4 模糊無母數迴歸

4.1 模糊無母數迴歸簡介

我們常使用迴歸來描述自變數 X 及應變數 Y 之間的因果關係，迴歸參數的估計是迴歸分析的主要研究課題。傳統的迴歸分析常假設觀察值是由實際自變數影響加上隨機誤差而產生的。也就是說，應變數是一種帶有不確定性的隨機變數，但在有些時候，觀察值並不能完全符合我們的假設。例如有些應戀數 Y 的觀察值不見得是一個確定值或不一定是以單一數值的型式存在，而是需要考慮多重隸屬的情況時，這樣的不確定性不再是來自非隨機現象，而是來自多重隸屬現象。

這樣從另一角度來看觀察值不確定性的問題，有別於傳統對統計資料的隨機觀點，這樣的情況下，我們考慮用模糊的觀點來取代隨機觀點。而如果用模糊的觀點套用迴歸分析來處理模糊資料問題，則爲模糊迴歸分析。模糊迴歸分析最早是由 Tanaka，Uejima & Asai（1980）提出的，一般模糊迴歸能處理的樣本型式有三種：一、是實數值樣本 (x_i, y_i)，其中 x_i 爲實數向量，y_i 爲觀測到的實數值；二、是模糊數樣本 (x_i, Y_i)，其中 Y_i 非屬實數的型式，而是觀測到的模糊數，三、是模糊樣本 (X_i, Y_i)，其中 X_i 、 Y_i 非屬實數的型式，而是觀測到的模糊數。之所以可採用實數值樣本 (x_i, y_i) 來推論模糊迴歸模式，是因爲使用者相信觀測值 (x_i, y_i) 內含有強烈的隸屬度資訊。

不過這樣的說法較不具有說服力，理由在於樣本 (x_i, y_i) 的資訊若是以實數值的型式存在，實數值能否將正確的隸屬度資訊含蓋進去，是有待商確的。若想真正從資訊中得要到好的估計，最好是搜集具有代表性的隸屬函數樣本，則這樣的樣本應該是以模糊數的形式存在。所以我們較傾向於以樣本樣糊數 (X_i, Y_i) 的型式來推論模糊無母數迴歸。

4.2 模糊無母數迴歸 Theil 法

假設我們有 n 筆模糊觀察值爲：

$$(X_1,Y_1),(X_2,Y_2),.....,\ (X_n,Y_n)$$

根據這 n 筆資料，我們希望能找到一條適合這筆資料的模糊迴歸直線，型式如下：

$$Y = a + bX$$

其中 a 是 y 的截距， b 是這條直線的斜率。

根據 Brown 和 Mood 建議，利用繪製樣本資料的散佈圖及中位數，來找出迴歸係數 b 後，Theil 提出另一種求算迴歸係數 b 的方法。

Theil 法的迴歸係數求算步驟：

1. 先將 n 筆模糊資料依 n 個相異的自變數 X_i ，按照大小做排列即

$$X_1 < X_2 < ... < X_n$$

則 n 對模糊觀察值排列爲

$$(X_1,Y_1),(X_2,Y_2),.....,\ (X_n,Y_n)$$

2. 根據 n 筆模糊資料，計算出所有可能的模糊樣本斜率 S_{ij} ，符合此要求之

S_{ij} 有 $N = \binom{n}{2}$ 個，其中 S_{ij} 的公式爲：

$$S_{ij} = \frac{Y_i - Y_j}{X_i - X_j} \qquad i < j$$

3. 決定迴歸係數 b 爲：

$$b = median\ \{S_{ij}\}$$

例 3.5：針對樣本觀察值 X 及 Y 爲連續型模糊數爲例

台灣農業推廣當局在極力宣傳台灣農業的同時，也關心著台灣農業的產量及品質。台南關廟鳳梨一直是台南市政府極力推廣的一項農產品，基於此，政府關心著明年是否會有好收成？希望能在未收成前做預測，以有利政府在今年可先做

配套的行銷計劃，並方便估計欲需求的經費支出。

於是針對過去 10 年來，由台南地區主要的鳳梨產區，採集種植面積及收成量的樣本資料如下：

表 4.1 前十年的連續型樣本資料

年度	種植面積 (公頃)	收成量 (百萬斤)
85	42~44	6.6~6.8
86	45~46	5.9~6.1
87	41~42	6.1~6.3
88	45~47	6.9~7.3
89	37~39	4.9~5.1
90	47~49	7.3~7.5
91	50~52	8.0~8.4
92	49~50	7.7~7.9
93	51~54	8.4~8.6
94	49~51	7.8~8.2

令種植面積爲自變數 X ，收成量爲應變數 Y 。分別取 X 及 Y 中每筆樣本區間值中點，做爲代表該筆資料的數據值如下：

表 4.2 模糊區間值樣本之中點

年度	種植面積 X (公頃)	收成量 Y (百萬斤)
85	43	6.7
86	45.5	7.2
87	41	6.2
88	46	7.1
89	38	5.0
90	48	7.4
91	51	8.2
92	49.5	7.8
93	52.5	8.5
94	50	8.0

根據 Theil 方法，先將模糊資料 X 由小到大排列爲 $X_1 < X_2 < ... < X_{10}$ ，則十個觀測值排列後如下表：

表 4.3 觀測值排序後之情形

(X_1, Y_1)	38	5.0
(X_2, Y_2)	41	6.2
(X_3, Y_3)	43	6.7
(X_4, Y_4)	45.5	7.2
(X_5, Y_5)	46	7.1
(X_6, Y_6)	48	7.4
(X_7, Y_7)	49.5	7.8
(X_8, Y_8)	50	8.0
(X_9, Y_9)	51	8.2
(X_{10}, Y_{10})	52.5	8.5

則 $N = \binom{10}{2}$ ＝45 個 S_{ij} 分別爲：

表 4.4　依據 Theil 方法計算 S_{ij} 值列表

0.4	0.34	0.29	0.26	0.24	0.24	0.25	0.25	0.24
0.25	0.22	0.18	0.17	0.19	0.2	0.2	0.2	0.2
0.13	0.14	0.17	0.19	0.19	0.19	-0.2	0.08	0.15
0.18	0.18	0.19	0.15	0.2	0.23	0.22	0.22	0.27
0.3	0.27	0.24	0.4	0.27	0.23	0.2	0.2	0.2

所以 S_{ij} 的中位數爲 0.2，亦即我們得到迴歸係數的估計值 b ＝0.2。