

112 年公務人員高等考試三級考試試題

類 科：統計

科 目：迴歸分析

王瑋老師

甲、申論題部分：

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	660.1763	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：

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)

【解題關鍵】

1. 《考題難易》★
2. 《破題關鍵》幾乎沒有任何難度相關係數計算基本題，110 年地特有類似考題，可參考王瑋，迴歸分析 P.1-11 相同例題。

【擬答】

$$\begin{aligned}
 \text{(一)} r_1 &= \frac{SS_{X_1Y_1}}{\sqrt{SS_{X_1}SS_{Y_1}}} = \frac{797.6 - \frac{99 \times 82.51}{11}}{\sqrt{(1001 - \frac{99^2}{11}) \times (660.1727 - \frac{82.51^2}{11})}} = 0.8164 \\
 \text{(二)} r_2 &= \frac{SS_{X_2Y_2}}{\sqrt{SS_{X_2}SS_{Y_2}}} = \frac{797.59 - \frac{99 \times 82.51}{11}}{\sqrt{(1001 - \frac{99^2}{11}) \times (660.1763 - \frac{82.51^2}{11})}} = 0.8162 \\
 \text{(三)} r_3 &= \frac{SS_{X_3Y_3}}{\sqrt{SS_{X_3}SS_{Y_3}}} = \frac{797.47 - \frac{99 \times 82.50}{11}}{\sqrt{(1001 - \frac{99^2}{11}) \times (659.9762 - \frac{82.80^2}{11})}} = 0.8163 \\
 \text{(四)} r_4 &= \frac{SS_{X_4Y_4}}{\sqrt{SS_{X_4}SS_{Y_4}}} = \frac{797.58 - \frac{99 \times 82.51}{11}}{\sqrt{(1001 - \frac{99^2}{11}) \times (660.1325 - \frac{82.51^2}{11})}} = 0.8165
 \end{aligned}$$

所以 $r_2 < r_3 < r_1 < r_4$

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)

【解題關鍵】

1. 《考題難易》★
2. 《破題關鍵》只需將相關係數平方即為決定係數，是相當簡單的基本題，過去考試幾乎沒有這樣單獨又容易的計算考題。

【擬答】

決定係數是相關係數的平方

1. $R_{21} = r_1^2 = 0.8164^2 = 0.6665$

2. $R_{22} = r_2^2 = 0.8162^2 = 0.6662$

3. $R_{23} = r_3^2 = 0.8163^2 = 0.6663$

4. $R_{24} = r_4^2 = 0.8165^2 = 0.6667$

所以 $R_{22} < R_{23} < R_{21} < R_{24}$



我們都在 **志光·保成·學儒** 成為公務員

商科上榜生一致的選擇

強 111高普商科 雙料金榜

高普考雙榜	財稅行政	吳○蓉	高普考雙榜	財稅行政	黃○萍	高普考雙榜	會計	黃○紅	高普考雙榜	會計	劉○彤
高普考雙榜	財稅行政	李○芸	高普考雙榜	財稅行政	黃○騰	高普考雙榜	會計	黃○璜	高普考雙榜	會計	賴○婷
高普考雙榜	財稅行政	林○弘	高普考雙榜	財稅行政	楊○聰	高普考雙榜	會計	楊○昇	高普考雙榜	會計	王○惠
高普考雙榜	財稅行政	花○廷	高普考雙榜	財稅行政	葉○君	高普考雙榜	會計	楊○芸	高普考雙榜	會計	歐○學
高普考雙榜	財稅行政	洪○懿	高普考雙榜	財稅行政	葉○瑜	高普考雙榜	會計	劉○均	高普考雙榜	會計	鄭○君
高普考雙榜	財稅行政	張○育	高普考雙榜	財稅行政	鄧○婷	高普考雙榜	會計	劉○嘉	高普考雙榜	會計	游○海
高普考雙榜	財稅行政	張○穎	高普考雙榜	財稅行政	賴○穎	高普考雙榜	會計	蔡○珊	高普考雙榜	會計	高○洋
高普考雙榜	財稅行政	張○穎	高普考雙榜	財稅行政	蘇○閱	高普考雙榜	會計	蔡○婷	高普考雙榜	會計	高○瑋
高普考雙榜	財稅行政	章○穎	高普考雙榜	財稅行政	鍾○凱	高普考雙榜	會計	蔡○葉	高普考雙榜	會計	傅○強
高普考雙榜	財稅行政	莊○瑜	高普考雙榜	財稅行政	蔡○穎	高普考雙榜	會計	謝○晉	高普考雙榜	會計	楊○深
高普考雙榜	財稅行政	郭○瑄	高普考雙榜	財稅行政	郭○瑄	高普考雙榜	會計	羅○芸	高普考雙榜	會計	蔡○泰
高普考雙榜	財稅行政	陳○維	高普考雙榜	會計	彭○娟	高普考雙榜	會計	陳○涵	版面有限 謹向未刊登者致歉		
高普考雙榜	財稅行政	黃○琪	高普考雙榜	會計		高普考雙榜	會計				

