

《迴歸分析》

一、兩位科學家試圖研究某一個解釋變數 X 與某一個反應變數 Y 之間的直線關係。數據如下，請根據提供的數據回答以下問題：

X	10	9	8	7	6	5	4	3
Y	45	20	34	58	70	57	55	44

- (一)兩位科學家委託一位統計學家協助計算他們有興趣的直線關係。為了徹底了解 X 與 Y 之間的直線關係，統計學家建議先試試「 $Y=A$ 」，也就是說， X 與 Y 之間沒關係。請問根據以上數據， A 的估計值等於多少？(5分)
- (二)接下來，統計學家假設的直線關係是「 $Y=A+B \times X$ 」，其中「 $B \times X$ 」意味著「『直線斜率』乘以『 X 』」。也就是說，統計學家第二次研究具有「直線截距」與「直線斜率」的直線關係。請問根據以上數據，直線斜率 B 的估計值等於多少？(10分)
- (三)最後，兩位科學家在收到上述分析報告之後，決定更動直線關係為「 $Y=B \times X$ 」。也就是說，兩位科學家第三次研究的議題是「無直線截距」的直線關係。請再一次根據以上數據，回答直線斜率 B 的估計值等於多少？(10分)

試題評析 本題是簡迴歸下三個常見模型之參數估計，屬於基本計算考題，要拿到分數並不難。

考點命中 《迴歸分析熱門題庫》高點出版，2016版，第二篇第二章，趙治勳編撰。

答：

$$\sum X_i = 52, \sum X_i^2 = 380, \sum Y_i = 383, \sum Y_i^2 = 20055, \sum X_i Y_i = 2365$$

$$SS_{XY} = -124.5 \quad SS_X = 42 \quad SS_Y = 1718.875$$

$$(一) \hat{A} = \bar{Y} = \frac{383}{8} = 47.875$$

$$(二) \hat{B} = \frac{SS_{XY}}{SS_X} = -2.9643$$

$$(三) \hat{B} = \frac{\sum X_i Y_i}{\sum X_i^2} = 6.2237$$

二、兩位科學家試圖研究某一個解釋變數 X 與某一個反應變數 Y 之間的直線關係。數據如下：

X	10	9	8	7	6	5	4	3
Y	45	20	34	58	70	57	55	44

這兩位科學家決定取得數據在「 $Y=A+B \times X + \varepsilon$ ， ε 為誤差項」這一項假設下的變異數分析表。請問：

- (一)總平方和 (total sum of squares) 等於多少？(10分)
- (二)迴歸平方和 (regression sum of squares) 等於多少？(10分)
- (三)誤差平方和 (residual sum of squares) 等於多少？(5分)

試題評析 本題是簡迴歸下配適度之三個平方和之基本計算考題，要拿到分數並不難。

考點命中 《迴歸分析熱門題庫》高點出版，2016版，第二篇第二章第四節，趙治勳編撰。

答：

$$(一) SST = SS_Y = 1718.875$$

$$(二) SSR = \hat{B}^2 SS_x = (-2.9643)^2 \times 42 = 369.0571$$

$$(三) SSE = SST - SSR = 1718.875 - 369.0571 = 1349.8179$$

三、假設一組38個樣本、三個變數的數據集，其中三個變數分別是一個反應變數、一個解釋變數，加上一個源自前述解釋變數的「兩水準虛擬變數(dummy variable)」。也就是說，這一個虛擬變數只會出現兩種數字，假設不是「0」就是「1」。請回答以下問題：

(一)如果數據科學家提出一個這樣的複迴歸模型「反應變數 $=A+B \times$ 解釋變數 $+C \times$ 虛擬變數 $+D \times$ 虛擬變數 $\times$ 解釋變數」。請寫下「虛擬變數等於1」的數學方程式？(5分)

(二)請說明(一)的迴歸係數A, B, C, D的數值以描述在「『虛擬變數等於0』與『虛擬變數等於1』」下的迴歸直線是同一條直線。」(10分)

