

$$\hat{\sigma}^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-3} = \frac{\sum e_i^2}{n-3} = \frac{SSE}{n-3} = MSE \quad (\text{不偏估計式})$$

$\hat{\sigma} = \sqrt{MSE}$ 稱為迴歸標準誤 (standard error of the regression)

證明：以下使用有常態母體假設下之證明方法

$$\begin{aligned} \frac{(n)\hat{\sigma}_{MLE}^2}{\sigma^2} &= \frac{\sum (Y_i - \hat{Y}_i)^2}{\sigma^2} = \frac{\sum e_i^2}{\sigma^2} \sim \chi^2_{(n-3)} \\ E\left(\frac{(n)\hat{\sigma}_{MLE}^2}{\sigma^2}\right) &= E\left(\frac{\sum (Y_i - \hat{Y}_i)^2}{\sigma^2}\right) = E\left(\frac{\sum e_i^2}{\sigma^2}\right) = n-3 \\ \Rightarrow E\left(\frac{\sum e_i^2}{n-3}\right) &= E(MSE) = \sigma^2 \end{aligned}$$

<註> 在 $\varepsilon_i \sim N$ 下， $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ 與 $e_i, \hat{\sigma}^2 = MSE$ 相互獨立。

<補充> 以下各點證明用到正規方程式可得

1. $\bar{\hat{Y}} = \bar{Y}$
2. $\sum e_i = 0$
3. e_i 與 \hat{Y}_i 無關 ($\sum e_i \hat{Y}_i = 0$)
4. e_i 與 X_{1i}, X_{2i} 無關 ($\sum e_i X_{1i} = 0$ 與 $\sum e_i X_{2i} = 0$)

第三節 偏相關係數與複判定係數

一、偏相關係數 (partial correlation coefficient)

在本章包含兩自變數之複迴歸模型中，研究者可以得到以下三個相關係數：

- r_{Y1} : Y 與 X_1 之相關係數
- r_{Y2} : Y 與 X_2 之相關係數
- r_{12} : X_1 與 X_2 之相關係數

問題：

若 X_2 與 Y 及 X_1 有相關時， r_{Y1} 就無法真實地反映出 Y 與 X_1 之關係，同理地，若 X_1 與 Y 及 X_2 有相關時， r_{Y2} 就無法真實地反映出 Y 與 X_2 之關係，因此統計學家發展出偏相關係數。

偏相關係數之意義：

把 X_2 對 Y 及 X_1 之影響淨化，從中求出一個能夠真實地反映出 Y 對 X_1 之相關係數，稱為偏相關係數 $r_{Y1\cdot 2}$ 。把 X_1 對 Y 及 X_2 之影響淨化，從中求出一個能夠真實地反映出 Y 對 X_2 之相關係數，稱為偏相關係數 $r_{Y2\cdot 1}$ 。

接下來，就要介紹如何計算偏相關係數：

(一) 利用淨化影響之觀念去推導偏相關係數

前述 $e = \hat{a}_0 + \hat{a}_1\eta$ 已淨化 X_2 對 Y 及 X_1 之影響

$$\text{因此 } r_{Y1\cdot 2} = r_{e\eta} = \frac{\sum (e_i - \bar{e})(\eta_i - \bar{\eta})}{\sqrt{\sum (e_i - \bar{e})^2} \sqrt{\sum (\eta_i - \bar{\eta})^2}} = \frac{\sum e_i \eta_i}{\sqrt{\sum e_i^2} \sqrt{\sum \eta_i^2}}$$

但實際上，研究者也可先計算出兩兩之相關係數後，再利用以下公式計算：

$$r_{Y1\cdot 2} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{(1-r_{Y2}^2)(1-r_{12}^2)}}, \quad r_{Y2\cdot 1} = \frac{r_{Y2} - r_{Y1}r_{12}}{\sqrt{(1-r_{Y1}^2)(1-r_{12}^2)}},$$

$$r_{12\cdot Y} = \frac{r_{12} - r_{Y1}r_{Y2}}{\sqrt{(1-r_{Y1}^2)(1-r_{Y2}^2)}}$$

