

香港統計學會考試



統計學高級證書，2004

試卷 III：統計應用和實踐

時間：3 小時

考生應解答五個題目
所有題目分數相等
每一題目的小題分數標明在括弧中

試場可提供畫圖紙和標準表格

考生可攜帶無聲、無電源線、不可編程電子計算器。
使用計算器的地方必須詳細說明計算的方法。

\log 代表以 e 為底的對數。
以其他數為底的對數會明確標示，例如 \log_{10} 。

注意： $\binom{n}{r}$ 等價於 ${}^n C_r$

1

高級證書試卷 III 2004

本試卷共有 12 頁，採用單面印刷

本頁為第一頁

題目 1 從第二頁開始

試卷共有八個題目

1. (i) 以例證或其他方式，說明你怎樣決定何時進行下列的顯著性檢驗。請說明在各情況下有關檢驗的零假設。

(a) 兩個獨立樣本 t 檢驗。

(b) 配對 t 檢驗。

(c) 曼-惠特尼(Mann-Whitney)檢驗，(亦稱威爾科克森(Wilcoxon)秩和檢驗)。

(d) 威爾科克森(Wilcoxon)符號秩檢驗。

(12 分)

(ii) 某一公司採用智能檢驗甲及智能檢驗乙作為決定哪些應徵者能獲得面試的程序之一。每位應徵者(以隨機分配方式)參與其中一項檢驗，並獲得一個分數。

七位申請某一職位的應徵者在參與檢驗甲後所得的分數為：

52, 61, 68, 50, 60, 58, 64。

其他七位申請同一職位的應徵者在參與檢驗乙後所得的分數

為： 62, 67, 69, 73, 72, 59, 71。

人事經理欲知道兩項智能檢驗是否得出相似的平均分數。請設定適當的零假設及備擇假設。在這情況下，請簡單解釋你會採用一項參數檢驗或非參數檢驗，並說明原因。請進行一項適當的檢驗及報告你的結論。

(8 分)

2. 一項研究比較某一特定電腦軟件應用程式在讀取不同格式檔案的資料時所需的時間。就三種檔案格式，各別隨機選取八個檔案樣本進行研究。以毫秒計，讀取各格式檔案所需時間如下：

標準格式	第一替換格式	第二替換格式
2.02	2.18	1.87
1.99	1.84	2.56
2.01	1.99	2.02
1.88	1.91	2.44
2.27	2.09	2.46
2.36	2.08	2.61
2.31	2.23	2.45
2.35	1.84	2.11

- (i) 在決定那一種分析是適合之前，請列舉你會向資料搜集者提出的兩個問題。 (2分)
- (ii) 請說出進行方差單向分析的模型及其有效性所需的假定。你會怎樣調查是否符合有關假定？ (5分)
- (iii) 使用以上數據進行方差單向分析，並說出有關的零假設及闡釋分析的結果。 (7分)
- (iv) 就零假設為「兩項可能替換格式的總體平均時間均是一樣」進行檢驗。並設定一備擇假設。計算兩項可能替換格式的平均時間差別的 95% 置信區間。 (6分)

3. 在兩項實驗中，一種相信會影響人體血漿膽固醇水平的藥物與一種安慰劑（安慰劑是一種化學惰性的物質，已知道它外表上與一種傳統的醫藥相似，但卻沒有實際效用）進行比較檢驗。

在實驗一中，隨機揀選了十位人士在某一天接受該藥物。一星期後，相信該藥物的所有效力已經消失的情況下，這些人士再接受安慰劑。在實驗二中，亦以隨機方式，揀選了二十位人士。當中隨機揀選十位給予該藥物，而其他十位同時給予安慰劑。在這些人士使用該藥物或安慰劑後兩小時，為他們每位都抽取血液樣本，並取得他們的膽固醇水平資料。

下表載列了四十個血液樣本的膽固醇水平（以每 100 毫升中的毫克計）。

實驗一			實驗二			
	反應			反應		反應
人士	藥物	安慰劑	人士	藥物	人士	安慰劑
A	196	192	K	203	U	168
B	190	187	L	197	V	174
C	155	149	M	210	W	205
D	199	200	N	153	X	251
E	190	183	O	197	Y	160
F	203	203	P	225	Z	173
G	237	242	Q	157	AA	162
H	202	194	R	236	BB	179

