

第 8 章

計量與質性預測變數之迴歸模型

Regression Models for Quantitative and Qualitative Predictors

8.1 多項式迴歸模型

Polynomial Regression Models

多項式迴歸模型的使用

- 多項式迴歸模型兩個基本的使用型態為：
 1. 真正的曲線反應函數的確為一個多項式函數。
 2. 真正的曲線反應函數未知或是非常複雜，但是透過多項式迴歸模型可以做出不錯的近似效果。
- 採用多項式迴歸模型應注意的是**外插**風險，尤其是在**高階的多項式**上，雖然多項式迴歸模型對手上的資料做出不錯的配適，不過一旦在資料範圍外進行外插時，很有可能會轉移到預期外的方向上。

單一預測變數的第二階模型

- 多項式迴歸模型可以包含一、二或更多個預測變數，並且每一個預測變數可以有**多種不同的乘冪**，我們先從具有**第一、二階乘冪的單一預測變數之多項式迴歸模型**開始：

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad (8.1)$$

其中， $x_i = X_i - \bar{X}$

- 多項式迴歸模型只有一個預測變數時，但模型中有第二階乘冪項，所以稱為**單一預測變數的第二階模型**（second-order model with one predictor variable）所以這裡的預測變數已經經過**置中**的程序 $x_i = X_i - \bar{X}$ ，也就是取了對 \bar{X} 的離差。
- 預測變數採用**置中**的理由是因為**X與X²經常有高度的相關性**，而造成利用標準方程式計算 $(X'X)^{-1}$ 的嚴重困難。

- 多項式迴歸模型主要是在反應指數的型態，在此形式為：

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_i \quad (8.2)$$

迴歸模型(8.2)之反應函數為：

$$E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 \quad (8.3)$$

此一反應函數為一個拋物線，習慣上稱為二次反應函數，圖8.1為兩個第二階多項式反應函數之例子。

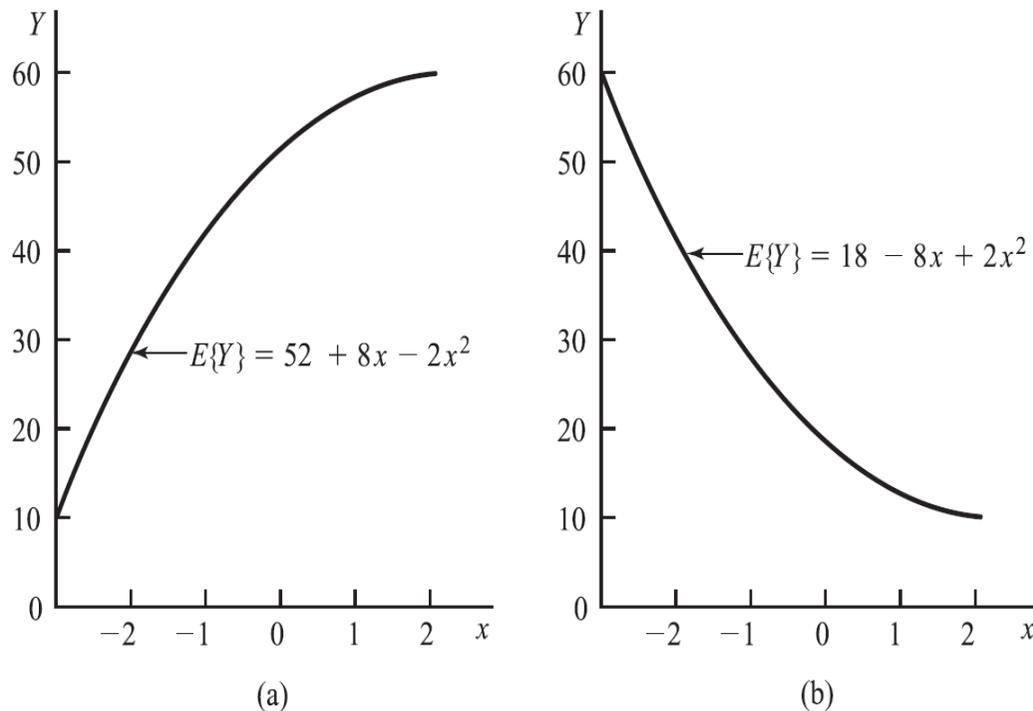
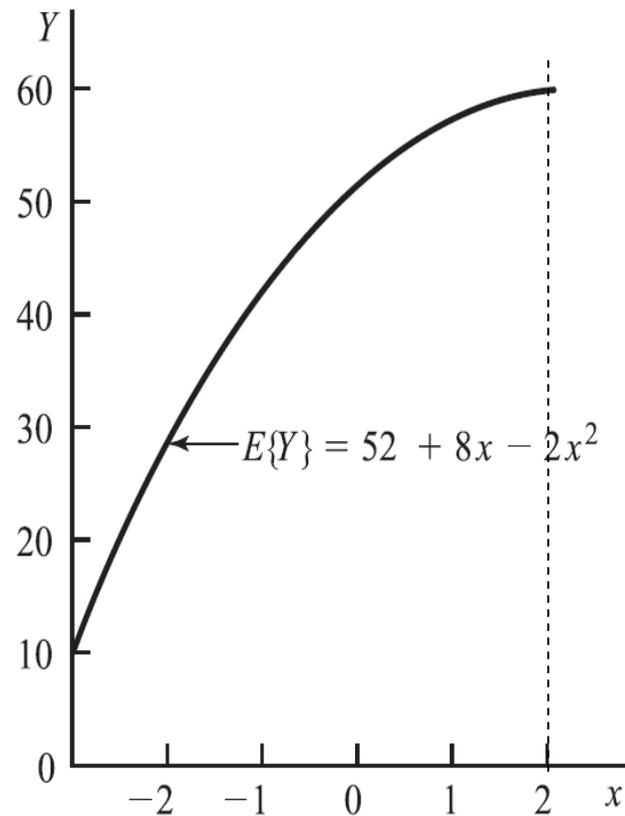


圖 8.1
第二階多項式反應
函數之例。

多項式迴歸模型 $Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_i$

- 這裡的迴歸係數 β_0 表示當 $x = 0$ 時， Y 的平均反應，迴歸係數 β_1 稱為線性效應係數（linear effect coefficient），迴歸係數 β_{11} 稱為二次效應係數（quadratic effect coefficient）。
- 多項式反應函數之外插風險可用圖8.1a的反應函數來說明，如果在超過 $x = 2$ 的地方做外插，則此時函數型態已經開始有向下的趨勢，而可能與原反應函數不配合。



(a)

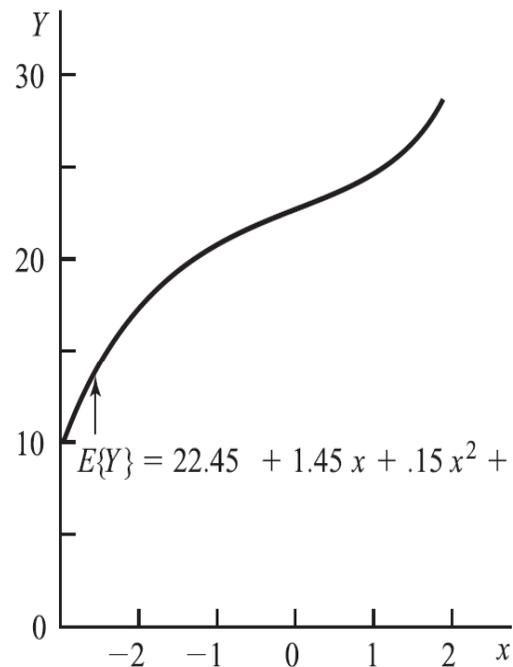
單一預測變數的第三階模型

$$\text{迴歸模型：} \quad Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i \quad (8.5)$$

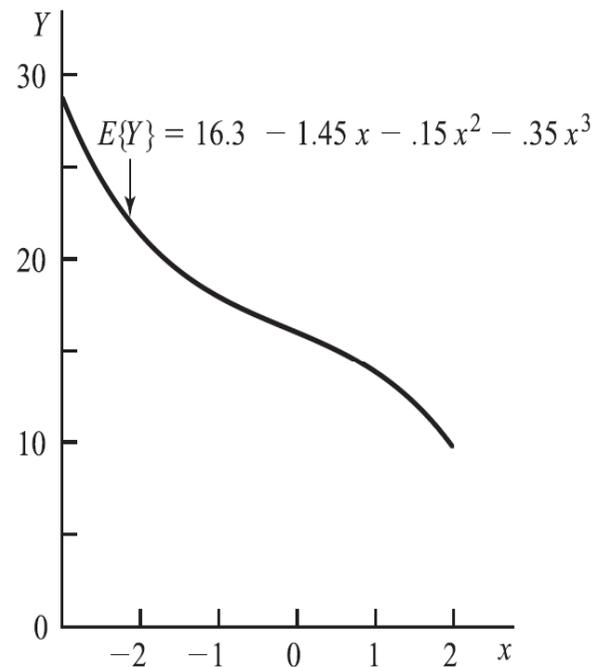
$$\text{其中，} \quad x_i = X_i - \bar{X}$$

$$\text{反應函數為：} \quad E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 + \beta_{111} x^3 \quad (8.6)$$

圖8.2是兩個第三階多項式反應函數的例子。



(a)



(b)

圖 8.2
第三階多項式反應
函數之例。

單一預測變數的更高階模型

- 採用乘幕高於第三階之預測變數的多項式迴歸模型需特別小心，由於這種迴歸係數不易解釋，同時在模型進行內插或外插時也可能有很意外的結果。
- 由於在無重複觀察的模型配適中，足夠高階的多項式迴歸模型將可能做到完全配適。

雙預測變數的第二階模型

迴歸模型：

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i \quad (8.7)$$

其中 $x_{i1} = X_{i1} - \bar{X}_1$

$$x_{i2} = X_{i2} - \bar{X}_2$$

迴歸模型之反應函數為：

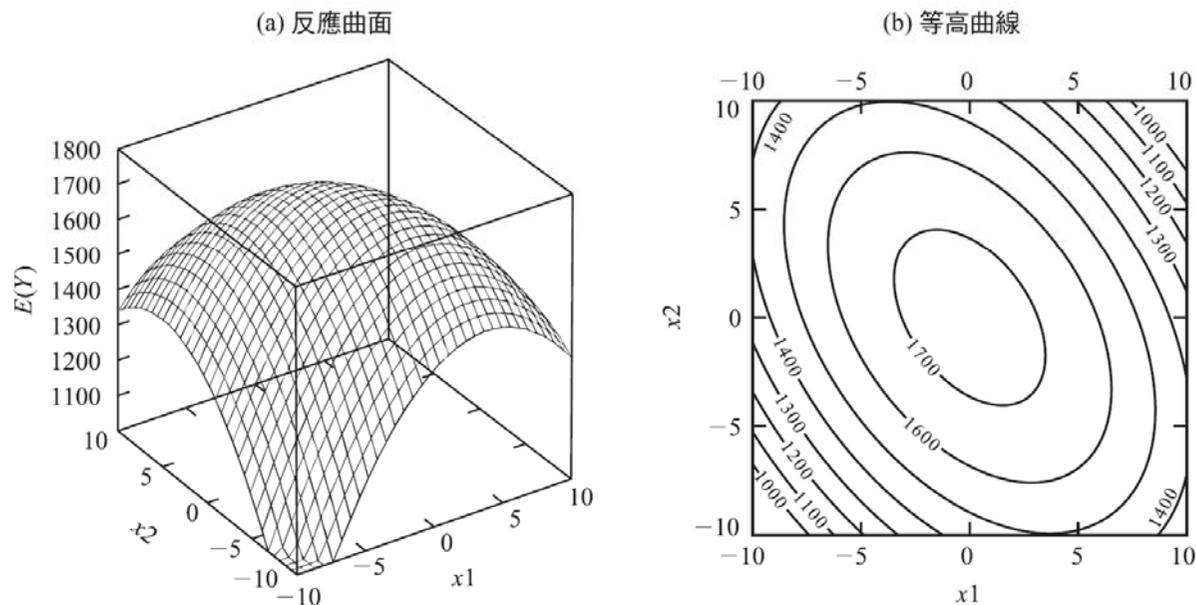
$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \quad (8.8)$$

