

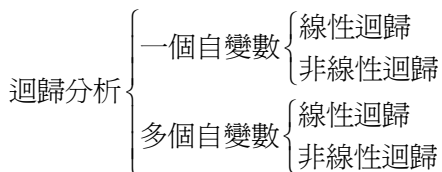
Chapter 12



簡單迴歸與相關分析

12.1 迴歸分析之概念

迴歸分析 (regression analysis) 是用來研究兩個或兩個以上變數間的關係，此方法的主要目的是要建立一個迴歸模式，然後根據一個或多個自變數 (independent variable) 之值，來預測相依變數 (dependent variable) 或稱反應變數 (response variable) 之值。因此迴歸分析的主要目的是在建立變數與變數間之統計關係，然後利用此統計模式去作預測。基本上迴歸分析可區分為：



本章主要內容將探討簡單線性迴歸模式，亦即利用單一自變數 X 來預測反應變數 Y 。至於多個自變數之問題將在下一章討論。



Remark

一般所謂線性迴歸中之“線性 (linear)”是指迴歸模式中所有迴歸參數皆為一次方的型態。

• 例題 1 •

Define the meaning of “linear” in the regression analysis.

(94 中原會計 12%)

Sol: 見課文內容。

一、兩個變數間之關係

兩變數間之關係，基本上可分為函數關係（function relationship）與統計關係（statistical relationship）兩大類。茲分述如下：

(一) 函數關係：

兩變數之間的函數關係，是指一般數學上的一對一及多對一之函數關係，它為一種確定性模式（deterministic model），可以下列數學公式表示，即

$$y = f(x)$$

若將上式函數的觀測值 (x, y) 點繪於圖中，則所有點 (x_i, y_i) 皆會落入線上（見圖 12-1(a)），亦即給一個 x 值，則函數對應到另一個確定 y 值。

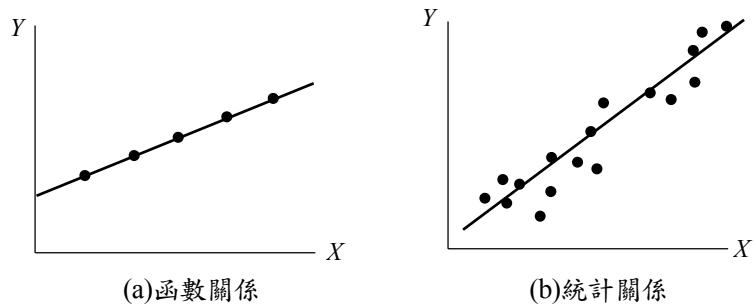


圖 12-1 兩變數之關係

(二) 統計關係：

統計關係與函數關係之不同處，主要在所有觀察值 (x_i, y_i) 並不會皆落在一條直線或曲線上（見圖 12-1(b)）。例如探討廣告費用（ X ）與銷售量（ Y ）之關係，通常不同廣告費，其銷售量亦不同。然而，第一季與第二季的廣告費用相同，並不能表示這兩季的銷售量會相同，因

為還有很多因素會影響銷售量，例如景氣好壞或競爭對手之多寡等。因此兩變數之關係似乎應該以下列方式表達較適當，即

$$y = f(x) + \varepsilon$$

二、簡單直線迴歸模式的建立

(一)直線迴歸模式之建立：

茲舉上例廣告費用（佰萬）與銷售量（千）之關係來說明迴歸模式如何建立。今觀察一組資料如下，且將此資料繪於下圖（見圖12-2），稱之為散布圖（scatter diagram）。即

| | | | | | | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| X (廣告費) | 5 | 5 | 6 | 6 | 6 | 7 | 8 | 8 | 8 | 9 | 9 | 10 | 10 |
| Y (銷售量) | 15 | 18 | 26 | 23 | 27 | 31 | 30 | 29 | 34 | 37 | 35 | 40 | 42 |

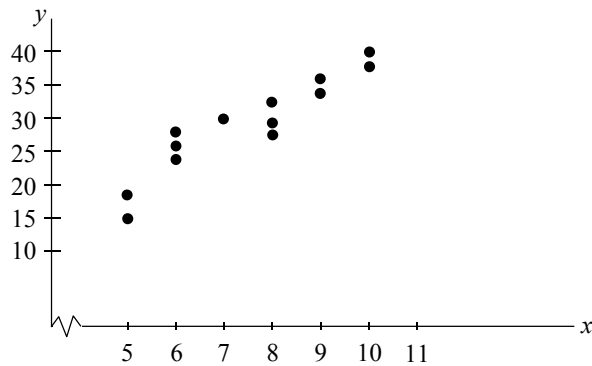


圖12-2 廣告費用與銷售量之散布圖

Remark •

一般所謂散布圖是指將一組資料 (x_i, y_i) 點繪在二維空間上所構成的圖形。

在圖12-2中，可看出廣告費用與銷售量之關係並非直線關係，亦即兩變數並非函數關係，因此銷售量多寡並非完全受廣告費之影響。但我們可粗略看出此兩變數關係接近直線，因此可將其關係建立為

12-4 統計學 (下)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \quad i=1,2,3,\dots,n$$

在上式中 ε_i 稱為隨機誤差項 (random error term) 或稱隨機干擾項 (random disturbance) , 而 β_0 與 β_1 則稱為迴歸係數 (regression coefficient) , 其中 β_0 表示截距項, 而 β_1 表示斜率。

(二) 模式及其基本假設：

在直線迴歸模式中, 是假設自變數 X 與反應變數 Y 接近直線關係, 因此可將模式寫為

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

其中

1. Y_i 是第 i 觀察值的相依變數 (反應變數) 。
2. X_i 為非隨機變數 (nonstochastically variable) 是已知常數, 為自變數的第 i 觀察值。
3. ε_i 是隨機誤差項, 且 $E(\varepsilon_i) = 0$, $\sigma^2(\varepsilon_i) = \sigma^2$, $i=1,2,3,\dots,n$
4. ε_i 和 ε_j 無相關 (uncorrelated) , 所以對所有 i 、 j 而言, 共變異數 $\sigma(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$

Remark •

1. 在迴歸模式中, 若參數及變數皆為線性的模型, 稱之為第一階模型 (first-order model) 。
2. 斜率 β_1 表示自變數 X 每變動一個單位, 所引起反應變數 Y 的平均變動量, 而截距項 β_0 表示在 Y 軸上的截點與原點之間的距離。

12.2 直線迴歸模式之參數估計

在直線迴歸模式中, 迴歸係數 β_0 與 β_1 皆未知, 因此必須利用樣本觀察資料 (x_i, y_i) 去估計參數 β_0 與 β_1 。一般在統計理論中, 尋找參數估計式之方法有許多種 (見第8章) , 在此處只討論普通最小平方法 (ordinary least squares method) 及最大概似法 (maximum likelihood method) 兩種方法。茲分述如下：