第十三、十四章: 複迴歸分析

Mei-Yuan Chen
Department of Finance
National Chung Hsing University

February 19, 2013

前言

在實際問題的分析上, 應變數或被解釋變數會受到不只一個變數的影響, 此時, 我們即要討論複迴歸模型(multiple regression) 的分析了, 當然複迴歸模型可為線型亦可為非線型, 在此, 為簡單起見, 我們以線型複迴歸 (multiple linear regression) 為討論的重點, 亦即我們所要討論的是一個應變數Y的實現值決定於許多個 (如k個) 解釋變數$X_1, X_2, \ldots, X_k$出現的數值, 而其間的關係是為線型的關係, 即

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i, i = 1, \cdots, n.$$

而複迴歸分析的主要工作即在以實際資料估計出模型中的參數$\beta_1, \beta_2, \ldots, \beta_k$, 以瞭解某一個解釋變數數值的變動將對應變數造成多少的影響。

最小平方估計方法

在線型複迴歸的討論中, 我們依然分為估計、檢定和預測三項主要工作進行說明; 而線型複迴歸模型參數的估計, 我們以最普遍使用的最小平方估計式加以討論, 而為能延續檢定與預測的討論, 如同在簡單線型模型的討論一般, 我們需要以下的假設條件:

(1) $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i, i = 1, \cdots, n$, 且$n > k$。

(2) 所有的解釋變數x_j, $j = 1, \ldots, k$為非隨機的且不為線性相依 (linearly dependent) 的。

(3) $E(\epsilon_i) = 0$。

(4) $E(\epsilon_i^2) = \sigma_0^2$且任何兩個不同的觀察值不存在相關性, 即 $E(\epsilon_i \epsilon_j) = 0$。

(5) 每一個ϵ_i為相互獨立且相等的常態分配, 即$\epsilon_i \sim$ i.i.d. $N(0, \sigma_0^2)$.

以上陳述的假設條件與在簡單線型模型中所假設的條件相似, 除了假設(1) 中的兩個變數間的線型關係改為多個變數間的線型關係外, 增加了假設 (2) 中解釋變數間不為線性相依的條件, 亦即解釋變數間為線性獨立 (linearly independent) 的; 所謂解釋變數$X_1, X_2, \ldots, X_k$為線性獨立[1] 表示其中的任何一個解釋變數無法由其他$k-1$個解釋變數的線性組合來描述, 也就是說, 我們所考慮的k個解釋變數中沒有一個是多餘的變數 (redundant variable); 而解釋變數$X_1, X_2, \ldots, X_k$為線性獨立的假設條件也使得最小平方估計式具有唯一解 (unique solution)。

[1] $X_1, X_2, \ldots, X_k$的線性組合$a_1X_1 + a_2X_2 + \cdots + a_kX_k = 0$, 若存在$(a_1, a_2, \ldots, a_k)$不全為零的參數解, 則稱$X_1, X_2, \ldots, X_k$為線性相依, 若僅存在全為零的參數解, 則稱$X_1, X_2, \ldots, X_k$為線性獨立

線型複迴歸模型參數的最小平方估計式依然是由平方誤差樣本平均數(sample average of squared errors) 的最小化求解所導出, 令 $(\hat{\beta}_{1n}, \hat{\beta}_{2n}, \ldots, \hat{\beta}_{kn})$ 表示線型複迴歸模型參數的最小平方估計式, 則

$$\begin{aligned}\{\hat{\beta}_{1n}, \ldots, \hat{\beta}_{kn}\} &= \arg \min_{\beta_1, \ldots, \beta_k} f(\beta_1, \cdots, \beta_k) \\ &= \arg \min_{\beta_1, \ldots, \beta_k} \frac{1}{n} \sum_{i=1} (y_i - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})^2.\end{aligned}$$

則極小化求解的一階條件為:

$$
\begin{aligned}
\frac{\partial}{\partial \beta_1} f(\beta_1, \cdots, \beta_k) &= -2\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})x_{1i} = 0, \\
\frac{\partial}{\partial \beta_2} f(\beta_1, \cdots, \beta_k) &= -2\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})x_{2i} = 0, \\
&\vdots \\
\frac{\partial}{\partial \beta_k} f(\beta_1, \cdots, \beta_k) &= -2\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})x_{ki} = 0.
\end{aligned}
$$

這個一階條件包含了k個方程式的聯立方程式, 而我們則要解出k個未知數$\beta_1, \ldots, \beta_k$的解, 此解即為最小平方估計式$\hat{\beta}_{1n}, \hat{\beta}_{2n}, \ldots, \hat{\beta}_{kn}$; 為能有$k$個唯一解, 則解釋變數$X_1, X_2, \ldots, X_k$為線性獨立的條件是必要的。至於求解的方法, 當然我們可以利用 Cramer rule 求解, 也可藉電腦軟體 (如 Excel、gretl 和 SAS 等) 求解, 在此我們不予討論。

在以實際資料$(y_i, x_i), i = 1, \ldots, n$代入由最小平方估計式$\hat{\beta}_{1n}, \hat{\beta}_{2n}, \ldots, \hat{\beta}_{kn}$中, 即可得到最小平方估計值 (ordinary least square estimates), 而方程式

$$\hat{y} = \hat{\beta}_{1n} x_1 + \cdots + \hat{\beta}_k x_k$$

即稱為迴歸平面 (regression hyperplane);
而$\hat{y}_i = \hat{\beta}_{1n} x_{1i} + \cdots + \hat{\beta}_k x_{ki}$為配適值 (fitted values),$e_i = y_i - \hat{y}_i$則為殘差值 (residuals)。至於, 誤差項變異數的最小平方估計式依然為

$$\hat{\sigma}_n^2 = \frac{1}{n-k} \sum_{i=1}^{n} e_i^2.$$

解釋變數X_j的參數估計值$\hat{\beta}_{jn}$的含意為: 在其他解釋變數不變下, 解釋變數X_j一個衡量單位的變動, 將使得被解釋變數$\hat{\beta}_{jn}$衡量單位的變動, 因為$\partial y/\partial x_j = \beta_j$; 至於兩個解釋變數迴歸參數估計值間是否進行大小的比較呢? 如果$\hat{\beta}_{jn} > \hat{\beta}_{hn}$, 是否隱含解釋變數$X_j$比解釋變數$X_h$對應變數具較大的影響效果呢? 答案當然是否定的, 因為迴歸參數估計值會因解釋變數的衡量單位的不同而有所不同, 假設解釋變數X_j為以公斤為衡量單位的體重變數, 而所得到的估計值為$\hat{\beta}_{jn}$, 若將衡量單位改為公克, 則將得到$1000 \times \hat{\beta}_{jn}$的估計值, 很顯然$|1000 \times \hat{\beta}_{jn}| > |\hat{\beta}_{jn}|$, 此外, 體重變動1公斤與變動1公克的的困難度也是不一樣的; 因此, 我們不能直接以估計值的大小比較解釋變數對應變數的影響效果。此時, 為能直接比較不同解釋變數迴歸估計值間的大小, 我們改以標準化迴歸 (standardized regression) 模型為分析對象。

所謂的標準化迴歸模型即是以原來迴歸模型的應變數與解釋變數的觀察資料, 分別減去個別的樣本平均數, 再除以個別的樣本標準差, 即計算Z-score 的標準化過程, 之後再以標準化後的應變數資料對標準化後的解釋變數資料進行迴歸, 即因

$$\frac{y_i - \bar{y}_n}{s_y} = \beta_2 \frac{x_{2i} - \bar{x}_2}{s_y} + \cdots + \beta_k \frac{x_{ki} - \bar{x}_k}{s_y} + \frac{\epsilon_i - \bar{\epsilon}}{s_y},$$

