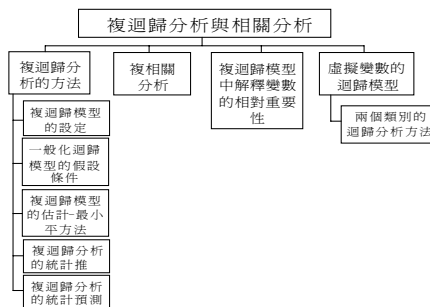


## 16 複迴歸分析與相關分析

### 學習目的

1. 了解複迴歸模型的設定方法。
2. 了解如何利用 OLS 估計複迴歸模型。
3. 了解複迴歸模型的統計推論與統計預測。
4. 比較複迴歸模型中解釋變數的相對重要性。
5. 解如何設定虛擬變數去衡量類別的解釋變數。
6. 學習如何利用 Excel 進行複迴歸分析。

### 本章結構



### 複迴歸分析與相關分析

#### ○ 複迴歸分析方法

研究兩個或兩個以上的自變數對依變數的影響的分析方法即是所謂的複迴歸分析方法 (multiple regression analysis)。複迴歸方程式又稱為多元迴歸方程式。

### 複迴歸分析與相關分析

#### ○ 複迴歸方程式

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

### 複迴歸分析的方法

表16.2 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 $Y$
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

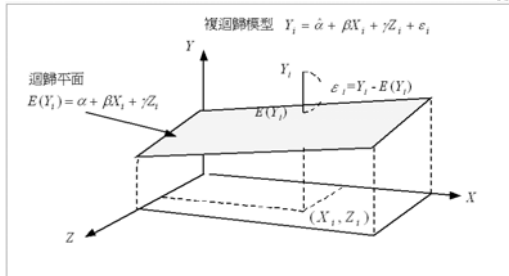
### 複迴歸分析的方法

#### ○ 兩個自變數的複迴歸方程式

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

複迴歸分析的方法

圖16.1 複迴歸模型



林島玲 陳正倉著 雙葉書局發行 2009

複迴歸分析的方法

○ 複迴歸模型的假設條件

- ① 依變數 $Y_i$ 為隨機變數，自變數 $(X_i, i=1, \dots, k)$ 為預先選定的變數。
- ②  $E(e_i) = 0 \quad i=1, \dots, n$ ，各組殘差項的條件平均數為0。條件平均數是指在 $X_{i1}, \dots, X_{ik}$ 的條件下，第組的平均數為0。
- ③  $V(e_i) = \sigma^2 \quad i=1, \dots, n$ ，殘差項的條件變異數均相同為 $\sigma^2$ ，具變異數齊一性 (homoscedasticity)， $\sigma^2$ 亦可表為 $\sigma^2_{Y_i, X_1, \dots, X_k}$ 。
- ④  $e_i \sim N$ ，殘差項為常態分配。
- ⑤  $Cov(e_i, e_j) = 0 \quad i \neq j, i, j=1, \dots, n$ ，任何兩組殘差項不相關，或共變數為0，或無自我相關 (autocorrelation) 或稱無序列相關 (serial correlation)。
- ⑥  $Cov(X_j, e_i) = 0 \quad j=1, \dots, k, i=1, \dots, n$ ，殘差項 $e_i$ 與 $k$ 個自變數無關。
- ⑦ 自變數彼此間無完全的線性關係。
- 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 $\alpha$ )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。