以上都是固定一個變數下求出來之偏相關係數，在統計上稱為一階偏相關係數。（固定 r 個變數，就稱 r 階）

(二) 利用偏平方和之觀念去推導偏相關係數（*常考）

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \qquad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

$$SSR(X_1) \qquad SSR(X_1, X_2)$$

$$SSE(X_1) \qquad SSE(X_1, X_2)$$

$$SST(X_1) \qquad SST(X_1, X_2)$$

其中：(1) $SST(X_1) = SST(X_1, X_2) = SS_Y$

(2) $SSR(X_1, X_2) \geq SSR(X_1)$ （當 X_1, X_2 無關時等號成立）

(3) $SSE(X_1, X_2) \leq SSE(X_1)$

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

因此， $r_{Y1\cdot 2}^2 = \frac{SSR(X_1 | X_2)}{SSE(X_2)}$ ，同理地， $r_{Y2\cdot 1}^2 = \frac{SSR(X_2 | X_1)}{SSE(X_1)}$

推廣至二階偏相關係數，例： $r_{Y12\cdot 34}^2 = \frac{SSR(X_1, X_2 | X_3, X_4)}{SSE(X_3, X_4)}$

- <註> 1. 總變異(SST)與自變數之個數無關
 2. 當 X_1, X_2 無關時, $SSR(X_1, X_2) = SSR(X_1)$ 與 $SSE(X_1, X_2) = SSE(X_1)$
 3. 電腦報表(SAS): 設應變數Y, 自變數 X_1, X_2, X_3

Source	df	Type I SS	Type III SS
X_1	1	$SSR(X_1)$	$SSR(X_1 X_2, X_3)$
X_2	1	$SSR(X_2 X_1)$	$SSR(X_2 X_1, X_3)$
X_3	1	$SSR(X_3 X_1, X_2)$	$SSR(X_3 X_1, X_2)$

二、複判定係數(multiple coefficient of determination)

定義：一個應變數及k個自變數之迴歸模型下，則稱 $R_{Y12...k}^2$ 為複判定係數，

$$R_{Y12...k}^2 = \frac{SSR(X_1, X_2, \dots, X_k)}{SST}$$

$$R_{Y1}^2 = \frac{SSR(X_1)}{SST} = \frac{\hat{\beta}_1^2 SS_1}{SS_Y} = \frac{\hat{\beta}_1 SS_{1Y}}{SS_Y}$$

$$R_{Y12}^2 = \frac{SSR(X_1, X_2)}{SST} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{SST} = \frac{\sum (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} - \bar{Y})^2}{SST}$$

$$= \frac{\sum (\bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} - \bar{Y})^2}{SST}$$

$$= \frac{\hat{\beta}_1^2 SS_1 + \hat{\beta}_2^2 SS_2 + 2\hat{\beta}_1 \hat{\beta}_2 SS_{12}}{SS_Y} = \frac{\hat{\beta}_1 SS_{1Y} + \hat{\beta}_2 SS_{2Y}}{SS_Y}$$

當迴歸模型中有一個自變數： $R_{Y1}^2 = \frac{\hat{\beta}_1 SS_{1Y}}{SS_Y}$

當迴歸模型中有兩個自變數： $R_{Y12}^2 = \frac{\hat{\beta}_1 SS_{1Y} + \hat{\beta}_2 SS_{2Y}}{SS_Y}$

⋮

當迴歸模型中有k個自變數： $R_{Y12...k}^2 = \frac{\hat{\beta}_1 SS_{1Y} + \hat{\beta}_2 SS_{2Y} + \dots + \hat{\beta}_k SS_{kY}}{SS_Y}$

由以上之整理可得知，當自變數個數增加時，複判定係數 R^2 之分子會增加，但分母則維持不變，如此會造成 R^2 必然會增加或相同之情況。其實，若研究者希望得到一個 R^2 解釋能力足夠之迴歸模型時，只要增加自變數即可，但加入之變數是否真正擁有顯著之解釋能力就不得而知。

例：研究所入學考試中統計學成績(應變數)