- 它是一個圓錐切面方程式，在迴歸模型(8.7)中包含了每一個預測變數的線性、二次式與一個雙預測變數的交叉乘積項，該交叉乘積項代表了這兩個預測變數的交互作用，而係數 β_{12} 稱為交互作用係數（interaction effect coefficient）
- 圖8.3是雙預測變數的第二階反應函數例子：

$$E\{Y\} = 1,740 - 4x_1^2 - 3x_2^2 - 3x_1x_2$$

圖 8.3

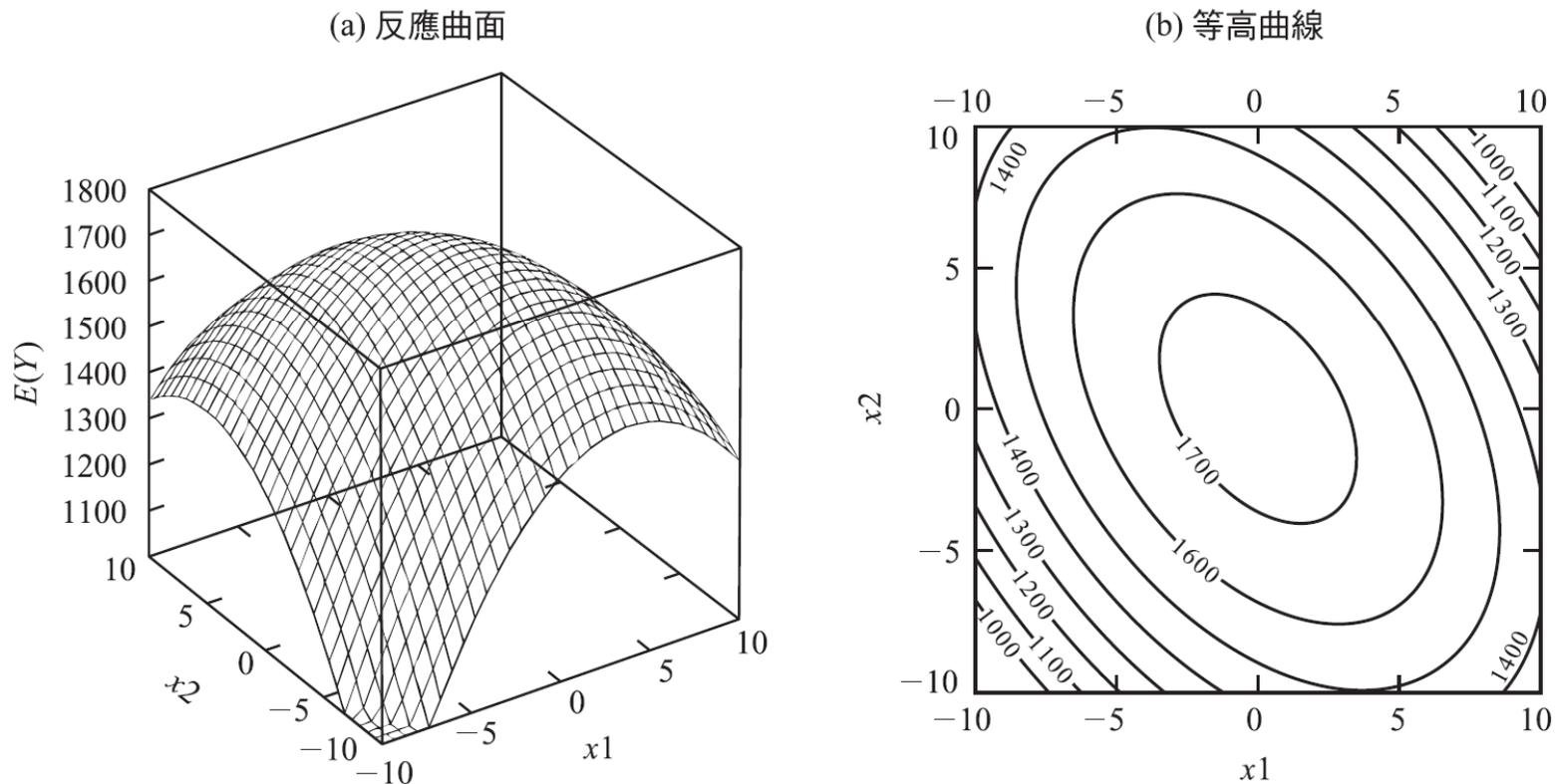
二次式反應曲面之例： $E[Y]=1,740-4x_1^2-3x_2^2-3x_1x_2$ 。



- 圖中所示為反應曲面與等高曲線，等高曲線對應於不同的預測變數水準下，將相同大小的反應值連起來，顯示出不同的預測變數組合，圖8.3a的反應曲面 $x_1=x_2=0$ 時達到最高點，該圖中呈現一個具有鞍點的特別型態。

圖 8.3

二次式反應曲面之例： $E[Y]=1,740-4x_1^2-3x_2^2-3x_1x_2$ 。



三預測變數的第二階模型

考慮下面的迴歸模型：

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{33} x_{i3}^2 \quad (8.9) \\ + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} + \varepsilon_i$$

其中 $x_{i1} = X_{i1} - \bar{X}_1$

$$x_{i2} = X_{i2} - \bar{X}_2$$

$$x_{i3} = X_{i3} - \bar{X}_3$$

稱為三預測變數的第二階模型，迴歸模型之反應函數為：

$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \quad (8.10) \\ + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

係數 β_{12} 、 β_{13} 、 β_{23} 稱為成對預測變數之交互作用係數。

多項式迴歸模型的實作

- 配適多項式模型

只是一般線性模型（6.7）的特例，因此前面與模型配適有關的推論結果在此均可適用。

- 階層式模型配適法

多項式迴歸模型近似一個真實的迴歸函數時，統計學家經常會先配適第二階或第三階模型，然後探索是否適當的較低階模型可用，例如先配適下面這個單一預測變數的模型：

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i$$

接下來看看是否可以去掉立方項，甚至是二次項，因此需檢定 $\beta_{111} = 0$ ，甚至是 $\beta_{11} = 0 = \beta_{111} = 0$ ，所以額外平方和 SSR 之分解依照下列順序進行：

SSR (x)

SSR (x²|x)

SSR (x³|x, x²)

- 額外平方和SSR (x³|x, x²) 適用於檢定 $\beta_{111}=0$ ，而檢定模型是否僅為線性，亦即檢定 $\beta_{11}=0=\beta_{111}=0$ ，則可以利用SSR (x², x³|x) = SSR (x²|x) + SSR (x³|x, x²) 進行檢定。
- 在階層式模型配適法中，當多項式中的某一個乘幕項被保留在模型中時，則所有相關的較低階項也應該保留在模型中，因此不可以保留三次項，卻剔除二次項。
- 階層式模型配適法中關於多個預測變數的檢定類似前面所介紹，例如保留一個二階交互作用項，則相關的一階預測變數項都必須保留在模型中。

- 以 X 表示的迴歸函數

當多項式迴歸模型產生後，通常我們會考慮使用原始變數，而非置中後的變數，例如以置中變數 $x = X - \bar{X}$ 表示配適的單預測變數第二階模型：

$$\hat{Y} = b_0 + b_1x + b_{11}x^2 \quad (8.11)$$

用原始變數表示為：

$$\hat{Y} = b'_0 + b'_1X + b'_{11}X^2 \quad (8.12)$$

其中，

$$b'_0 = b_0 - b_1\bar{X} + b_{11}\bar{X}^2 \quad (8.12a)$$

$$b'_1 = b_1 - 2b_{11}\bar{X} \quad (8.12b)$$

$$b'_{11} = b_{11} \quad (8.12c)$$

- 以原始變數 X 表示的迴歸函數其配適值與殘差，跟用置中變數 x 表示的迴歸函數其配適值與殘差相同，如同前面曾經提過的，採用置中變數的目的僅在於減輕多項式迴歸模型中，預測變數間存在多重共線性問題所造成的計算負擔。

個案實例

案例介紹

- 一位研究人員在新研發電池的初步小型實驗中，正在進行關於充電率與溫度對於新研發電池的效應影響之研究，充電率 (x_1) 控制在三個水準下 (0.6, 1.0, 1.4) (安培)，周圍溫度 (x_2) 也控制在三個水準下 (10, 20, 30) ($^{\circ}\text{C}$)，同時控制影響電池放電的相關水準在固定的條件之下，電池壽命 (Y) 用失效前的放電充電循環次數表示，表8.1的第一、二、三欄為實驗進行結果的資料。

這位研究人員不能確定在各因子的研究範圍內，反應函數的性質，因此他先配適了一個第二階多項式迴歸模型

(8.7)：

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i \quad (8.13)$$

而其反 $E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$

$$(8.14)$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i$$

表 8.1

電池案例的資料。

實驗電池 編號	(1)	(2)	(3)	(4) (5)		(6)	(7)	(8)
	循環次數	充電率	溫度	編碼值				
i	Y_i	X_{i1}	X_{i2}	x_{i1}	x_{i2}	x_{i1}^2	x_{i2}^2	$x_{i1}x_{i2}$
1	150	.6	10	-1	-1	1	1	1
2	86	1.0	10	0	-1	0	1	0
3	49	1.4	10	1	-1	1	1	-1
4	288	.6	20	-1	0	1	0	0
5	157	1.0	20	0	0	0	0	0
6	131	1.0	20	0	0	0	0	0
7	184	1.0	20	0	0	0	0	0
8	109	1.4	20	1	0	1	0	0
9	279	.6	30	-1	1	1	1	-1
10	235	1.0	30	0	1	0	1	0
11	224	1.4	30	1	1	1	1	1
		$\bar{X}_1 = 1.0$	$\bar{X}_2 = 20$					

資料來源：S. M. Sidik, H. F. Leibecki, and J. M. Bozek, *Cycles Till Failure of Silver-Zinc Cells with Competing Failure Modes—Preliminary Data Analysis*, NASA Technical Memorandum 815-56, 1980.

為了顧及 X_1 與 X_2 所取水準之平衡性，研究人員將預測變數 X_1 與 X_2 置中於各自的平均數後，再將尺度化為方便比較的單位：

$$\begin{aligned}x_{i1} &= \frac{X_{i1} - \bar{X}_1}{.4} = \frac{X_{i1} - 1.0}{.4} \\x_{i2} &= \frac{X_{i2} - \bar{X}_2}{10} = \frac{X_{i2} - 20}{10}\end{aligned}\quad (8.15)$$

- 上式所用的分母是採用各預測變數水準間的差距，轉換後的預測變數資料在表8.1的第四、五欄中，透過(8.15)的轉換後，預測變數將產生新的編碼值： -1 、 0 、 1 ，其平方與交叉乘積在表8.1的第六、七、八欄中。

- 透過編碼轉換程序後，第一階與第二階乘幕第相關性明顯降低許多：

相關係數	相關係數
x_1 與 x_1^2 : 0.991	x_2 與 x_2^2 : 0.986
x_1 與 x_1^2 : 0.000	x_2 與 x_2^2 : 0.000

由於兩個解釋變數之實驗水準採取平衡設計，所以透過編碼轉換程序後相關係數為零，類似的情形，交叉乘積 x_1x_2 與 x_1 、 x_1^2 、 x_2 、 x_2^2 的相關係數也變為零，降低了多重共線性之程序將有助於避免計算上的不精確問題。

•配適模型

圖8.4是利用統計套裝軟體SAS的迴歸程序，進行模型（8.13）配適的輸出結果，可以得到估計的迴歸函數為：

$$\hat{Y} = 162.84 - 55.83x_1 + 75.50x_2 + 27.39x_1^2 - 10.61x_2^2 + 11.50x_1x_2 \quad (8.16)$$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	162.842105	16.60760542	9.805	0.0002
X1	1	-55.833333	13.21670483	-4.224	0.0083
X2	1	75.500000	13.21670483	5.712	0.0023
X1SQ	1	27.394737	20.34007956	1.347	0.2359
X2SQ	1	-10.605263	20.34007956	-0.521	0.6244
X1X2	1	11.500000	16.18709146	0.710	0.5092

Variable	DF	Type I SS
INTERCEP	1	325424
X1	1	18704
X2	1	34202
X1SQ	1	1645.966667
X2SQ	1	284.928070
X1X2	1	529.000000

