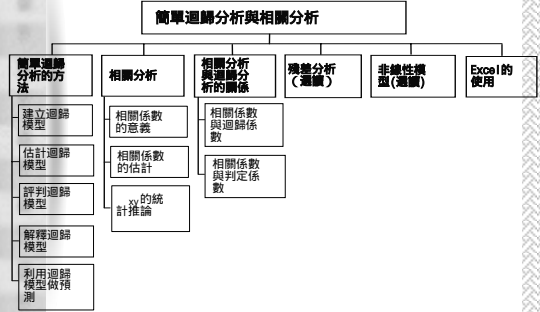


14 簡單迴歸分析與相關分析

學習目的

1. 了解迴歸分析及相關分析的意義、重要性、估計方法及步驟。
2. 了解如何利用最小平方方法來估計迴歸方程式。
3. 了解判定係數及檢定方法 (t 檢定、 z 檢定與 F 檢定)，及如何評判簡單迴歸方程是否可以接受及迴歸係數是否顯著，以及如何解釋與預測。
4. 了解如何估計與計算簡單相關係數，並檢定相關程度。
5. 了解如何做複迴歸分析包括模型的設定、估計複迴歸方程式。如何使用 F 檢定與 t 檢定來檢定整條迴歸方程式與迴歸係數及如何利用複迴歸模型來做預測。
6. 了解迴歸分析與相關分析在日常經濟活動、企業管理、政府政策等方面的應用。
7. 利用 Excel 來做迴歸分析與相關分析。

本章結構



迴歸分析與相關分析的意義

○ 迴歸分析的意義

迴歸分析是用來分析一個或一個以上自變數與依變數間的数量關係，以了解當自變數為某一水準或數量時，依變數反應的數量或水準。

○ 相關分析的意義

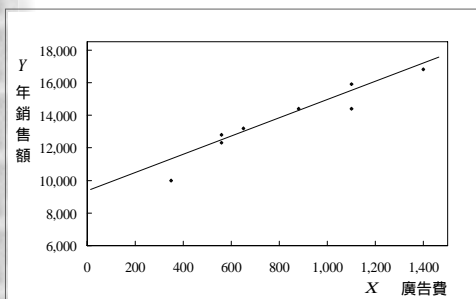
相關分析(correlation analysis)是分析變數間關係的方向與程度大小的統計方法。

表14.1 廣告支出與銷售額

	A	B	C
1	營業單位	廣告支出	年銷售額
2	台北	1400	16800
3	桃園	1100	14400
4	新苗	560	12300
5	台中	650	13200
6	彰雲	560	12800
7	台南	880	14400
8	高雄	1100	15900
9	花東	350	10000

• 單位：新台幣萬元。

圖14.4 廣告支出與銷售額的關係

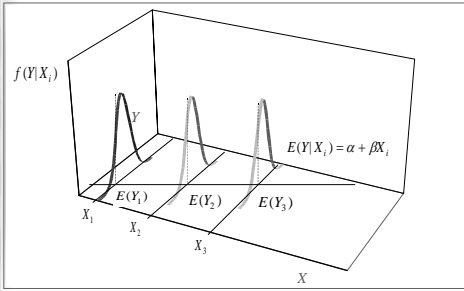


建立迴歸模型

○ 線性迴歸模型

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, \dots, n$$

圖14.6 簡單迴歸模型



簡單迴歸模型的假設條件

- ① $E(\epsilon_i) = 0$
是指 $X = X_i$ 條件下，每一組的殘差項的平均數為0。
- ② $V(\epsilon_i) = \sigma^2$ 變異數齊一性
每一組的殘差項的變異數均相等，我們稱為變異數齊一性。
- ③ $Cov(\epsilon_i, \epsilon_j) = 0 \quad i \neq j \quad i, j = 1, \dots, n$
任何一組 ϵ_i 與 ϵ_j 的共變數為0，即 ϵ_i 與 ϵ_j 間無關。
- ④ $Cov(\epsilon_i, X) = 0$ 或 $E(\epsilon_i, X) = 0 \quad i = 1, \dots, n$
即 ϵ_i 與 X 無關。
- ⑤ X 為固定變數或事前決定的變數， Y_i 為隨機變數。
- ⑥ ϵ_i 為常態分配及 Y_i 為常態分配
各組殘差項為常態分配，或各組之依變數為常態分配。

