

08 線性迴歸

陳欣得

靜宜大學企管系

1

08 線性迴歸

迴歸分析 (*regression analysis*) 可以視為變異數分析的進階版。對於目標變數觀察值之總變異 (均差平方和) 的來源, 單因子 ANOVA 將此變異歸因於某因子之組間平均不完全相等, 雙因子 ANOVA 則延伸歸因於兩種因子分組以及因子間可能的交互作用。迴歸分析直接假設目標變數與因子 (解釋變數) 有一個函數關係; 若該函數為線性函數, 則稱為**線性迴歸分析** (*linear regression analysis*)。找出迴歸函數並做必要的假設檢定是迴歸分析的主要工作。

章節安排：

- 8.1 線性迴歸概論
- 8.2 簡單線性迴歸分析
- 8.3 迴歸模型係數的推導
- 8.4 考慮質性資料

2

§8.1 線性迴歸概論

迴歸分析與變異數分析中，我們有一個關心的主要變數，計算其觀察值的變異，並分析該變異的來源；此主要變數稱為**依變數**（*dependent variables*）。而用來解釋變異來源的因子或變數稱為**獨立變數**（*independent variables*）；相對於依變數只有一個，獨立變數可能會有一個以上。

3

迴歸分析嘗試建立依變數與獨立變數之間的函數關係（稱為**迴歸方程式**），如：

$$y = f(x_1, x_2, \dots, x_k)$$

其中， y 為依變數， $x_1, x_2, \dots, x_k$ 為獨立變數。以下是幾個常見的迴歸方程式：

$$(1) \quad y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k ;$$

$$(2) \quad y = \beta_0 x_1^{\beta_1} \dots x_k^{\beta_k} ;$$

$$(3) \quad y = \beta_0 e^{\beta_1 x_1 + \dots + \beta_k x_k} ;$$

$$(4) \quad y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}} .$$

其中，(1)為線性函數，(2)~(4)為非線性函數。另外，方程式中的參數， $\beta_0, \beta_1, \dots, \beta_k$ ，稱為**迴歸參數**（*regression parameters*），請留意，迴歸參數的數目比獨立變數的數目多一個。

4

若所建立的迴歸方程式為線性函數，則該迴歸分析稱為**線性迴歸分析**（*linear regression analysis*），否則稱為**非線性迴歸分析**（*nonlinear regression analysis*）；只有一個獨立變數的迴歸分析稱為**簡單迴歸分析**（*simple regression analysis*），若有多個獨立變數則稱為**複迴歸分析**（*multiple regression analysis*）。

第一年統計學的重點是**簡單線性迴歸分析**（*simple linear regression analysis*）。

最後，迴歸分析涉及的變數因分析角色的不同而有不同的名稱。依變數也稱為**Y 變數**、**因變數**、**反應變數**；獨立變數也稱為**X 變數**、**解釋變數**。

§8.2 簡單線性迴歸分析

只有一個獨立變數的線性迴歸稱為**簡單線性迴歸**（*simple linear regression*），其迴歸方程式如下：

$$y = \alpha + \beta x + \varepsilon \quad (8.1)$$

其中， α 、 β 為迴歸參數，或稱為**迴歸係數**（*regression coefficient*）， y 為依變數、 x 為獨立變數，誤差項 ε 為屬於 $N(0, \sigma)$ 的隨機變數。請留意，在迴歸分析中，依變數 y 與誤差項 ε 為隨機變數，而獨立變數 x 為一般變數。

假設有觀察值 (x_i, y_i) ， $i=1, \dots, n$ ，來驗證(8.1)的迴歸方程式，也就是

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i=1, \dots, n. \quad (8.2)$$

對於 ε_i 、 x_i 、 y_i ， $i=1, \dots, n$ ，我們有以下迴歸分析的假設：

- (1) $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \stackrel{i.i.d.}{\sim} N(0, \sigma)$ ；
- (2) $y_1, y_2, \dots, y_n \sim N(\alpha + \beta x_i, \sigma)$ ；
- (3) 所有 x_i ， $i=1, \dots, n$ ，都不是隨機變數。

請注意，這些假設隱含： $\text{cov}(y_i, y_j) = 0$ ， $\forall i \neq j$ ，以及 $\text{cov}(\varepsilon_j, x_i) = 0$ ， $\forall 1 \leq i, j \leq n$ 。

另外，可以很容易推導出以下結果：

$$\sqrt{\text{var}(y_i - (\alpha + \beta x_i))} = \sqrt{\text{var}(\varepsilon_i)} = \sigma,$$

這稱為依變數 y 的**測量標準誤** (*Standard Error of Estimate*)。

§8.2.1 迴歸參數的估計

迴歸分析的首要工作是以(8.2)的資料來估計(8.1)的母體參數 α 、 β ，相對應的估計量分別標示為 $\hat{\alpha}$ 與 $\hat{\beta}$ 。寫上估計量的迴歸方程式如下

$$(0.1) \quad \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i, \quad i=1, \dots, n.$$

這裡的 $\hat{y}_i$ 只是方便寫法，不可以說成 y_i 的估計量；母體參數才有估計量。

常用於產生迴歸參數估計量的方法有兩種：

- (1) 讓 y 之觀察值 y_i 與估計值 $\hat{y}_i$ ， $i=1, \dots, n$ ，間的誤差平方和最小為條件，來求取估計量 $\hat{\alpha}$ 與 $\hat{\beta}$ ，此方法稱為**最小平方法** (*Least Square method*)。
- (2) 讓觀察值 (x_i, y_i) ， $i=1, \dots, n$ ，出現機率最大之下，來求取估計量 $\hat{\alpha}$ 與 $\hat{\beta}$ ，此方法稱為**最大概似法** (*Maximum Likelihood Method*)。