圖 8.4
第二階多項式模型
(8.13)的 SAS 迴歸
程序輸出結果－電
池案例。

Model: MODEL1
 Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	55365.56140	11073.11228	10.565	0.0109
Error	5	5240.43860	1048.08772		
C Total	10	60606.00000			
Root MSE		32.37418	R-square	0.9135	
Dep Mean		172.00000	Adj R-sq	0.8271	
C.V.		18.82220			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	162.842105	16.60760542	9.805	0.0002
X1	1	-55.833333	13.21670483	-4.224	0.0083
X2	1	75.500000	13.21670483	5.712	0.0023
X1SQ	1	27.394737	20.34007956	1.347	0.2359
X2SQ	1	-10.605263	20.34007956	-0.521	0.6244
X1X2	1	11.500000	16.18709146	0.710	0.5092

Variable	DF	Type I SS
INTERCEP	1	325424
X1	1	18704
X2	1	34202
X1SQ	1	1645.966667
X2SQ	1	284.928070
X1X2	1	529.000000

圖 8.4

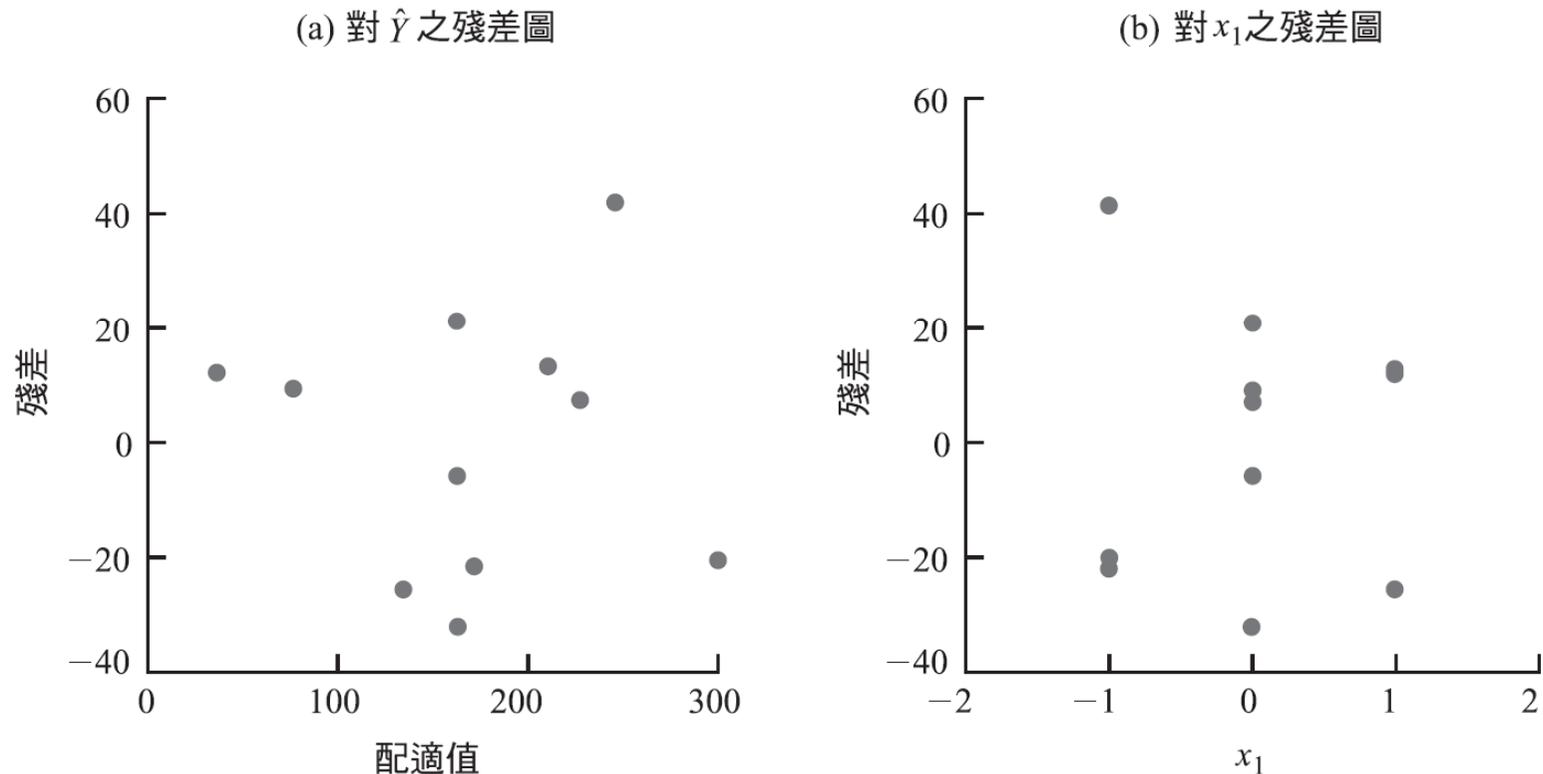
第二階多項式模型
 (8.13)的 SAS 迴歸
 程序輸出結果－電
 池案例。

殘差圖

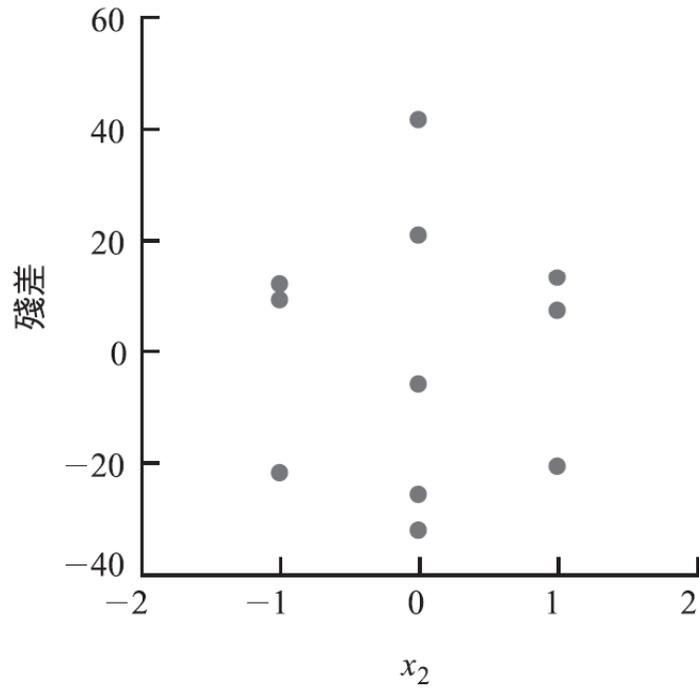
檢查迴歸模型 (8.13) 中對 \hat{Y} 、 x_1 、 x_2 之殘差圖及常態機率圖如圖 8.5，並無發現配適不佳之處，排序後的殘差與其常態期望值的相關係數達到 0.974，表 B6 支持誤差項服從常態分配之假設

圖 8.5

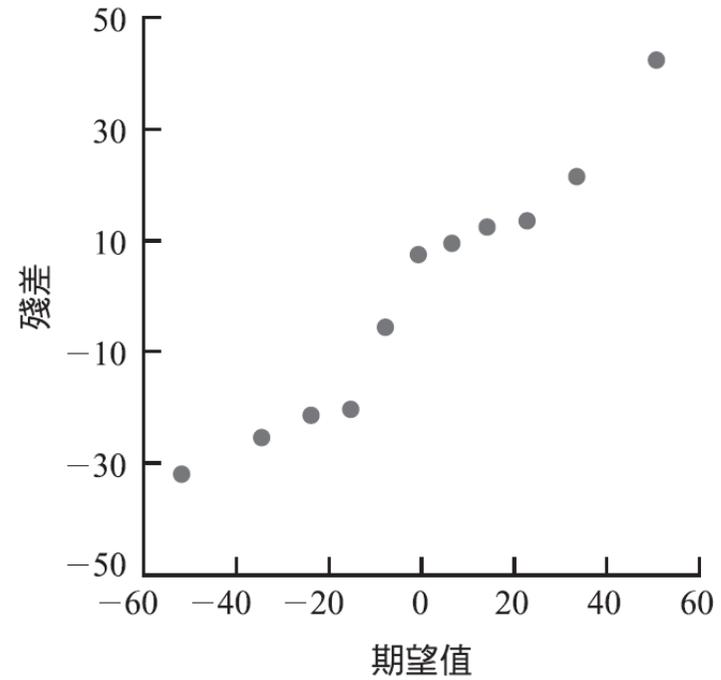
診斷用之殘差圖－電池案例。



(c) 對 x_2 之殘差圖



(d) 常態機率圖



配適度檢定

由於在 $x_1 = x_2 = 0$ 的地方有三個重複的觀測值，純誤差很容易得到：

$$SSPE=(157-157.33)^2+(131-157.33)^2+(181-157.33)^2=1,404.67$$

因為共有 $c=9$ 種不同的預測變數水準組合，所以 SSPE 有 $n-c=11-9=2$ 個自由度。

從圖8.4中可以得到 $SSE = 5,240.44$ ，所以 (3.24) 的配適不良平方和為：

$$SSLF = SSE-SSPE = 5,240.44-1,404.67 = 3,835.77$$

有 $c-p = 9-6 = 3$ 個自由度，（此處 $p = 6$ 是指模型(8.13)需估計的迴歸係數），所以對於迴歸函數(8.14)配適不良之檢定，檢定統計量 (6.68b)：

$$F = \frac{SSLF}{c-p} \div \frac{SSPE}{n-c} = \frac{3,835.77}{3} \div \frac{1,404.67}{2} = 1.82$$

在 $\alpha=.05$ 下， $F(0.95; 3,2) = 19.2 \geq F^* = 1.82$ ，根據 (6.68c) 的檢定法則可以接受迴歸函數 (8.14) 的配適。

複判定係數

圖8.4在標示為R-square的地方為複判定係數， $R^2 = 0.9135$ ，所以新研發之電池壽命由充電率與周圍溫度的第一、二階乘幕解釋後，可以縮減變異91%，調整的複判定係數在圖8.4在標示為AdjR-sq的地方， $R_a^2 = 0.8271$ ，此係數低於複判定係數很多，是因為雙預測變數多項式迴歸函數有相對較多之參數。

部份F檢定

第一階模型便足以解釋新研發電池壽命的效應影響，於是檢定的假說為：

$$H_0 : \beta_{11} = \beta_{22} = \beta_{12} = 0$$

$$H_\alpha : H_0 \text{ 之 } \beta_s \text{ 非全為 } 0$$

在(2.27)的部份F檢定統計量為：

$$F = \frac{SSR(x_1^2, x_2^2, x_1x_2 \mid x_1, x_2)}{3} \div MSE$$

Variable	DF	Type I SS
INTERCEP	1	325424
X1	1	18704
X2	1	34202
X1SQ	1	1645.966667
X2SQ	1	284.928070
X1X2	1	529.000000

圖 8.4

第二階多項式模型
(8.13)的 SAS 迴歸
程序輸出結果－電
池案例。

- 如圖8.4最下面的順序，額外平方和為 $SSR(x_1) = 18,704$ ，第三項平方和為 $SSR(x_2|x_1) = 34,202$ ，以下類推，因此所需之額外平方和 $SSR(x_1^2, x_2^2, x_1x_2|x_1, x_2)$ 可以計算如下：

$$SSR(x_1^2, x_2^2, x_1x_2|x_1, x_2) = SSR(x_1^2|x_1, x_2) + SSR(x_2^2|x_1, x_2, x_1^2) + SSR(x_1x_2|x_1, x_2, x_1^2, x_2^2) = 1,646.0 + 284.9 + 529.0 = 2,459.9$$

- 另外誤差均方可以從圖8.4中得到 $MSE = 1,048.1$ ，所以檢定統計量為：
$$F^* = \frac{2,459.9}{3} \div 1,048.1 = .78$$

在 $\alpha = .05$ 下， $F(0.95; 3, 5) = 5.41 \geq F^* = 0.78$ ，證據不足以拒絕虛無假設，亦即模型可以暫時不需要用到曲線與交互作用項，僅第一階模型便足夠了。

第一階模型

- 基於上面的種種分析，研究人員現在決定下面的第一階模型：

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad (8.17)$$

模型配適後所

$$\hat{Y} = 172.00 - 55.83x_1 + 75.50x_2 \quad (8.18)$$

(12.67) (12.67)

$$\hat{Y} = 162.84 - 55.83x_1 + 75.50x_2 + 27.39x_1^2 - 10.61x_2^2 + 11.50x_1x_2 \quad (8.16)$$

