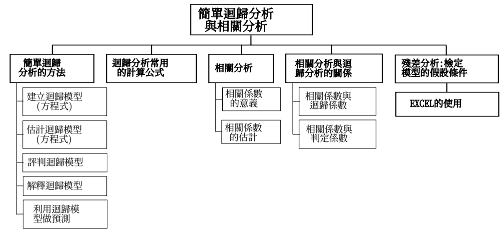
15　迴歸分析與相關分析

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數量關係，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。

圖15.1　變數間的線性關係

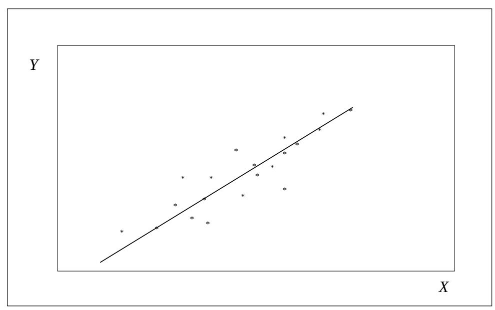
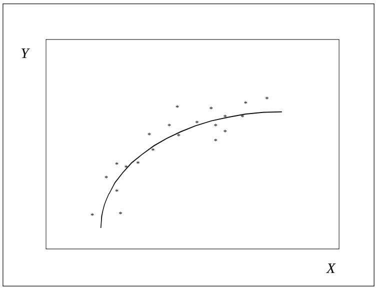


圖15.2　變數間的非線性關係



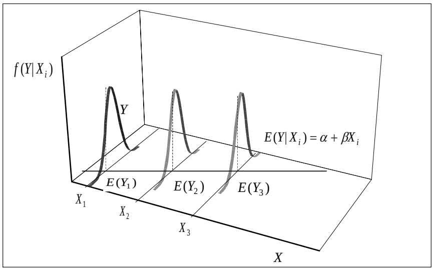
建立迴歸模型

- **簡單線性迴歸模型**

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad i = 1, \cdots, n$$

式中α、β為迴歸模型的參數，α稱為截距，β稱為迴歸係數(regression coefficient)或斜率。

圖15.6　簡單迴歸模型



簡單迴歸模型的假設條件

❍ 迴歸模型的假設條件

① $E(\varepsilon_i)=0$

是指每一組的殘差項的平均數為0，即

② $V(\varepsilon_i)=\sigma^2$ 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 $\varepsilon_i \sim N$ 或 $Y_i \sim N$

各組殘差項為一常態分配，或各組之依變數為一常態分配。

④ $Cov(\varepsilon_i,\varepsilon_j)=0 \quad i\neq j \quad i,j=1,\cdots,n$

ε_i 與 ε_j 的共變數為0，即任何兩組殘差項 ε_i 與 ε_j 間無關。

⑤ $Cov(\varepsilon_i,X)=0$ 或 $E(\varepsilon_i X)=0$

即任何一組殘差項 ε_i 與 X 無關。

⑥ X 為一固定變數或事前決定之變數，Y 為一隨機變數。

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

❍ 估計的迴歸方程式

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

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 $\hat{\beta}$

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

❍ α 的估計式 $\hat{\alpha}$

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

圖15.8　估計的迴歸方程式

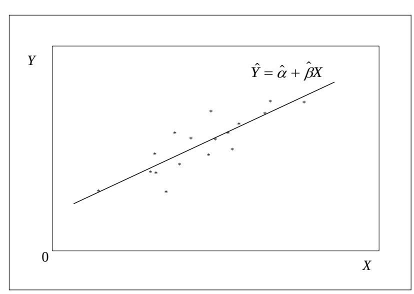


圖15.9 轉換原點的迴歸方程式

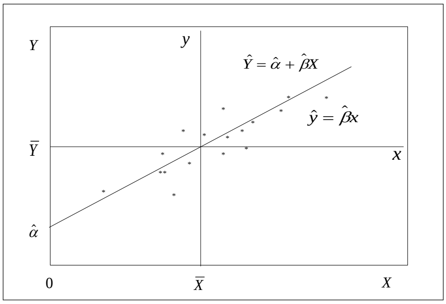
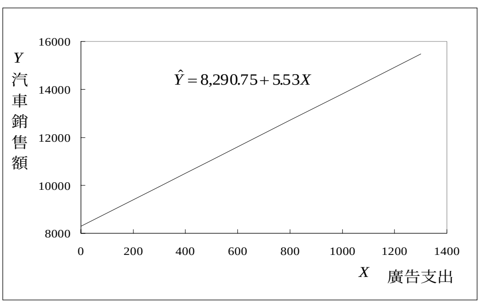


表15.4 廣告支出與營業收入的迴歸結果

	A	B	C	D	E
16		係數	標準誤	t 統計	P-值
17	截距	8289.7143	480.37831	17.256637	2.438E-06
18	廣告費用X	5.5314286	0.6028323	9.1757333	9.438E-05

圖15.10 汽車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$V(\hat{\alpha})=\dfrac{\sum X^2}{n\sum x^2}\sigma^2$，$V(\hat{\beta})=\dfrac{\sigma^2}{\sum x^2}$。

