

連續30年人氣爆棚，題點超過10,000名上榜生

高普考 高點名師

線上解題講座

給你最快
最精準的詳解!



高凱 (高凱傑)

行政學/
公共政策

f 高點行政學院



7/17 (一) 首播



張政 (張家瑋)

財政學/
經濟學

f 高點高上
高普特考公職



7/17 (一) 首播



陳世華 (邱垂炎)

會計學/
中會

f 高點會人會語



7/18 (二) 首播



曾榮耀 (蘇偉強)

土法/土登/
土經

f 高點來勝
不動產專班



7/18 (二) 首播



王致強 (蕭立人)

資料結構

高點線上
影音學習



7/19 (三) 首播



何昀峯

考銓

高點線上
影音學習



7/19 (三) 首播



陳熙哲

行政法

高點線上
影音學習



7/19 (三) 首播



葉哲瑋

抽樣/
迴歸

高點線上
影音學習



7/19 (三) 首播

《迴歸分析》

試題評析	本考卷所有考題都圍繞在簡迴歸而已，皆屬於基礎計算題型，包括相關係數與判定係數之關係，截距項與斜率項之估計量計算題，考生只要熟悉簡迴歸之計算題，本卷應該可以輕鬆獲得接近滿分之成績。
考點命中	《迴歸分析申論題完全制霸》，高點文化出版，趙治勳編著，第八章第一節。

Anscombe (1973) 發表了四組數據集，每一組都只有一個解釋變數和一個反應變數，這四組數據集經常被用來示範散佈圖 (scatter plot) 在迴歸分析的重要性。數據如下所示：

	x1	x2	x3	x4	y1	y2	y3	y4
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.10	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.10	5.39	12.50
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89

其中「x1, x2, x3, x4」依序分別表示第一組數據集的解釋變數、第二組數據集的解釋變數、第三組數據集的解釋變數、第四組數據集的解釋變數。「y1, y2, y3, y4」依序分別表示第一組數據集的反應變數、第二組數據集的反應變數、第三組數據集的反應變數、第四組數據集的反應變數。有了數據之後，在假設簡單線性迴歸模型之前，通常會先繪製散佈圖，但這一回我們先配適這個模型「 $y = A + Bx$ 」，其中「y」是反應變數、「A」是截距、「B」是斜率、「x」是解釋變數。

答題時，請用「A1」代表第一組數據集上述模型「A」的估計值、「B1」代表第一組數據集上述模型「B」的估計值；用「A2」代表第二組數據集上述模型「A」的估計值、「B2」代表第二組數據集上述模型「B」的估計值；用「A3」代表第三組數據集上述模型「A」的估計值、「B3」代表第三組數據集上述模型「B」的估計值；用「A4」代表第四組數據集上述模型「A」的估計值、「B4」代表第四組數據集上述模型「B」的估計值。

為了估計「A (截距)」和「B (斜率)」，有下列4組數據集。（請注意，欄位名稱x和y是原始數據、xx表示解釋變數的平方、yy表示反應變數的平方、xy表示解釋變數乘以反應變數。「總和」為其上方11個數字的加總）

1. 第一組數據集：

	x	y	xx	yy	xy
1	10	8.04	100	64.6416	80.40
2	8	6.95	64	48.3025	55.60
3	13	7.58	169	57.4564	98.54
4	9	8.81	81	77.6161	79.29
5	11	8.33	121	69.3889	91.63
6	14	9.96	196	99.2016	139.44

7	6	7.24	36	52.4176	43.44
8	4	4.26	16	18.1476	17.04
9	12	10.84	144	117.5056	130.08
10	7	4.82	49	23.2324	33.74
11	5	5.68	25	32.2624	28.40
總和	99	82.51	1001	660.1727	797.60

