



《迴歸分析》

試題評析

今年迴歸分析之考題如往常一樣題目簡單且容易計算，而且只偏重在簡單迴歸與複迴歸二單元，同學只要有基本觀念，且計算過程無錯誤必能考90分以上，而程度好者必能考95分以上，在此份考題中除了填充題最後一題2分外大致皆屬於上課之內容，如計算題第一題可在第一回講義中，例10及例13找到類化題，而計算題第二題可在第二回中之例11找到類化題型。

一、計算題：（每小題10分，共80分。請按題號依序作答，否則不予計分。）

注意事項：所有檢定都必須建立統計假設，寫出決策法則，並作結論。

(一)欲探究觀光收入與觀光客人數的關係，乃登錄最近三十個月來本地區之觀光客人數（萬人）與收入（千萬元）分別為：

人數X	96	45	89	88	68
收入Y	53	58	73	45	58

經Microsoft Excel求算而得下列部分電腦報表：

電腦報表

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	9013.7915	9013.7915	43.7221	3.5825E-07
殘差	28	5772.5085	206.1610		

	係數	t統計	P-值
截距	4.9776	-0.6610	0.5140
人數X	0.7290	6.6123	0.0000

假設此資料適合作有母數統計分析，試以顯著水準 $\alpha=0.05$

- 檢定迴歸直線 $\beta_0 + \beta_1 X$ 是否與X軸平行？
- 檢定觀光收入與觀光客人數是否有正相關？
- 求算迴歸係數 β_1 之95%信賴度的區間估計。
- 若本月預估觀光客人數為64萬人，正巧與最近三十個月來本地區之觀光客平均人數相等，則觀光收入 Y_h 之95%信賴度的預測區間為何？

【註： $t_{(0.025, 28)}=2.048$ 】

(二)欲了解本區之平均房價Y的影響因素：平均家庭收入X及犯罪率W對房價的影響，得近二十年來的資料為：

Y	48	58	66	84	89	260
X	37	30	40	55	56	85
W	2.4	1.9	2.1	1.9	1.6	0.1

經Microsoft Excel求算而得下列部分電腦報表：

電腦報表

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	2	80415.9186	40207.9593	174.4143	4.688E-12
殘差	17	3919.0314	230.5313		

	係數	標準誤	t統計	P-值
截距	134.5471	32.8442	4.0965	0.0008
X	1.3288	0.4062	3.2715	0.0045
W	-62.5654	8.3113	-7.5277	0.0000

相關係數

	Y	X	W
Y	1		
X	0.8937	1	
W	-0.9614	-0.8304	1

Durbin-Watson Statistic

d=1.1535

假設此資料適合作有母數統計分析，試以顯著水準 $\alpha=0.05$

(1)檢定：若犯罪率W已在迴歸模型內，平均家庭收入X是否值得引入？

【註： $F_{(0.05,1,17)}=4.45$ 】(2)檢定平均房價與平均家庭收入及犯罪率之複相關係數 ρ 是否為0？(3)檢定偏迴歸係數 β_1 、 β_2 是否皆不為0？(4)檢定複迴歸模型 $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \varepsilon_i$ 之殘差項 ε_i 是否具有隨機性？【註：Durbin-Watson檢定的臨界值分別為： $d_L=1.100$ 、 $d_U=1.537$ 】

答：

(一)1.(1) $H_0: \beta_1 = 0$ (2) $H_1: \beta_1 \neq 0$ (3) $\alpha = 0.05$ (4) $C = \{F | F > F_{0.05}(1,28) = 4.196\}$

(5)計算：由ANOVA表知

$$F = \frac{MSR}{MSE} = 43.7221 \in C$$

(亦可用p-value < 0.05)。

(6)結論：拒絕 H_0 ，即表示迴歸直線不與X軸平行。2.(1) $H_0: \rho \leq 0$ (2) $H_1: \rho > 0$ (3) $\alpha = 0.05$ (4) $C = \{t | t > t_{0.05}(28) = 1.701\}$ (5)計算： $\therefore t = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} = 6.6123 \in C$ (6)結論：拒絕 H_0 ，即表示觀光收入與觀光人數為正相關。

$$3. \therefore t \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} \Rightarrow S(\hat{\beta}_1) = \frac{0.7290}{6.6123} = 0.11025$$

 $\therefore \beta_1$ 之95%信賴區間為

$$\Rightarrow (\hat{\beta}_1 - t_{0.025}(28) \cdot S(\hat{\beta}_1), \hat{\beta}_1 + t_{0.025}(28) \cdot S(\hat{\beta}_1))$$

$$\Rightarrow (0.7290 - 2.048 \times 0.11025, 0.7290 + 2.048 \times 0.11025)$$

$$\Rightarrow (0.50321, 0.95479)$$

4. $\therefore \hat{Y} = 4.9776 + 0.729x$

$$\therefore \hat{Y}_{x=64} = 4.9776 + 0.729 \times 64 = 51.6336$$

故 Y_h 之95%信賴區間為

$$\Rightarrow (\hat{Y}_x \mp t_{0.025}(28) \cdot \sqrt{MSE} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}})$$

$$\Rightarrow (51.6336 \mp 2.048 \sqrt{206.161} \sqrt{1 + \frac{1}{30} + \frac{(0-0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}})$$

$$\Rightarrow (21.7417, 81.5255)$$

(二)1.迴歸模式為 $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \varepsilon_i$

- (1) $H_0: \rho_{YX \cdot W} = 0$ \longleftrightarrow 相對於 $H_0: \beta_1 = 0$
 (2) $H_1: \rho_{YX \cdot W} \neq 0$ $H_1: \beta_1 \neq 0$
 (3) $\alpha = 0.05$

