測變數間多重共線性，並未妨礙衡量誤差項變異性的均方誤差，呈現穩定下降的趨勢：

| 模型中的變數 | MSE |
|---|---|
| $X_1$ | 7.95 |
| $X_{1.}$ $X_2$ | 6.47 |
| $X_{1,}$ $X_{2,}$ $X_3$ | 6.15 |

- 此外，只要在預測變數的觀測範圍內，關於配適值的精確度，並不會因為模型中加入了具有相關性的預測變數，受到影響。

- 在人體脂肪量的例子中，考慮只採用三頭肌外表厚度（$X_1$）作為預測變數，在此對於 $X_{h1}=25.0$ 下人體脂肪量配適值的估計，以及估計的標準差分別為：

$$\hat{Y}_h=19.36，s\{\hat{Y}_h\}=0.632$$

當具有高度相關性的預測變數也一併加入模型中時，例如大腿周長（$X_2$），對應 $X_{h1}=25.0$、$X_{h2}=50.0$ 下，人體脂肪量配適值地估計，以及估計地標準差分別為：

$$\hat{Y}_h=19.36，s\{\hat{Y}_h\}=0.624$$

- 加入的預測變數與原先的預測變數具有高度相關性，估計的平均反映其精確度，跟原來的模型相差不大， 如果三個預測變數一併加入模型中時，對應 $X_{h1}=25.0$、$X_{h2}=50.0$與 $X_{h3}=29.0$ 下，人體脂肪量配適值的估計，以及估計的標準差分別為：

$$\hat{Y}_h=19.19，s\{\hat{Y}_h\}=0.621$$

- 因此，在模型中加入了具有相關性的第三個預測變數，也不會使預測值精確度受到影響。

## 對 $\beta_k$ 同時檢定的效果

- 在複迴歸的模型中，有一個經常被誤用的分析就是對於每一個迴歸係數，透過（6.15b）的 t*統計量：

$$t^* = \frac{b_k}{s\{b_k\}}$$

逐一檢定 $\beta_k=0$，k=1,….p-1。就算是推論的程序同時進行，如果預測變數間具有高度相關性，則推論還是可能有問題。

1. 若是多重共線性情形嚴重，就會使X`X行列式值接近零，因此當發生多重共線性情形嚴重時，應該於標準方程式的計算中，先特別進行過（7.44）變數相關轉換 。
2. 預測變數間的高度相關性，導致估計的迴歸係數不精確。
3. 當$X_1$與$X_2$相關性提高時，變異數也會隨之提高。
4. 當某一組預測變數與反應變數間具有相關性，但是關於個別迴歸係數的檢定，可能會因預測變數間多重共線性的問題，而得到迴歸係數均為零的結論。

## 對於多重共線性所需的更有效診斷方法

- 當預測變數間的多重共線性的問題，在解釋上使用迴歸模型的配適時，會產生重要的作用，通常使用預測變數間的成對簡單相關係數，可以有助於我們辨識多重共線性，不過有時嚴重的多重共線性的問題卻未能在成對簡單相關係數中表現出來。
- 於10章中介紹一種更為有效的工具（變異數膨脹因子$VIF_k = 1/(1-R_k^2) < 10$），以助於我們便是嚴重的多重共線性。