注意這裡的迴歸係數估計量 b_1 與 b_2 ，跟（8.16）估計的結果相同，但 b_0 (162.842)不同，這是因為對 X_1 與 X_2 選擇了適當的水準所致，在迴歸係數估計量底下的括號，是估計量標準差，研究人員進一步畫了一些與模型相關之殘差圖（未列出）並經過分析後，確認了第一階模型（8.17）的適當性。

以X表示所配適的第一階模型 配適的第一階迴歸函數 (8.18) 可以利用 (8.15) 轉回原始預測變數，結果為：

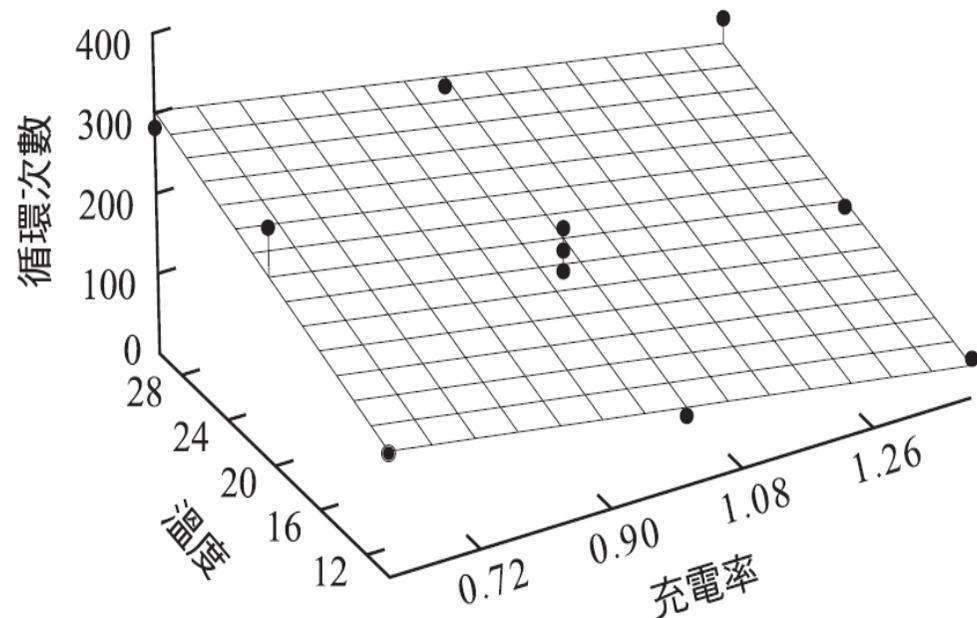
$$x_{i1} = \frac{X_{i1} - \bar{X}_1}{.4} = \frac{X_{i1} - 1.0}{.4} \quad \hat{Y} = 160.58 - 139.58X_1 + 7.55X_2 \quad (8.19)$$

$$x_{i2} = \frac{X_{i2} - \bar{X}_2}{10} = \frac{X_{i2} - 20}{10}$$

圖8.6是統計套裝軟體S-Plus關於配適的反應平面與資料的散佈圖輸出

圖 8.6

S-Plus 之配適反應平面圖(8.19)－電池案例。



迴歸係數的估計

利用 Bonferroni 程序，以全族信賴係數 90%，估計的第一階迴歸模型的雙預測變數之線性效應，這裡有 $g=2$ 個陳述：

$$B = t [1-0.10/2(2)] = t(0.975,8) = 2.306$$

在(8.18)中， b_1 與 b_2 的標準差適用編碼變數的模型，由於所配適的模型只有第一階項，所以配適出的模型(8.19)其迴歸係數估計量為 b'_1 與 b'_2 之標準差為： $\hat{Y} = 160.58 - 139.58X_1 + 7.55X_2$

$$s\{b'_1\} = (1/0.4)s\{b_1\} = \frac{12.67}{.4} = 31.68$$

$$s\{b'_2\} = (1/1.0)s\{b_2\} = \frac{12.67}{10} = 1.267$$

在應用(6.52)的 Bonferroni 信賴界限，可以得到 $-139.58 \pm 2.306(31.68)$ 及 $7.55 \pm 2.306(1.267)$ ，亦即：

$$-212.6 \leq \beta_1 \leq -66.5$$

$$4.6 \leq \beta_2 \leq 10.5$$

多項式迴歸模型的進一步評論

- 這種模型往往比起其他非線性模型，或經過變數轉換後線性模型，要浪費許多的自由度，而另外一個潛在的缺點就是即使預測變數已經經過置中程序，還是可能存在有嚴重的多重共線性問題。
- 在多項式迴歸模型中，可以利用正交多項式（orthogonal polynomials）來代替預測變數置中程序，正交多項的設計是無相關性的

- 有時配適二次反應函數的本意就是為了當沒有重複觀測值可用時，可以直接進行線性反應函數的檢驗，以便確定線性反應函數的合理性，不過配適二次模型：

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_i \quad (8.20)$$

並檢定 $\beta_{11}=0$ ，其結果仍然不能確定線性反應函數適當，圖 8.2a 便是一例，如果從圖 8.2a 的反應函數取出一組觀測值來配適模型 (8.20)，然後檢定 $\beta_{11}=0$ ，可能會得到 $\beta_{11}=0$ 的結論，然而在此線性反應函數確實不適當，透過殘差的檢驗將可以發現不良配適的情形，而必須進行多項式迴歸係數之檢定。

8.2 交互作用迴歸模型

- 交互作用效應

一個含有 $(p - 1)$ 個預測變數的迴歸模型，若其反應變數可以表示成：

$$E\{Y\} = f_1(X_1) + f_2(X_2) + \cdots + f_{p-1}(X_{p-1}) \quad (8.21)$$

- 線性效應之迴歸模型交互作用之解釋

- 迴歸係數的解釋 計量的雙預測變數 X_1 與 X_2 對於反應變數 Y 之影響為線性時，且 X_1 與 X_2 對於 Y 具有交互作用效應，其模型如下：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \quad (8.22)$$

我們可以推導出當 X_2 固定時， X_1 每增加一單位，平均反應的改變量為：

$$\beta_1 + \beta_3 X_2 \quad (8.23)$$

類似的情形，當 X_1 固定時， X_2 每增加一單位，平均反應的改變量為：

$$\beta_2 + \beta_3 X_1 \quad (8.24)$$

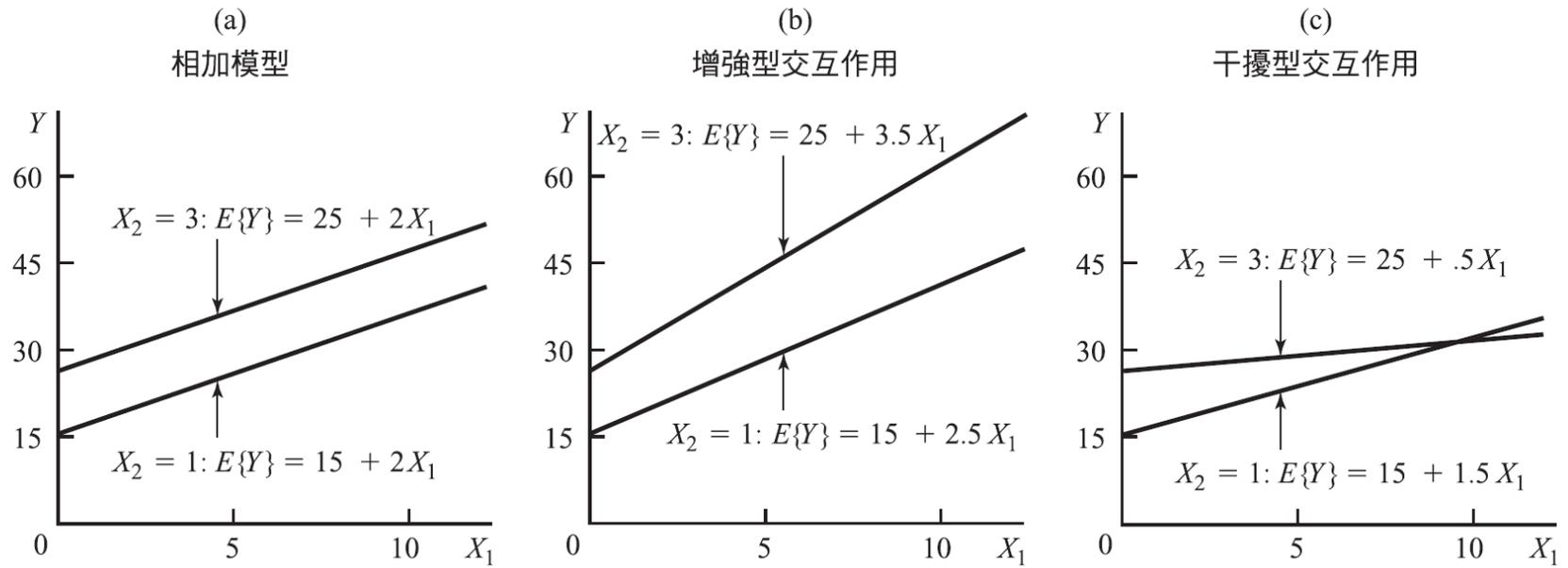
一個預測變數的效應如何與另一個預測變數有關係。

在此案例中，市場銷售量（ Y ，萬元）對於銷售點的促銷支出（ X_1 ，千元）與電視廣告支出（ X_2 ，千元）的反應函數(6.3)是具有相加性的：

$$E\{Y\} = 10 + 2X_1 + 5X_2 \quad (8.25)$$

圖 8.7

增強與干擾效應圖例－促銷案例。



接下來我們來看圖8.7b，它在原來的反應函數中增加了一個交叉乘積項 $.5 X_1 X_2$ ，該交叉乘積項的意義代表了兩個預測變數對市場銷售量的交互作用，反應函數為：