連過3榜 楊○芸 111關務特考關稅會計 狀元
111高普會計 榜眼
111普考會計 狀元

我是選擇面授上課，與老師面對面能讓我提高專注力，對課程內容有疑問時，也能即時向老師詢問，雖然往返補習班的時間不少，但對於時間規劃不是特別有自信的我來說，面授能幫助我跟上老師的進度。

一年考取

任○宜 111高普財稅行政 榜眼

我自己覺得理想的讀書方式是上完課之後在上下一堂之前，先把上課講過的題目自己再做一次，如果有相關的練習題也一併寫完。補習班每個月會出申論題，寫好後會有老師批改回饋，以檢討修正及釐清觀念。

一年考取

公職王歷屆試題 (112 高考三級)

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)

【解題關鍵】

1. 《考題難易》★
2. 《破題關鍵》簡單迴歸方程式的截距項計算是課內相當基本的題目，104 年地特有類似的考法，可參考王瑋，迴歸分析 P.2-38 與 P.3-12 相當類似的題型。

【擬答】

截距估計值為 $\hat{A} = \bar{y} - \hat{B}\bar{x}$

$$1. \hat{A}_1 = \bar{y}_1 - \hat{B}_1\bar{x}_1 = \frac{82.51}{11} - 0.5001 \times \frac{99}{11} = 3.0000$$

$$2. \hat{A}_2 = \bar{y}_2 - \hat{B}_2\bar{x}_2 = \frac{82.51}{11} - 0.5000 \times \frac{99}{11} = 3.0009$$

$$3. \hat{A}_3 = \bar{y}_3 - \hat{B}_3\bar{x}_3 = \frac{82.5}{11} - 0.4997 \times \frac{99}{11} = 3.0036$$

$$4. \hat{A}_4 = \bar{y}_4 - \hat{B}_4\bar{x}_4 = \frac{82.51}{11} - 0.4999 \times \frac{99}{11} = 3.0018$$

所以 $\hat{A}_1 < \hat{A}_2 < \hat{A}_4 < \hat{A}_3$

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)

【解題關鍵】

1. 《考題難易》★
2. 《破題關鍵》簡單迴歸方程式的斜率項計算是課內相當基本的題目，104 年地特有類似的考法，可參考王瑋，迴歸分析 P.2-38 與 P.3-12 相當類似的題型。

