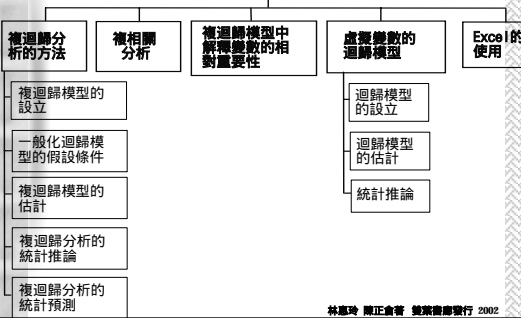


15 複迴歸分析與相關分析

15 複迴歸分析與相關分析

學習目的

複迴歸分析與相關分析



複迴歸模型的設定

○ 複迴歸方程式

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i$$

○ 兩個自變數的複迴歸方程式

$$Y_i = \alpha + \beta X_i + \gamma Z_i + \epsilon_i$$

表15.1 教育年數、工作年數與個人年所得

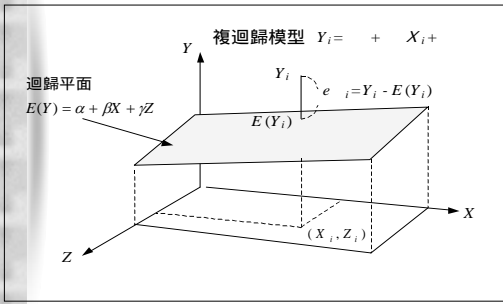
	A	B	C	D
1	樣本	教育年數 X	工作年數 Z	年所得 Y (萬元)
2	1	6	7	38
3	2	9	5	40
4	3	12	14	53
5	4	16	8	50
6	5	18	6	55
7	總計	61	40	236
8	平均數	12.2	8	47.2

複迴歸模型的設定

○ 複迴歸模型的假設條件

- ① $E(\epsilon_i) = 0 \quad i = 1, \dots, n$ ，各組殘差項的條件平均數為0。條件平均數是指在 X_{1i}, \dots, X_{ki} 的條件下，第 i 組的平均數為0。
- ② $V(\epsilon_i) = \sigma^2 \quad i = 1, \dots, n$ ，殘差項的條件變異數均相同為 σ^2 ，具變異數齊一性， σ^2 亦可寫為 $\sigma_{\epsilon_i | X_{1i}, \dots, X_{ki}}$ 。
- ③ $\epsilon_i \sim N$ ，殘差項為常態分配。
- ④ $Cov(\epsilon_i, \epsilon_j) = 0 \quad i \neq j, i, j = 1, \dots, n$ ，任何兩組殘差項不相關，或共變數為0，或無自我相關或稱無序列相關。
- ⑤ $Cov(X_j, \epsilon_i) = 0 \quad j = 1, \dots, k, i = 1, \dots, n$ ，殘差項 ϵ_i 與 k 個自變數無關。
- ⑥ 自變數 $(X_j, j = 1, \dots, k)$ 為預先選定的變數，依變數 Y 為隨機變數。
- ⑦ 自變數彼此間無完全的線性關係，即 $r_{X_i X_j} \neq \pm 1 \quad i \neq j, i, j = 1, \dots, k$ 。
- ⑧ 樣本數 $n > k + 1$ ，在複迴歸模型若有 k 個自變數，則有 $k + 1$ (包括截距項 α) 個迴歸參數，此時利用樣本來估計迴歸參數時，樣本數必須大於 $k + 1$ 個。

圖15.1 複迴歸模型



複迴歸模型的估計 普通最小平方

- 觀察值與估計值差的平方和

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

$$= \sum_{i=1}^n e_i^2$$

- 迴歸係數的估計

$$\hat{\beta} = \frac{\sum z^2 \sum xy - \sum xz \sum zy}{\sum x^2 \sum z^2 - (\sum xz)^2}$$

$$\hat{\gamma} = \frac{\sum x^2 \sum zy - \sum xz \sum xy}{\sum x^2 \sum z^2 - (\sum xz)^2}$$