因此, 得到以下的標準化迴歸模型

$$y_i^* = \beta_2^* x_{2i}^* + \cdots + \beta_k^* x_{ki}^* + \epsilon_i^*,$$

其中

$$y_i^* = \frac{y_i - \bar{y}}{s_y}, \qquad x_{ji}^* = \frac{x_{ji} - \bar{x}_j}{s_{x_j}}, \qquad \beta_j^* = \frac{\beta_j s_{x_j}}{s_y}.$$

標準化迴歸模型中的參數通常被稱為β係數 (beta coefficients), 其估計值的大小即可直接用來進行比較, 而標準化迴歸模型中的解釋變數均為標準化後的Z-score, 因此, 是以樣本標準差為衡量單位, 是以每一個標準化後的解釋變數變動1單位 (即個樣本標準差) 的困難度是相同的; 而解釋變數X_j的 beta 係數估計值$\hat{\beta}^*_{jn}$的含意為: 在其他解釋變數不變下, 解釋變數X_j變動1個樣本標準差將造成應變數$\hat{\beta}^*_{jn}$個樣本標準差 ($\hat{\beta}^*_{jn} \times s_y$) 的變動。

如同在簡單線型迴歸模型的討論, 我們可以用判定係數來衡量所估計的迴歸模型與實際資料間的配適度; 對於線型複迴歸估計結果的配適度衡量, 判定係數(R^2) 的定義為

$$R^2 = \frac{\text{RSS}}{\text{TSS}} = 1 - \left(\frac{\text{ESS}}{\text{TSS}}\right),$$

其中,

$$\begin{aligned}
\text{TSS} &= \sum_{i=1}^{n}(y_i - \bar{y}_n)^2 \\
\text{RSS} &= \sum_{i=1}^{2}(\hat{y}_i - \bar{y}_n)^2 \\
\text{ESS} &= \sum_{i=1}^{2}(y_i - \hat{y}_i)^2.
\end{aligned}$$

依據定義, 最小平方估計式是在所觀察到的樣本資料中使平方誤差最小的參數解, 亦即沒有其他的參數值會比最小平方估計值$(\hat{\beta}_1, \ldots, \hat{\beta}_k)$使平方誤差$\sum_{i=1}(y_i - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})^2$更小, 所以,$\sum_{i=1}^{n}(y_i - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki})^2 =$ ESS為最小的數值, 相對地, 最小平方估計值亦是使RSS最大的解, 此時以R_k^2為此迴歸模型的判定係數;

令$(\tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1})$為迴歸模型$y_i = \beta_1 x_{1i} + \cdots + \beta_{k-1} x_{(k-1)i} + \epsilon_i$的最小平方估計式, 此結果亦可寫為$(\tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1}, 0)$為迴歸模型$y_i = \beta_1 x_{1i} + \cdots + \beta_{k-1} x_{(k-1)i} + \beta_k x_{ki} + \epsilon_i$的最小平方估計式, 此時以$R^2_{k-1}$為此迴歸模型的判定係數 顯而易見, 因$\sum_{i=1}^{n}(y_i - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki})^2$必定比$\sum_{i=1}^{n}(y_i - \tilde{\beta}_1 x_{1i} - \cdots - \tilde{\beta}_{k-1} x_{(k-1)i} - 0 x_{ki})^2$來得小, 而兩個模型具有相同的 TSS, 所以,$R^2_k$必定大於$R_{k-1}$, 因此我們得到結論: 判定係數$R^2$為解釋變數個數$k$的非遞減函數 (nondecreasing function)。

雖然在實證分析時，我們希望有更佳的配適度迴歸模型，亦即我們迴歸模型的判定係數越大越好，但根據R^2為解釋變數個數k的非遞減函數的特性，我們只要儘可能的加入新的解釋變數於迴歸模型中，即可得到更大的判定係數值，表示我們得到更好的迴歸模型；但是，加入新的解釋變數雖可提高判定係數值，但是迴歸模型中也增加了待估計的參數，因而，在固定的樣本觀察資料下，使得估計式的自由度因而降低，這是增加解釋變數的負面效果；從誤差項變異數的樣本估計式:$\hat{\sigma}_n^2 = \sum_{i=1}^n e_i^2/(n-k)$可知，解釋變數資加將使$\sum_{i=1}^n e_i^2$變小，卻也使自由度$n-k$變小，如果$\sum_{i=1}^n e_i^2$變小的幅度小於$n-k$變小的幅度，則$\hat{\sigma}_n^2$會變大，進而使得迴歸參數估計式的變異數變大，也就是越不準確，如此解釋變數的增加反而不利。

因此, 一個好的配適度判定準則, 不但能考量$\sum_{i=1}^{n} e_i^2$的變動效果, 也應能考慮到對自由度的影響效果; 因此, 經自由度調整後的判定係數, 即所謂調整後的判定係數 (adjusted R^2) 定義為:

$$\bar{R}^2 = 1 - \frac{\mathsf{ESS}/(n-k)}{\mathsf{TSS}/(n-1)}.$$

由這個定義式, 可看出若解釋變數增加使 ESS 降低的幅度大於$(n-k)$降低的幅度, 將使得ESS$/(n-k)$變小, 而TSS$/(n-1)$為固定的數, 因而使$\bar{R}^2$變大, 此時, 解釋變數的增加是適當的; 反之, 若解釋變數增加使 ESS 降低的幅度小於$(n-k)$降低的幅度, 反將使得ESS$/(n-k)$變大, 因而使$\bar{R}^2$變小, 此時, 解釋變數的增加是不適當的。

調整後的判定係數$\bar{R}^2$可進一步改寫為

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2).$$

因此, 上式等號右邊的第二項為因解釋變數增加而使自由度減少的處罰項 (penalty)。此外, 由上式可知,$\bar{R}^2$將永遠小於$\bar{R}^2$, 且將永遠小於1; 雖然$\bar{R}^2$必定大於零, 但$\bar{R}^2$卻有可能小於零。

除了以調整後的判定係數($\bar{R}^2$) 外, 尚有以下的準則作為模型選擇 (model selection) 的依據; 令$\tilde{\sigma}_n^2 = \sum_{i=1}^n e_i^2/n$,

1. AIC (Akaike Information Criterion):
   $\tilde{\sigma}_n^2\, e^{2k/n}$或$\log \tilde{\sigma}_n^2 + 2k/n$。
2. SIC (Schwarz Information Criterion):
   $\tilde{\sigma}_n^2\, n^{k/n}$或$\log \tilde{\sigma}_n^2 + (k/n)(\log n)$。
3. HQ (Hannan and Quinn Criterion):
   $\tilde{\sigma}_n^2\, (\log n)^{2k/n}$或$\log \tilde{\sigma}_n^2 + (2k/n)(\log \log n)$.