2. 第二組數據集：

	x	y	xx	yy	xy
1	10	9.14	100	83.5396	91.40
2	8	8.14	64	66.2596	65.12
3	13	8.74	169	76.3876	113.62
4	9	8.77	81	76.9129	78.93
5	11	9.26	121	85.7476	101.86
6	14	8.10	196	65.6100	113.40
7	6	6.13	36	37.5769	36.78
8	4	3.10	16	9.6100	12.40
9	12	9.13	144	83.3569	109.56
10	7	7.26	49	52.7076	50.82
11	5	4.74	25	22.4676	23.70
總和	99	82.51	1001	660.1736	797.59

3. 第三組數據集：

	x	y	xx	yy	xy
1	10	7.46	100	55.6516	74.60
2	8	6.77	64	45.8329	54.16
3	13	12.74	169	162.3076	165.62
4	9	7.11	81	50.5521	63.99
5	11	7.81	121	60.9961	85.91
6	14	8.84	196	78.1456	123.76
7	6	6.08	36	36.9664	36.48
8	4	5.39	16	29.0521	21.56
9	12	8.15	144	66.4225	97.80
10	7	6.42	49	41.2164	44.94
11	5	5.73	25	32.8329	28.65
總和	99	82.50	1001	659.9762	797.47

4. 第四組數據集：

	x	y	xx	yy	xy
1	8	6.58	64	43.2964	52.64
2	8	5.76	64	33.1776	46.08
3	8	7.71	64	59.4441	61.68
4	8	8.84	64	78.1456	70.72
5	8	8.47	64	71.7409	67.76
6	8	7.04	64	49.5616	56.32

7	8	5.25	64	27.5625	42.00
8	19	12.50	361	156.2500	237.50
9	8	5.56	64	30.9136	44.48
10	8	7.91	64	62.5681	63.28
11	8	6.89	64	47.4721	55.12
總和	99	82.51	1001	660.1325	797.58

請回答下列問題：

答：

由題目資訊可得各組數據之

$$SS_X = \sum X_i^2 - \frac{(\sum X_i)^2}{n}, SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \text{ 與 } SS_{XY} = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} \text{ 如下：}$$

第一組數據集：

$$\begin{aligned} SS_{X1} &= \sum X_{1i}^2 - \frac{(\sum X_{1i})^2}{n} = 1001 - \frac{99^2}{11} = 110 \\ SS_{Y1} &= \sum Y_{1i}^2 - \frac{(\sum Y_{1i})^2}{n} = 660.1727 - \frac{(82.51)^2}{11} = 41.2727 \\ SS_{X1Y1} &= \sum X_{1i} Y_{1i} - \frac{\sum X_{1i} \sum Y_{1i}}{n} = 797.6 - \frac{(99)(82.51)}{11} = 55.01 \end{aligned}$$

第二組數據集：

$$\begin{aligned} SS_{X2} &= \sum X_{2i}^2 - \frac{(\sum X_{2i})^2}{n} = 1001 - \frac{99^2}{11} = 110 \\ SS_{Y2} &= \sum Y_{2i}^2 - \frac{(\sum Y_{2i})^2}{n} = 660.1763 - \frac{(82.51)^2}{11} = 41.2763 \\ SS_{X2Y2} &= \sum X_{2i} Y_{2i} - \frac{\sum X_{2i} \sum Y_{2i}}{n} = 797.59 - \frac{(99)(82.51)}{11} = 55 \end{aligned}$$

第三組數據集：

$$\begin{aligned} SS_{X3} &= \sum X_{3i}^2 - \frac{(\sum X_{3i})^2}{n} = 1001 - \frac{99^2}{11} = 110 \\ SS_{Y3} &= \sum Y_{3i}^2 - \frac{(\sum Y_{3i})^2}{n} = 659.9762 - \frac{(82.50)^2}{11} = 41.2262 \\ SS_{X3Y3} &= \sum X_{3i} Y_{3i} - \frac{\sum X_{3i} \sum Y_{3i}}{n} = 797.47 - \frac{(99)(82.50)}{11} = 54.97 \end{aligned}$$