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

圖15.2 估計迴歸平面

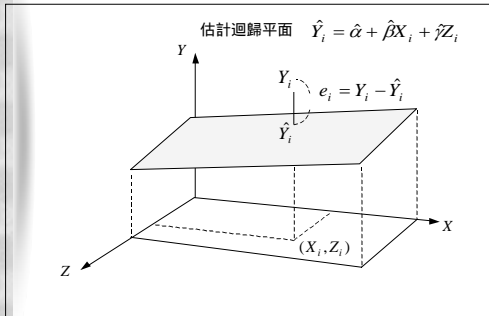


表15.2 $\hat{\beta}, \hat{\gamma}$ 計算表

X	Z	Y	x	z	y	x ²	z ²	y ²	xz	xy	zy
6	7	38	-6.2	-1	-9.2	38.44	1	84.64	6.2	57.04	9.2
9	5	40	-3.2	-3	-7.2	10.24	9	51.84	9.6	23.04	21.6
12	14	53	-0.2	6	5.8	0.04	36	33.64	-1.2	-1.16	34.8
16	8	50	3.8	0	2.8	14.44	0	7.84	0	10.64	0
18	6	55	5.8	-2	7.8	33.64	4	60.84	-11.6	45.24	-15.6
總和	61	40	236	0	0	96.80	50	238.80	3	134.80	50.0

圖15.3 教育、工作年數與所得

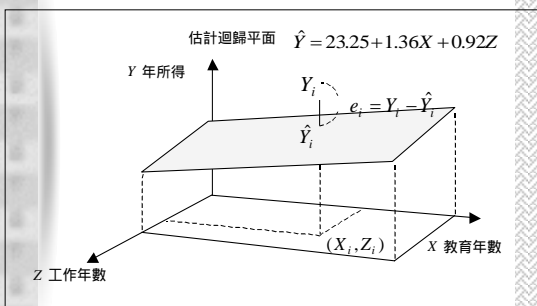


表15.3 教育、工作年數與所得的複迴歸結果

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	數距	23.21267	3.615649	6.420056	0.023413
18	教育年數	1.364107	0.215939	6.31709	0.024155
19	工作年數	0.918154	0.300458	3.055845	0.092473

複迴歸分析的統計推論

(1) 迴歸方程式的配合度

○ 依變數的總變異

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

① 判定係數

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SSR}{SST}$$

複迴歸分析的統計推論

○ 調整的判定係數

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / n - k - 1}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / n - 1} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n Y_i^2} \times \frac{n-1}{n-k-1}$$

式中：n為樣本數，k為自變數的個數。

表15.4 判定係數

	A	B	C	D
3	迴歸統計			
4	R 的指數	0.980952	複相關係數 $r_{Y\hat{Y}}$	
5	R 平方	0.962267	R^2	
6	調整的 R 平方	0.924533	\bar{R}^2	
7	F 檢定值	2.122584	$F_{Y\hat{Y}}$	
8	觀察值個數	5		

複迴歸分析的統計推論

② 迴歸方程式有無解釋能力 - F 檢定

○ F 檢定統計量

$$F = \frac{MSR}{MSE} \sim F_{k, n-k-1, \alpha}$$

○ 檢定法則

- ① 若 $F > F_{k, n-k-1, \alpha}$ ，則拒絕 H_0 。
- ② 若 $F \leq F_{k, n-k-1, \alpha}$ ，則接受 H_0 。

表15.6 所得對教育年數、工作年數迴歸的變異數分析表

變數來源	平方和	自由度	平均平方和	F	p 值
	SS	d.f.	MS		
迴歸	229.33	2	114.67	24.2	0.038
隨機	9.47	2	4.74		
總和	238.80	4			

圖15.4 教育、工作年數對所得影響的檢定

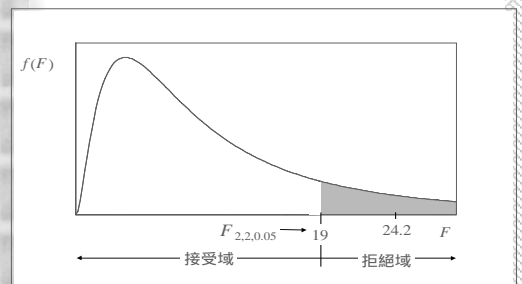


表15.7 教育、工作年數對所得影響的F檢定

	A	B	C	D	E	F
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	回歸	2	229.7893	114.8946	25.50176	0.037733
13	殘差	2	9.010722	4.505361		
14	總和	4	238.8			

複迴歸分析的統計推論

◎個別迴歸參數的檢定

表15.8 個別迴歸參數的t值

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	截距	23.21267	3.615649	6.420056	0.023413
18	教育年數	1.364107	0.215939	6.31709	0.024155
19	工作年數	0.918154	0.300458	3.055845	0.092473