這些準則均是由$\log \tilde{\sigma}_n^2$再加上一個處罰項; 我們選擇的模型是具有最小 AIC(SIC, HQ) 數值的模型, 至於哪個準則比較好, 則沒有定論。

為對線型複迴歸模型參數進行假設檢定, 我們必須先討論最小平方估計式的分配, 在對誤差項分配的假設(5):$\epsilon_i \sim$ i.i.d.$N(0, \sigma_0^2)$下, 應變數觀察值y_i的分配為獨立的常態分配$N(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}, \sigma_0^2)$, 因為

$$\begin{aligned} E(y_i) &= E(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i) \\ &= \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + E(\epsilon_i) \quad \text{因}\beta_j\text{為常數而}x_{ji}\text{非隨機的} \\ &= \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}, \text{因}\epsilon_i) = 0 \end{aligned}$$

且

$$\begin{aligned} \text{var}(y_i) &= \text{var}(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i) \\ &= \text{var}(\epsilon_i) \quad \text{因}\beta_j\text{為常數而}x_{ji}\text{非隨機的} \\ &= \sigma_0^2. \end{aligned}$$

在常態分配的假設(5) 下, 不但y_i為常態分配, 我們也可得到以下結果:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2_{\hat{\beta}_j}), \quad j = 1, \cdots, k$$

即迴歸模型參數之最小平方估計式具有常態分配的抽樣分配; 則經過標準化後, 得到

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0, 1), \quad j = 1, \cdots, k.$$

再者, 如同在簡單線型迴歸模型中的討論一般, 誤差項變異數σ_0^2的最小平方估計式$\hat{\sigma}_n^2$的抽樣分配為

$$\sum_{i=1}^{n} e_t^2/\sigma_0^2 = (n-k)\hat{\sigma}^2/\sigma_0^2 \sim \chi^2(n-k)$$

亦如同在簡單線型迴歸模型中的結果:

$$\frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \sim t(n-k), j = 1, \cdots, k$$

根據這個抽樣分配結果, 我們就可以進行迴歸模型參數的假設檢定了。

由於在線型複迴歸模型中, 存在許多個解釋變數, 因此, 對這些解釋變數的迴歸參數的假設檢定, 可分為單獨對單一迴歸參數和同時對多個迴歸參數進行假設檢定。對於單一虛無假設 (single null hypothesis) $H_0 : \beta_j = b$的統計檢定量為

$$t_{\beta_j} = \frac{\hat{\beta}_j - b}{s_{\hat{\beta}_j}} \sim t(n-k). \tag{1}$$

此即為所謂的t檢定; 若統計檢定量$t_{\beta_j} \geq t_{\alpha/2}(n-k)$或$t_{\beta_j} \leq -t_{\alpha/2}(n-k)$, 則棄卻虛無假設$H_0 : \beta_j = b$; 另者, 對於單尾檢定而言, 若統計檢定量$t_{\beta_j} \geq t_\alpha(n-k)$, 則棄卻虛無假設$H_0 : \beta_j < b$, 或若統計檢定量$t_{\beta_j} lgeq t_\alpha(n-k)$, 則棄卻虛無假設$H_0 : \beta_j > b$。通常在實證分析上, 我們想要知道的是解釋變數$X_j$是否對應變數具有顯著的影響效果, 亦即其所對應的迴歸參數是否顯著異於零, 因此, 虛無假設設定為$H_0 : \beta_j = 0$。

此外, 單一虛無假設也可包含多個迴歸參數, 例如, 虛無假設為 $H_0: a\beta_h + b\beta_j = c$, 在這個虛無假設下, 表示迴歸參數存在一個限制式, 亦即在虛無假設中僅存在一個等號, 若令 $Z = a\beta_h + b\beta_j$, 則虛無假設 $H_0: a\beta_h + b\beta_j = c$ 即成為 $H_0: Z = c$, 這個虛無假設與 $H_0: \beta_j = b$ 相類似, 所以其統計檢定量如同式 (1) 一般, 為

$$t_Z = \frac{\hat{Z} - c}{s_{\hat{Z}}} \sim t(n-k).$$

由於$Z = a\beta_h + b\beta_j$, 則其最小平方估計式為$\hat{Z} = a\hat{\beta}_h + b\hat{\beta}_j$, 而其變異數為

$$\begin{aligned} \text{var}\hat{Z} &= \text{var}(a\hat{\beta}_h + b\hat{\beta}_j) \\ &= a^2\text{var}(\hat{\beta}_h) + b^2\text{var}(\hat{\beta}_j) + 2ab\text{cov}(\hat{\beta}_h, \hat{\beta}_j) \end{aligned}$$

其樣本估計式為

$$s^2_{\hat{Z}} = a^2 s^2_{\hat{\beta}_h} + b^2 s^2_{\hat{\beta}_j} + 2abs_{\hat{\beta}_h, \hat{\beta}_j}.$$

利用同樣的過程, 我們可以對包含更多個迴歸參數的單一虛無假設進行檢定。

對於包含多個迴歸參數及兩個以上等式虛無假設檢定, 我們稱之為聯合檢定, 因此, 單一檢定與聯合檢定之區別不在於所包含迴歸參數的多寡, 而在於所包含限制式(等式) 的個數, 若僅包含一個限制式, 即稱為單一虛無假設, 而若包含一個以上的限制式, 即稱為聯合虛無假設; 假設我們想要知道兩個解釋變數X_j和X_h是否對應變數具有顯著的影響效果, 則所對應的聯合虛無假設為

$$\begin{aligned} H_0 : \beta_j &= 0 \\ \beta_h &= 0 \end{aligned}$$

對於這個聯合假設檢定的統計檢定量, 可藉由兩個線型迴歸所估計之判定係數 (R^2) 計算而得;

令R_u^2表示來自未受限迴歸模型 (unconstrained regression)$y_i = \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$的估計結果, 而$R_c^2$表示來自受限迴歸模型 (constrained regression)

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{2i} + \cdots + \beta_{j-1} x_{(j-1)i} + \beta_{j+1} x_{(j+1)i} \\ &\quad + \cdots + \beta_{h-1} x_{(h-1)i} + \beta_{h+1} x_{(h+1)i} + \cdots + \beta_k x_{ki} + \epsilon_i \end{aligned}$$

的估計結果, 此所謂受限迴歸模型乃是在原來未受限迴歸模型中將解釋變數X_j和X_h剔除, 因為在$\beta_j = 0$和$\beta_h = 0$的假設下,$\beta_j x_{ji} = 0$和$\beta_h x_{hi} = 0$, 所以在虛無假設下的受限迴歸模型, 即等於應變數對剔除X_j和X_h後的其他解釋變數的迴歸模型。

而虛無假設的檢定統計量則為

$$\psi = \frac{(R_u^2 - R_c^2)/2}{(1 - R_u^2)/(n-k)} \sim F(2, n-k).$$

若統計檢定量$\psi \geq F_\alpha(2, n-k)$, 則在顯著水準$\alpha$下棄卻虛無假設$H_0 : \beta_j = \beta_h = 0$; 統計量$\psi$的邏輯在於比較$R_u^2$和$R_c^2$的差距是否顯著, 因為理論上,$R_u^2$必定大於或等於$R_c^2$, 若解釋變數$X_j$和$X_h$對應變數完全沒有影響效果, 即虛無假設$H_0 : \beta_j = \beta_h = 0$成立, 則$R_u^2$將等於$R_c^2$, 而統計量$\psi$將等於零; 而若解釋變數$X_j$和$X_h$對應變數的確有影響效果, 即虛無假設$H_0 : \beta_j = \beta_h = 0$不成立, 則$R_u^2$將恆大於$R_c^2$, 而統計量$\psi$亦將顯著大於零; 因此, 越大的$\psi$值表示越有可能棄卻虛無假設。

在一般的統計軟體或計量軟體所提供的迴歸模型估計結果, 其中包括了一個變異數分析表, 提供了一個F-統計量, 這個統計量即是對所有解釋變數的迴歸參數全為零的虛無假設的檢定統計量, 即虛無假設

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

的檢定統計量; 在此虛無假設下, 受限的迴歸模型成為$y_i = \beta_1 + \epsilon_i$, 則$\beta_1$的最小平方估計式為$\tilde{\beta}_1 = \sum_{i=1}^n y_i/n = \bar{y}_n$而配適值$\hat{y}_i = \bar{y}_n$, 所以$\text{RSS} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 = 0$而$\text{ESS} = \text{TSS}$, 因此$R_c^2 = 0$, 則$\psi$統計量成為

$$\psi = \frac{R_u^2/(k-1)}{(1-R_u^2)/(n-k)} \sim F(k-1, n-k).$$

再者, 對於比較一般化的聯合虛無假設:

$$
\begin{aligned}
H_0 : \beta_j &= a \\
\beta_h &= b
\end{aligned}
$$

的假設檢定, 也可利用比較受限與未受限迴歸的判定係數值進行檢定,
令R_u^2仍為未受限迴歸的判定係數, 而在虛無假設下的受限迴歸模型則為

$$
\begin{aligned}
y_i - a x_{ji} - b x_{hi} &= \beta_1 x_{1i} + \cdots + \beta_{j-1} x_{(j-1)i} + \beta_{j+1} x_{(j+1)i} \\
&\quad + \cdots + \beta_{h-1} x_{(h-1)i} + \beta_{h+1} x_{(h+1)i} + \cdots + \beta_k x_{ki} + \epsilon_i
\end{aligned}
$$