我們接下來會介紹如何以最小平方法求取迴歸參數估計量。以最小平方法求得的估計量稱為**最小平方估計量** (*Ordinary Least Squares Estimator, OLSE*)。

求解下列最小值問題：

$$\min_{\hat{\alpha}, \hat{\beta}} SS = \sum_{i=1, \dots, n} \varepsilon_i^2 = \sum_{i=1, \dots, n} (y_i - \hat{y}_i)^2 = \sum_{i=1, \dots, n} (y_i - \hat{\alpha} - \hat{\beta}x_i)^2,$$

其解如下

$$\begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix}.$$

也就是

$$\hat{\beta} = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma x & \Sigma xy \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2};$$

$$\hat{\alpha} = \frac{\begin{vmatrix} \Sigma y & \Sigma x \\ \Sigma xy & \Sigma x^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \bar{y} - \hat{\beta}\bar{x}.$$

在實務紙筆計算的場合，先求出 $\hat{\beta}$ ，然後再代入公式求 $\hat{\alpha}$ ；這樣比較不容易出錯。另外，也請注意 $\hat{\beta}$ 與共變異數 s_{xy} 和相關係數 r_{xy} 間的關係：

$$\hat{\beta} = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x}.$$

©陳欣得 2025/8/25

(113-4 統計學(二)) 08線性迴歸

9

9

更多資訊8.1 (最小平方法的求解過程)

極值位於函數一次微分等於零的位置：

$$\begin{cases} \frac{d}{d\hat{\alpha}} SS = \sum_{i=1, \dots, n} 2(y_i - \hat{\alpha} - \hat{\beta}x_i) \times (-1) = -2(\Sigma y - n\hat{\alpha} - \hat{\beta}\Sigma x) = 0 \\ \frac{d}{d\hat{\beta}} SS = \sum_{i=1, \dots, n} 2(y_i - \hat{\alpha} - \hat{\beta}x_i) \times (-x_i) = -2(\Sigma xy - \hat{\alpha}\Sigma x - \hat{\beta}\Sigma x^2) = 0 \end{cases}$$

也就是 $(\hat{\alpha}, \hat{\beta})$ 為以下聯立方程式的解：

$$\begin{aligned} n \times \hat{\alpha} + \Sigma x \times \hat{\beta} &= \Sigma y \\ \Sigma x \times \hat{\alpha} + \Sigma x^2 \times \hat{\beta} &= \Sigma xy \end{aligned}$$

其解為：

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix} = \begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix} \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix} \Bigg/ \begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}.$$

©陳欣得 2025/8/25

(113-4 統計學(二)) 08線性迴歸

10

10

範例8.2 (迴歸參數的估計)

請就以下資料，計算迴歸方程式 $y = \alpha + \beta x$ 的參數估計量。

x	4	5	9	12
y	9	8	6	3

【解】

首先，整理原始資料如下：

n	Σx	Σy	Σx^2	Σy^2	Σxy	$ \Sigma $
4	30	26	266	190	166	164

$$\text{其中，} |\Sigma| = \begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix} = \begin{vmatrix} 4 & 30 \\ 30 & 266 \end{vmatrix} = 164。$$

則

$$\hat{\beta} = \frac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = \frac{4 \times 166 - 30 \times 26}{164} = -0.707；$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = \frac{26}{4} - (-0.707) \times \frac{30}{4} = 11.805。$$

迴歸估計量為

$$\hat{y} = \hat{\alpha} + \hat{\beta}x = 11.805 - 0.707x。$$

§8.2.2 迴歸模型的檢定

若以變異數分析的立場來看迴歸分析的檢定，則

$$\text{總變異 } SST = \Sigma (y_i - \bar{y})^2；$$

$$\text{組間變異 } SSB = \Sigma (\hat{y}_i - \bar{y})^2；$$

$$\text{隨機變異 } SSE = \Sigma (y_i - \hat{y}_i)^2。$$

在迴歸分析領域，組間變異改稱為迴歸變異，也就是

$$\text{組間變異 } SSB = \Sigma (\hat{y}_i - \bar{y})^2 \Rightarrow \text{迴歸變異 } SSR = \Sigma (\hat{y}_i - \bar{y})^2。$$

以上變異，在實務紙筆計算場合的計算順序如下：

$$\begin{aligned} SST &= \sum y^2 - \frac{(\sum y)^2}{n} ; \\ SSR &= \sum \hat{y}^2 - \frac{(\sum y)^2}{n} = \left(\hat{\alpha}^2 \times n + 2\hat{\alpha}\hat{\beta} \times \sum x + \hat{\beta}^2 \times \sum x^2 \right) - \frac{(\sum y)^2}{n} ; \\ SSE &= SST - SSR \circ \end{aligned}$$

其中， $\sum \hat{y}^2$ 可以直接由各 x_i ， $i=1,2,\dots,n$ ，計算出 $\hat{y}_i^2$ 後加總；但因計算個別 $\hat{y}_i^2$ 會涉及小數位數取捨問題，代入公式可解決該問題：

$$\sum \hat{y}^2 = \sum \left(\hat{\alpha} + \hat{\beta}x \right)^2 = \sum \left(\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}x + \hat{\beta}^2 x^2 \right) = \hat{\alpha}^2 \times n + 2\hat{\alpha}\hat{\beta} \times \sum x + \hat{\beta}^2 \times \sum x^2 \circ$$

迴歸分析的變異數分析表如表 0.1 所示，其中， k 為迴歸參數的數目，在目前的簡單線性迴歸， $k = 2$ 。這個變異數分析表所檢定的虛無假設為

$$H_0 : \alpha = \beta = 0 \circ$$

表 0.1 迴歸變異數分析表

變異來源	平方和	自由度	均方	F 值
迴歸變異	SSR	k-1	MSR = SSR / k-1	F = MSR / MSE
隨機變異	SSE	n-k	MSE = SSE / n-k	
總和	SST	n-1		