估計迴歸模型

○ 普通最小平方方法

普通最小平方方法是使樣本觀察值與估計值的差異之平方和為最小的估計方法。即是使得 $SSE = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$ 為最小而求取 $\hat{\alpha}$ ， $\hat{\beta}$ 的方法。利用此一估計方法所得到的估計式稱為普通最小平方估計式 (ordinary least squares estimator, OLSE)。

估計迴歸模型

○ 迴歸估計式

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

式中， \hat{Y}_i 為 $E(Y_i)$ 的估計式， $\hat{\alpha}$ 、 $\hat{\beta}$ 分別為 α 、 β 的估計式。

○ 觀察值與估計值之差的平方和

$$SSE = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$$

○ 標準方程式

$$\sum Y = n\hat{\alpha} + \hat{\beta}\sum X$$

$$\sum XY = \hat{\alpha}\sum X + \hat{\beta}\sum X^2$$

估計迴歸模型

○ β 的估計式 $\hat{\beta}$

$$\hat{\beta} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{\sum xy}{\sum x^2}$$

○ α 的估計式 $\hat{\alpha}$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

圖14.7 估計的迴歸方程式

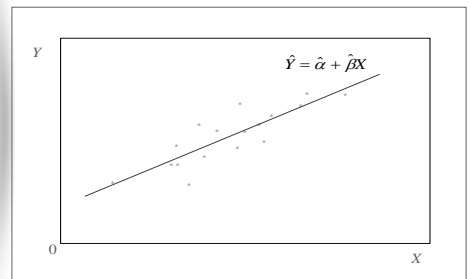


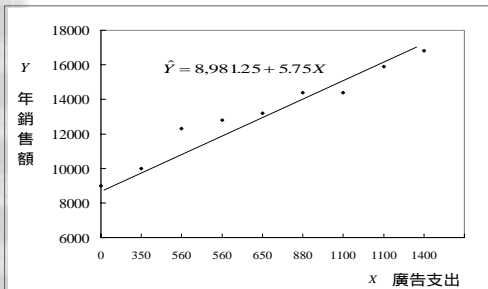
表14.2 最小平方法的計算

X_i	Y_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(X_i - \bar{X}) \times (Y_i - \bar{Y})$	
350	10,000	-475	-3,725	225,625	1,769,375	
560	12,300	-265	-1,425	70,225	377,625	
560	12,800	-265	-925	70,225	245,125	
650	13,200	-175	-525	30,625	91,875	
880	14,400	55	675	3,025	37,125	
1,100	14,400	275	675	75,625	185,625	
1,100	15,900	275	2,175	75,625	598,125	
1,400	16,800	575	3,075	330,625	1,768,125	
總和	6,600	109,800	0	0	881,600	5,073,000
平均數	825	13,725	0	0	110,200	634,125

表14.3 廣告支出與營業收入的迴歸結果

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	截距	8977.694	663.3645	13.53358	1.01E-05
18	廣告支出	5.75431	0.745954	7.714029	0.000249

圖14.8 飲料及果汁銷售額的迴歸估計式



估計迴歸模型

○ 最小平方估計式的性質

① $\hat{\alpha}$ 、 $\hat{\beta}$ 為 α 、 β 的不偏估計式，即 $E(\hat{\alpha}) = \alpha$ ， $E(\hat{\beta}) = \beta$ 。

② $V(\hat{\alpha}) = \frac{\sum X_i^2}{n \sum x^2} \sigma^2$ ， $V(\hat{\beta}) = \frac{\sigma^2}{\sum x^2}$ 。

③ $\hat{\alpha}$ 、 $\hat{\beta}$ 的抽樣分配為：

$$\hat{\alpha} \sim N\left(\alpha, \frac{\sum X_i^2}{n \sum x^2} \sigma^2\right), \hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum x^2}\right)$$