【擬答】

斜率估計值為 $\hat{B} = \frac{SS_{XY}}{SS_X}$

$$1. \hat{B}_1 = \frac{SS_{X_1Y_1}}{SS_{X_1}} = \frac{55.01}{110} = 0.50009$$

$$2. \hat{B}_2 = \frac{SS_{X_2Y_2}}{SS_{X_2}} = \frac{55}{110} = 0.50000$$

$$3. \hat{B}_3 = \frac{SS_{X_3Y_3}}{SS_{X_3}} = \frac{54.97}{110} = 0.49973$$

$$4. \hat{B}_4 = \frac{SS_{X_4Y_4}}{SS_{X_4}} = \frac{54.99}{110} = 0.49991$$

所以 $\hat{B}_3 < \hat{B}_4 < \hat{B}_2 < \hat{B}_1$

志光·保成·學儒

解鎖高普商科上榜VIP 贏家的聰明選擇

聚焦前三名菁英群

狀元 楊○芸 111普考會計	狀元 姜○佑 110高考經建行政	狀元 李○宜 110高考公平交易管理	狀元 邱○文 110高考財經廉政	狀元 陳○宏 110普考統計	狀元 黃○慧 110普考經建行政
狀元 蔡○ 109高考經建行政	狀元 陳○玄 109高考財務審計	狀元 蔣○涵 109普考會計	狀元 鄭○賢 109普考經建行政	狀元 徐○晟 108高考財稅行政	狀元 朱○宇 108高考績效審計
狀元 徐○晟 108普考財稅行政	狀元 廖○雅 108普考經建行政	狀元 王○文 107高考財務審計	狀元 張○齡 107高考績效審計	狀元 林○吟 107普考會計	榜眼 楊○芸 111高考會計
榜眼 任○宜 111高考財稅行政	榜眼 李○萱 111高考金融保險	榜眼 黃○慧 110高考經建行政	榜眼 董○學 110高考商業行政	榜眼 卓○慧 110高考統計	榜眼 劉○瑾 109高考績效審計
榜眼 曾○ 109普考經建行政	榜眼 蔡○諭 109普考會計	榜眼 廖○雅 109高考經建行政	榜眼 黃○玟 108高考績效審計	榜眼 楊○名 108普考會計	榜眼 曾○郁 107高考績效審計
榜眼 林○雅 107高考財稅行政	探花 黃○菘 111普考會計	探花 李○丞 111高考經建行政	探花 莊○家 110高考統計	探花 林○俊 110高考經建行政	探花 卓○仁 110高考統計
探花 姜○佑 110高考經建行政	探花 曾○ 109高考財稅行政	探花 潘○廷 109高考財務審計	探花 鄭○賢 109高考經建行政	探花 王○慧 109普考經建行政	探花 陳○毅 109普考金融保險
探花 江○怡 108高考財稅行政	探花 高○鈞 108高考財務審計	探花 謝○謙 108高考績效審計	探花 鮑○萱 108高考經建行政	探花 林○涵 107高考統計	探花 詹○樞 107高考財務審計

因版面有限無法一一列出，詳盡榜單請上公職王查詢

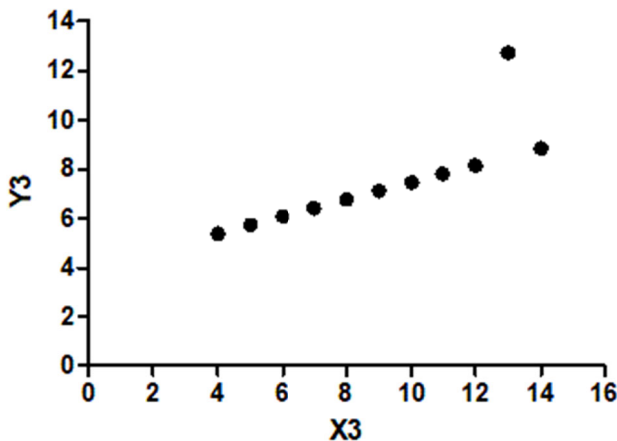
五、針對第三組數據集提出決定係數 (coefficient of determination) 最接近 1.0 的模型。(提示：繪製這一組數據集的散佈圖，並刪除一個影響點。)(20分)

【解題關鍵】

1. 《考題難易》★
2. 《破題關鍵》透過散布圖找出離群值並排除後，重新計算決定係數，103 年地特有考過非常類似的考題，可參考王瑋，迴歸分析 P.7-130 完全相同例題。

【擬答】

(一)第三組數據散佈圖如下：



所以應刪除第 3 筆數據

(二)第三組數據刪除第 3 筆數據後，再計算其決定係數

$$R_{23} = \frac{SS_{X_3 Y_3}^2}{SS_{X_3} SS_{Y_3}} = \frac{(631.85 - \frac{86 \times 69.76}{12})^2}{(497.6686 - \frac{86^2}{10}) \times (49491.1415 - \frac{69.76^2}{10})} = 0.99999$$