對於整體迴歸模型的解釋能力，除了變異數分析，我們也會以迴歸變異佔總變異的比例來衡量。該比例稱為**判定係數**（*Coefficient of Determination*， R^2 ），判定係數越大表示迴歸模型的解釋力越強。判定係數數學定義如下：

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}。$$

也請留意判定係數 R^2 以下關係：

$$R^2 = \frac{SSR}{SST} = \frac{\text{var}(\hat{y})}{\text{var}(y)} = \frac{\hat{\beta}^2 \text{var}(x)}{\text{var}(y)} = \hat{\beta}^2 \frac{s_x^2}{s_y^2} = \left(r_{xy} \frac{s_y}{s_x} \right)^2 \frac{s_x^2}{s_y^2} = (r_{xy})^2。$$

因為定義判定係數的迴歸變異 SSR 對自由度很敏感，實務上多使用以自由度調整後的判定係數，稱為**調整判定係數**（*Adjusted Coefficient of Determination*， R_{adj}^2 ），定義如下：

$$R_{adj}^2 = 1 - \frac{SSE/(n-k)}{SST/(n-1)} = 1 - \frac{MSE}{s_y^2}。$$

請注意， $R_{adj}^2 < R^2$ 。另外，雖然判定係數 R^2 的數值確實等於相關係數 r 的平方，在描述整體迴歸模型解釋力時需用 R^2 ，不要投機取巧地使用相關係數 r 或者 R 來誤導讀者。例如， $R^2 = 0.36$ 表示迴歸模型的解釋力為 36%；但是若寫成 $R = 0.6$ 或 $r = 0.6$ ，會讓讀者誤以為迴歸模型的解釋力為 60%。

範例8.2 (續) (迴歸分析之變異數分析表)

請寫出變異數分析表，並對整體迴歸模型作檢定。

【解】

已知

n	Σx	Σy	Σx^2	Σy^2	Σxy	$ \Sigma $
4	30	26	266	190	166	164

以及 $k=1$ ， $\hat{\alpha}=11.805$ ， $\hat{\beta}=-0.707$ 。

則

$$SST = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 21 ;$$

$$SSR = \left(11.805^2 \times 4 + 2 \times 11.805 \times (-0.707) \times 30 + (-0.707)^2 \times 266 \right) - \frac{26^2}{4} = 20.512 .$$

變異數分析表如下：

變異來源	平方和	自由度	均方	F
迴歸	20.512	1	20.512	84.100
隨機	0.488	2	0.244	
總和	21	3		

©陳欣得 2025/8/25

(113-4 統計學(二)) 08線性迴歸

17

17

範例8.2 (續) (迴歸分析之變異數分析表) (續)

變異數分析表如下：

變異來源	平方和	自由度	均方	F
迴歸	20.512	1	20.512	84.100
隨機	0.488	2	0.244	
總和	21	3		

檢定結果：

$$F^* = 84.1 > F_{\alpha=0.05, df=(1,2)} = 18.5128, \text{ 拒絕 } H_0 .$$

迴歸模型 $y = 11.805 - 0.707x$ 有顯著的解釋力。

其判定係數與調整判定係數分別如

$$R = \frac{SSR}{SST} = \frac{20.512}{21} = 0.9768, \quad R_{adj}^2 = 1 - \frac{SSE/n-k}{SST/n-1} = 1 - \frac{0.488/2}{21/3} = 0.9652 .$$

©陳欣得 2025/8/25

(113-4 統計學(二)) 08線性迴歸

18

18

範例8.2 (續) (迴歸分析之變異數分析表) (續2)

另外，在樣本數不大的場合，也可以直接計算 $\Sigma \hat{y}^2$ ，如下：

x	4	5	9	12	
y	9	8	6	3	總和
$\hat{y}$	8.976	8.268	5.439	3.317	26.000
$\hat{y}^2$	80.562	68.365	29.583	11.003	189.512

其中， $\hat{y}_1 = \hat{\alpha} + \hat{\beta}x_1 = 11.805 - 0.707 \times 4 = 8.976$ 。

則

$$SSR = \Sigma \hat{y}^2 - \frac{(\Sigma y)^2}{n} = 189.512 - \frac{26^2}{4} = 20.512。$$

§8.2.3 迴歸參數的檢定

迴歸參數的檢定如同變異數分析的事後分析，在整體模型分析拒絕 H_0 後，接者檢定個別參數是否等於零，即

$$H_0: \alpha = 0 \quad \text{以及} \quad H_0: \beta = 0。$$

接下來我們需要參數估計量的期望值與變異數。我們列出結果如下：

$$E(\hat{\alpha}) = \alpha, \quad \text{var}(\hat{\alpha}) = \frac{\Sigma x^2}{n\Sigma x^2 - (\Sigma x)^2} = \frac{\Sigma x^2}{|\Sigma|} \sigma^2;$$

$$E(\hat{\beta}) = \beta, \quad \text{var}(\hat{\beta}) = \frac{n}{n\Sigma x^2 - (\Sigma x)^2} \sigma^2 = \frac{n}{|\Sigma|} \sigma^2。$$

也就是 $\hat{\alpha}$ 與 $\hat{\beta}$ 的機率分配分別為

$$\hat{\alpha} \sim N\left(\alpha, \sqrt{\frac{\Sigma x^2}{|\Sigma|}} \sigma\right), \quad \hat{\beta} \sim N\left(\beta, \sqrt{\frac{n}{|\Sigma|}} \sigma\right)。$$