I	228	223	S	171	CC	202
J	212	207	T	222	DD	199

(i) 你認為以上哪個實驗設計較好，並說明原因。(3 分)

(ii) 根據實驗一的結果，計算採用藥物及安慰劑人士的膽固醇平均水平差別的 95% 置信區間。說明你所採用的任何假定。

(7 分)

(iii) 根據實驗二的結果，計算採用藥物及安慰劑人士的膽固醇平均水平差別的 95% 置信區間。說明你所採用的任何假定。

(7 分)

(iv) 請說明該藥物是否會增加膽固醇水平。請說明你答案的論據。

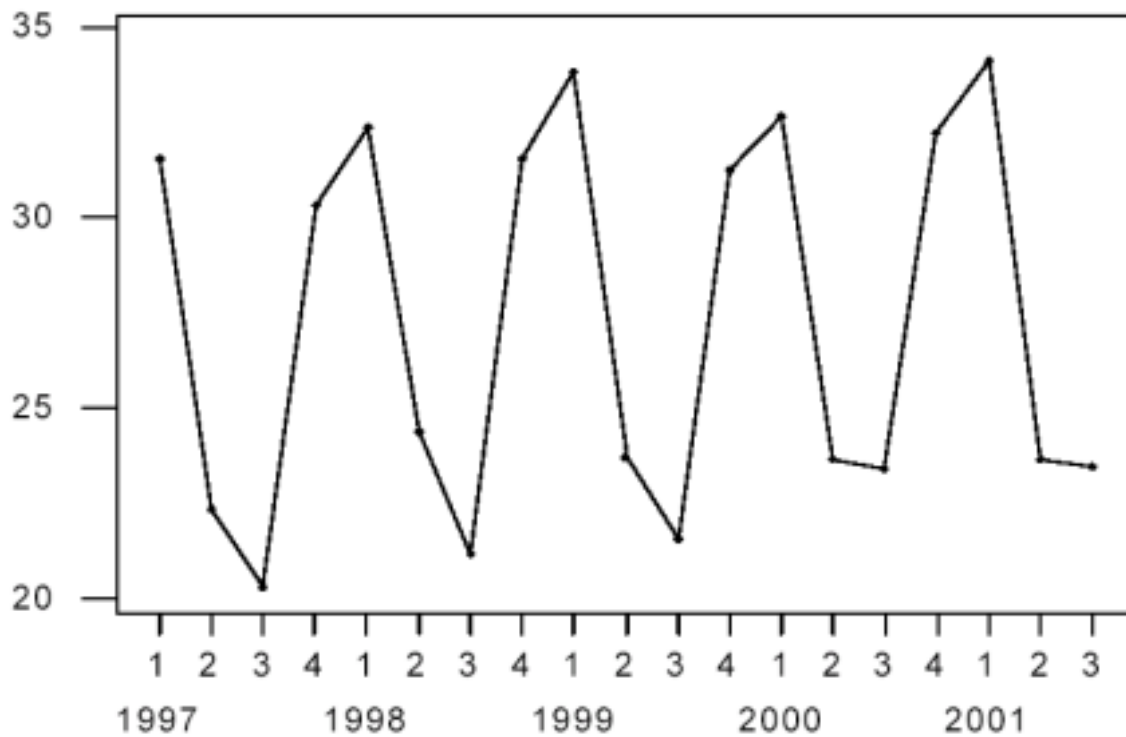
(3 分)

4. 以下圖表展示了英國由 1997 年第一季至 2001 年第三季公共電力供應系統以兆瓦特計的按季本地銷售（簡稱銷售）、集中四季移動平均數（簡稱移動平均）及「銷售減去移動平均」的差異數值。

年	季度	銷售	移動平均	銷售減去移動平均
1997	1	31.54	*	*
1997	2	22.33	*	*
1997	3	20.29	26.22	-5.93
1997	4	30.30	26.57	3.73
1998	1	32/35	26.93	5.42
1998	2	24.36	27.20	-2.84
1998	3	21.16	27.54	-6.38
1998	4	31.54	27.64	3.90
1999	1	33.85	27.61	6.24
1999	2	23.69	27.62	-3.93
1999	3	21.55	27.43	-5.88