③$\hat{\alpha}$、$\hat{\beta}$的抽樣分配為：

$$\hat{\alpha}\sim N(\alpha\,,\,\frac{\sum X^2}{n\sum x^2}\sigma^2),\quad \hat{\beta}\sim N(\beta\,,\,\frac{\sigma^2}{\sum x^2})$$

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。

- **Gauss-Markov定理**

在下列假設條件下：

① $E(\varepsilon_i)=0$，殘差項為0

② $V(\varepsilon_i)=\sigma^2$，殘差項具變異數齊一性

③ $Cov(\varepsilon_i,\varepsilon_j)=0$，$i\neq j$，無自我相關

④ $Cov(\varepsilon_i,X)=0$，X 與ε_i無相關

利用最小平方法求得之估計式(OLSE)為一最佳線性不偏估計式(BLUE)。證明請參閱附錄。

評判迴歸模型-統計上的檢定

(1)迴歸方程式的配適度

- **依變數的總差異**

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

總差異 = 可解釋的差異 + 不可解釋之差異

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

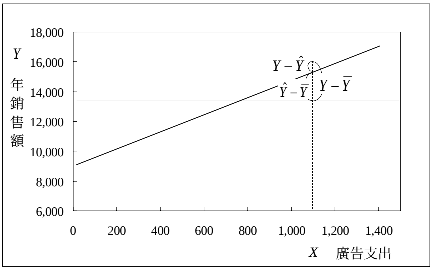
總變異 = 可解釋的變異 + 不可解釋的變異

① **樣本迴歸方程式的判定係數** R^2

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

即 $R^2=\dfrac{\Sigma(\hat{Y}-\bar{Y})^2}{\Sigma(Y-\bar{Y})^2}=1-\dfrac{\Sigma(Y-\hat{Y})^2}{\Sigma(Y-\bar{Y})^2}$

圖13.12　可解釋之差異與不可解釋之差異



評判迴歸模型-統計上的檢定

②*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 $F\le F_{1,n-2,\alpha}$ 時，則接受 H_0。

評判迴歸模型-統計上的檢定

對個別迴歸係數做統計推論-對 β 的檢定

- Z **檢定統計量**(σ^2 **已知**)

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

- t **檢定統計量**(σ^2 **未知**)

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

簡單迴歸分析—評判迴歸模型

對個別迴歸係數做統計推論-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**)

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

簡單迴歸分析-評判迴歸模型

對個別迴歸係數做統計推論- β 的區間估計

- **β 的信賴區間**(σ^2 **已知**)

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

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

- **β 的信賴區間**(σ^2 **未知**)

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

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

利用迴歸模型做預測

(1)在既定 X_0 下預測母體迴歸線的平均數 $E(Y|X_0)$

- $E(Y|X_0)$ **的信賴區間(母體變異數已知)**

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

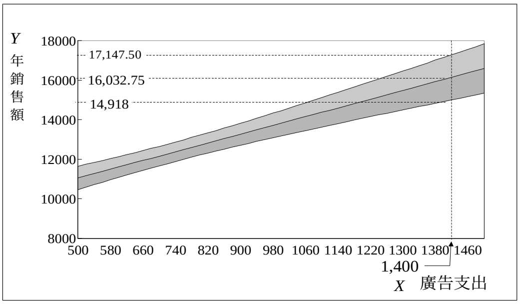
$\hat{Y}_0$ **的估計變異數**

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

- $E(Y|X_0)$ **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15.18 $E(Y \mid X_0)$ 的信賴區間



利用迴歸模型做預測

- t檢定統計量

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

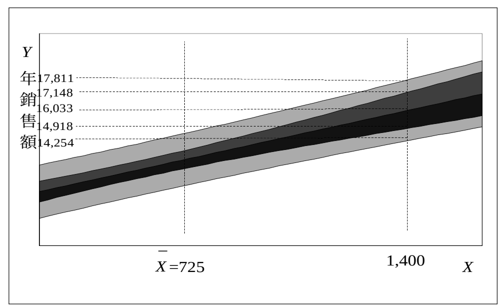
利用迴歸模型做預測

(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15.19 $E(Y \mid X_0)$與 Y_0 的信賴區間



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)。

相關分析- ρ_{xy} 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

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

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

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

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
- 檢查模型否為線性模型
- 檢查殘差是否為常態分配

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

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

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$Cov(\varepsilon_i,\varepsilon_j)=0 \ i\ne j,\ i,j=1,\cdots,n$，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$r_{X_iX_j} \ne \pm 1 \quad i\ne j \quad i,j=1\cdots k$。

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

- 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

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

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

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

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 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。

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。

複迴歸分析的統計推論

③個別迴歸參數的檢定

複迴歸分析的統計預測

(1) $E(Y_0)$ 的信賴區間

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

其中

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

- **$E(Y_0)$ 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：$S_{\hat{Y}_0}^2=S_{Y/XZ}^2\cdot\left[\frac{x_0^2\sum z^2+z_0^2\sum x^2-2x_0z_0\sum xz}{\sum x^2\sum z^2-(\sum xz)^2}+\frac{1}{n}\right]$

複迴歸分析的統計預測

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

- **Y_0 的信賴區間(σ^2 已知)**

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

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

複相關分析

- Y與$\hat{Y}$ 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

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