$$E\{Y\} = 10 + 2X_1 + 5X_2 + .5X_1X_2 \quad (8.26)$$

假如迴歸模型(8.26)中， $\beta_3 < 0$ ，

$$E\{Y\} = 10 + 2X_1 + 5X_2 - .5X_1X_2 \quad (8.27)$$

它代表的意義將是兩個預測變數之交互作用造成電視廣告支出高檔時，促銷活動每增加一單位，可以產生的效應比低檔時可以產生的效應小

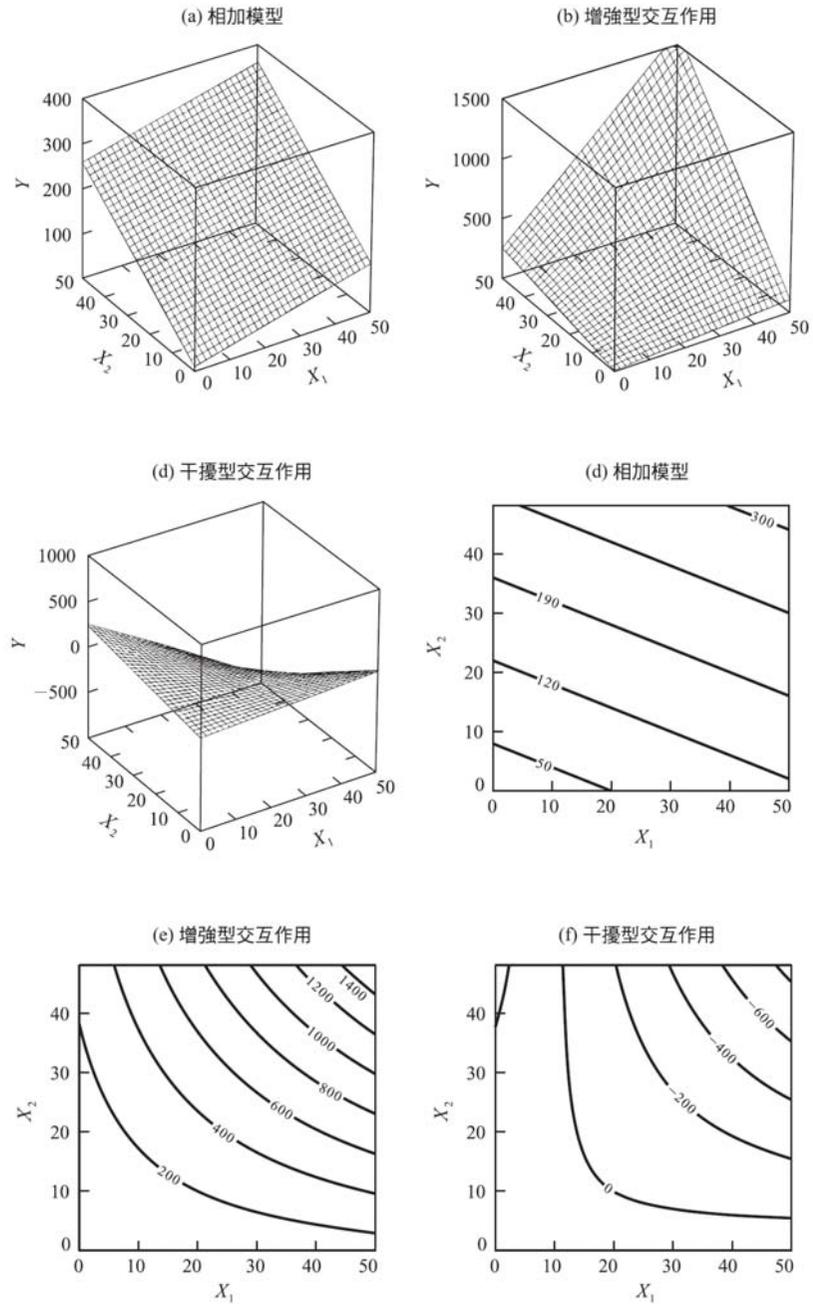


圖 8.8
 相加型與交互作用
 迴歸模型的反應曲
 面與等高線圖－促
 銷案例。

圖 8.9

具交互作用之曲線型迴歸模型的反應曲面與等高線圖－速發麵包案例。

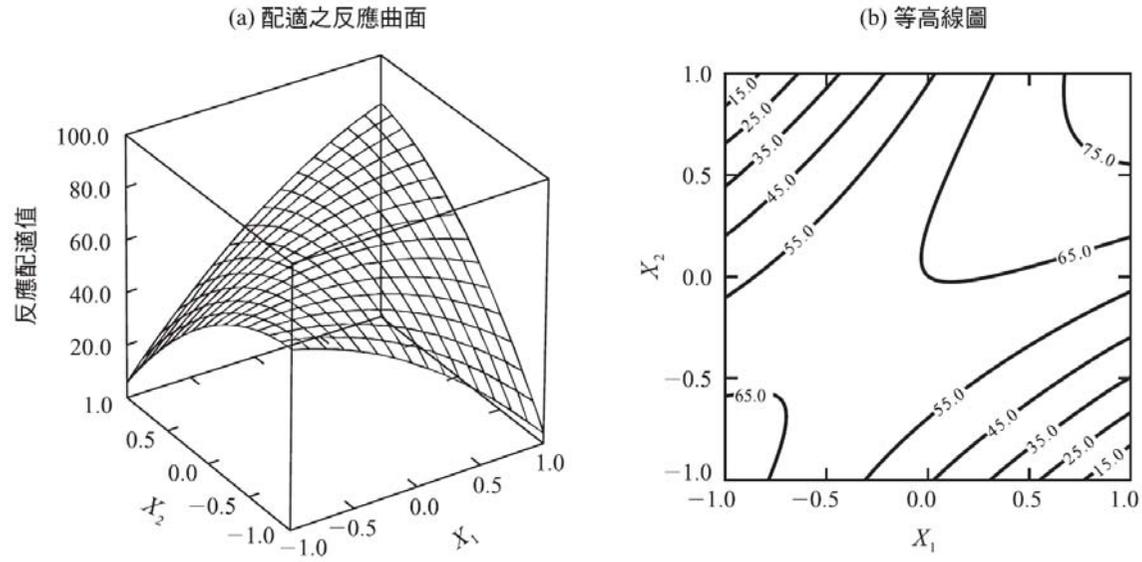
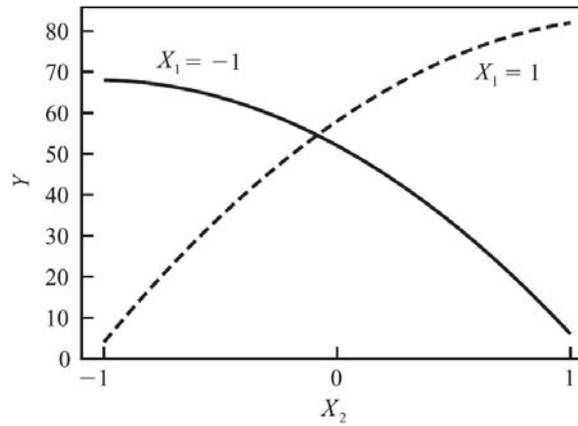


圖 8.10

具交互作用效應之
曲線迴歸模型的條
件效應圖－速發麵
包案例。



- 反應函數的形狀
- 曲線效應之迴歸模型交互作用之解釋
- 迴歸模型交互作用的實作

現在我們考慮表 7.1 的人體脂肪量案例中的迴歸模型是否需加入三個預測變數間的交互作用項，並進行檢定，我們考慮下面的迴歸模型：

例題

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1} X_{i3} + \beta_6 X_{i2} X_{i3} + \varepsilon_i \quad (8.28)$$

此一模型引進了新變數 $X_1 X_2$ 、 $X_1 X_3$ 、 $X_2 X_3$ ，將這些新變數 X 加至表 7.1 中後檢查這些新變數，發現有一些預測變數與這些交互作用項有高度相關。同時這些交互作用項間彼此也是高度相關，例如 X_1 與 $X_1 X_2$ 的相關係數達 .989， $X_1 X_3$ 與 $X_2 X_3$ 的相關係數達 .998。

因此在迴歸模型中我們先採用置中變數：

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i1} x_{i2} + \beta_5 x_{i1} x_{i3} + \beta_6 x_{i2} x_{i3} + \varepsilon_i \quad (8.29)$$

其中，

$$x_{i1} = X_{i1} - \bar{X}_1 = X_{i1} - 25.305$$

$$x_{i2} = X_{i2} - \bar{X}_2 = X_{i2} - 51.170$$

$$x_{i3} = X_{i3} - \bar{X}_3 = X_{i3} - 27.620$$

採用置中變數後計算其交叉乘積項，將發現此時與交叉乘積項間的相關性將比原來低，例如相關性最大的 $X_1 X_3$ 與 $X_2 X_3$ ，其相關係數從 .998 降低至 .891，而其他變項間的相關性縮減幅度更大。

配適迴歸模型(8.29)後，可以得到估計的迴歸函數、均方誤差及額外平方和如下：

$$\hat{Y} = 20.53 + 3.438x_1 - 2.095x_2 - 1.616x_3 + .00888x_1x_2 - .08479x_1x_3 + .09042x_2x_3$$

$$MSE = 6.745$$

變數	額外平方和
x_1	$SSR(x_1) = 352.270$
x_2	$SSR(x_2 x_1) = 33.169$
x_3	$SSR(x_3 x_1, x_2) = 11.546$
x_1x_2	$SSR(x_1x_2 x_1, x_2, x_3) = 1.496$
x_1x_3	$SSR(x_1x_3 x_1, x_2, x_3, x_1x_2) = 2.704$
x_2x_3	$SSR(x_2x_3 x_1, x_2, x_3, x_1x_2, x_1x_3) = 6.515$

現在要檢定是否需要交互作用項：

$$H_0: \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_a: H_0 \text{ 之 } \beta_s \text{ 非全為 } 0$$

部分 F 檢定統計量(7.27)需要下面的額外平方和：

$$SSR(x_1x_2, x_1x_3, x_2x_3 | x_1, x_2, x_3) = 1.496 + 2.704 + 6.515 = 10.715$$

得到檢定統計量為：

$$F^* = \frac{SSR(x_1x_2, x_1x_3, x_2x_3 | x_1, x_2, x_3)}{3} \div MSE$$

$$= \frac{10.715}{3} \div 6.745 = .53$$

在 $\alpha = .05$ 下， $F(.95; 3, 13) = 3.41 \geq F^* = .53$ ，檢定之 P -值為 .67，所以結論無法拒絕虛無假設，表示此一迴歸模型不需加入交互作用項。

8.3 質性預測變數

- 二類別的質性預測變數

在保險新制的案例中，質性預測變數分成兩類，假設我們定義 X_2 與 X_3 如下：

$$\begin{aligned} X_2 &= \begin{cases} 1 & \text{股份公司} \\ 0 & \text{其他} \end{cases} \\ X_3 &= \begin{cases} 1 & \text{合作社} \\ 0 & \text{其他} \end{cases} \end{aligned} \quad (8.30)$$

而第一階模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (8.31)$$

解決此問題的一個簡單辦法就是剔除一個指標變數，在這裡我們剔除 X_3 ，注意剔除一個指標變數並非解決此問題的唯一辦法，不過這種方法使得參數的解釋簡單多了。一般採用如下原則：

具有 c 種類別的質性變數可以用 $c - 1$ 個指標變數表示，每一個指標變數均採用0與1的值。

(8.32)

- 迴歸係數之解釋

在保險新制的案例中，剔除迴歸模型(8.31)中的指標變數 X_3 後，迴歸模型成為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (8.33)$$

其中，
 X_{i1} = 公司規模
 $X_{i2} = \begin{cases} 1 & \text{股份公司} \\ 0 & \text{合作社} \end{cases}$

該迴歸模型之反應函數為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (8.34)$$

接下來解釋迴歸模型中之係數的意義。首先對於公司類別為合作社的情形， $X_2 = 0$ ，所以反應函數(8.34)成為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (0) = \beta_0 + \beta_1 X_1 \quad \text{合作社} \quad (8.34a)$$

對於公司類別為股份公司的情形， $X_2 = 1$ ，所以反應函數(8.34)成為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (1) = (\beta_0 + \beta_2) + \beta_1 X_1 \quad \text{股份公司} \quad (8.34b)$$

例題

在保險新制的案例中，該位經濟學家分別研究了10家合作社與10家股份公司，其基本資料在表8.2的第一、二、三欄，公司類型的指標編碼在第四欄。注意這裡的股份公司 $X_2 = 1$ ，而合作社 $X_2 = 0$ 。

現在開始配適迴歸模型(8.33)，表8.3中有 Y 對 X_1 與 X_2 迴歸的主要結果，所配適的反應函數為：

$$\hat{Y} = 33.87407 - .10174X_1 + 8.05547X_2$$

圖8.12有每一類公司的反應函數與實際觀測值。

該位經濟學家目前最想知道的是公司的類別對於採用新制的速度效應，並希望能得到 β_2 的95%信賴區間，我們在估計時需要有 $t(.975; 17) = 2.110$ ，然後利用表8.3的結果，可以得到 β_2 的95%信賴區間為：

$$4.98 \leq \beta_2 \leq 11.13$$

因此在95%的信心水準下，我們推論在不同的公司規模下，股份公司將比合作社平均約慢了5到11個月才採用新制。

當我們考慮在顯著水準.05下檢定：

$$H_0: \beta_2 = 0$$

$$H_a: \beta_2 \neq 0$$

結論將拒絕虛無假設，亦即公司類別對採用新制的速度有效應，這點我們可從上面的95%信賴區間不包含零看出。

這位經濟學家同時也進行其他的分析工作，我們留待後面再做討論。

說明

讀者或許會懷疑為何不直接對股份公司與合作社分別配適兩個迴歸模型，卻採用含有指標變數的單一模型？其原因有二，首先是模型假設相等的斜率，且誤差項變異數為相同的常數，所以估計共同的斜率 β_1 最好是合併兩種公司。其次，迴歸模型的一些相關推論，例如對 β_0 與 β_1 的推論，使用含有指標變數的單一模型，因為MSE的自由度較大，所以推論之結果會較為精確。 ■