亦即以應變數資料 (y_i) 減去$a x_{ji} + b x_{hi}$後再對其他解釋變數進行迴歸。

令R_c^2為由此受限迴歸模型所估計的判定係數，則虛無假設的檢定統計量仍為

$$\psi = \frac{(R_u^2 - R_c^2)/2}{(1 - R_u^2)/(n-k)} \sim F(2, n-k).$$

當然，上述的ψ統計檢定量可應用到更多個別迴歸參數限制的虛無假設。然而，對於以下的虛無假設，就無法利用比較受限與未受限迴歸的判定係數值進行檢定，如虛無假設為

$$\begin{aligned} H_0 : a_1\beta_h + b_1\beta_j &= c_1 \\ a_2\beta_j - b_2\beta_l &= c_2 \end{aligned}$$

在這樣的虛無假設下, 我們無法建構受限迴歸模型, 而無法得到R_c^2, 以計算統計量ψ; 此時, 我們可藉助於電腦軟體進行統計量的計算, 如在 gretl 的「模型 (M)」、「一般最小平方法 (OLS)」的估計結果輸出視窗中, 選擇「檢定 (T)」、「係數線型限制檢定 (L)」進入「線型限制式」視窗, 在視窗中鍵入$b[2] = 0$, 換行再鍵入$b[4] = 1$, 以進行線型虛無假設$H_0 : \beta_2 = 0, \beta_4 = 1$的檢定, 一般所算得的統計量, 通常具有$F(q, n-k)$的抽樣分配, 其中, q為虛無假設中限制式 (等式) 的個數, 例子中$q = 2$。

在迴歸分析中, 有時被解釋變數(y) 與解釋變數 $(x_1, x_2, \ldots, x_k)$ 間的關係, 會因研究對象某種特徵的不同, 而有不同的迴歸模型參數; 例如身高和體重間的直線關係會因研究對象性別上的差異而不同, 因此, 對於男性研究對象身高 (y) 與體重 (x) 間的線性迴歸模型為

$$y_i = \alpha_0 + \beta_0 x_i + \epsilon_i,$$

而女性研究對象身高與體重間的線性迴歸模型為

$$y_i = \alpha_1 + \beta_1 x_i + \epsilon_i.$$

所以, 如果$\alpha_0 \neq \alpha_1$或$\beta_0 \neq \beta_1$均代表身高與體重間的線性關係存在性別間的差異。對於上述男、女性的兩個線性迴歸模型, 可以藉由虛擬變數 (dummy variable) 的使用, 將兩個迴歸模型合而為一; 其做法為定義一個性別虛擬變數:

$$
\begin{aligned}
d_i &= 1, \qquad \text{若第}i\text{個研究對象為男性} \\
&= 0, \qquad \text{若第}i\text{個研究對象為女性}
\end{aligned}
$$

考慮以下複迴歸模型:

$$y_i = \alpha + \gamma_0 d_i + \beta x_i + \gamma_1 (d_i \times x_i) + \epsilon_i. \tag{2}$$

在這個模型中, 當第i個研究對象為男性 ($d_i = 1$) 時, 此模型成為

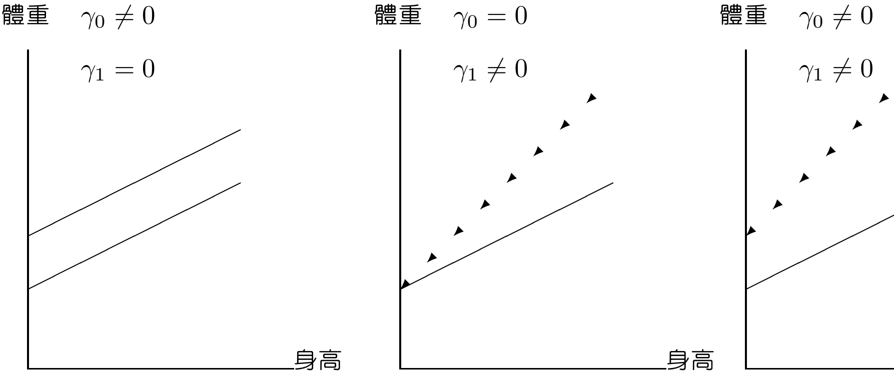
$$y_i = (\alpha + \gamma_0) + (\beta + \gamma_1)x_i + \epsilon_i,$$

而當第i個研究對象為女性 ($d_i = 0$) 時, 此模型成為

$$y_i = \alpha + \beta x_i + \epsilon_i.$$

所以, 如果男、女性研究對象身高與體重間的線性關係存在截距項 (intercept) 的差異, 則$\gamma_0 \neq 0$; 若男、女性研究對象身高與體重間的線性關係存在斜率 (slope) 的差異, 則$\gamma_1 \neq 0$; 當$\gamma_0 \neq 0$且$\gamma_1 \neq 0$時, 表示身高與體重間的線性關係在性別間存在截距項與斜率的差異; 這三種情形可以用圖 13.1 加以說明。

性別虛擬變數在迴歸模型中的說明



由以上的討論知道, 利用性別虛擬變數的設計, 我們可以用一個複迴歸模型檢定身高與體重是否存在性別差異; 若要檢定是否存在截距項差異的虛無假設為$H_0 : \gamma_0 = 0$, 若要檢定是否存在斜率差異的虛無假設為$H_0 : \gamma_1 = 0$, 至於是否存在二者差異的虛無假設則為$H_0 : \gamma_0 = 0, \gamma_1 = 0$。

此外，純為了好玩，我們想了解台灣管理學院學生身高與體重間的線性關係是否在不同科系間有所不同，也可以設計以下4個科系虛擬變數：

$$
\begin{aligned}
d1_i &= 1, && \text{若第}i\text{個學生屬經濟系} \\
&= 0, && \text{若第}i\text{個學生屬其他科系} \\
d2_i &= 1, && \text{若第}i\text{個學生屬財金系} \\
&= 0, && \text{若第}i\text{個學生屬其他科系} \\
d3_i &= 1, && \text{若第}i\text{個學生屬企管系} \\
&= 0, && \text{若第}i\text{個學生屬其他科系} \\
d4_i &= 1, && \text{若第}i\text{個學生屬會計系} \\
&= 0, && \text{若第}i\text{個學生屬其他科系}
\end{aligned}
$$

考慮以下線性迴歸模型:

$$
\begin{aligned}
y_i &= \alpha + \gamma_1 d1_i + \gamma_2 d2_i + \gamma_3 d3_i + \gamma_4 d4_i + \beta x_i + \\
&\quad \lambda_1(d1_i \times x_i) + \lambda_2(d2_i \times x_i) + \lambda_3(d3_i \times x_i) + \lambda_4(d4_i \times x_i) + \epsilon_i
\end{aligned}
$$

在模型中, 若$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$表示身高與體重間的線性關係不存在截距項的差異, 若$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$表示身高與體重間的線性關係不存在斜率的差異, 因此, 身高與體重間的線性關係不存在科系別差異的虛無假設為:

$$
\begin{aligned}
H_0 : \gamma_1 &= \gamma_2 = \gamma_3 = \gamma_4 = 0 \\
\lambda_1 &= \lambda_2 = \lambda_3 = \lambda_4 = 0.
\end{aligned}
$$