林島玲 陳正倉著 雙葉書局發行 2009

複迴歸分析的方法

○ 複迴歸方程式的估計式

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

林島玲 陳正倉著 雙葉書局發行 2009

複迴歸分析的方法

○ 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

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

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

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

林島玲 陳正倉著 雙葉書局發行 2009

複迴歸分析的方法

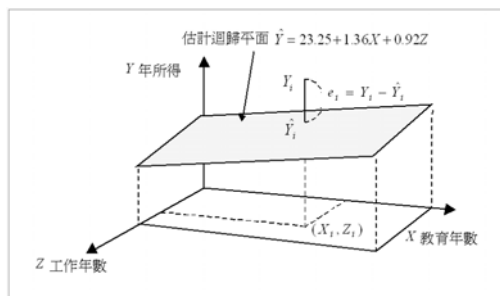
表16.3  $\hat{\beta}, \hat{\gamma}$  計算表

$\rho$	$X_i$	$Z_i$	$Y_i$	$x_i$	$z_i$	$y_i$	$x_i^2$	$z_i^2$	$y_i^2$	$x_i z_i$	$x_i y_i$	$z_i y_i$
$\rho$	6	7	38	-6.2	-1	-9.2	38.44	1	84.64	6.2	57.04	9.2
$\rho$	9	5	40	-3.2	-3	-7.2	10.24	9	51.84	9.6	23.04	21.6
$\rho$	12	14	53	-0.2	6	5.8	0.04	36	33.64	-1.2	-1.16	34.8
$\rho$	16	8	50	3.8	0	2.8	14.44	0	7.84	0	10.64	0
$\rho$	18	6	55	5.8	-2	7.8	33.64	4	60.84	-11.6	45.24	-15.6
合計	61	40	236	0	0	0	96.80	50	238.80	3	134.80	50.0

林島玲 陳正倉著 雙葉書局發行 2009

複迴歸分析的方法

圖16.2 教育、工作年數與所得



林島玲 陳正倉著 雙葉書局發行 2009

複迴歸分析的方法

表16.4 EXCEL的迴歸結果

	A	B	C	D	E
16		係數	標準誤	t 統計	P-值
17	截距	23.2126682	3.61564863	6.42005642	0.02341306
18	教育年數	1.36410681	0.2159391	6.31709031	0.02415487
19	工作年數	0.91815359	0.30045814	3.05584534	0.09247313

複迴歸分析的方法

○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

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

總變異=迴歸解釋的變異+隨機變異  
表為：

$$SST = SSR + SSE$$

○ 判定係數

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

複迴歸分析的方法

○ 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。

複迴歸分析的方法

表16.5 判定係數

	A	B	C
3	迴歸統計		
4	R 的係數	0.980951913	複相關係數
5	R 平方	0.962266657	判定係數
6	調整的 R 平方	0.924533313	調整的判定係數
7	標準誤	2.122583616	
8	觀察值個數	5	

複迴歸分析的方法

表16.6 複迴歸變異數分析表

變異來源	平方和 SS	自由度 df	平均平方和 MS	F 值
迴歸	$SSR = \sum \hat{y}^2 = \sum (\hat{Y} - \bar{Y})^2$	k	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
隨機	$SSE = \sum e^2 = \sum (Y - \hat{Y})^2$	n - k - 1	$MSE = \frac{SSE}{n - k - 1}$	○
總和	$SST = \sum y^2 = \sum (Y - \bar{Y})^2$	n - 1	○	○

複迴歸分析的統計推論

○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  $H_0$ 。

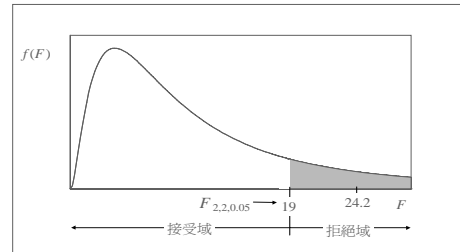
複迴歸分析的方法

表16.7 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 SS	自由度 df	平均平方和 MS	F 值	P 值
迴歸	229.33	2	114.67	24.2	0.038
隨機	9.47	2	4.74		
總合	238.80	4			

複迴歸分析的方法

圖16.3 教育、工作年數對所得影響的檢定



複迴歸分析的方法

表16.8 教育年數、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	迴歸	2	229.78928	114.89464	25.50176	0.037733
13	殘差	2	9.0107224	4.5053612		
14	總和	4	238.8			

複迴歸分析的方法

③個別迴歸參數的檢定

複迴歸分析的方法

表16.9 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	23.2126682	3.61564863	6.42005642	0.02341306
18	教育年數	1.36410681	0.2159391	6.31709031	0.02415487
19	工作年數	0.91815359	0.30045814	3.05584534	0.09247313