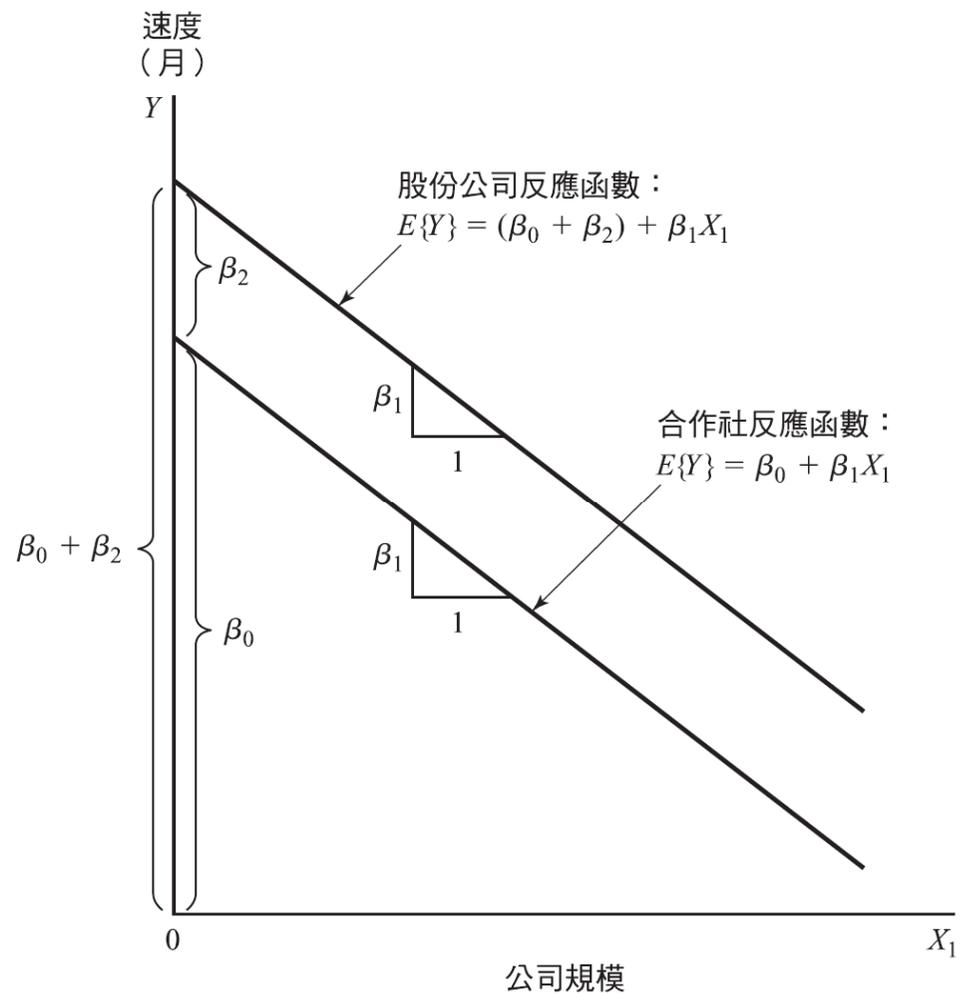


圖 8.11

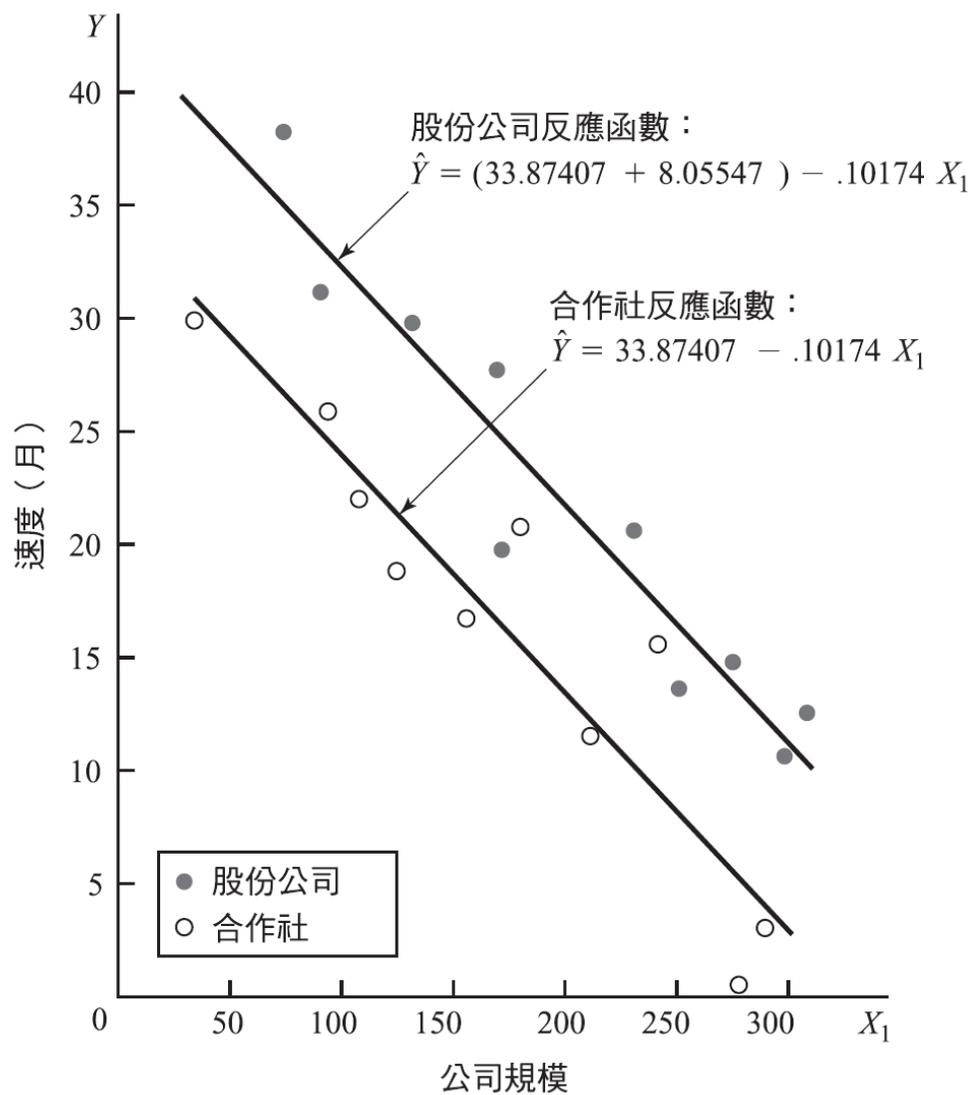
含有指標變數 X_2 的迴歸模型(8.33)其迴歸係數之意義圖—保險新制案例。

公司個案 編號	(1) 速度 (月)	(2) 公司規模 (百萬元)	(3) 公司類別	(4) 指標編碼 X_{i2}	(5) $X_{i1}X_{i2}$
i	Y_i	X_{i1}			
1	17	151	合作社	0	0
2	26	92	合作社	0	0
3	21	175	合作社	0	0
4	30	31	合作社	0	0
5	22	104	合作社	0	0
6	0	277	合作社	0	0
7	12	210	合作社	0	0
8	19	120	合作社	0	0
9	4	290	合作社	0	0
10	16	238	合作社	0	0
11	28	164	股份公司	1	164
12	15	272	股份公司	1	272
13	11	295	股份公司	1	295
14	38	68	股份公司	1	68
15	31	85	股份公司	1	85
16	21	224	股份公司	1	224
17	20	166	股份公司	1	166
18	13	305	股份公司	1	305
19	30	124	股份公司	1	124
20	14	246	股份公司	1	246

表 8.2
資料與指標編碼－
保險新制案例。

圖 8.12

迴歸模型(8.33)配
適的迴歸函數－保
險新制案例。



(a) 迴歸係數			
迴歸係數	迴歸係數 的估計量	估計量 的標準差	t^*
β_0	33.87407	1.81386	18.68
β_1	-.10174	.00889	-11.44
β_2	8.05547	1.45911	5.52

(b) 變異數分析			
變異來源	平方和	自由度	均方
迴歸	1,504.41	2	752.20
誤差	176.39	17	10.38
總合	1,680.80	19	

表 8.3

配適迴歸模型
(8.33)之結果－保
險新制案例。

- 超過二類別的質性預測變數

當質性預測變數之分類超過兩種時，則需要在迴歸模型中增加指標變數。下面的例子是關於工具耐用度 Y 對於工具速度 X_1 與工具樣式的迴歸分析，其中工具樣式是一個四種分類(M1、M2、M3、M4)的質性預測變數，所以需要三個指標變數，指標變數之定義如下：

$$\begin{aligned} X_2 &= \begin{cases} 1 & \text{工具樣式M1} \\ 0 & \text{其他} \end{cases} \\ X_3 &= \begin{cases} 1 & \text{工具樣式M2} \\ 0 & \text{其他} \end{cases} \\ X_4 &= \begin{cases} 1 & \text{工具樣式M3} \\ 0 & \text{其他} \end{cases} \end{aligned} \tag{8.35}$$

第一階模型 第一階迴歸模型如下：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \quad (8.36)$$

迴歸模型(8.36)之反應函數為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (8.37)$$

關於模型中迴歸係數之意義，我們先考慮工具樣式為M4時的反應函數，亦即 $X_2 = X_3 = X_4 = 0$ 時的反應函數：

$$E\{Y\} = \beta_0 + \beta_1 X_1 \quad \text{工具樣式M4 (8.37a)}$$

同理當 $X_2 = 1, X_3 = X_4 = 0$ 時，代表工具樣式為M1的反應函數：

$$E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1 \quad \text{工具樣式M1 (8.37b)}$$

類似的情形，代表工具樣式為M2、M3的反應函數分別為：

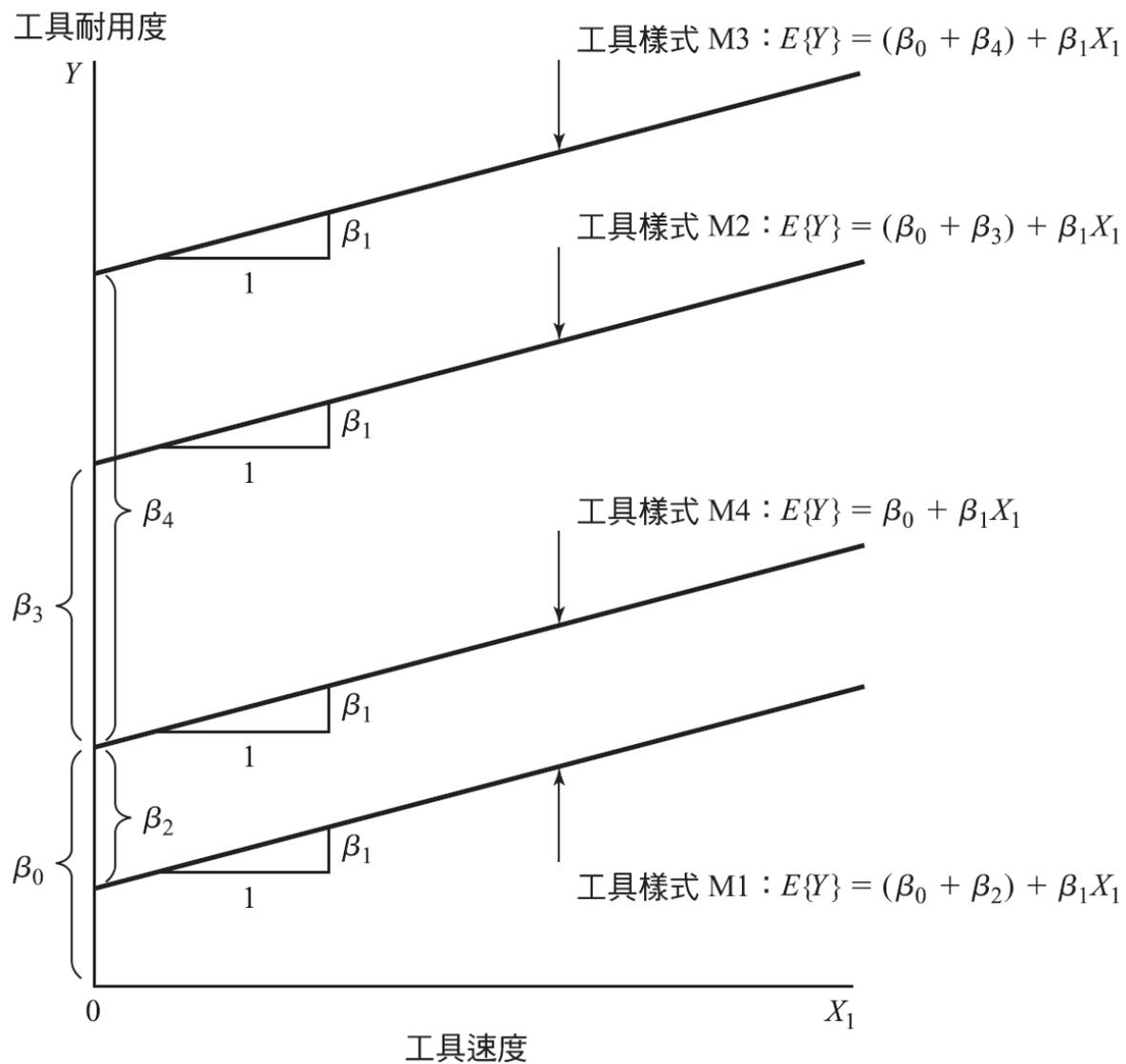
$$E\{Y\} = (\beta_0 + \beta_3) + \beta_1 X_1 \quad \text{工具樣式M2 (8.37c)}$$

$$E\{Y\} = (\beta_0 + \beta_4) + \beta_1 X_1 \quad \text{工具樣式M3 (8.37d)}$$

經由反應函數(8.37c)與(8.37d)可以知道的點估計量為，而此估計量之變異數為：

$$s^2 \{b_4 - b_3\} = s^2 \{b_4\} + s^2 \{b_3\} - 2s \{b_4, b_3\} \quad (8.38)$$

圖 8.13
 迴歸模型(8.36)之
 圖示－工具耐用度
 案例。



- 時間數列應用

指標變數在時間數列資料中的迴歸分析，經常是很有用的工具。例如儲蓄 Y 對所得 X 所進行的迴歸分析，其中儲蓄與所得都是連續多年的年度資料，可能考慮模型為：

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad t = 1, \dots, n \quad (8.39)$$