在此值得注意的是, 在設計性別虛擬變數時, 因性別特徵只有2個類項, 所以只設計一個虛擬變數; 假定我們多設計了一個虛擬變數

$$
\begin{aligned}
s_i &= 1, \qquad \text{若第}i\text{個研究對象為女性} \\
&= 0, \qquad \text{若第}i\text{個研究對象為男性}
\end{aligned}
$$

並將這個虛擬變數加入迴歸模型 (2) 中即成為

$$y_i = \alpha + \gamma_0 d_i + \delta_0 s_i + \beta x_i + \gamma_1 (d_i \times x_i) + \delta_1 (s_i \times x_i) + \epsilon_i. \quad (3)$$

由於對所有的i而言, 若$d_i = 1$則$s_i = 0$, 或若$d_i = 0$則$s_i = 1$, 因此$d_i + s_i = 1$且$d_i \times x_i + s_i \times x_i = x_i$, 這樣在迴歸模型 (3) 中的解釋變數將不為線性獨立, 違反了假設條件 (2), 將使得最小平方估計式不存在唯一解 (unique solution)。同樣道理, 在科系別虛擬變數的設計上, 因科系別有5個, 因此, 我們只設計了4個虛擬變數。

如果我們想了解台灣的大學生的身高與體重間的關係, 而假定其間的關係為線形, 並且, 我們會懷疑其間關係在不同性別間存在不同的線性關係, 因此, 我們所考慮的迴歸模型為$Y = \alpha + \beta X + \gamma D + \epsilon$, 其中, X和Y分別代表台灣的大學生身高和體重的隨機變數,D為性別特徵虛擬變數, 而ϵ為迴歸誤差項, 在這個迴歸模型中, 迴歸係數γ表示身高與體重間的線性關係在性別間存在截距項的不同;

假定在統計學課堂上的80位同學具有台灣大學生的代表性, 則其身高、體重及性別虛擬變數的資料$\{(y_i, x_i, d_i)\}_{i=1}^{80}$即為一組樣本資料, 假定其儲存在 H-W-Sex.xls 的 Excel 檔案中, A1–A80 為身高資料、B1–B80 為體重資料而 C1–C80 為性別虛擬變數的資料, 點選『工具 (T)』中的『資料分析 (D)』, 在「資料分析」的對話框中的「分析工具 (A)」選取『迴歸』, 即進入「迴歸」對話框, 在「輸入 Y 範圍 (Y)」中輸入 A1–A80, 在「輸入 X 範圍 (X)」中輸入 B1–C80, 選擇「輸出範圍 (O)」為 D1, 並勾選「殘差」框中的「殘差圖 (D)」及「樣本迴歸線圖 (I)」, 按下『確定』即可得到輸出結果; 以下為其摘要輸出結果:

摘要輸出

迴歸統計	
R的倍數	0.772805982
R平方	0.597229086
調整的 R 平方	0.586767504
標準誤	4.208759844
觀察值個數	80

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	2	2022.469	1011.234	57.088	6.23357E-16
殘差	77	1363.952	17.71365		
總和	79	3386.421			

	係數	標準誤	t 統計	P-值	下限 95 %	上限
截距	112.3194	4.998147	22.47221	1.47101E-35	102.36683	122
X變數 1	0.867835	0.081474	10.65171	8.38787E-17	0.7056002	1.03
X變數 2	0.271989	0.946441	0.287381	0.774592385	-1.612617	2.15

在最上層表中, 判定係數$R^2 = 0.597229086$, 經自由度調整後的判定係數為

$$\begin{aligned}\bar{R}^2 &= R^2 - \frac{k-1}{n-k}(1-R^2)\\ &= 0.597229086 - \frac{3-1}{80-3}(1-0.597229086) = 0.586767504\end{aligned}$$

至於在中間的 ANOVA 表中, 將「迴歸」列的「MS」除以「殘差」列中的「MS」, 即得到ψ統計量, 因其抽樣分配為F-分配, 故又稱為F-統計量, 即

$$\psi = \frac{1011.234}{17.71365} = 57.088.$$

而表中的「顯著值」即為ψ統計量的p-值,
即$P(F(1,78) \geq 115.451) = 4.76886E - 17$, 此數值小於0.05, 因此, 在顯著水準$\alpha = 5\ \%$下棄卻虛無假設$H_0 : \beta = \gamma = 0$。此外,$\psi$統計量亦可由下式計算而得:

$$\begin{aligned}
\psi &= \frac{R_u^2/(k-1)}{(1-R_u^2)/(n-k)} \\
&= \frac{0.597229086/(3-1)}{(1-0.597229086)/(80-3)} = 57.088.
\end{aligned}$$

在最下層的迴歸結果輸出表中,「X 變數 1」的「係數」為$\hat{\beta}_n = 0.867835$, 其「標準誤」為$s_{\hat{\beta}_n} = 0.081474$, 而其針對於$H_0 : \beta = 0$的$t$-統計量為$\hat{\beta}_n / s_{\hat{\beta}_n} = 0.867835/0.081474 = 10.65171$, 在此統計量下, 其$p$-值為 8.38787E-17 小於 0.025, 或在虛無假設下的數值 0 不在 95 % 的下限 0.7056002、上限 1.030071 的範圍內, 故在顯著水準$\alpha = 5$ %下應棄卻虛無假設, 身高與體重間存在顯著的關係, 而「X 變數 2」的「係數」為$\hat{\gamma}_n = 0.271989$, 其「標準誤」為$s_{\hat{\gamma}_n} = 0.946441$, 而其針對於$H_0 : \gamma = 0$的$t$-統計量為$\hat{\gamma}_n / s_{\hat{\gamma}_n} = 0.271989/0.946441 = 0.287381$, 在此統計量下, 其$p$-值為 0.774592385 大於 0.025, 或在虛無假設下的數值 0 不在 95 % 的下限-1.612617、上限 2.156596 的範圍內, 故在顯著水準$\alpha = 5$ %下不應棄卻虛無假設, 身高與體重間的線性關係, 在性別間不存在截距項上的顯著差異。

由於判定係數為解釋變數個數k的非遞減函數 (non-decreasing function), 因此, 相較於前一章中身高對體重的簡單迴歸結果, 此迴歸模型的判定係數將變大, 簡單迴歸的$R^2 = 0.596797085$, 的確大於此迴歸的$R^2 = 0.597229086$; 然由於性別虛擬變數的t-統計量小於1, 因此, 相較於前一章中身高對體重的簡單迴歸結果, 此迴歸模型的調整後判定係數將變小, 對照於簡單迴歸的$\bar{R}^2 = 0.591627817$, 的確大於此迴歸的$\bar{R}^2 = 0.586767504$。

如果我們想了解台灣的大學生的身高與體重間的關係, 而假定其間的關係為線形, 並且, 我們會懷疑其間關係在不同性別間存在不同的線性關係, 因此, 我們所考慮的迴歸模型為$Y = \alpha + \beta X + \gamma D + \lambda(D \times X) + \epsilon$, 其中, X和Y分別代表台灣的大學生身高和體重的隨機變數,D為性別特徵虛擬變數, 而ϵ為迴歸誤差項, 在這個迴歸模型中, 迴歸係數γ表示身高與體重間的線性關係在性別間存在截距項的不同, 而迴歸係數λ表示身高與體重間的線性關係在性別間存在斜率項的不同, 因此, 身高與體重間的線性關係在性別間是否存在差異的虛無假設為$H_0 : \gamma = \lambda = 0$;