(三)請說明(一)的迴歸係數A, B, C, D的數值以描述在「『虛擬變數等於0』與『虛擬變數等於1』」下的迴歸直線是兩平行直線。」(10分)

試題評析	本題是有關虛擬變數，且所有觀念都在課本中詳述過，只要熟讀課本內容，要拿到分數並不難。
考點命中	《迴歸分析熱門題庫》高點出版，2016版，第二篇第五章，趙治勳編撰。

答：

(一)反應變數 $= (A+C) + (B+D) \times$ 解釋變數

(二) $(A, B) \in R, C=0$ 且 $D=0$

(三) $(A, B, C) \in R, D=0$

四、統計學家為一位有興趣採用「迴歸分析」實踐「檢定三個處理平均數是否相等」的科學家寫下這樣的設計矩陣X，

1	1	0
1	1	0
1	1	0
1	0	1
1	0	1
1	0	1
1	0	0
1	0	0
1	0	0

請回答以下問題：

(一)請計算H矩陣(hat matrix)。(5分)

(二)請用(一)結果證明這時候 $(I-H)(I-H)=(I-H)$ ，其中I是一個對角線上是1、其他都是0的方陣。(10分)

(三)根據迴歸分析理論的計算，統計學家發現檢定「三個處理平均數相等」這一項虛無假設的檢定統計量是兩項「迴歸平方和」之間的「差」。請問是那兩項迴歸平方和？回答問題時，請根據題意自行定義相關的符號。(10分)

試題評析	本題實為變異數分析，只不過利用迴歸分析，在(一)(二)中雖然不難，但計算繁瑣，考生務必要小心計算，在(三)中考生容易被「迴歸平方和的差」所困惑。
考點命中	《迴歸分析熱門題庫》高點出版，2016版，第二篇第三章，趙治勳編撰。

答：
(一)

$$H = X(X^T X)^{-1} X^T = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{bmatrix}$$

(二) $(I - H)(I - H) = II - IH - HI + HH = I - H - H + H = I - H$ 得證
(但題目有要求利用(一)去驗證)

$$I - H = \begin{bmatrix} 2/3 & -1/3 & -1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/3 & 2/3 & -1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/3 & -1/3 & 2/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & 2/3 & -1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & -1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/3 & 2/3 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/3 & -1/3 & 2/3 \end{bmatrix}$$

$$(I - H)(I - H) = \begin{bmatrix} 2/3 & -1/3 & -1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/3 & 2/3 & -1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/3 & -1/3 & 2/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & 2/3 & -1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & -1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/3 & 2/3 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/3 & -1/3 & 2/3 \end{bmatrix} = (I - H)$$

(三) 假設：應變數為 Y ，自變數為 X_1 及 X_2

模型一： $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), i = 1, 2, \dots, 9$

此模型下之迴歸平方和為 $SSR(\beta_0 \beta_1 \beta_2)$ 及誤差平方和為 $SSE(\beta_0 \beta_1 \beta_2)$

模型二： $Y_i = \beta_0 + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), i = 1, 2, \dots, 9$

此模型下之迴歸平方和為 $SSR(\beta_0)$ 及誤差平方和為 $SSE(\beta_0)$

檢定 $H_0: \mu_1 = \mu_2 = \mu_3$ vs H_0 : 至少一個 $\mu_i \neq \mu_j, i \neq j$

等價於檢定 $H_0: \beta_1 = \beta_2 = 0$ vs H_0 : 至少一個 $\beta_i \neq 0$

$$\text{T.S.: } F = \frac{SSR(\beta_0\beta_1\beta_2) / 2}{SSE(\beta_0\beta_1\beta_2) / 9 - 2 - 1} = 3 \times \frac{SSR(\beta_0\beta_1\beta_2) - SSR(\beta_0)}{SSE(\beta_0\beta_1\beta_2)} \sim F_{(2,6)}$$

高點
·
高上

【版權所有，重製必究！】