最後，如同 ANOVA 的結果，隨機項變異數 σ^2 的估計量為

$$\hat{\sigma}^2 = MSE \text{。}$$

因此，我們有檢定迴歸參數的檢定統計量，分別為

$$t_{df=n-k}^{\hat{\alpha}} = \frac{\hat{\alpha}}{s_{\hat{\alpha}}} = \frac{\hat{\alpha}}{\sqrt{\frac{\sum x^2}{|\Sigma|} \times MSE}} \quad \text{以及} \quad t_{df=n-k}^{\hat{\beta}} = \frac{\hat{\beta}}{s_{\hat{\beta}}} = \frac{\hat{\beta}}{\sqrt{\frac{n}{|\Sigma|} \times MSE}} \text{。}$$

其中， $|\Sigma| = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix} = n\sum x^2 + (\sum x)^2$ 。

模仿變異數分析表，我們也會做成如表 0.1 所示的迴歸分析表，有助於事後對個別參數的檢定工作。

表 0.1 迴歸分析表

參數	期望值	標準差	自由度	t 值
常數	$\hat{\alpha}$	$\sqrt{MSE \times \sum x^2 / \Sigma }$	n - k	$\hat{\alpha} / \sqrt{MSE \times \sum x^2 / \Sigma }$
x	$\hat{\beta}$	$\sqrt{MSE \times n / \Sigma }$	n - k	$\hat{\beta} / \sqrt{MSE \times n / \Sigma }$
R^2	SSR/SST			
R^2_{adj}	$1 - SSE /_{n-k} / SST /_{n-1}$			

更多資訊8.3 （迴歸參數估計量的變異數）

就 $\hat{\alpha}$ 與 $\hat{\beta}$ 之變異數，目前可以這樣記：

$$\text{var}\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = \begin{bmatrix} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\beta}) \\ \text{cov}(\hat{\alpha}, \hat{\beta}) & \text{var}(\hat{\beta}) \end{bmatrix} = \Sigma^{-1} \sigma^2 = \frac{1}{|\Sigma|} \begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix} \sigma^2 ,$$

也就是

$$\text{var}(\hat{\alpha}) = \frac{\Sigma x^2}{|\Sigma|} \sigma^2 , \quad \text{var}(\hat{\beta}) = \frac{n}{|\Sigma|} \sigma^2 ,$$

其中，

$$\Sigma = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix} , \quad |\Sigma| = n \Sigma x^2 - (\Sigma x)^2 .$$

在複迴歸章有完整的推導。

範例8.2 (續2) （迴歸分析之個別參數檢定）

請寫出迴歸分析表，並個別對參數作檢定。

【解】

已知

n	Σx	Σy	Σx^2	Σy^2	Σxy	$ \Sigma $
4	30	26	266	190	166	164

以及 $k=1$ ， $\hat{\alpha}=11.805$ ， $\hat{\beta}=-0.707$ ， $MSE=0.244$ 。

則

$$s_{\hat{\alpha}} = \sqrt{MSE \times \Sigma x^2 / |\Sigma|} = \sqrt{0.244 \times 266 / 164} = 0.629 ;$$

$$s_{\hat{\beta}} = \sqrt{MSE \times n / |\Sigma|} = \sqrt{0.244 \times 4 / 164} = 0.077 .$$

範例8.2 (續2) (迴歸分析之個別參數檢定) (續)

迴歸分析表如下：

參數	期望值	標準差	自由度	t
常數	11.805	0.629	2	18.77
x	-0.707	0.077	2	-9.171
R ²	0.9768			
R ² -adj	0.9652			
迴歸式	$y = 11.8 - 0.71x$			

檢定結果， $\hat{\alpha}$ 與 $\hat{\beta}$ 都顯著地不會是零：

對於 $\hat{\alpha}$ 的檢定： $t^* = 18.77 > t_{\alpha/2=0.025, df=2} = 4.3027$ ，拒絕 H_0 。

對於 $\hat{\beta}$ 的檢定： $t^* = |-9.171| > t_{\alpha/2=0.025, df=2} = 4.3027$ ，拒絕 H_0 。

§8.2.4 以迴歸方程式做預測

得到迴歸方程式後，如

$$\hat{y} = \hat{\alpha} + \hat{\beta}x,$$

代入某獨立變數，如 $x = x^*$ ，就可以知道依變數數值的估計量 $\hat{y}(x^*)$ ；這使得迴歸方程式成為一個很好的預測工具。

以上 $\hat{y}(x^*) = \hat{\alpha} + \hat{\beta}x^*$ 是點估計，若要作信賴區間，我們需要估計量的變異數 $\text{var}(\hat{y})$ 。這個變異數與所帶入之獨立變數數值的性質有關係；獨立變數數值可能是單一個體的數值， $x = x_d$ ，或是一個群體的平均值， $x = x_g$ 。例如，假設迴歸方程式的依變數是年薪，獨立變數是年資，則我們可能想預測年資 $x_d = 6$ 年之張三的年薪，也可能試圖預測平均年資 $x_g = 6$ 年之某群人的(平均)年薪。

若作個別預測，給單一個體的獨立變數值 $x = x_d$ ，代入迴歸估計式得到的 y 值為

$$y'_d = \hat{\alpha} + \hat{\beta}x_d + \varepsilon ;$$

若作群體預測，給單一個體的獨立變數值 $x = x_g$ ，代入迴歸估計式得到的 y 值為

$$y'_g = \hat{\alpha} + \hat{\beta}x_g \circ$$

其中，群體預測的測量誤差 ε 因取平均而消失。 y'_d 與 y'_g 的機率分配分別如下：

$$y'_d \sim N \left(\alpha + \beta x_d, \sigma \sqrt{\frac{1}{n} + \frac{n(\bar{x} - x_d)^2}{n \sum x^2 - (\sum x)^2} + 1} \right),$$

$$y'_g \sim N \left(\alpha + \beta x_g, \sigma \sqrt{\frac{1}{n} + \frac{n(\bar{x} - x_g)^2}{n \sum x^2 - (\sum x)^2}} \right) \circ$$