假定在統計學課堂上的80位同學具有台灣大學生的代表性, 則其身高、體重及性別虛擬變數的資料$\{(y_i, x_i, d_i, d_i \times x_i)\}_{i=1}^{80}$即為一組樣本資料, 假定其儲存在 H-W-Sex1.xls 的 Excel 檔案中, A1–A80 為身高資料、B1–B80 為體重資料、C1–C80 為性別虛擬變數的資料、而 D1–D80 為性別虛擬變數乘上體重的資料, 點選『工具 (T)』中的『資料分析 (D)』, 在「資料分析」的對話框中的「分析工具 (A)」選取『迴歸』, 即進入「迴歸」對話框, 在「輸入 Y 範圍 (Y)」中輸入 A1–A80, 在「輸入 X 範圍 (X)」中輸入 B1–D80, 選擇「輸出範圍 (O)」為 E1, 並勾選「殘差」框中的「殘差圖 (D)」及「樣本迴歸線圖 (I)」, 按下『確定』即可得到輸出結果; 以下為其摘要輸出結果:

摘要輸出

迴歸統計	
R的倍數	0.776669151
R平方	0.60321497
調整的 R 平方	0.587552403
標準誤	4.204760859
觀察值個數	80

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	3	2042.739	680.913	38.513	3.05879E-15
殘差	76	1343.681	17.6800		
總和	79	3386.420			

	係數	標準誤	t 統計	P-值	下限 95 %	上限 95 %
截距	106.149	7.625	13.921	1.32249E-22	90.962	121.335
X變數 1	0.96932	0.124	7.7586	3.1854E-11	0.7204	1.21814
X變數 2	10.9011	9.971	1.0932	0.277752715	-8.959	130.761
X變數 3	-0.1763	0.164	-1.070	0.287664812	-0.504	0.15165

在最上層表中, 判定係數$R^2 = 0.60321497$, 此數值較前一範例中之R^2為大, 因迴歸模型中之解釋變數增加所造成, 經自由度調整後的判定係數為

$$
\begin{aligned}
\bar{R}^2 &= R^2 - \frac{k-1}{n-k}(1-R^2) \\
&= 0.60321497 - \frac{4-1}{80-4}(1-0.60321497) = 0.587552403
\end{aligned}
$$

至於在中間的 ANOVA 表中, 將「迴歸」列的「MS」除以「殘差」列中的「MS」, 即得到ψ統計量, 因其抽樣分配為F-分配, 故又稱為F-統計量, 即

$$
\psi = \frac{680.913}{17.6800} = 38.513.
$$

而表中的「顯著值」即為ψ統計量的p-值,

即$P(F(3,76) \geq 38.513) = 3.05879E - 15$, 此數值小於0.05,

因此, 在顯著水準$\alpha = 5\ \%$下棄卻虛無假

設$H_0 : \beta = \gamma = \lambda = 0$。此外,$\psi$統計量亦可由下式計算而得:

$$\begin{aligned}\psi &= \frac{R_u^2/(k-1)}{(1-R_u^2)/(n-k)}\\ &= \frac{0.60321497/(4-1)}{(1-0.60321497)/(80-4)} = 38.513.\end{aligned}$$

在最下層的迴歸結果輸出表中,「X變數1」的「係數」為$\hat{\beta}_n = 0.96932$, 其「標準誤」為$s_{\hat{\beta}_n} = 0.124$, 而其針對於$H_0 : \beta = 0$的$t$-統計量為$\hat{\beta}_n / s_{\hat{\beta}_n} = 0.96932/0.124 = 7.7586$, 在此統計量下, 其$p$-值為3.1854E-11小於0.025, 或在虛無假設下的數值0不在95 %的下限0.7204、上限1.21814的範圍內, 故在顯著水準$\alpha = 5$ %下應棄卻虛無假設, 身高與體重間存在顯著的關係;「X 變數2」的「係數」為$\hat{\gamma}_n = 10.9011$, 其「標準誤」為$s_{\hat{\gamma}_n} = 9.971$, 而其針對於$H_0 : \gamma = 0$的$t$-統計量為$\hat{\gamma}_n / s_{\hat{\gamma}_n} = 10.9011/9.971 = 1.0932$, 在此統計量下, 其$p$-值為0.277752715大於0.025, 或在虛無假設下的數值0不在95 %的下限-8.959、上限130.761的範圍內, 故在顯著水準$\alpha = 5$ %下不應棄卻虛無假設, 身高與體重間的線性關係, 在性別間不存在截距項上的顯著差異;

「X變數3的「係數」為$\hat{\lambda}_n = -0.1763$, 其「標準誤」為$s_{\hat{\lambda}_n} = 0.164$, 而其針對於$H_0 : \lambda = 0$的$t$-統計量為$\hat{\lambda}_n / s_{\hat{\lambda}_n} = -0.1763/0.164 = -1.070$, 在此統計量下, 其$p$-值為0.287664812大於0.025, 或在虛無假設下的數值0不在95 %的下限-0.504、上限0.15165的範圍內, 故在顯著水準$\alpha = 5$ %下不應棄卻虛無假設, 身高與體重間的線性關係, 在性別間不存在斜率項上的顯著差異;

至於在檢定身高與體重間的線性關係是否在性別間有所不同, 此時, 我們需同時檢定在截距項與斜率項是否存在顯著差異, 故所對應的虛無假設為 $H_0 : \gamma = \lambda = 0$, 這個聯合檢定可藉由受限迴歸的$R_c^2$與非受限迴歸的$R_u^2$的比較進行, 由於在虛無假設下之受限迴歸即為前一章中的簡單迴歸, 其$R^2 = R_c^2 = 0.596797085$, 而非受限迴歸之$R^2 = R_u^2 = 0.60321497$, 因此, ψ-統計量為

$$\psi = \frac{(R_u^2 - R_c^2)/2}{1 - R_u^2} = \frac{(0.60321497 - 0.596797085)/2}{1 - 0.60321497} = .$$

令矩陣$\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2, \cdots, \mathbf{a}_k]$是由$n$個向量$\mathbf{a}_1\ \mathbf{a}_2, \cdots, \mathbf{a}_k$所組成, 而每一個向量包含$n$元素 (element), 故$\mathbf{A}$為一個$n \times k$的矩陣, 其可表現如下:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nk} \end{bmatrix},$$

而此矩陣亦可以$[a_{ij}]$表示, 即此矩陣內所包含第i列 (row)、第j行 (column) 的元素為a_{ij}的數值。

考慮n個向量$\mathbf{a}_1\ \mathbf{a}_2, \cdots, \mathbf{a}_k$的線性組合 (linear combination) 並令其為0, 即成為下式的聯立方程式 (system equations):

$$\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \cdots + \alpha_k\mathbf{a}_k = \mathbf{0},$$

或

$$\alpha_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix} + \alpha_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{n2} \end{bmatrix} + \cdots + \alpha_n \begin{bmatrix} a_{1k} \\ a_{2k} \\ a_{3k} \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

這個聯立方程式因等號右邊為$\mathbf{0}$向量, 故稱為均質聯立方程式 (homogeneous system), 若等號右邊不為$\mathbf{0}$向量, 則稱為非均質聯立方程式 (non-homogeneous system)。

一個均質聯立方程式必定存在一個解,$(\alpha_1, \alpha_2, \cdots, \alpha_k) = (0, 0, \cdots, 0)$, 此均為0的一組解稱為必然解 (trivial solution); 而若此均質聯立方程式存在一組非全為0的解$(\alpha_1^*, \alpha_2^*, \cdots, \alpha_k^*)$, 則此解稱為非必然解 (non-trivial solution)。假設此均質聯立方程式存在一組非必然解, 並假定$\alpha_2^* \neq 0$, 則

$$\mathbf{a}_2 = -\frac{\alpha_1^*}{\alpha_2^*}\mathbf{a}_1 - \frac{\alpha_3^*}{\alpha_2^*}\mathbf{a}_3 - \cdots - \frac{\alpha_k^*}{\alpha_2^*}\mathbf{a}_k,$$