第四組數據集：

$$\begin{aligned} SS_{X4} &= \sum X_{4i}^2 - \frac{(\sum X_{4i})^2}{n} = 1001 - \frac{99^2}{11} = 110 \\ SS_{Y4} &= \sum Y_{4i}^2 - \frac{(\sum Y_{4i})^2}{n} = 660.1325 - \frac{(82.51)^2}{11} = 41.2325 \\ SS_{X4Y4} &= \sum X_{4i} Y_{4i} - \frac{\sum X_{4i} \sum Y_{4i}}{n} = 797.58 - \frac{(99)(82.51)}{11} = 54.99 \end{aligned}$$

一、請利用上述表格計算解釋變數與反應變數之間的相關係數到小數點第四位，並且由小到大排序計算結果。（假設第一組數據集的相關係數為「r1」、第二組數據集的相關係數為「r2」、第三組數據集的相關係數為「r3」、第四組數據集的相關係數為「r4」。）（20分）

答：

$$r_1 = \frac{SS_{X1Y1}}{\sqrt{SS_{X1}}\sqrt{SS_{Y1}}} = \frac{55.01}{\sqrt{110}\sqrt{41.2727}} = 0.8164$$

$$r_2 = \frac{SS_{X2Y2}}{\sqrt{SS_{X2}}\sqrt{SS_{Y2}}} = \frac{55}{\sqrt{110}\sqrt{41.2763}} = 0.8162$$

$$r_3 = \frac{SS_{X3Y3}}{\sqrt{SS_{X3}}\sqrt{SS_{Y3}}} = \frac{54.97}{\sqrt{110}\sqrt{41.2262}} = 0.8163$$

$$r_4 = \frac{SS_{X4Y4}}{\sqrt{SS_{X4}}\sqrt{SS_{Y4}}} = \frac{54.99}{\sqrt{110}\sqrt{41.2325}} = 0.8165$$

比較後可得 $r_2 < r_3 < r_1 < r_4$

二、請利用上述表格針對模型「 $y = A + Bx$ 」，計算各個數據集估計方程式的決定係數（coefficient of determination）到小數點第四位，並且由小到大排序計算結果。（假設第一組數據集的決定係數為「R21」、第二組數據集的決定係數為「R22」、第三組數據集的決定係數為「R23」、第四組數據集的決定係數為「R24」。）（20分）

答：

在簡迴歸下，判定係數等於相關係數之平方（ $R^2 = r^2$ ）

$$R21 = r_1^2 = 0.8164^2 = 0.6665 \quad R22 = r_2^2 = 0.8162^2 = 0.6662$$

$$R23 = r_3^2 = 0.8163^2 = 0.6663 \quad R24 = r_4^2 = 0.8165^2 = 0.6667$$

比較後可得 $R22 < R23 < R21 < R24$

三、請利用上述表格計算各個數據集估計模型「 $y = A + Bx$ 」的截距到小數點第四位，並且由小到大排序計算結果。（20分）

答：

$$\text{在簡迴歸下，截距項之估計量為 } \hat{A} = \bar{Y} - \hat{B}\bar{X} = \frac{\sum Y_i}{n} - r \frac{\sqrt{SS_Y}}{\sqrt{SS_X}} \frac{\sum X_i}{n}$$

$$\hat{A}_1 = \frac{\sum Y_{1i}}{n} - r_1 \frac{\sqrt{SS_{Y1}}}{\sqrt{SS_{X1}}} \frac{\sum X_{1i}}{n} = \frac{82.51}{11} - 0.8164 \frac{\sqrt{41.2727}}{\sqrt{110}} \frac{99}{11} = 3.0002$$

$$\hat{A}_2 = \frac{\sum Y_{2i}}{n} - r_2 \frac{\sqrt{SS_{Y2}}}{\sqrt{SS_{X2}}} \frac{\sum X_{2i}}{n} = \frac{82.51}{11} - 0.8162 \frac{\sqrt{41.2763}}{\sqrt{110}} \frac{99}{11} = 3.0011$$