若以 $\sqrt{MSE}$ 代替 σ ，則兩者都適用自由度 $df = n - k$ 的 t 分配。

更多資訊8.4 （迴歸預測值的變異數）

就 $y'_g = \hat{\alpha} + \hat{\beta}x_g$ 之變異數，推導如下：

$$\begin{aligned} \text{var}(\hat{\alpha} + \hat{\beta}x_g) &= \text{var}(\hat{\alpha}) + 2x_g \text{cov}(\hat{\alpha}, \hat{\beta}) + x_g^2 \text{var}(\hat{\beta}) = \frac{\sum x^2 - 2x_g \sum x + nx_g^2}{|\Sigma|} \sigma^2 \\ &= \frac{\sum x^2 - (\sum x)^2/n + n(\bar{x}^2 - 2x_g \bar{x} + x_g^2)}{|\Sigma|} \sigma^2 = \left(\frac{1}{n} + \frac{n(\bar{x} - x_g)^2}{|\Sigma|} \right) \sigma^2 \end{aligned}$$

而誤差項 ε 與 $\hat{\alpha}$ 、 $\hat{\beta}$ 獨立，因此

$$\text{var}(\hat{\alpha} + \hat{\beta}x_d + \varepsilon) = \text{var}(\hat{\alpha} + \hat{\beta}x_d) + \text{var}(\varepsilon) = \left(\frac{1}{n} + \frac{n(\bar{x} - x_d)^2}{|\Sigma|} + 1 \right) \sigma^2 \circ$$

整理以上結果，得到

$$\begin{aligned} \text{var}(y'_d) &= \left(\frac{1}{n} + \frac{n(\bar{x} - x_d)^2}{n \sum x^2 - (\sum x)^2} + 1 \right) \sigma^2, \\ \text{var}(y'_g) &= \left(\frac{1}{n} + \frac{n(\bar{x} - x_g)^2}{n \sum x^2 - (\sum x)^2} \right) \sigma^2. \end{aligned}$$

範例8.2 (續3) (群體預測值的信賴區間)

假設有某團體的 x 平均值為 5，請寫出該團體之 y 值在 信賴度下的信賴區間。

【解】

已知

n	Sx	Sy	Sx ²	Sy ²	Sxy	S
4	30	26	266	190	166	164

以及 $k=1$ ， $\hat{\alpha}=11.805$ ， $\hat{\beta}=-0.707$ ， $MSE=0.244$ ， $SST=21$ ， $\bar{x}=7.5$ ， $x_g=5$ ， $1-\alpha=95\%$ 。

則，預測值之期望值與標準差分別為

$$E(y'_g) = \hat{\alpha} + \hat{\beta}x_g = 8.268；$$

$$s_{y'_g} = \sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_g)^2}{SST}\right)MSE} = 0.313。$$

最大容忍誤差與信賴區間分別為

$$\varepsilon = t_{\alpha/2=0.025, df=2} \times s_{y'_g} = 4.3027 \times 0.313 = 1.348；$$

$$CI_{y'_g} = \{E(y'_g) \pm \varepsilon\} = \{8.268 \pm 1.348\} = \{6.920 \leq y'_g \leq 9.616\}。$$

§8.2.5 可以轉換為線性的非線性模型

有些看起來是非線性的迴歸函數，經由簡單的變數變換，可以轉換成可以由線性迴歸技術處理的線性函數。

(1) $y = \alpha + \beta x^2$

令 $z = x^2$ ，則原式可以轉換為 $y = \alpha + \beta z$ ，這是一個線性函數。

(2) $y = \alpha e^{\beta x}$

等號兩端取對數，則原式成為 $\ln y = \ln(\alpha e^{\beta x}) = \ln \alpha + \beta x$ 。再令 $z = \ln y$ ， $a = \ln \alpha$ ，

則我們有 $z = a + \beta x$ 這個線性方程式。

$$(3) y = \alpha x^\beta$$

等號兩端取對數，則原式成為 $\ln y = \ln(\alpha x^\beta) = \ln \alpha + \beta \ln x$ 。再令 $z = \ln y$ ，
 $a = \ln \alpha$ ， $w = \ln x$ ，則我們有 $z = a + \beta w$ 這個線性方程式。

$$(4) e^y = \alpha x^\beta$$

等號兩端取對數，則原式成為 $y = \ln(\alpha x^\beta) = \ln \alpha + \beta \ln x$ 。再令 $a = \ln \alpha$ ， $w = \ln x$ ，
 則我們有 $y = a + \beta w$ 這個線性函數。

§8.2.6 殘差分析

在迴歸分析的模型中，觀察值 y_i 與其對應預測值 $\hat{y}_i$ 之差誤， $e_i = y_i - \hat{\alpha} - \hat{\beta}x_i$ ，稱為**殘差 (residuals)**；殘差可以視為誤差項 ε_i 的觀測值。以下是迴歸模型的假設：

$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 為 $N(0, \sigma)$ 的一致且獨立分配。

殘差分析藉由分析 e_i ， $i = 1, 2, \dots, n$ ，的分佈情形，來瞭解以上假設是否成立。另外，我們也想了解迴歸模型為線性函數的假設是否成立。總結，殘差分析重點有以下三項：

- (1) 線性假設是否成立？
- (2) $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 的一致假設， $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ ，是否成立？
- (3) $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 的獨立假設， $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$ ，是否成立？