其中 Y_t 與 X_t 分別代表的儲蓄與所得。假設資料使用期間涵蓋了平時與戰時，由於戰時可能有較高之儲蓄傾向，因此可以利用下面的迴歸模型：

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \varepsilon_t \quad (8.40)$$

其中，

$$X_{t1} = \text{所得} \quad X_{t2} = \begin{cases} 1 & \text{第 } t \text{ 期為平時} \\ 0 & \text{其他} \end{cases}$$

如果每季的銷貨量會受到季別的影響，則合併季節影響效應之第一階迴歸模型為：

$$Y_i = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \beta_4 X_{t4} + \beta_5 X_{t5} + \varepsilon_t \quad (8.41)$$

其中，

X_{t1} = 每季的廣告支出

X_{t2} = 每季的個人可支配所得

$X_{t3} = \begin{cases} 1 & \text{第 } t \text{ 期為第一季} \\ 0 & \text{其他} \end{cases}$

$X_{t4} = \begin{cases} 1 & \text{第 } t \text{ 期為第二季} \\ 0 & \text{其他} \end{cases}$

$X_{t5} = \begin{cases} 1 & \text{第 } t \text{ 期為第三季} \\ 0 & \text{其他} \end{cases}$

8.4 使用指標變數之考慮

- 指標變數與配置碼

配置碼可以是任意的數值，在上例中假設沒有其他的預測變數，則使用配置碼的第一階迴歸模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \quad (8.42)$$

指標變數對於分組的間距並沒有任何假設前提，完全是依據資料間的差別效應，則第一階迴歸模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (8.43)$$

此處 β_1 的意義在於衡量差別效應：

$$E\{Y|\text{經常使用者}\} - E\{Y|\text{從不使用者}\}$$

而 β_2 的意義在於衡量差別效應：

$$E\{Y|\text{偶而使用者}\} - E\{Y|\text{從不使用者}\}$$

- 指標變數與計量變數
- 指標變數的其它編碼法
第一種編碼方式為：

$$X_2 = \begin{cases} 1 & \text{股份公司} \\ -1 & \text{合作社} \end{cases} \quad (8.44)$$

其第一階線性迴歸模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (8.45)$$

具有反應函數：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (8.46)$$

此一反應函數依照不同的公司類別，可以拆開為：

$$E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1 \quad \text{股份公司} \quad (8.46a)$$

$$E\{Y\} = (\beta_0 - \beta_2) + \beta_1 X_1 \quad \text{合作社} \quad (8.46b)$$

第二種編碼方式為去掉迴歸模型的截距項，然後對 c 種類別的質性變數，分別用編碼值0與1的方法來定義不同的類別，在前面的保險新制案例中，模型為：

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (8.47)$$

其中，

$$X_{i1} = \text{公司規模}$$

$$X_{i2} = \begin{cases} 1 & \text{股份有限公司} \\ 0 & \text{其他} \end{cases}$$

$$X_{i3} = \begin{cases} 1 & \text{合作社} \\ 0 & \text{其他} \end{cases}$$

而兩個反應函數分別為：

$$E\{Y\} = \beta_2 + \beta_1 X_1 \quad \text{股份有限公司} \quad (8.48a)$$

$$E\{Y\} = \beta_3 + \beta_1 X_1 \quad \text{合作社} \quad (8.48b)$$

8.5 計量與質性預測變數之交互作用

- 公司規模與公司類別很可能有交互作用的存在，雖然模型中有一個質性的預測變數，交互作用項仍然可以跟過去相同，利用交叉乘積項的形式引進模型中，本案例中含有交互作用的第一階迴歸模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \quad (8.49)$$

其中，

X_{i1} = 公司規模

$X_{i2} = \begin{cases} 1 & \text{股份公司} \\ 0 & \text{其他} \end{cases}$

而反應函數為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (8.50)$$

- 迴歸係數之意義

首先在合作社的部份，由於 $X_2 = 0$ ，所以 $X_1 X_2 = 0$ ，因此合作社下的反應函數(8.50)成為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (1) + \beta_3 X_1 \quad (8.50a)$$

其次，在股份公司方面，因為 $X_2 = 1$ ，所以 $X_1 X_2 = X_1$ ，因此在股份公司下的反應函數(8.50)成為：

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (1) + \beta_3 X_1$$

或是：

$$E\{Y\} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 \quad \text{股份公司} \quad (8.50b)$$

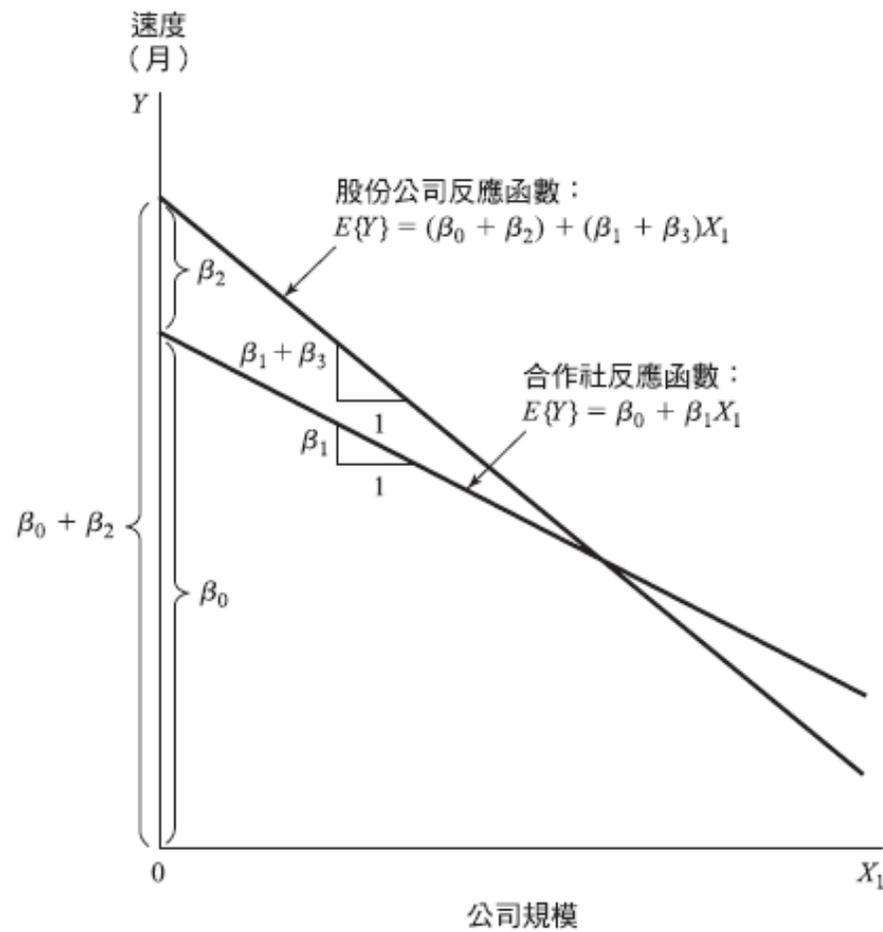
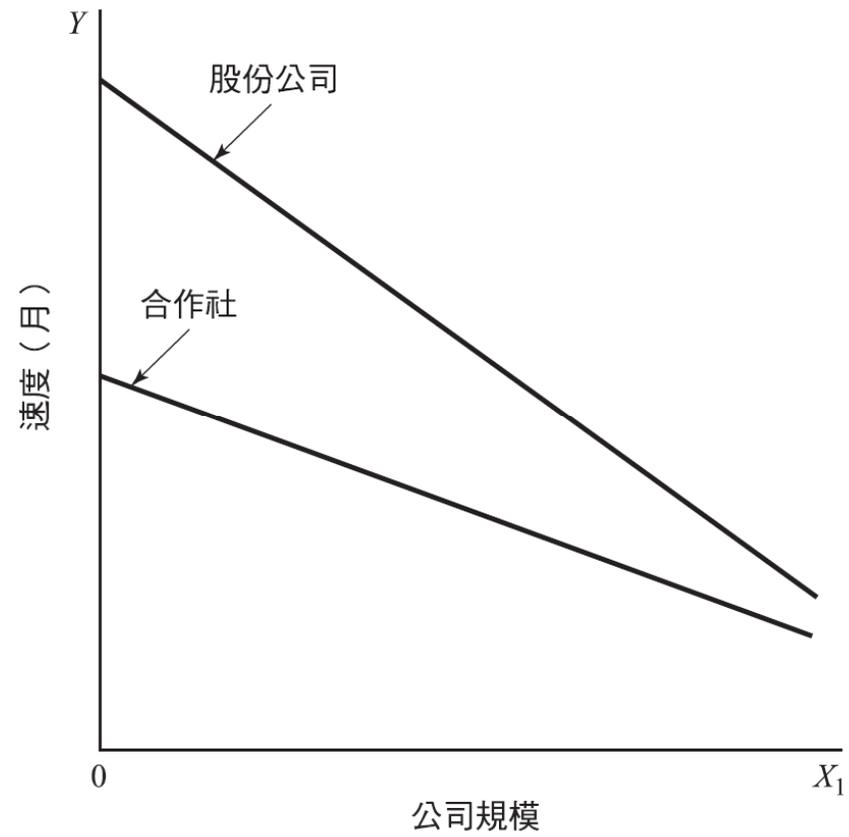


圖 8.14

具有指標變數 X_2 與
交互作用項的迴歸
模型(8.49)其迴歸
係數之圖示－保險
新制案例。

圖 8.15

具有指標變數 X_2 與
交互作用項的迴歸
模型(8.49)之另一
圖示－保險新制案
例。



(a) 迴歸係數			
迴歸係數	迴歸係數 的估計量	估計量 的標準差	t^*
β_0	33.83837	2.44065	13.86
β_1	-.10153	.01305	-7.78
β_2	8.13125	3.65405	2.23
β_3	-.0004171	.01833	-.02

(b) 變異數分析			
變異來源	平方和	自由度	均方
迴歸	1,504.42	3	501.47
誤差	176.38	16	11.02
總合	1,680.80	19	

表 8.4
配適迴歸模型
(8.49)之結果－保
險新制案例。

例題

因為這位經濟學家關心的是公司規模與公司類型存在交互作用，而最初配適的模型為(8.49)：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

在表 8.2 的第五欄中顯示了交互作用項 $X_1 X_2$ 的值。注意合作社在此欄位中為 0，而股份公司則為 X_{i1} 。

接下來仍舊是進行模型配適，利用電腦進行 Y 對 X_1 、 X_2 、 $X_1 X_2$ 的基本迴歸，其輸出結果在表 8.4 中，再來就是檢定交互作用的存在與否：

$$H_0: \beta_3 = 0$$

$$H_a: \beta_3 \neq 0$$

利用表 8.4a 的 t^* 統計量：

$$t^* = \frac{b_3}{s\{b_3\}} = \frac{-0.0004171}{.01833} = -.02$$

在 $\alpha = .05$ 下， $t(.975;16) = 2.120 \geq |t^*| = .02$ ，雙邊檢定之 P -值為 .98，所以結論無法拒絕虛無假設，表示此一迴歸模型不需加入交互作用項，也就是該經濟學家選擇前面討論過的迴歸模型(8.33)的理由。