(4) $C = \{F | F > F_{0.05}(1,17) = 4.45\}$

(5) 計算：(解法一)

由電腦報表知自變數X之t值為

$$t = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} = 3.2715, \text{ 故}$$

$$F = t^2 = (3.2715)^2 = 10.7027 \in C$$

(解法二)

$$\text{因 } r_{XY \cdot W}^2 = \frac{(r_{YX} - r_{YW} \cdot r_{XW})^2}{(1 - r_{YW}^2)(1 - r_{XW}^2)} = \frac{[(0.8937 - (-0.9614)(-0.8304))]^2}{[1 - (-0.9614)^2][1 - (-0.8304)^2]} = 0.38685$$

$$\therefore F = \frac{r_{YX \cdot W}^2(n-3)}{(1 - r_{YX \cdot W}^2)} = \frac{(0.38685)(20-3)}{(1 - 0.38685)} = 10.73 \in C$$

(6) 結論：拒絕 H_0 ，即表示X值得引入。

2. (1) $H_0: \rho_{YXW} = 0$

(2) $H_1: \rho_{YXW} \neq 0$

(3) $\alpha = 0.05$

(4) $C = \{F | F > F_{0.05}(2,17) = 3.5915\}$

(5) 計算：由ANOVA表知

$$F = \frac{MSR}{MSE} = 174.4143 \in C$$

(亦可由p-value < $\alpha = 0.05$ 判斷)

(6) 結論：拒絕 H_0 ，即表示 ρ 不為0。

3. $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$

決策規則為p-value < α ，則拒絕 H_0

因p-value = 0.0045 < $\alpha = 0.05$

故拒絕 H_0 ，即表示 β_1 顯著不為0

$$\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$$

決策規則為p-value < α ，則拒絕 H_0

因p-value = 0.0000 < $\alpha = 0.05$

故拒絕 H_0 ，即表示 β_2 顯著不為0

4. $\begin{cases} H_0: \varepsilon_i \text{ 具有隨機性} \\ H_1: \varepsilon_i \text{ 不具有隨機性} \end{cases}$

決策法則為

當 $D-W < d_L$ 及 $D-W > 4-d_L$ 時拒絕 H_0 ，而

若 $D-W$ 介於 (d_L, d_U) 間及 $(4-d_U, 4-d_L)$ 間則無法下結論

又由電腦報表知 $D-W = 1.1535$ 介於 $(d_L = 1.1, d_U = 1.537)$ 間，故無法下結論。

二、填充題：(每格2分，共20分。請按空格編號依序作答，否則不予計分。)

(一) 設二個連續隨機變數X與Y之聯合機率函數 $f(x,y) = e^{-(x+y)}$ ， $x \geq 0, y \geq 0$ ，可求得X與Y的共變數 $Cov(X,Y) = \underline{\text{①}}$ ；若條件機率分配之期望值為 $E(Y|X) = \beta_0 + \beta_1 X$ ，則可得Y對X的迴歸係數 $\beta_1 = \underline{\text{②}}$ 。

(二) 設二項隨機變數X之期望值 $E(X) = \mu$ 、變異數 $V(X) = \sigma^2$ ，令新的隨機變數 $Y = (X - \mu)/\sigma$ ，則可求得X與Y的積差相關係數

$$\rho_{XY} =$$

$$\underline{\text{③}} \text{ 及 Y對X的迴歸係數 } \beta_1 = \underline{\text{④}}。$$

(三) 隨機抽訪一行業的30家廠商，得其去年之廣告費與營收的Spearman等級相關係數 $r_s = 0.8$ ，則可求得檢定統計量 $t =$

$$\underline{\text{⑤}}$$

與臨界值 $t_{(0.025,28)} = 2.048$ 作比較，而 $\underline{\text{⑥}}$ 虛無假設 H_0 ：廣告費與營收無關。

(四) 欲了解四種品牌甲、乙、丙、丁櫃員機的操作效率，乃由銀行行員中隨機抽出5人進行實驗設計，得各提領一萬元

的操作時間分別為 Y_{ij} 。此一因子集區實驗設計的樣本資料可以加⑦個Dummy Variables而改以迴歸分析進行。此迴歸分析之F檢定的自由度分別為⑧。

(五)欲探究本地區之家庭支出與家庭收入、每戶人口數的關係，乃隨機抽訪三十個家庭，得樣本資料後，經電腦求算而得White檢定之檢定統計量 $nR^2=15.99$ ，與臨界值 $\chi^2_{(0.05)}=11.071$ 作比較，而⑨虛無假設 H_0 ：等變異性成立。此 χ^2 檢定的自由度為⑩。

答：

(一)①0；②0

$$\because f(x, y) = e^{-(x+y)}, \quad x \geq 0, \quad y \geq 0$$

故知 x 與 y 為獨立隨機變數

$$\therefore \text{Cov}(x, y) = 0, \quad \text{且 } \rho = 0, \quad \text{即 } \beta_1 = 0$$

(二)③1；④ $\frac{1}{\sigma}$

$$\therefore Y = \frac{x - \mu}{\sigma} = \frac{1}{\sigma}x - \frac{\mu}{\sigma}$$

故知 x 與 y 為直線關係其相關係數為1，且斜率為 $\frac{1}{\sigma}$

(三)⑤7.0553；⑥拒絕

$$\therefore t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{0.8 \sqrt{30-2}}{\sqrt{1-(0.8)^2}} = 7.0553 > t_{0.025}(28) = 2.048$$

(四)⑦7；⑧3；12

因有4個處理，5個集區，故需要 $(4-1) + (5-1) = 7$ 個虛擬變數

(五)⑨拒絕；⑩5