以下列出常見殘差分析的名稱和它們處理的重點：

- (1) Lack of Fit test，作線性假設是否成立的檢定；
- (2) Kolmogorov – Smirnov test，作殘差項是否為常態分配的檢定；
- (3) Durbin-Watson test，作殘差項之獨立性是否成立的檢定；
- (4) Bartlett test，為同質性檢定，即殘差項之變數數是否相等的檢定。

33

在執行殘差分析之前，我們會畫出 (x_i, e_i) ， $i=1,2,\dots,n$ ，的點陣圖（*dot plot*）來作初步分析。圖 0.1 陳列幾個不同狀況的殘差點陣圖。

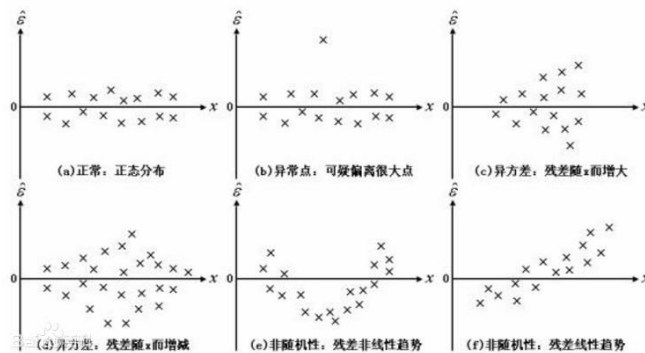


圖 0.1 殘差的點陣圖

34

§8.3 迴歸模型係數的推導

假設變數 y 、 x_1 、 x_2 、 $\dots$ 、 x_k 間有以下關係：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

其中，參數 $\beta_0, \beta_1, \beta_2, \dots, \beta_k \in R$ 。若這些變數的觀測值為 $(y_i, x_{i1}, x_{i2}, \dots, x_{ik})$ ， $i = 1, 2, \dots, n$ ，各次的測量誤差分別為 ε_i ， $i = 1, 2, \dots, n$ ，即

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (8.4)$$

其中， $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 假設為 $N(0, \sigma)$ 的 *i.i.d.*（一致且獨立分配）。也就是 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 有以下性質：

- (1) $\varepsilon_i, i = 1, 2, \dots, n$ 為常態分配；
- (2) $E(\varepsilon_1) = E(\varepsilon_2) = \dots = E(\varepsilon_n) = 0$ ；
- (3) $\text{var}(\varepsilon_1) = \text{var}(\varepsilon_2) = \dots = \text{var}(\varepsilon_n) = \sigma^2$ ；
- (4) $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ ， $\forall i \neq j$ 。

更進一步假設，測量誤差完全來自 y_i ，亦即假設 y_i 為隨機變數，且

- (1) $\text{var}(y_1) = \text{var}(y_2) = \dots = \text{var}(y_n) = \sigma^2$ ；
- (2) $\text{cov}(y_i, y_j) = 0$ ， $\forall i \neq j$ 。

獨立變數 $x_{i1}, x_{i2}, \dots, x_{ik}$ 只是一般變數（不是隨機變數）。以上描述的是**多變量線性迴歸模型**（*multivariable linear regression*），也稱為**多元線性迴歸**。

將(8.4)之觀測資料迴歸方程式寫成矩陣形式，如下

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} , \quad (8.5)$$

其中，

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} .$$

由假設得知隨機變數 $\mathbf{y}$ 與 $\boldsymbol{\varepsilon}$ 的變異共變異矩陣：

$$\text{var}(\mathbf{y}) = \text{var}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n ,$$

37

其中， $\mathbf{I}_n$ 為 n 階的單位矩陣。而且

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} ,$$

其中，

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} .$$

38

以最小平方方法估計 β

誤差項的平方和為

$$SSE = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) ,$$

令其微分值為零來取極值，

$$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = -2\mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 2(\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} - \mathbf{X}^T \mathbf{y}) = 0$$

此方程式的解即為估計量：

$$\mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}^T \mathbf{y} = \mathbf{0} \Rightarrow \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y} \Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} .$$

迴歸參數估計量的期望值與變異數

參數估計量 $\hat{\boldsymbol{\beta}}$ 的期望值推導如下：

$$E(\hat{\boldsymbol{\beta}}) = E\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}\right) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} ,$$

因此， $\hat{\boldsymbol{\beta}}$ 是一個不偏估計量。其變異數為

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}) &= V\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}\right) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T V(\mathbf{y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\sigma^2 \mathbf{I}_n) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned}$$

預測值的期望值與變異數

令 $\mathbf{x}_d$ 與 $\mathbf{x}_g$ 分別為某一個體與某一群體之 $x_1, x_2, \dots, x_k$ 觀測值，假設則依變數 y 的估計值分別為 y_d 與 y_g ，則

$$y_d = \mathbf{x}_d^T \hat{\boldsymbol{\beta}} + \varepsilon, \quad y_g = \mathbf{x}_g^T \hat{\boldsymbol{\beta}}.$$

個體觀測值尚有測量誤差，群體觀測值因以取平均，故沒有測量誤差項。 y_d 與 y_g 之期望值與變異數分別如下：

$$E(y_d) = E(\mathbf{x}_d^T \hat{\boldsymbol{\beta}} + \varepsilon) = \mathbf{x}_d^T E(\hat{\boldsymbol{\beta}}) = \mathbf{x}_d^T \boldsymbol{\beta};$$

$$\text{var}(y_d) = \text{var}(\mathbf{x}_d^T \hat{\boldsymbol{\beta}} + \varepsilon) = \mathbf{x}_d^T \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_d + \sigma^2 = \sigma^2 \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d + \sigma^2;$$

$$E(y_g) = E(\mathbf{x}_g^T \hat{\boldsymbol{\beta}}) = \mathbf{x}_g^T E(\hat{\boldsymbol{\beta}}) = \mathbf{x}_g^T \boldsymbol{\beta};$$

$$\text{var}(y_g) = \text{var}(\mathbf{x}_g^T \hat{\boldsymbol{\beta}}) = \mathbf{x}_g^T \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_g = \sigma^2 \mathbf{x}_g^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_g.$$

