國立雲林科技大學 工業工程與管理所
National Yunlin University of Science & Technology Industrial Engineering and Management

品質與可靠度工程實驗室
Quality and Reliability Engineering Lab.

迴歸分析

主講人：童超塵

實驗室網址 永久：http://campusweb.yuntech.edu.tw/~qre/index.htm
目前：http://140.125.88.116/QRE

大綱

1. 迴歸模型建構
2. 迴歸模型檢定
3. 迴歸模型診斷
4. 特殊型態的迴歸分析

簡介

- 迴歸分析是建構經驗模式(empirical model)的主要工具。
- 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。
- 區分簡單迴歸分析(simple regression)與複迴歸分析 (multiple regression)。前者為一個自變數與一個依變數。後者為二個或以上的自變數與一個依變數。

迴歸模型之建構

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$

  β_i = 迴歸係數
- 以最小平方法求迴歸係數β_i的估計式b_i

  $y = X\beta + \varepsilon$

  $b = (X'X)^{-1}X'y$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\hat{y} = b_0 + b_1 x_1 + \ldots + b_k x_k$$

- 迴歸係數之顯著性檢定: t 檢定

H_0: $\beta_j = 0$

H_1: $\beta_j \neq 0$

$$t_0 = \frac{b_j}{se(b_j)} = \frac{b_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

If $t > t_{\alpha/2,n-p}$ or $t < -t_{\alpha/2,n-p}$, then reject H_0 .

- 迴歸係數之信賴區間

$$b_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le b_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}}$$

迴歸模型之顯著性檢定

- 總變異=已解釋變異+未解釋變異

$S_{yy} = SS_R + SS_E$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- 判定係數與調整判定係數

判定係數越大表示模型對變異的解釋能力越強

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}}$$

$$R_{adj}^{\ 2} = 1 - \frac{SS_E/(n-p)}{S_{yy}/(n-1)} = 1 - \left(\frac{n-1}{n-p}\right)(1-R^2)$$

迴歸模型之顯著性檢定

- $H_0: \beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$（不具線性關係）

  $H_1: \beta_j \neq 0$ for at least one j

  If $F_0 > F_{\alpha,v1,v2}$, then model is significant.

$$F_0 = \frac{SS_R / k}{SS_E /(n-k-1)} = \frac{MS_R}{MS_E}$$

- 變異數分析表(ANOVA)

	df	SS	MS	F	Significance
迴歸	k	SS_R	MS_R	MS_R/MS_E	p
殘差	n-k-1	SS_E	MS_E		
總和	n-1	S_{yy}			

迴歸模型之充分性檢定

- 在相同的自變數進行重複試驗

  $SS_E = SS_{PE(pure\ error)} + SS_{LOF(lack\ of\ fit)}$

  = 隨機誤差 + 配適誤差

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{m} n_i(\bar{y}_i - \hat{y}_i)^2$$

- If $F_0 > F_{\alpha,v1,v2}$, then LOF is significant, 模式未充分解釋數據

$$F_0 = \frac{SS_{LOF} /(m-p)}{SS_{PE} /(n-m)} = \frac{MS_{LOF}}{MS_{PE}}$$

迴歸模型之診斷

- 迴歸分析的基本假設

 1.殘差變異常態假設

 2.殘差變異常數假設

 3.殘差變異獨立假設

 4.因果線性關係假設

- 利用常態機率圖判定數目是否呈常態分佈
- 利用殘差圖判定2.&3.是否成立

殘差分析(residual analysis)

- 殘差 $e_i = y_i - \hat{y}_i$
- 標準化殘差

$$d_i = \frac{e_i}{\hat{\sigma}} = \frac{e_i}{\sqrt{SS_E/(n-p)}}$$

- 如果di大於3或小於-3 ，表示數據可疑，考慮刪除或重作該實驗
- 如果x殘差圖及y殘差圖中，點的分佈與橫軸無關，則符合殘差變異常數假設
- 如果時序殘差圖中，點的分佈與橫軸無關，則符合殘差變異獨立假設

多項式函數之迴歸分析

- 一階模型(first-order)

$$y=\beta_0+\sum_{i=1}^{k}\beta_i x_i+\varepsilon$$

- 具有交互作用之一階模型(first-order with interaction)

$$y=\beta_0+\sum_{i=1}^{k}\beta_i x_i+\sum_{i=1}^{k}\sum_{j\rangle i}^{k}\beta_{ij}x_i x_j+\varepsilon$$

- 二階模型(second-order)

$$y=\beta_0+\sum_{i=1}^{k}\beta_i x_i+\sum_{i=1}^{k}\sum_{j\rangle i}^{k}\beta_{ij}x_i x_j+\sum_{i=1}^{k}\beta_{ii}x_i^2+\varepsilon$$

- 並非較複雜的模式就較準確可靠，可用F檢定顯著值p大小為參考，p值小者較準確可靠