複迴歸分析的方法

表16.10 個別迴歸參數的信賴區間

	A	B	C
16		下限 95%	上限 95%
17	截距	7.655787748	38.76954862
18	教育年數	0.434995843	2.293217777
19	工作年數	-0.374613425	2.210920607

複迴歸分析的方法

表16.11 汽車銷售量對所得、大眾運輸支出、原油價格的迴歸結果

	A	B	C	D	E
16		係數	標準誤	t統計	P-值
17	截距	-46690.5	12664.54	-3.68671	0.006159
18	每人所得	0.532619	0.193211	2.756679	0.024802
19	搭乘大眾交通工具支出	10.89377	2.195103	4.96276	0.001103
20	國際原油平均價格	-2.67151	0.531482	-5.02653	0.001019

複迴歸分析的方法

表16.12 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	0.892817				
5	R平方	0.797123				
6	0.721044					
7	標準誤	582.5498				
8	觀察值個數	12				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	3	1.1E+07	3555716	10.4776	0.00381
13	殘差	8	2714914	339364		
14	總和	11	1.3E+07			

複迴歸分析的方法

○複迴歸分析的統計預測

(1)  $E(Y | X_0, Z_0)$  的信賴區間

○  $\hat{Y}_0$  的抽樣分配

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

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 ( $\sigma^2$  已知)

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

○  $E(Y_0)$  的信賴區間 ( $\sigma^2$  未知)

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

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

複迴歸分析的方法

(2)  $Y_0$  的信賴區間

○  $Y_0 - \hat{Y}_0 = e_0$  的抽樣分配

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

其中： $\sigma_{e_0}^2 = \sigma^2 \left[ \frac{x_0^2 \sum z^2 + z_0^2 \sum x^2 - 2x_0 z_0 \sum xz}{\sum x^2 \sum z^2 - (\sum xz)^2} + \frac{1}{n} + 1 \right]$

○  $Y_0$  的信賴區間 ( $\sigma^2$  已知)

$$\hat{Y}_0 \pm Z_{\alpha/2} \sigma_{e_0}$$

○  $Y_0$  的信賴區間 ( $\sigma^2$  未知)

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

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

複相關分析

○複相關分析

複相關分析是分析依變數Y與一群自變數 $X_1, X_2, \dots, X_k$ 間的相關程度，以複相關係數來衡量。

複相關分析

○Y與 $X_1, X_2, \dots, X_k$ 的複相關係數

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

式中 $\hat{Y}$ 為Y對 $X_1, X_2, \dots, X_k$ 迴歸的估計式。

○t檢定統計量

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

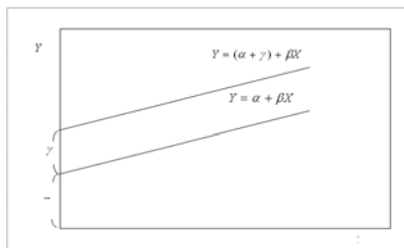
虛擬變數的迴歸模型

○兩個類別的迴歸分析方法

- (1) 迴歸模型的設立
- (2) 迴歸模型的估計
- (3) 統計推論

虛擬變數的迴歸模型

圖16.4 虛擬變數的二條迴歸線



虛擬變數的迴歸模型

表16.13 都市與鄉村汽車消費支出資料

	A	B	C	D
1	標本	平均年所得 $X$	虛擬變數 $D$	汽車消費支出 $Y$
2	1	100	1	10
3	2	80	0	6
4	3	120	1	14
5	4	150	1	16
6	5	70	0	5
7	6	110	1	11
8	7	130	1	13
9	8	140	1	12
10	9	90	0	8
11	10	105	0	10
12	11	110	0	7
13	12	115	0	9
14	13	90	1	10
15	14	80	1	8
16	15	75	0	6

虛擬變數的迴歸模型

表16.14 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	標準誤	t 統計	P-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	虛擬變數	2.433721	0.695516	3.499159	0.004389