迴歸的變異數分析

在 $\boldsymbol{\beta}$ 之估計值 $\hat{\boldsymbol{\beta}}$ 計算出來後，我們有兩組應變數：觀測值 $\mathbf{y}$ 、以及其估計值 $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ 。由觀測值計算來的均差平方和為總變異(SST)，而從估計值計算來的均差平方和為迴歸變異(SSR)，此兩者的差值為無法解釋的變異(SSE)。以下是變異數分析需要的相關資料。

$$SST = (\mathbf{y} - \bar{\mathbf{y}})^T (\mathbf{y} - \bar{\mathbf{y}}), \quad \text{自由度 } df = n - 1;$$

$$SSR = (\hat{\mathbf{y}} - \bar{\mathbf{y}})^T (\hat{\mathbf{y}} - \bar{\mathbf{y}}), \quad \text{自由度 } df = (k+1) - 1;$$

$$SSE = SST - SSR = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}), \quad \text{自由度 } df = n - (k+1);$$

$$MSR = \frac{SSR}{(k+1)-1}, \quad MSE = \frac{SSE}{n-(k+1)}, \quad F = \frac{MSR}{MSE}.$$

請注意，我們一共有 $(k+1)$ 個需要估計的參數 $\boldsymbol{\beta}$ (k 個 x 變數之係數和一個常數項)。

判定係數

判定係數是迴歸變異（可解釋變異）佔總變異的百分比：

$$R^2 = \frac{SSR}{SST} \times 100\% = \left(1 - \frac{SSE}{SST}\right) \times 100\%。$$

修正判定係數是判定係數中的 SSE、SST 要作自由度的修正：

$$\hat{R}^2 = \left(1 - \frac{SSE/n - (k+1)}{SST/n - 1}\right) \times 100\%。$$

43

σ^2 的估計量

我們以 MSE 作為 σ^2 的估計量，即

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n - (k+1)} = \frac{(\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}})}{n - (k+1)} = \frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}}{n - (k+1)}。$$

44

§8.3.1 再看簡單線性迴歸的參數估計

簡單線性迴歸， $y = \alpha + \beta x$ ，的資料寫成矩陣形式為

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}。$$

則

$$\mathbf{X}^T \mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix}；$$

$$\Sigma = \mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix}。$$

45

因此 α 、 β 的估計量為

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix} = \frac{\begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} \begin{bmatrix} \Sigma y \\ \Sigma xy \end{bmatrix}。$$

$\hat{\alpha}$ 、 $\hat{\beta}$ 的變異數分別為

$$\text{var} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\beta}) \\ \text{cov}(\hat{\alpha}, \hat{\beta}) & \text{var}(\hat{\beta}) \end{bmatrix} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 \frac{\begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}}。$$

46

群體預測值， $y'_g = \hat{\alpha} + \hat{\beta}x_g = \begin{bmatrix} 1 & x_g \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$ ，的變異數為

$$\text{var}(y'_g) = \text{var}(\hat{\alpha} + \hat{\beta}x_g) = \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x} = \sigma^2 \frac{\begin{bmatrix} 1 & x_g \end{bmatrix} \begin{bmatrix} \Sigma x^2 & -\Sigma x \\ -\Sigma x & n \end{bmatrix} \begin{bmatrix} 1 \\ x_g \end{bmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}}。$$

個別預測值， $y'_d = \hat{\alpha} + \hat{\beta}x_d + \varepsilon = \begin{bmatrix} 1 & x_d \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} + \varepsilon$ ，的變異數為

$$\text{var}(y'_d) = \text{var}(\hat{\alpha} + \hat{\beta}x_d + \varepsilon) = \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x} + \sigma^2。$$

47

§8.4 考慮質性資料

迴歸模式中如有質性資料，須以**虛擬變數** (*dummy variables*) 替換。

48

範例8.5 （處理質性資料）

請就以下資料找迴歸方程式：

y	x
9	3
8	3
5	2
3	1
6	2
4	1

其中，獨立變數 x 為名目尺度資料。

【解】

我們需將 x 變數轉換成三個虛擬變數： $x_{D1}, x_{D2}, x_{D3} \in \{0,1\}$ 。其規則為

$$x=1 \rightarrow (x_{D1}, x_{D2}, x_{D3}) = (1, 0, 0) ;$$

$$x=2 \rightarrow (x_{D1}, x_{D2}, x_{D3}) = (0, 1, 0) ;$$

$$x=3 \rightarrow (x_{D1}, x_{D2}, x_{D3}) = (0, 0, 1) 。$$

49

範例8.5 （處理質性資料）(續)

資料轉換的結果如下表

y	x_{D1}	x_{D2}	x_{D3}
9	0	0	1
8	0	0	1
5	0	1	0
3	1	0	0
6	0	1	0
4	1	0	0

新的迴歸模型為

$$y = \beta_0 + \beta_1 x_{D1} + \beta_2 x_{D2} + \beta_3 x_{D3} 。$$

然而，因為 $x \neq 0$ ，故常數項係數 β_0 沒有意義。實務上的作法是去掉第一個虛擬變數 x_{D1} ，也就是說，我們以下列資料作迴歸分析：

50

範例8.5 （處理質性資料）(續2)