$$\hat{A}_3 = \frac{\sum Y_{3i}}{n} - r_3 \frac{\sqrt{SS_{Y3}}}{\sqrt{SS_{X3}}} \frac{\sum X_{3i}}{n} = \frac{82.50}{11} - 0.8163 \frac{\sqrt{41.2262}}{\sqrt{110}} \frac{99}{11} = 3.0024$$

$$\hat{A}_4 = \frac{\sum Y_{4i}}{n} - r_4 \frac{\sqrt{SS_{Y4}}}{\sqrt{SS_{X4}}} \frac{\sum X_{4i}}{n} = \frac{82.51}{11} - 0.8165 \frac{\sqrt{41.2325}}{\sqrt{110}} \frac{99}{11} = 3.0018$$

比較後可得 $\hat{A}_1 < \hat{A}_2 < \hat{A}_4 < \hat{A}_3$

四、請利用上述表格計算各個數據集估計模型「 $y = A + Bx$ 」的斜率到小數點第五位，並且由小到大排序計算結果。（20分）

答：

在簡迴歸下，斜率項之估計量為 $\hat{B} = \frac{SS_{xy}}{SS_x}$

$$\hat{B}_1 = \frac{SS_{x1y1}}{SS_{x1}} = \frac{55.01}{110} = 0.50009$$

$$\hat{B}_2 = \frac{SS_{x2y2}}{SS_{x2}} = \frac{55}{110} = 0.50000$$

$$\hat{B}_3 = \frac{SS_{x3y3}}{SS_{x3}} = \frac{54.97}{110} = 0.49973$$

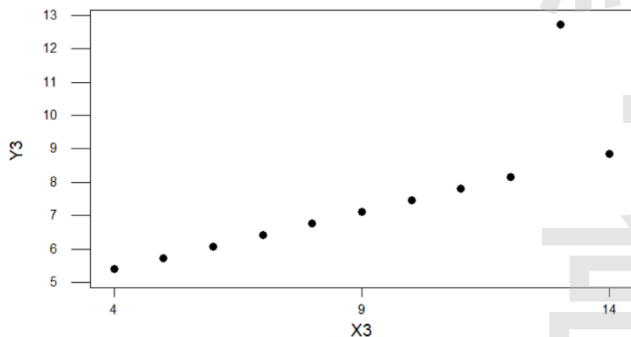
$$\hat{B}_4 = \frac{SS_{x4y4}}{SS_{x4}} = \frac{54.99}{110} = 0.49991$$

比較後可得 $\hat{B}_3 < \hat{B}_4 < \hat{B}_2 < \hat{B}_1$

五、針對第三組數據集提出決定係數（coefficient of determination）最接近1.0的模型。（提示：繪製這一組數據集的散佈圖，並刪除一個影響點。）（20分）

答：

第三組數據集之散佈圖如下：



由以上散佈圖發現有一筆疑似離群值為第3筆數據（ $X_3 = 13, Y_3 = 12.74$ ）

刪除第3筆後，可得如下：

$$SS_{X_3} = \sum X_3^2 - \frac{(\sum X_3)^2}{n} = 832 - \frac{86^2}{10} = 92.4$$

$$SS_{Y_3} = \sum Y_3^2 - \frac{(\sum Y_3)^2}{n} = 497.6686 - \frac{(69.76)^2}{10} = 11.02284$$

$$SS_{X_3Y_3} = \sum X_3Y_3 - \frac{\sum X_3 \sum Y_3}{n} = 631.85 - \frac{(86)(69.76)}{10} = 31.914$$

計算判定係數為 $R_{23} = \frac{SS_{X_3Y_3}^2}{SS_{X_3}SS_{Y_3}} = \frac{31.914^2}{92.4 \times 11.02284} = 0.999993107$ ，發現此模型之判定係數更接近於1

【版權所有，重製必究！】