複迴歸分析的統計預測

(1) $E(Y_0)$ 的信賴區間

$$\hat{Y}_0 \sim N(E(Y_0), \sigma_{\hat{Y}_0}^2)$$

其中

$$\sigma_{\hat{Y}_0}^2 = \sigma^2 \left[\frac{x_0^2 \sum z^2 + z_0^2 \sum x^2 - 2x_0 z_0 \sum xz}{\sum x^2 \sum z^2 - (\sum xz)^2} + \frac{1}{n} \right]$$

○ $E(Y_0)$ 的信賴區間 (σ^2 已知)

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

○ $E(Y_0)$ 的信賴區間 (σ^2 未知)

$$\hat{Y}_0 \pm t_{n-3, \alpha/2} S_{\hat{Y}_0}$$

$$\text{式中： } S_{\hat{Y}_0}^2 = S_{Y/XZ}^2 \cdot \left[\frac{x_0^2 \sum z^2 + z_0^2 \sum x^2 - 2x_0 z_0 \sum xz}{\sum x^2 \sum z^2 - (\sum xz)^2} + \frac{1}{n} \right]$$

複迴歸分析的統計預測

(2) Y_0 的信賴區間

○ $Y_0 - \hat{Y}_0 = e_0$ 的抽樣分配

$$Y_0 - \hat{Y}_0 = e_0 \sim N(0, \sigma_e^2)$$

$$\text{其中： } \sigma_e^2 = \sigma^2 \left[\frac{x_0^2 \sum z^2 + z_0^2 \sum x^2 - 2x_0 z_0 \sum xz}{\sum x^2 \sum z^2 - (\sum xz)^2} + \frac{1}{n} + 1 \right]$$

○ Y_0 的信賴區間 (σ^2 已知)

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

○ Y_0 的信賴區間 (σ^2 未知)

$$\hat{Y}_0 \pm t_{n-3, \alpha/2} S_{e_0}$$

$$\text{其中： } S_{e_0}^2 = S_{Y/XZ}^2 \cdot \left[\frac{x_0^2 \sum z^2 + z_0^2 \sum x^2 - 2x_0 z_0 \sum xz}{\sum x^2 \sum z^2 - (\sum xz)^2} + \frac{1}{n} + 1 \right]$$

表15.10 房屋價格、面積與衛浴設備的資料

房價	830	885	950	950	930	980	990	1080	1150	1250	1350	1475
面積	25.5	47.8	36.2	26.9	26.9	37.4	31.5	32	35	38.5	39.2	51
衛浴	1.5	2.5	2	1.5	1.5	1.5	1.5	2	2	2	2.5	1.5
房價	1475	1670	1680	1880	550	610	660	750	888	1148	1058	1050
面積	51	48	47.2	51.2	23.8	24.4	33.3	23.2	26.2	36.4	30.4	29.6
衛浴	1.5	2	2	3	1	1	1	1	1	2	2	2
房價	1150	950	1220	1250	1350	1250	1380	1298	1300	1320	1360	1400
面積	37.5	22.5	39.6	37.3	41.4	33	48	45.4	41.3	36.3	39.5	45.2
衛浴	1.5	2	1.5	2	2	1.5	1.5	2	2	2	1.5	2
房價	1450	1480	1520	1900	2200	1850	2100	2300	2480	2780	3685	2750
面積	40.5	36.9	43.9	60.8	65.2	43.8	59.7	60	67.8	48.5	82.9	61.4
衛浴	2	2	2	2	2	3	3	3	3	3	3	1

表15.11 房價對房屋面積、衛浴設備數的迴歸結果

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	截距	-483.767	151.0281	-3.20316	0.002498
18	面積	37.59355	3.771905	9.966728	5.77E-13
19	衛浴設備	177.6134	83.68053	2.122517	0.039331

表15.12 複迴歸方程式的判定係數與ANOVA表及F值

	A	B	C	D	E	F	
3	迴歸統計						
4	R的係數	0.900309					
5	R平方	0.810377					
6	調整的R平方	0.801949					
7	標準誤	274.7726					
8	觀察值個數	48					
9							
10	ANOVA						
11		自由度	88	145	F	顯著值	
12	迴歸		1	14519620	7299810	96.15643	5.66E-17
13	殘差		45	3397500	75489.99		
14	總和		47	17917119			

複相關分析

- Y與 \hat{y} 的複相關係數

$$r_{\hat{y}Y} = \frac{\sum y\hat{y}}{\sqrt{\sum y^2} \sqrt{\sum \hat{y}^2}} \text{ 或 } \sqrt{R^2}$$