④ $\hat{\alpha}$ 、 $\hat{\beta}$ 均為最小變異線性不偏估計式 (best linear unbiased estimator BLUE)，亦即 $\hat{\alpha}$ 、 $\hat{\beta}$ 是所有線性不偏估計式中變異數最小的估計式，我們簡稱為 BLUE。

評判迴歸模型-統計上的檢定

(1) 迴歸方程式的配適度

○ 依變數的總變異

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

總變異 = 可解釋的變異 + 隨機變異

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

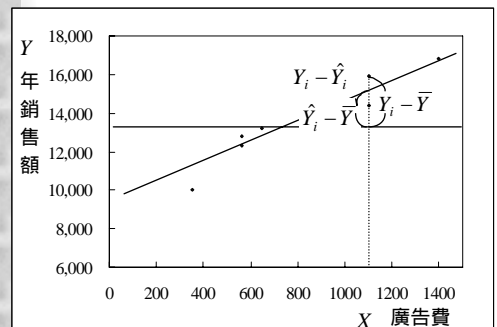
總變異 = 可解釋的變異 + 隨機變異

○ 判定係數 R^2

$$R^2 = \frac{\text{可解釋的變異}}{\text{總變異}} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$\text{即 } R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum e^2}{\sum (Y_i - \bar{Y})^2}$$

圖14.12 可解釋之差異與不可解釋之差異



第14章 簡單迴歸分析與相關分析 應用統計學

表14.5 迴歸式的判定係數

	A	B	C
3	迴歸統計		
4	R 的倍數	0.953103182	
5	R 平方	0.908405675	R^2
6	調整的 R 平方	0.893139955	
7	標準誤	700.4026962	$S_{Y/X}$
8	觀察值個數		8

林惠玲 陳正倉著 雙葉書局發行 2002

第14章 簡單迴歸分析與相關分析 應用統計學

評判迴歸模型-統計上的檢定

◎F檢定

○F檢定統計量

$$F = \frac{\sum(\hat{Y} - \bar{Y})^2 / 1}{\sum(Y - \hat{Y}) / n - 2} = \frac{\sum \hat{y}^2 / 1}{\sum e^2 / n - 2} = \frac{MSR}{MSE} \sim F_{1, n-2}$$

○決策法則

① $F > F_{1, n-2, \alpha}$ 時，則拒絕 H_0 。

② $F \leq F_{1, n-2, \alpha}$ 時，則接受 H_0 。

林惠玲 陳正倉著 雙葉書局發行 2002

第14章 簡單迴歸分析與相關分析 應用統計學

表14.8 迴歸式的ANOVA表及F值

	A	B	C	D	E	F
10	ANOVA					
11		自由度	SS	MS	F	顯著性
12	迴歸	1	29191616	29191616	59.50624	0.000249
13	殘差	6	2943384	490563.9		
14	總和	7	32135000			

林惠玲 陳正倉著 雙葉書局發行 2002

第14章 簡單迴歸分析與相關分析 應用統計學

評判迴歸模型-統計上的檢定

(2)對個別迴歸係數作統計推論

◎對 β 的檢定

○Z檢定統計量(σ^2 已知)

$$\frac{\hat{\beta} - \beta}{\frac{\sigma}{\sqrt{\sum x^2}}} \sim Z$$

○t檢定統計量(σ^2 未知)

$$\frac{\hat{\beta} - \beta}{\frac{S_{Y/X}}{\sqrt{\sum x^2}}} \sim t_{n-2}$$

林惠玲 陳正倉著 雙葉書局發行 2002

第14章 簡單迴歸分析與相關分析 應用統計學

表14.9 迴歸式的 β 值的標準誤 $S_{\hat{\beta}}$

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	截距	8977.694	663.3645	13.53358	1.01E-05
18	廣告支出	5.75431	0.745954	7.714029	0.000249