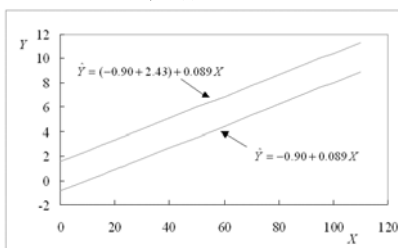
虛擬變數的迴歸模型

表16.15 汽車消費支出與所得的判定係數與ANOVA表

	A	B	C	D	E	F
3	迴歸統計					
4	R 的係數	0.93912				
5	R 平方	0.881946				
6	調整的 R 平方	0.862271				
7	標準誤	1.170785				
8	觀察值個數	15				
9						
10	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	2	122.884	61.4422	44.8243	2.7E-06
13	殘差	12	16.4488	1.37074		
14	總和	14	139.333			

虛擬變數的迴歸模型

圖16.5 汽車消費支出與所得的關係



虛擬變數的迴歸模型

表16.16 房屋價格、面積與衛浴設備的資料

房價	825	660	1,200	500	1,050	480	1,810	1,350	980	630	1,050
坪數	22.6	14.26	46.11	20.22	33.39	15.6	53.56	45.22	26.51	21.73	47.28
衛浴數	2	1	2	1	1.5	1	1	2	1	2	2
樓別	4	4	5	4	4	1	1	4	4	3	3
地區別 (D)	1	1	1	0	1	0	1	0	1	0	0
房價	1,580	500	730	1,180	480	2,200	1,280	687	620	1,380	920
坪數	56.66	20.22	21.76	30.85	15.6	36.58	30.46	28.59	21.5	34.5	31.68
衛浴數	2	1	1	2	1	2	2	2	1	2	2
樓別	1	4	3	4	1	1	1	3	4	3	4
地區別 (D)	0	0	1	1	0	0	0	0	0	1	0
房價	1,200	580	1,580	1,150	580	680	1,300	1,480	790		
坪數	26.95	22.52	35.32	39.87	22.52	28.86	36.77	36.51	23.54		
衛浴數	2	1	1	2	1	1.5	2	2	1		
樓別	1	2	1	4	2	5	4	4	4		
地區別 (D)	0	0	1	0	0	0	1	1	0		

林島玲 陳正倉著 雙葉書房發行 2009

虛擬變數的迴歸模型

表16.17 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	232.056687	171.339048	1.35437129	0.18727233
18	坪數	23.8962451	4.61035548	5.1831676	2.0737E-05
19	衛浴數	194.671235	102.6792	1.89591694	0.06913753
20	樓別	-116.76216	32.4626462	-3.5968158	0.00132531
21	D1	267.123073	89.269383	2.99232574	0.00599609

林島玲 陳正倉著 雙葉書房發行 2009

虛擬變數的迴歸模型

表16.18 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	0.8770502				
5	R 平方	0.7692171				
6	調整的 R 平方	0.733712				
7	標準誤	227.61325				
8	觀察值個數	31				
9	ANOVA					
11		自由度	SS	MS	F	顯著值
12	迴歸	4	4489662	1122415	21.66499	5.79E-08
13	殘差	26	1347003	51807.79		
14	總和	30	5836664			

林島玲 陳正倉著 雙葉書房發行 2009

虛擬變數的迴歸模型

表16.19 身高與薪資

身高	180.1	156.7	172.7	167.2	182.2	168.9	169.3	171.4	169.9	165.2	175.7
薪資	43,300	49,200	44,100	32,000	36,000	40,900	34,000	57,800	42,500	46,400	35,000
性別	1	1	1	1	1	1	1	1	1	1	1
身高	158.8	163.8	169.4	178.2	167.5	175	171.7	158.4	167.7	158.5	153.9
薪資	41,000	43,100	50,500	22,700	26,300	42,800	40,900	41,900	41,100	23,800	40,600
性別	1	1	1	1	1	1	1	1	1	0	0
身高	158.2	164.7	164.7	146.8	153.8	161.8	168.1	160.6	158.1	157.3	159.2
薪資	29,500	31,600	30,300	39,500	34,400	26,600	29,900	34,100	39,000	25,600	33,300
性別	0	0	0	0	0	0	0	0	0	0	0

林島玲 陳正倉著 雙葉書房發行 2009