- t檢定統計量

$$\frac{r_{\hat{y}Y}}{\sqrt{1-r_{\hat{y}Y}^2}} \sim t_{n-k-1}$$

表15.13 複相關係數

	A	B	C
3	迴歸統計		複相關係數 $r_{\hat{y}Y}$
4	R的係數	0.980951913	
5	R平方	0.962266657	R^2
6	調整的R平方	0.924533313	\bar{R}^2
7	標準誤	2.122583616	
8	觀察值個數	5	

圖15.5 虛擬變數的二條迴歸線

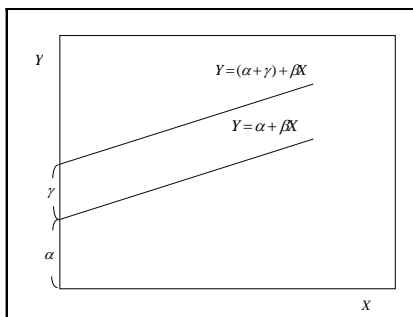


表15.14 都市與鄉村汽車消費支出資料

	A	B	C	D
1	樣本	平均年所得Y	虛擬變數D	汽車消費支出Y'
2	1	100	1	10
3	2	88	0	6
4	3	120	1	14
5	4	150	1	16
6	5	78	0	5
7	6	110	1	11
8	7	130	1	15
9	8	140	1	12
10	9	98	0	8
11	10	105	0	10
12	11	110	0	7
13	12	115	0	9
14	13	98	1	10
15	14	88	1	8
16	15	75	0	6

資料來源：本資料為虛擬。單位：萬元。

表15.15 汽車消費支出與所得的關係的迴歸結果

	A	B	C	D	E
16		係數	標準差	t統計	F-值
17	截距	-0.9	1.445812	-0.62249	0.545269
18	平均年所得X	0.088837	0.014938	5.947071	6.74E-05
19	總平方變數D	2.433721	0.695516	3.499159	0.004389

圖15.6 汽車消費支出與所得的關係

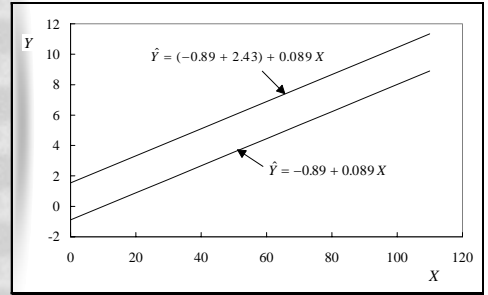


圖1.1 母體與樣本間的關係

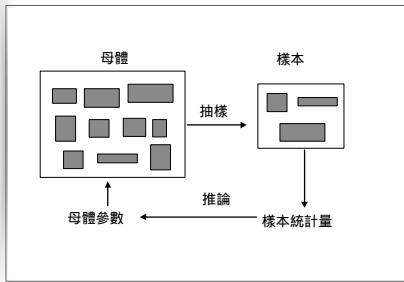


表1.1 景豐電腦經銷商的進貨

公司名稱	營業收入(萬元)	變數
經技	5,850	觀察值
盛聲	7,605	
全華	15,860	
博捷	48,900	
安文	69,200	

圖1.2 歸納法

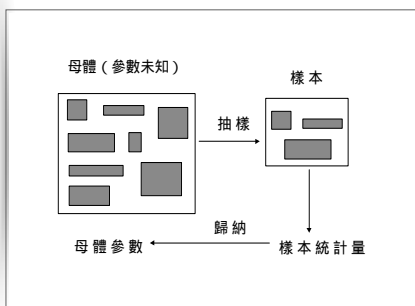


圖1.3 演繹法

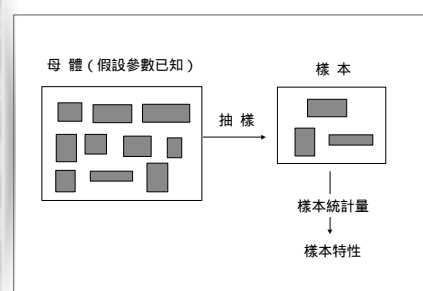


圖1.4 統計方法

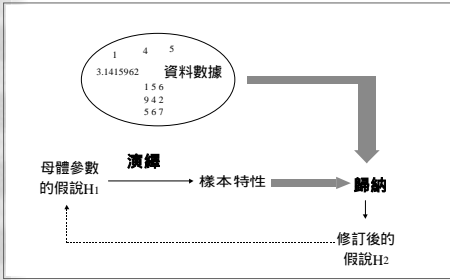


圖1.5 統計方法的實施步驟