也就是說，我們以下列資料作迴歸分析：

y	x _{D2}	x _{D3}
9	0	1
8	0	1
5	1	0
3	0	0
6	1	0
4	0	0

迴歸模型為

$$y = \beta_0 + \beta_2 x_{D2} + \beta_3 x_{D3} \circ$$

51

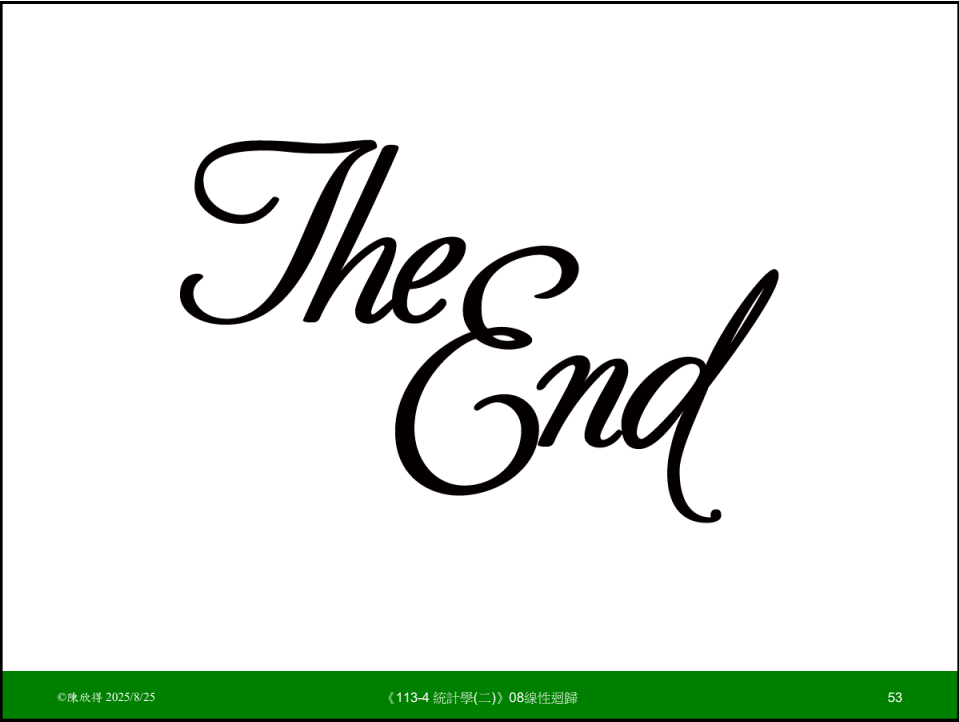
範例8.5 （處理質性資料）(續3)

將計算後，整理成迴歸分析表與迴歸變異數分析表分別如下：

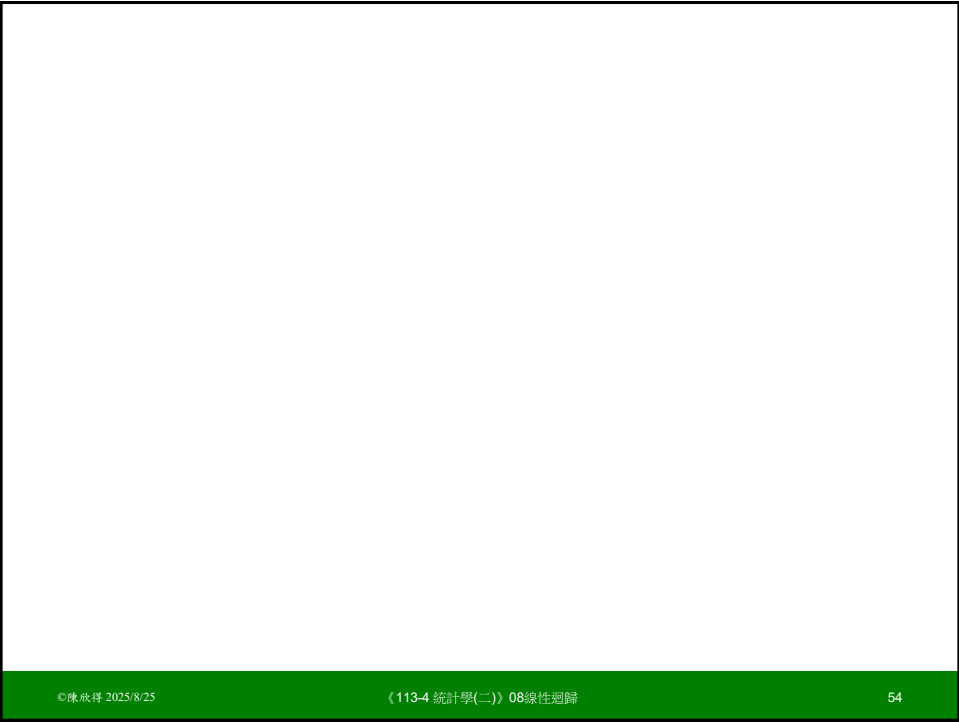
參數	期望值	標準差	自由度	t	p	t*
常數	3.500	0.500	3	7.000	0.006	3.1824
X _{d2}	2.000	0.707	3	2.828	0.066	3.1824
X _{d3}	5.000	0.707	3	7.071	0.006	3.1824
R ²	0.944					
R ² -adj	0.907				α =	0.05
迴歸式	$y = 3.5 + 2 \cdot X_{d2} + 5 \cdot X_{d3}$					
變異來源	平方和	自由度	均方	F	p	F*
迴歸	25.333	2	12.667	25.333	0.013	9.5521
隨機	1.500	3	0.500			
總和	26.833	5			α =	0.05

當 $x = 1$ 時， y 的估計值為 $y_{x=1} = \beta_0 = 3.5$ ； $x = 2$ 時， y 的估計值為 $y_{x=2} = \beta_0 + \beta_2 = 5.5$ ；
 $x = 3$ 時， y 的估計值為 $y_{x=3} = \beta_0 + \beta_3 = 8.5$ 。

52



53



54