此表示向量$\mathbf{a}_2$可由其他的$k-1$個向量的線性組合來表示, 亦即向量$\mathbf{a}_2$依附於其他的$k-1$個向量, 因此, 包含於矩陣$\mathbf{A}$內的k個向量$\mathbf{a}_1$ $\mathbf{a}_2, \cdots, \mathbf{a}_k$是為線性相依 (linearly dependent)。

假若向量$\mathbf{a}_1\ \mathbf{a}_2, \cdots, \mathbf{a}_k$所組成的均質聯立方程式僅存在必然解而不存在非必然解, 則向量$\mathbf{a}_1\ \mathbf{a}_2, \cdots, \mathbf{a}_k$是為線性獨立 (linearly independent)。假若向量$\mathbf{a}_1\ \mathbf{a}_2, \cdots, \mathbf{a}_k$為線性獨立, 則其所組成的矩陣$\mathbf{A}$, 其$\mathbf{A}'\mathbf{A}$不存在反矩陣 (inverse matrix), 即$(\mathbf{A}'\mathbf{A})^{-1}$不存在。

在複迴歸中, 假設條件(1):$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i, i = 1, \cdots, n$, 可以以向量與矩陣表示為$\mathbf{y} = \mathbf{X}\beta + \epsilon$, 其中

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \cdots\, \mathbf{x}_k] = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2k} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

由於最小平方估計式$\hat{\beta}_n$是以下極小值問題的解:

$$\min_{\hat{\beta}} \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2,$$

而此極小值問題可用矩陣表示成

$$\min_{\hat{\beta}} \frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta).$$

其極小值求解的一階條件為

$$\hat{\beta}: \frac{-2}{n}\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}_n) = \mathbf{0}.$$

由此一階條件即可得到正規方程式 (normal equation):

$$\mathbf{X}'\mathbf{X}\hat{\beta}_n = \mathbf{X}'\mathbf{y}.$$

因為假設條件 (2) 假設解釋變數$\mathbf{x}_j, j = 1, \ldots, k$為線性獨立, 因而$(\mathbf{X}'\mathbf{X})^{-1}$存在, 因而我們可以在上式等號兩邊前乘以$(\mathbf{X}'\mathbf{X})^{-1}$, 即

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

進而得到最小平方估計式

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

而其二階條件為$(2/n)\mathbf{X}'\mathbf{X}$, 此矩陣為正定 (positive definite), 故一階條件的解為極小值解。也由於$(\mathbf{X}'\mathbf{X})^{-1}$的存在, $\hat{\beta}_n$是最小平方誤差的唯一解。

再者, 定義$\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n]' = X\hat{\beta}_n$為$n$配適值所組成的向量, 而$\hat{\epsilon} = [\hat{\epsilon}_1, \hat{\epsilon}_2, \cdots, \hat{\epsilon}_n]' = \mathbf{y} - \hat{\mathbf{y}}$為$n$個殘差值所組成的向量, 其極小值求解的一階條件可改為

$$
\begin{aligned}
& \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}_n) \\
&= \mathbf{X}'(\mathbf{y} - \hat{\mathbf{y}}) \\
&= \mathbf{X}'\mathbf{e} = [\mathbf{x}_1 \cdot \mathbf{e}\, \mathbf{x}_2 \cdot \mathbf{e} \cdots \mathbf{x}_k \cdot \mathbf{e}]' = \mathbf{0}.
\end{aligned}
$$

其中$\mathbf{x}_1 \cdot \mathbf{e}$表$\mathbf{x}_1$和$\mathbf{e}$兩個向量的內積 (inner product), 其值為0, 表示這兩個向量相互垂直 (orthogonal);

因此, 一階條件的正規方程式隱含殘差值向量$\mathbf{e}$與$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$均相互垂直, 而且

$$
\begin{aligned}
& \mathbf{y} \cdot \mathbf{e} \\
&= [\hat{\beta}_{1n}\mathbf{x}_1 + \hat{\beta}_{2n}\mathbf{x}_2 + \cdots + \hat{\beta}_{kn}\mathbf{x}_k] \cdot \mathbf{e} \\
&= [\hat{\beta}_{1n}\mathbf{x}_1 \cdot \mathbf{e} + \hat{\beta}_{2n}\mathbf{x}_2 \cdot \mathbf{e} + \cdots + \hat{\beta}_{kn}\mathbf{x}_k \cdot \mathbf{e}] \\
&= 0.
\end{aligned}
$$

即殘差值向量$\mathbf{e}$與$\hat{\mathbf{y}}$相互垂直, 且$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, 因此,$\hat{\mathbf{y}}$為向量$\mathbf{y}$垂直映射 (orthogonal project) 在由$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$向量所擴展的空間上。

由於

$$
\begin{aligned}
\hat{\beta}_n &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon,
\end{aligned}
$$

因此,$\hat{\beta}_n$的期望值為

$$
\begin{aligned}
E(\hat{\beta}_n) &= E[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon] \\
&= \beta + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon] \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon), \quad \text{因假設條件(2):} X \text{為非隨機} \\
&= \beta, \quad \text{因假設條件(3): } E(\epsilon_i) = 0
\end{aligned}
$$

所以,$\hat{\beta}_n$在假設條件 (1)、(2) 和 (3) 下, 不但是最小平方誤差的解, 也是真正參數β的不偏估計式

此外,$\hat{\beta}_n$的變異-共變異數矩陣為

$$
\begin{aligned}
\mathsf{VAR}(\hat{\beta}_n) &= \left[\begin{array}{ccccc}
\mathsf{var}(\hat{\beta}_{1n}) & \mathsf{cov}(\hat{\beta}_{1n},\hat{\beta}_{2n}) & \mathsf{cov}(\hat{\beta}_{1n},\hat{\beta}_{3n}) & \cdots & \mathsf{cov}(\hat{\beta}_{1n}, \\
\mathsf{cov}(\hat{\beta}_{2n},\hat{\beta}_{1n}) & \mathsf{var}(\hat{\beta}_{2n}) & \mathsf{cov}(\hat{\beta}_{2n}\hat{\beta}_{3n}) & \cdots & \mathsf{cov}(\hat{\beta}_{2n}, \\
\mathsf{cov}(\hat{\beta}_{3n},\hat{\beta}_{1n}) & \mathsf{cov}(\hat{\beta}_{3n},\hat{\beta}_{2n}) & \mathsf{var}(\hat{\beta}_{3n}) & \cdots & \mathsf{cov}(\hat{\beta}_{3n}, \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
\mathsf{cov}(\hat{\beta}_{kn},\hat{\beta}_{1n}) & \mathsf{cov}(\hat{\beta}_{kn},\hat{\beta}_{2n}) & \mathsf{cov}(\hat{\beta}_{kn},\hat{\beta}_{3n}) & \cdots & \mathsf{var}(\hat{\beta}_{
\end{array}\right. \\
&= \begin{bmatrix}
\sigma^2_{\hat{\beta}_{1n}} & \sigma_{\hat{\beta}_{1n},\hat{\beta}_{2n}} & \sigma_{\hat{\beta}_{1n},\hat{\beta}_{3n}} & \cdots & \sigma_{\hat{\beta}_{1n},\hat{\beta}_{kn}} \\
\sigma_{\hat{\beta}_{2n},\hat{\beta}_{1n}} & \sigma^2_{\hat{\beta}_{2n}} & \sigma_{\hat{\beta}_{2n}\hat{\beta}_{3n}} & \cdots & \sigma_{\hat{\beta}_{2n},\hat{\beta}_{kn}} \\
\sigma_{\hat{\beta}_{3n},\hat{\beta}_{1n}} & \sigma_{\hat{\beta}_{3n},\hat{\beta}_{2n}} & \sigma^2_{\hat{\beta}_{3n}} & \cdots & \sigma_{\hat{\beta}_{3n},\hat{\beta}_{kn}} \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
\sigma_{\hat{\beta}_{kn},\hat{\beta}_{1n}} & \sigma_{\hat{\beta}_{kn},\hat{\beta}_{2n}} & \sigma_{\hat{\beta}_{kn},\hat{\beta}_{3n}} & \cdots & \sigma^2_{\hat{\beta}_{kn}}
\end{bmatrix} \\
&= E[(\hat{\beta}_n - E(\hat{\beta}_n))(\hat{\beta}_n - E(\hat{\beta}_n))'] \\
&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon)((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon)'] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

