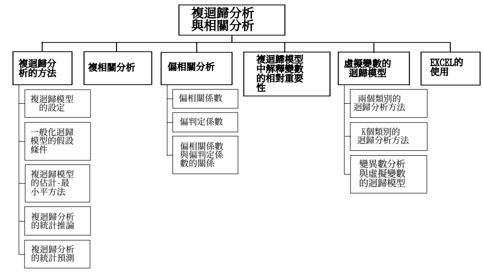
16 複迴歸分析與相關分析

學習目的

1. 了解複迴歸分析及複相關分析的意義、重要性、估計方法及步驟。
2. 了解如何估計複迴歸方程式。
3. 了解如何做複迴歸的統計推論。
4. 了解如何使用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。
5. 了解複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。
6. 利用Excel來做複迴歸分析與複相關分析。

本章結構



複迴歸模型的設定

- **複迴歸方程式**

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

- **兩個自變數的複迴歸方程式**

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

表16.1 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
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 $E(\varepsilon_i)=0 \quad i=1,\cdots,n$，各組殘差項的條件平均數為0。條件平均數是指在 $X_{1i},\cdots,X_{Ki}$ 的條件下，第i組的平均數為0。

② $V(\varepsilon_i)=\sigma^2 \quad i=1,\cdots,n$，殘差項的條件變異數均相同為$\sigma^2$，具變異數齊一性，$\sigma^2$ 亦可寫為 $\sigma^2_{Y/X_1X_2\cdots X_k}$。

③ $\varepsilon_i \sim N$，殘差項為常態分配。

④ $Cov(\varepsilon_i,\varepsilon_j)=0 \;\; i\neq j,\; i,j=1,\cdots,n$，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 $Cov(X_j,\varepsilon_i)=0 \quad j=1,\cdots,k,\; i=1,\cdots,n$，殘差項$\varepsilon_i$與$k$個自變數無關。

⑥自變數($X_i, i=1,\cdots,k$)為預先選定的變數，依變數Y_i為隨機變數。

⑦自變數彼此間無完全的線性關係，即 $r_{X_iX_j}\neq \pm 1 \quad i\neq j \;\; i,j=1\cdots k$。

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。

複迴歸模型的估計—普通最小平方法

- **複迴歸方程式的估計式**

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

複迴歸模型的估計─普通最小平方法

- 觀察值與估計值差的平方和

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

- 迴歸係數的估計

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

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

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

表16.3 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	23.212668	3.615648628	6.4200564	0.0234131
18	教育年數X	1.3641068	0.215936102	6.3170903	0.0241549
19	工作年數Z	0.9181536	0.300458135	3.0558453	0.0024731

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

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

① 複判定係數

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

複迴歸分析的統計推論

- 調整的複判定係數

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

式中：n為樣本數，k為自變數的個數。

表16.3 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	23.212668	3.615648628	6.4200564	0.0234131
18	教育年數X	1.3641068	0.215936102	6.3170903	0.0241549
19	工作年數Z	0.9181536	0.300458135	3.0558453	0.0024731

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

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

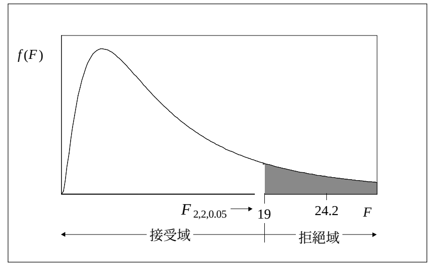
- 檢定法則

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

表16.5 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*
迴歸	$SSR=\sum\hat{y}^2=\Sigma(\hat{Y}-\bar{Y})^2$	k	$MSR=\frac{SSR}{k}$	$F=\frac{MSR}{MSE}$
隨機	$SSE=\sum e^2=\Sigma(Y-\hat{Y})^2$	$n-k-1$	$MSE=\frac{SSE}{n-k-1}$	
總和	$SST=\sum y^2=\Sigma(Y-\bar{Y})^2$	$n-1$		

圖16.5 教育、工作年數對所得影響的檢定



複迴歸分析的統計推論

(2)部分迴歸係數的檢定-F 檢定

- **F 檢定-檢定部份迴歸係數**

$$F=\frac{(\Sigma\hat{y}_Q^2-\Sigma\hat{y}_k^2)/(Q-k)}{\Sigma e_Q^2/(n-Q-1)}\sim F_{Q-k,n-Q-1}$$