1999	4	31.22	27.27	3.95
2000	1	32.64	27.49	5.15
2000	2	23.64	27.84	-4.20
2000	3	23.37	28.15	-4.78
2000	4	32.20	28.32	3.88
2001	1	34.10	28.33	5.77
2001	2	23.61	*	*
2001	3	23.46	*	*

本地電力銷售



年份和季度

(i) 請展示 1997 年第三季移動平均值為 26.22 是如何計算的。

(4 分)

(ii) 根據「銷售減去移動平均」的差異值，估計銷售的季節性變動的型態。

(8 分)

(iii) 為甚麼在估計前描繪數據及其移動平均趨勢數列是一個好的做法？

(4 分)

(iv) 根據(ii)部所得的估計，將 2000 年銷售數字扣除季節性的影響。

你是否能同樣地更正 2001 年數字？請解釋你的答案。

(4 分)

5. 下表載列南非在四年間道路交通撞車的數目及其所引致的傷亡數字的詳細資料。

- (i) 審查有關數字隨着時間的變化，特別是意外的嚴重性是否逐漸減輕。以任何你認為合適的圖表及運算證實你的答案。
- (ii) 請指出哪些其他背景資料會對分析表中的數字有用，並說明原因。

	數目			
	1994	1995	1996	1997
<u>撞車意外</u>				
致命	8140	8335	7850	7790
主要	22594	23988	22707	23059
輕微	60199	61716	54190	57931
沒有受傷	377064	406194	436027	417748

總計	467997	500233	520774	505988
傷亡				
死亡	9981	10256	9848	9691
嚴重受傷	36548	39780	38473	39302
輕微受傷	91887	96689	86728	91760
總計	138416	146725	135049	140753

資料來源：南非統計局

6. 一個鑄造物在七個不同的溫度 t （以攝氏度數計）下的體積 v （立方厘米）的計量數據如下：

溫度	18	19	20	21	22	23	24
體積	10.10	10.80	11.45	12.18	12.80	13.35	15.90

$$\sum t^2 = 3115 \quad \sum v^2 = 1092.9774 \quad \sum tv = 1842.03$$

根據數據估計兩組綫性回歸。一組是根據所有七對計量得出，而另外一組則在計算時省略最後一組計量（即溫度為 24°C ）。以下展示經審核的結果；有關第一組回歸的一些數字已故意地被刪除。

回歸分析：體積對溫度—所有計量

回歸方程式是 體積 = ** + ** 溫度

預估量	系數	標準差系數	T	P
常數	**	2.386	-2.31	0.069
溫度	**	0.1131	7.53	0.001

S = 0.5986 R-Sq = **%

回歸分析：體積對溫度—省略 24°C 的計量

回歸方程式是 體積 = -1.68 + 0.657 溫度

預估量	系數	標準差系數	T	P
常數	-1.6797	0.2803	-5.99	0.004
溫度	0.65657	0.01362	48.19	0.000

S = 0.5700 R-Sq = 99.8%

- (i) 描繪數據及評論。 (4 分)
- (ii) 計算第一組回歸中遺漏的估計系數，並找出其可決系數 R^2 。 (5 分)
- (iii) 哪組回歸較好？請說出原因。 (4 分)
- (iv) 解釋你所選擇的回歸中的估計系數及其可決系數 R^2 。請說明以上展示的「標準差系數」、「T」及「P」數值的意思。為甚麼這些數值有用？ (7 分)

7. 從「1985年美國當前人口統計調查」中隨機抽選了250人。以下載列有關他們的工資（以每小時美元計）及行業（0是其他，1是製造業，2是建築業）變量的一些電腦輸出資料。
- (i) 電腦輸出資料中的「調整平均數」是經刪除了最大5%及最小5%（取最近的整數）的數值所得的平均數。採用調整平均數較常規平均數有何優點？ (4分)
- (ii) 分別繪製三個行業的工資箱形圖。請在箱形圖上指出任何可能被認為是離群值的數值。 (6分)
- (iii) 根據電腦輸出的資料和你在(ii)部的答案，請撰寫一份報告，比較各行業的工資。 (10分)