其中,$E(\epsilon\epsilon')$為誤差向量的變異-共變異數矩陣, 其定義為

$$
\begin{aligned}
E(\epsilon\epsilon') &= \text{VAR}(\epsilon) \\
&= E\begin{bmatrix} \epsilon_1^2 & \epsilon_1\epsilon_2 & \epsilon_1\epsilon_3 & \cdots & \epsilon_1\epsilon_n \\ \epsilon_2\epsilon_1 & \epsilon_2^2 & \epsilon_2\epsilon_3 & \cdots & \epsilon_2\epsilon_n \\ \epsilon_3\epsilon_1 & \epsilon_3\epsilon_2 & \epsilon_3^2 & \cdots & \epsilon_3\epsilon_n \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \epsilon_n\epsilon_1 & \epsilon_n\epsilon_2 & \epsilon_n\epsilon_3 & \cdots & \epsilon_n^2 \end{bmatrix} \\
&= \begin{bmatrix} E(\epsilon_1^2) & E(\epsilon_1\epsilon_2) & E(\epsilon_1\epsilon_3) & \cdots & E(\epsilon_1\epsilon_n) \\ E(\epsilon_2\epsilon_1) & E(\epsilon_2^2) & E(\epsilon_2\epsilon_3) & \cdots & E(\epsilon_2\epsilon_n) \\ E(\epsilon_3\epsilon_1) & E(\epsilon_3\epsilon_2) & E(\epsilon_3^2) & \cdots & E(\epsilon_3\epsilon_n) \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ E(\epsilon_n\epsilon_1) & E(\epsilon_n\epsilon_2) & E(\epsilon_n\epsilon_3) & \cdots & E(\epsilon_n^2) \end{bmatrix} \\
&= \begin{bmatrix} \text{var}(\epsilon_1) & \text{cov}(\epsilon_1,\epsilon_2) & \text{cov}(\epsilon_1,\epsilon_3) & \cdots & \text{cov}(\epsilon_1,\epsilon_n) \\ \text{cov}(\epsilon_2,\epsilon_1) & \text{var}(\epsilon_2) & \text{cov}(\epsilon_2,\epsilon_3) & \cdots & \text{cov}(\epsilon_2,\epsilon_n) \\ \text{cov}(\epsilon_3,\epsilon_1) & \text{cov}(\epsilon_3,\epsilon_2) & \text{var}(\epsilon_3) & \cdots & \text{cov}(\epsilon_3,\epsilon_n) \end{bmatrix}
\end{aligned}
$$

又如同在簡單線性迴歸中,$\hat{\sigma}_n^2 = \sum_{i=1}^n e_i^2/(n-k) = \mathbf{e}'\mathbf{e}/(n-k)$是$\sigma_0^2$的不偏估計式, 因此,$\widehat{\text{VAR}}(\hat{\beta}_n) = \hat{\sigma}_n^2(\mathbf{X}'\mathbf{X})^{-1}$也是$\text{VAR}(\hat{\beta}_n) = \sigma_0^2(\mathbf{X}'\mathbf{X})^{-1}$的不偏估計式, 且定義

$$
\begin{aligned}
\widehat{\text{VAR}}(\hat{\beta}_n) &= \hat{\sigma}_n^2(\mathbf{X}'\mathbf{X})^{-1} \\
&= \left[\begin{array}{ccccc}
\widehat{\text{var}}(\hat{\beta}_{1n}) & \widehat{\text{cov}}(\hat{\beta}_{1n}, \hat{\beta}_{2n}) & \widehat{\text{cov}}(\hat{\beta}_{1n}, \hat{\beta}_{3n}) & \cdots & \widehat{\text{cov}}(\hat{\beta}_{1n}, \\
\widehat{\text{cov}}(\hat{\beta}_{2n}, \hat{\beta}_{1n}) & \widehat{\text{var}}(\hat{\beta}_{2n}) & \widehat{\text{cov}}(\hat{\beta}_{2n}\hat{\beta}_{3n}) & \cdots & \widehat{\text{cov}}(\hat{\beta}_{2n}, \\
\widehat{\text{cov}}(\hat{\beta}_{3n}, \hat{\beta}_{1n}) & \widehat{\text{cov}}(\hat{\beta}_{3n}, \hat{\beta}_{2n}) & \widehat{\text{var}}(\hat{\beta}_{3n}) & \cdots & \widehat{\text{cov}}(\hat{\beta}_{3n}, \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
\widehat{\text{cov}}(\hat{\beta}_{kn}, \hat{\beta}_{1n}) & \widehat{\text{cov}}(\hat{\beta}_{kn}, \hat{\beta}_{2n}) & \widehat{\text{cov}}(\hat{\beta}_{kn}, \hat{\beta}_{3n}) & \cdots & \widehat{\text{var}}(\hat{\beta}
\end{array}\right. \\
&= \begin{bmatrix}
s_{\hat{\beta}_{1n}}^2 & s_{\hat{\beta}_{1n},\hat{\beta}_{2n}} & s_{\hat{\beta}_{1n},\hat{\beta}_{3n}} & \cdots & s_{\hat{\beta}_{1n},\hat{\beta}_{kn}} \\
s_{\hat{\beta}_{2n},\hat{\beta}_{1n}} & s_{\hat{\beta}_{2n}}^2 & s_{\hat{\beta}_{2n}\hat{\beta}_{3n}} & \cdots & s_{\hat{\beta}_{2n},\hat{\beta}_{kn}} \\
s_{\hat{\beta}_{3n},\hat{\beta}_{1n}} & s_{\hat{\beta}_{3n},\hat{\beta}_{2n}} & s_{\hat{\beta}_{3n}}^2 & \cdots & s_{\hat{\beta}_{3n},\hat{\beta}_{kn}} \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
s_{\hat{\beta}_{kn},\hat{\beta}_{1n}} & s_{\hat{\beta}_{kn},\hat{\beta}_{2n}} & s_{\hat{\beta}_{kn},\hat{\beta}_{3n}} & \cdots & s_{\hat{\beta}_{kn}}^2
\end{bmatrix}.
\end{aligned}
$$

如前所述,$\hat{\beta}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$為$\mathbf{y}$的線性估計式, 且

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma_0^2 I_n),$$

因此,

$$\hat{\beta}_n \sim N(\beta, \sigma_0^2(\mathbf{X}'\mathbf{X})^{-1}).$$

或

$$[\sigma_0^2(\mathbf{X}'\mathbf{X})^{-1}]^{-1/2}(\hat{\beta}_n - \beta) \sim N(\mathbf{0}, I_n).$$

而就單一個迴歸參數估計式$\hat{\beta}_{in}$而言，

$$\hat{\beta}_{in} \sim N(\beta_i, \sigma^2_{\hat{\beta}_{in}}),$$

或

$$\frac{\hat{\beta}_{in} - \beta_i}{\sigma_{\hat{\beta}_{in}}} \sim N(0, 1).$$

再者，因

$$\frac{(n-k)\hat{\sigma}_n^2}{\sigma_0^2} \sim \chi^2(n-k),$$

因此，

$$\frac{\hat{\beta}_{in} - \beta_i}{s_{\hat{\beta}_{in}}} \sim t(n-k).$$