

類 科：統計
科 目：迴歸分析
考試時間：2小時

座號：_____

※注意：(一)可以使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

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)
- 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)
- 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)
- 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)
- 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)