林惠玲 陳正倉著 雙葉書局發行 2002

第14章 簡單迴歸分析與相關分析 應用統計學

簡單迴歸分析 評判迴歸模型

◎對 α 的檢定

○Z檢定統計量(σ^2 已知)

$$\frac{\hat{\alpha} - \alpha}{\sigma \sqrt{\frac{\sum X^2}{n \sum x^2}}} \sim Z$$

○t檢定統計量(σ^2 未知)

$$\frac{\hat{\alpha} - \alpha}{S_{Y/X} \sqrt{\frac{\sum X^2}{n \sum x^2}}} \sim t_{n-2}$$

林惠玲 陳正倉著 雙葉書局發行 2002

簡單迴歸分析-評判迴歸模型

③ 的區間估計

- 的信賴區間(σ^2 已知)

$$\hat{\beta} - Z_{\alpha/2} \sigma_{\hat{\beta}} \leq \beta \leq \hat{\beta} + Z_{\alpha/2} \sigma_{\hat{\beta}}$$

其中： $\sigma_{\hat{\beta}} = \frac{\sigma}{\sqrt{\sum x^2}}$

- 的信賴區間(σ^2 未知)

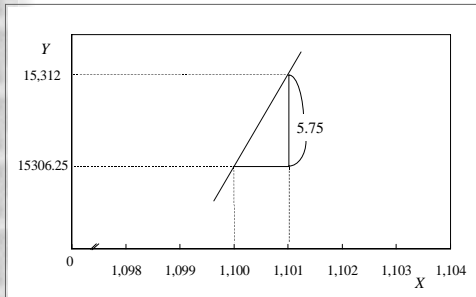
$$\hat{\beta} - t_{n-2, \alpha/2} S_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{n-2, \alpha/2} S_{\hat{\beta}}$$

其中 $S_{\hat{\beta}} = \frac{S_{Y/X}}{\sqrt{\sum x^2}}$

表14.10 β 的 95%的信賴區間

	A	B	C
16		下限 95%	上限 95%
17	截距	7354.498371	10600.88956
18	廣告支出	3.929025447	7.579595242

圖 14.16 迴歸線的斜率



利用迴歸模型做預測

- (1) 在既定 X_0 下預測母體平均數 $E(Y|X_0)$

- $E(Y|X_0)$ 的信賴區間(母體變異數已知)

$$\hat{Y}_0 \pm Z_{\alpha/2} \sigma_{\hat{Y}_0}$$

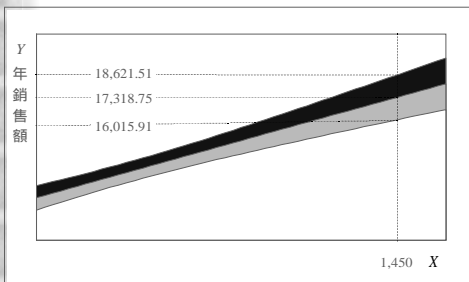
樣本變異數

$$S_{\hat{Y}_0}^2 = S_{Y/X}^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x^2} \right]$$

- $E(Y|X_0)$ 的信賴區間(母體變異數未知)