描述統計：按行業的工資

變量	行業	數目	平均數	中位數	調整	
					平均數	標準差
工資	0	188	9.609	8.500	9.055	5.920
	1	48	9.492	9.245	9.293	4.244
	2	14	9.507	10.375	9.529	3.633
			平均數的		第一	第三
變量	行業	標準誤差	最小值	最大值	四分位數	四分位數
工資	0	0.432	1.750	44.500	5.250	12.120

1	0.613	3.000	20.400	5.890	11.658
2	0.971	3.750	15.000	6.500	11.808

數據展示—行業 0 樣本的工資（遞升次序）

1.75	3.35	3.35	3.40	3.45	3.50	3.50	3.50	3.50
3.50	3.51	3.55	3.56	3.65	3.65	3.75	3.75	3.80
3.84	4.00	4.00	4.13	4.17	4.25	4.25	4.28	4.35
4.35	4.50	4.50	4.50	4.50	4.55	4.55	4.55	4.85
5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.13	5.20
5.21	5.25	5.25	5.30	5.50	5.50	5.50	5.50	5.71
5.71	5.75	5.75	5.80	5.80	5.85	5.87	6.00	6.00
6.25	6.25	6.25	6.25	6.25	6.25	6.50	6.67	6.75
6.85	7.00	7.00	7.00	7.00	7.14	7.50	7.50	7.50
7.53	7.67	7.69	7.75	7.80	7.88	7.96	8.00	8.00
8.20	8.49	8.50	8.50	8.50	8.56	8.63	8.75	8.75
8.75	8.75	8.75	8.80	8.85	8.89	8.90	8.93	9.00
9.00	9.00	9.00	9.00	9.15	9.33	9.37	9.42	9.50
9.50	9.60	9.60	9.86	10.00	10.00	10.00	10.00	10.00
10.00	10.00	10.20	10.25	10.50	10.81	11.11	11.25	11.25
11.35	11.79	11.84	12.00	12.00	12.00	12.16	12.50	12.50
12.65	12.67	13.00	13.12	13.16	13.20	13.33	13.45	13.45
13.45	13.65	13.95	14.00	14.29	14.53	14.67	15.00	15.00
15.00	15.56	15.79	15.95	16.00	16.14	16.65	17.25	18.00
19.98	19.98	20.00	20.00	20.50	20.55	22.20	22.20	22.50
22.50	22.50	24.98	24.98	24.98	24.98	44.50	44.50	

數據展示—行業 1 樣本的工資（遞升次序）

3.00	3.35	4.00	4.50	4.62	4.80	4.85	4.95	5.10
5.40	5.65	5.77	6.25	6.50	6.67	6.75	6.80	7.00
7.78	8.40	8.50	8.89	9.00	9.24	9.25	10.00	10.00
10.50	10.53	10.58	10.62	11.00	11.00	11.25	11.32	11.50
11.71	12.00	12.00	12.00	12.50	13.89	15.38	15.38	16.42
19.00	19.98	20.40						

數據展示—行業 2 樣本的工資（遞升次序）

3.75	4.30	5.00	7.00	7.30	8.90	10.00	10.75	10.78
11.43	11.67	12.22	15.00	15.00				

8. 下表展示在過去五年持有不同學位級別的男女大學畢業生進入某一機構的數目。這些數據可視作從一個大總體中隨機抽選的一個樣本。

學位級別	男	女
第一級	18	17
第二級	90	50
第三級	12	18
合格	61	32

- (i) 請檢驗性別與學位級別是否存在相聯性。請撰寫一份簡短報告給該機構的人事經理，以非技術性文字描述你的結果。

(9分)

- (ii) 在過去五年全國所有畢業生中獲得第一級學位的百分比為

8.3。假定持有第一級學位進入這機構的畢業生數目服從二項分布，請研究這機構在這期間所僱用的第一級學位畢業生總比率是否較平均比率為高。為甚麼二項分布的假定未必是嚴格有效？

(6分)

- (iii) 用過去五年的摘要數據，研究進入這機構而持有第二級學位的男畢業生及女畢業生的總體比率是否相同。(5分)