- **檢定法則**

①若$F>F_{Q-k,n-Q-1,\alpha}$，則拒絕H_0。

②若$F\le F_{Q-k,n-Q-1,\alpha}$，則接受$H_0$。

複迴歸分析的統計預測

在既定的 X_0, Z_0 下預測複$E(Y\mid X_0,Z_0)$的信賴區間

(1)$E(Y_0)$的信賴區間

$$\hat{Y}_0\sim N\left(E(Y_0),\sigma^2_{\hat{Y}_0}\right)$$

其中

$$\sigma^2_{\hat{Y}_0}=\sigma^2\left[\frac{x_0^2\sum z^2+z_0^2\sum x^2-2x_0z_0\sum xz}{\sum x^2\sum z^2-(\sum xz)^2}+\frac{1}{n}\right]$$

- **$E(Y_0)$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)**

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

式中：$S^2_{\hat{Y}_0}=S^2_{Y/XZ}\cdot\left[\frac{x_0{}^2\sum z^2+z_0{}^2\sum x^2-2x_0z_0\sum xz}{\sum x^2\sum z^2-(\sum xz)^2}+\frac{1}{n}\right]$

複迴歸分析的統計預測

在既定的 X_0, Z_0 下預測新的觀察值 Y_0 的信賴區間

(2) Y_0的信賴區間

- **$Y_0-\hat{Y}_0=e_0$的抽樣分配**

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

其中：$\sigma^2_{e_0}=\sigma^2\left[\frac{x_0^2\sum z^2+z_0^2\sum x^2-2x_0z_0\sum xz}{\sum x^2\sum z^2-(\sum xz)^2}+\frac{1}{n}+1\right]$

- **Y_0的信賴區間(σ^2 已知)**

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

- **Y_0的信賴區間(σ^2 未知)**

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

其中：$S^2_{e_0}=S^2_{Y/XZ}\cdot\left[\frac{x_0^2\sum z^2+z_0^2\sum x^2-2x_0z_0\sum xz}{\sum x^2\sum z^2-(\sum xz)^2}+\frac{1}{n}+1\right]$

複相關分析

- **Y與$X_1,\cdots,X_k$的複相關係數**

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

- **t檢定統計量**

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

偏相關分析

- X 與 Y 在固定條件下的偏相關係數

$$r_{XY\cdot Z}=\frac{r_{XY}-r_{YZ}\cdot r_{XZ}}{\sqrt{1-r_{YZ}^2}\sqrt{1-r_{XZ}^2}}$$

偏相關分析

- X 對 Y 的偏判定係數

$$r_{XY\cdot Z}^2=\frac{SSR_{XZ}-SSR_{Z}}{SSE_{Z}}$$

式中：SSR_{XZ} 為 Y 對 X、Z 兩個自變數迴歸的解釋變異；SSR_{Z} 為 Y 對 Z 一個自變數迴歸的解釋變異；SSE_{Z} 為 Y 對 Z 一個自變數迴歸的未解釋變異。

- Z 對 Y 的偏判定係數

$$r_{ZY\cdot X}^2=\frac{SSR_{XZ}-SSR_{X}}{SSE_{X}}$$

式中：SSR_{X} 為 Y 對 X 自變數迴歸的解釋變異；SSE_{X} 為 Y 對 X 自變數迴歸的未解釋變異。

偏相關分析

- 偏相關係數與偏判定係數

$$r_{XY\cdot Z}^2=\frac{SSR_{XZ}-SSR_{Z}}{SSE_{Z}}=\left[\frac{r_{YX}-r_{YZ}\cdot r_{XZ}}{\sqrt{1-r_{YZ}^2}\sqrt{1-r_{XZ}^2}}\right]^2$$

虛擬變數的迴歸模型

圖16.6 虛擬變數的二條迴歸線

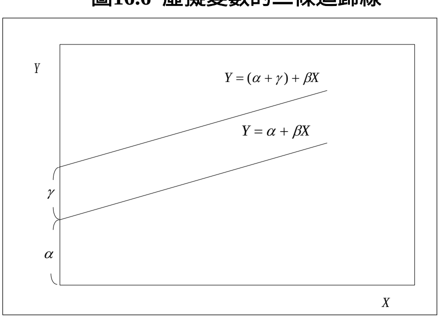


圖16.8 未加入虛擬變數的偏誤

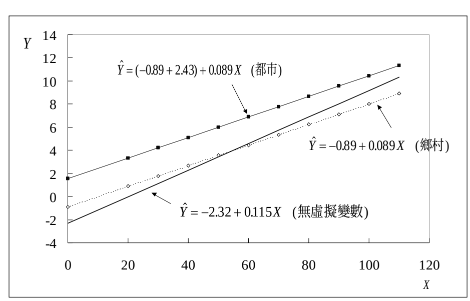


圖16.9 類別變數交叉效果