$$\hat{Y}_0 \pm t_{n-2, \alpha/2} S_{\hat{Y}_0}$$

圖 14.17 $E(Y|X_0)$ 的信賴區間



利用迴歸模型做預測

- t 檢定統計量

$$\frac{\hat{Y} - E_0}{S_{\hat{Y}_0}} \sim t_{n-2}$$

利用迴歸模型做預測

(2)在既定 X_0 預測新觀察值 Y_0

○ Y_0 的信賴區間

$$\hat{Y}_0 \pm t_{n-2, \alpha/2} S_{e_0}$$

圖14.18 $E(Y|X_0)$ 與 Y_0 信賴區間

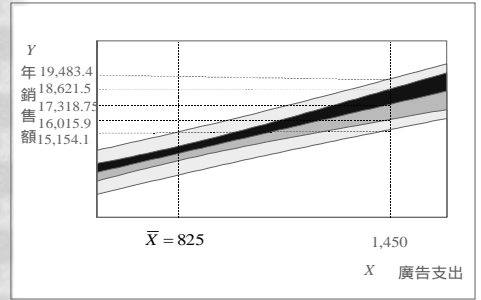


表14.11 自用小客車數與加油站數

縣市別	自用小客車數量 (千輛)	加油站 (千輛)	縣市別	自用小客車數量 (千輛)	加油站 (千輛)
台北縣	617.601	125	高雄縣	238.585	118
宜蘭縣	82.679	38	屏東縣	155.594	80
桃園縣	377.060	164	台東縣	39.086	28
新竹縣	101.866	55	花蓮縣	67.619	37
苗栗縣	120.742	59	澎湖縣	10.506	11
台中縣	329.824	138	基隆市	55.722	19
彰化縣	258.970	110	新竹市	83.646	23
南投縣	114.185	65	台中市	233.638	68
雲林縣	135.319	73	嘉義市	54.236	20
嘉義縣	101.102	63	台南市	148.419	45
台南縣	222.034	113			

表14.13 迴歸式的判定係數與F值

	A	B	C	D	E	F
3	迴歸統計					
4	R-平方	0.837552743				
5	R-平方	0.781484587				
6	調整的 R-平方	0.685783787				
7	標準誤	24.44941516				
8	調整標準誤	23				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著性
12	迴歸	1	24690.8673	24690.8673	44.65043929	2.37E-06
13	殘差	19	11357.79413	597.7739016		
14	總和	20	36048.66143			

相關分析

○ 相關係數

$$\rho_{XY} = E\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right)$$

$$= \frac{E(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}$$

$$= \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

式中： ρ_{XY} 為 X、Y 隨機變數的相關係數， μ_X 、 σ_X 為 X 的平均數與標準差， μ_Y 、 σ_Y 為 Y 的平均數與標準差， σ_{XY} 為 X 與 Y 的共變數。

相關分析

○ 樣本的相關係數

$$r_{XY} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}}$$

$$= \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}} = \frac{S_{XY}}{S_X S_Y}$$

式中： $x = X - \bar{X}$ $y = Y - \bar{Y}$ ， $S_{XY} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{n-1}$

(樣本共變數)， $S_X = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$ (X 的樣本標準差)，

$S_Y = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n-1}}$ (Y 的樣本標準差)。

表14.15 廣告與銷售額的相係數

	A	B	C
1		廣告支出	年銷售額
2	廣告支出	1	
3	年銷售額	0.953103	1

相關分析- 的統計推論

- t 檢定統計量

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim t_{n-2}$$

- Z_ρ 的信賴區間

$$Z_r - Z_{\alpha/2} \sqrt{\frac{1}{n-3}} \leq Z_\rho \leq Z_r + Z_{\alpha/2} \sqrt{\frac{1}{n-3}}$$

相關分析與迴歸分析的關係

- $\hat{\beta}$ 與 r_{XY} 間的關係

$$\hat{\beta} = r_{XY} \frac{S_Y}{S_X}$$

- 相關係數與判定係數

$$r_{XY}^2 = R^2$$

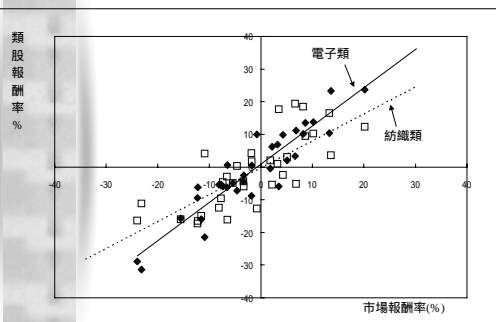
表14.16 相關係數與判定係數

	A	B	C	D
3	迴歸統計			
4	R 的指數	0.953103182	相關係數	r_{XY}
5	R 平方	0.908405675	判定係數	R^2
6	調整的 R 平方	0.893139955		
7	標準誤	700.4026962		S_{MY}
8	觀察直數	3		

表14.18 電子類股票報酬率的迴歸結果

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	截距	0.870396	0.870449	0.99994	0.325081
18	大盤報酬率	1.170958	0.085364	13.71726	1.04E-14