說明

配適迴歸模型(8.49)所得到的反應函數，相同於股份公司與合作社分別配適的反應函數，而使用指標變數的好處就是在於只需進行一次迴歸程序，就可以同時得到兩者的迴歸配適。

另一個好處就是比較質性變數的兩個不同類別之迴歸函數檢定，就是一般線性迴歸模型中的迴歸係數檢定，例如在保險新制案例中，檢定兩迴歸模型是否具有相同之斜率，其實就是在檢定：

$$H_0: \beta_3 = 0$$

$$H_a: \beta_3 \neq 0$$

類似的情形，圖 8.14 顯示兩迴歸函數是否相等之檢定，相當於檢定：

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

$$H_a: \beta_2 \text{ 與 } \beta_3 \text{ 不全為 } 0$$



8.6 更為複雜之模型

- 質性預測變數多於一個

考慮廣告支出 Y 對銷售量 X_1 、公司類型（有限公司、非有限公司）、業務管理品質（高、低），定義質性預測變數如下：

$$X_2 = \begin{cases} 1 & \text{有限公司} \\ 0 & \text{其他} \end{cases} \quad (8.51)$$

$$X_3 = \begin{cases} 1 & \text{高品質} \\ 0 & \text{其他} \end{cases}$$

- 第一階模型 上例中的第一階迴歸模型為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (8.52)$$

- 加入交互作用後的第一階模型

在上面的廣告支出案例中，第一階迴歸模型如果增加了兩兩預測變數的交互作用項，模型成為：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1} X_{i3} + \beta_6 X_{i2} X_{i3} + \varepsilon_i \quad (8.53)$$

- 僅有質性預測變數

迴歸模型也允許僅含有質性之預測變數。在上面的廣告支出案例中，廣告支出可以只針對公司類型與管理品質進行迴歸分析。第一階迴歸模型為：

$$Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (8.54)$$

8.7 兩個或兩個以上迴歸函數之比較

表 8.5

肥皂生產線案例的資料（所有資料均已經過編碼）。

第1條生產線				第2條生產線			
個案	碎屑量	速率		個案	碎屑量	速率	
i	Y_i	X_{i1}	X_{i2}	i	Y_i	X_{i1}	X_{i2}
1	218	100	1	16	140	105	0
2	248	125	1	17	277	215	0
3	360	220	1	18	384	270	0
4	351	205	1	19	341	255	0
5	470	300	1	20	215	175	0
6	394	255	1	21	180	135	0
7	332	225	1	22	260	200	0
8	321	175	1	23	361	275	0
9	410	270	1	24	252	155	0
10	260	170	1	25	422	320	0
11	241	155	1	26	273	190	0
12	331	190	1	27	410	295	0
13	275	140	1				
14	425	290	1				
15	367	265	1				

- 肥皂生產線案例
- 暫時性模型

此模型假設了碎屑量 Y 與生產線速率 X_1 的迴歸關係，在兩條生產線上的表現均為直線，且有相同的誤差項變異數，僅可能有不同的迴歸斜率與截距。

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \quad (8.55)$$

其中，

X_{i1} = 生產線速率

$X_{i2} = \begin{cases} 1 & \text{第1條生產線} \\ 0 & \text{第2條生產線} \end{cases}$

$i = 1, 2, \dots, 27$

- 診斷

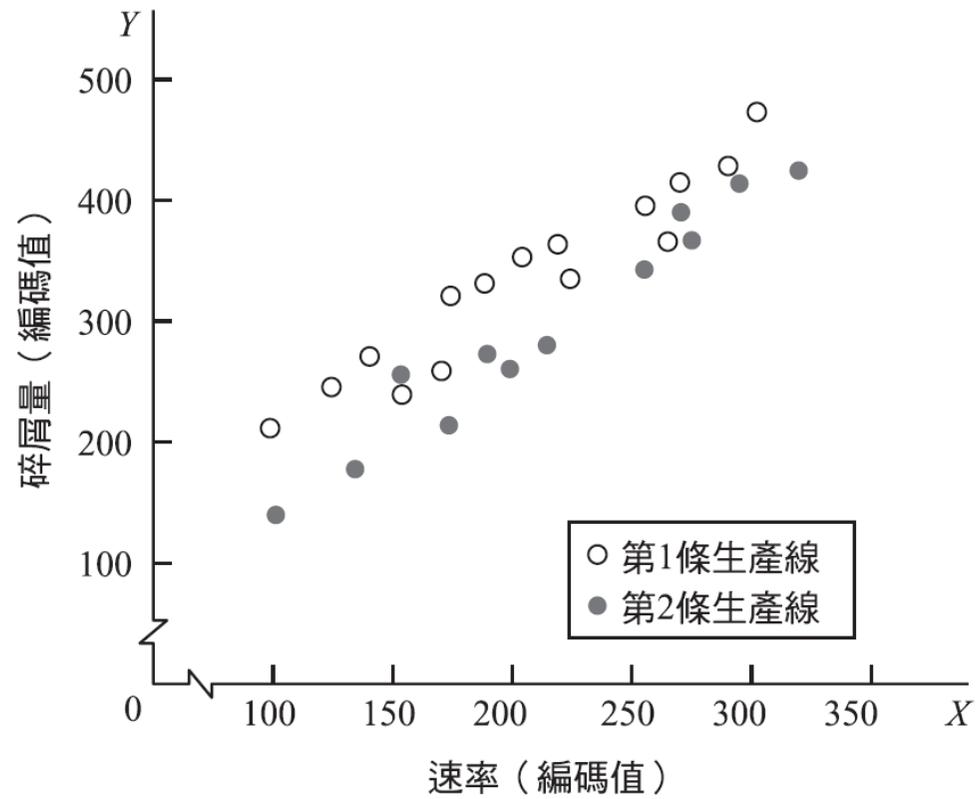


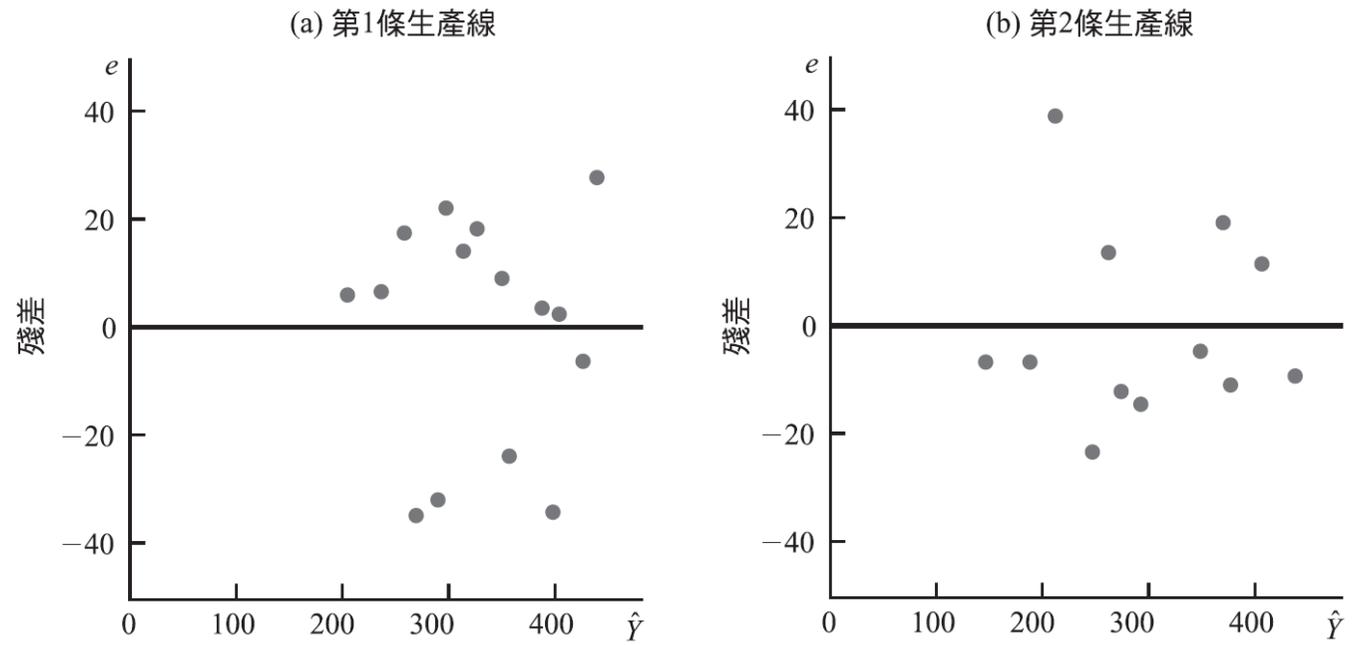
圖 8.16
符號散佈圖－肥皂
生產線案例。

(a) 迴歸係數			
迴歸係數	迴歸係數 的估計量	估計量 的標準差	t^*
β_0	7.57	20.87	.36
β_1	1.322	.09262	14.27
β_2	90.39	28.35	3.19
β_3	-.1767	.1288	-1.37
(b) 變異數分析			
變異來源	平方和	自由度	均方
迴歸	169,165	3	56,388
X_1	149,661	1	149,661
$X_2 X_1$	18,694	1	18,694
$X_1 X_2 X_1, X_2$	810	1	810
誤差	9,904	23	430.6
總合	179,069	26	

表 8.6
 配適迴歸模型
 (8.55)之結果—肥
 皂生產線案例。

圖 8.17

對 \hat{Y} 之殘差圖－肥皂生產線案例。



- 對兩條迴歸直線的推論

兩條生產線之迴歸函數是否相同的問題，可以考慮下面的假設檢定：

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

$$H_a : \beta_2 \text{ 與 } \beta_3 \text{ 非全為 } 0 \quad (8.56)$$

適當的檢定統計量為(7.27)：

$$F^* = \frac{SSR(X_2, X_1X_2 | X_1)}{2} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \quad (8.56a)$$

其中 n 為兩母體之樣本合併後之個案數，利用表

$$\begin{aligned} 8.6 \text{ 之結果 : } SSR(X_2, X_1X_2 | X_1) &= SSR(X_2 | X_1) + SSR(X_1X_2 | X_1, X_2) \\ &= 18,694 + 810 = 19,504 \end{aligned}$$

$$F^* = \frac{19,504}{2} \div \frac{9,904}{23} = 22.65$$

其次，分析人員想知道兩條生產線之迴歸函數是否有相同的直線斜率，可以考慮下面的假設檢定：

$$\begin{aligned}H_0 : \beta_3 &= 0 \\H_a : \beta_3 &\neq 0\end{aligned}\tag{8.57}$$

適當的檢定統計量為(7.25)的 t^* 統計量或是(7.24)的部份 F^* 統計量：

$$F^* = \frac{SSR(X_1X_2 | X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4}\tag{8.57a}$$

利用表8.6以及部份 F^* 統計量之結果：

$$F^* = \frac{810}{1} \div \frac{9,904}{23} = 1.88$$

- 儀器測定案例

兩套儀器的迴歸函數可能不同，因此採用下面的模型

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1}^2 X_{i2} + \varepsilon_i \quad (8.58)$$

其中，

$$x_{i1} = X_{i1} - \bar{X}_1 = \text{置中後的真實壓力}$$
$$X_{i2} = \begin{cases} 1 & \text{儀器B} \\ 0 & \text{其他} \end{cases}$$

對於儀器A而言， $X_2 = 0$ ，反應函數為：

$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \quad \text{儀器A} \quad (8.59a)$$

對於儀器B而言， $X_2 = 1$ ，反應函數為：

$$E\{Y\} = (\beta_0 + \beta_3) + (\beta_1 + \beta_4)x_1 + (\beta_2 + \beta_5)x_1^2 \quad \text{儀器B (8.59b)}$$

因此檢定兩個反應函數是否相等，所涉及之假說為：

$$\begin{aligned} H_0 &: \beta_3 = \beta_4 = \beta_5 = 0 \\ H_a &: H_0 \text{ 之 } \beta_k \text{ 非全為 } 0 \end{aligned} \quad (8.60)$$

適當的檢定統計量為(7.27)：

$$F^* = \frac{SSR(X_2, x_1 X_2, x_1^2 X_2 | x_1, x_1^2)}{3} \div \frac{SSE(x_1, x_1^2, X_2, x_1 X_2, x_1^2 X_2)}{n-6} \quad (8.60a)$$

其中 n 為兩母體之樣本合併後之樣本大小。