圖14.25 股票的報酬率



殘差分析:檢定模型的假設條件

- 檢查變異齊一性
- 檢查序列相關
- 檢查模型是否為線性模型
- 檢查模型是否為常態分配

表14.21 銷售額的預測值與殘差值

	A	B	C	D
24	觀測值	年銷售額預測值	殘差	標準化殘差
25	1	17033.72845	-233.728	-0.33370581
26	2	15307.43534	-907.435	-1.29559088
27	3	12200.10776	99.89224	0.142621155
28	4	12717.99569	482.0043	0.683181689
29	5	12200.10776	599.8922	0.856496191
30	6	14041.48707	358.5129	0.511866863
31	7	15307.43534	592.5647	0.846034229
32	8	10991.70259	-991.703	-1.41590344

圖14.26 對x的殘差圖

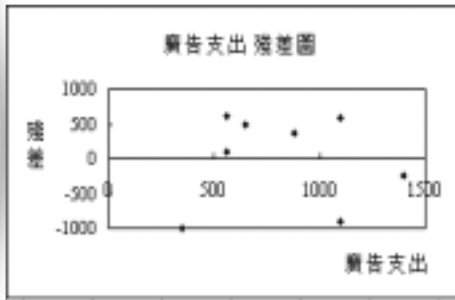


圖14.27 對ŷ的殘差圖

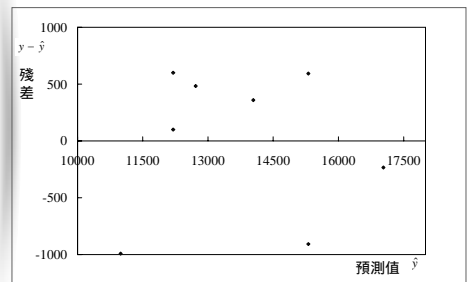
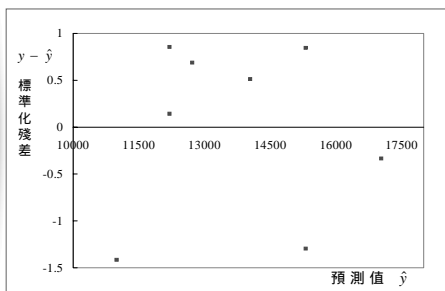


圖14.28 對ŷ的標準殘差圖



非線性模型

- 二次式模型
 $Y_i = \alpha + \beta_1 X + \beta_2 X^2 + \epsilon$
- 指數模型
 $Y = Ae^{\beta X} \epsilon$
- 雙對數模型
 $Y = AX^\beta \epsilon$
- 對數模型
 $e^Y = \alpha X^\beta \epsilon$

圖14.37 成本函數

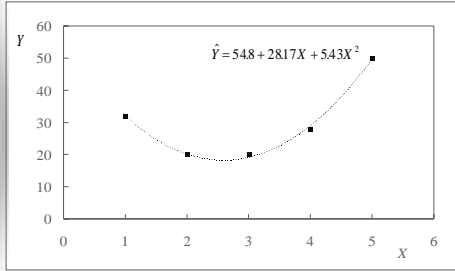


圖14.38 生產函數

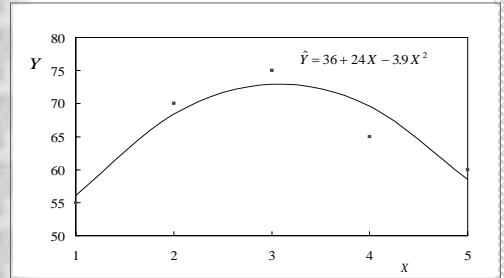


圖14.39 指數模型 $b > 0$

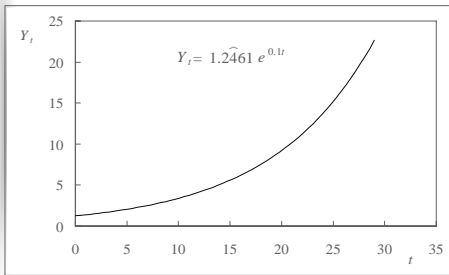


圖14.42 對數模型

