

# 香港統計學會考試



## 統計學高級證書，2005

### 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

考生可按照香港統計學會列載於考試指引的規例 (文件 Ex1)，使用計算器。

$\log$  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 2005

---

本試卷共有 9 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. 就影響白兔血糖水平的兩項因素進行了一項實驗。其中一項因素為胰島素的兩種類別（即 I 及 II 類）。其他因素則為每種胰島素類別所採用的藥劑份量（即「低」或「高」）。該實驗共用了 4 頭白兔，每頭白兔均接受 4 項不同的處理組合。處理組合是以隨機方式決定次序，並以適當的相間時段進行，以避免一項處理組合對下一項的延後影響。

下表展示血糖水平（mg/100 c.c.）的量度數字：

	胰島素類別			
	I		II	
	藥劑份量		藥劑份量	
	低	高	低	高
白兔 A	61	47	76	63
白兔 B	89	61	74	46
白兔 C	86	63	79	58
白兔 D	69	59	66	50
處理組合總和	305	230	295	217

- (i) 計算 4 項處理組合的平均回應，並以圖表示有關數字。根據圖表所示，兩項因素是否存在交互影響？ (4)
- (ii) 抄下及完成以下的方差分析表，並將處理細分為各有自由度為一的部分，並且計算處理平均數的標準誤差。 (8)

來源	自由度	平方和	均方	F 比率
白兔		297.19		
處理		1496.69		
剩餘				
總和		2463.94		

- (iii) 就該項實驗的結果寫一份簡單報告。 (4)
- (iv) 另外一位實驗員認為，若以 4 組每組 4 頭不同白兔，以隨機方式，每組進行一處理組合的實驗會較為節省時間。請簡單討論以這方式進行實驗的優點和缺點。 (4)

2. (i) 某汽油公司欲測試一種他們認為會減低(並相信不會增加)汽油消耗量的新添加劑。為此進行一項實驗，測試每公升汽油在不同型號的同一類汽車所能行走的哩數。現選取了 12 輛汽車，並隨機分成兩組，每組 6 輛。A 組的汽車使用加了添加劑的汽油，而 B 組的汽車則使用沒有添加劑的汽油，結果如下：

A	6.55	8.78	10.8	9.05	7.35	9.38
B	9.23	10.55	6.73	7.55	8.23	7.72

- (a) 請檢驗是否有證據證明該添加劑能夠增加每公升汽油的平均行走哩數。說明你所作出的任何假定。 (6)
- (b) 現在設想以下情況，若每樣本為  $n$  輛汽車，而  $n$  為大數，得出的樣本平均數為  $\bar{x}_A$  及  $\bar{x}_B$ 。該公司欲想有 95% 信心能察覺使用添加劑後每公升平均行走哩數的每 0.5 公哩的升幅。以  $\mu_A$  為加了添加劑的平均汽油消耗量， $\mu_B$  為沒有添加劑汽油的平均消耗量。請寫出一概率陳述，以表達該公司的目的。

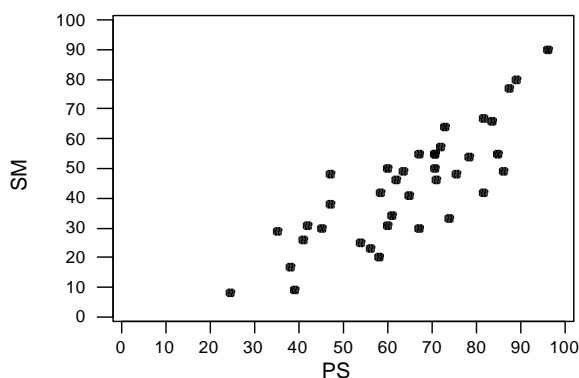
假定現有數據能提供汽油消耗量抽樣方差的一個合理估算，計算  $n$  必須有多大才能達到該公司的目的。 (6)

- (ii) 汽油公司研究及發展主管並不滿意 (i) 部的實驗設計。他決定進行一項新的實驗。在實驗中，上述 A 組同樣的 6 輛汽車使用沒有添加劑的汽油再次進行測試。所得的結果如下：

有添加劑	6.55	8.78	10.80	9.05	7.35	9.38
沒有添加劑	6.15	7.73	10.34	8.16	7.27	8.02

- (a) 請解釋這類實驗設計為何較 (i) 部所採用的好。若現有計劃進行另外一項實驗，請描述你或會作出甚麼其他的改善。 (3)
- (b) 確定是否有證據證明添加劑增加每公升的平均哩數。 (5)

3. 一項有關數學學位課程學生進度的研究中，以隨機方式，選取了 38 位學生，觀察他們在二年級「概率及統計 (PS)」及三年級「統計模型 (SM)」的考試成績。以下的散布圖展示了有關結果。



現建議採用回歸模型，以 PS 課程所得結果  $x$  的線性函數，描述 SM 課程結果  $y$ 。

該模型為  $y_i = \alpha + \beta x_i + e_i$ ，而  $\{e_i\}$  為獨立的  $N(0, \sigma^2)$  分布。

- (i) 根據以下的摘要資料，計算  $\alpha$  及  $\beta$  的最小平方估計數  $\hat{\alpha}$  及  $\hat{\beta}$ 。

$$\sum_{i=1}^{38} x_i = 2436 \quad \sum_{i=1}^{38} y_i = 1670 \quad \sum_{i=1}^{38} x_i y_i = 116888 \quad \sum_{i=1}^{38} x_i^2 = 166991 \quad \sum_{i=1}^{38} y_i^2 = 86402 \quad (5)$$

- (ii) 用實際觀點，分析所配合的模型，並評論有關模型的合適性。 (4)

- (iii) 計算  $\sigma^2$  的估計數及  $\hat{\alpha}$  及  $\hat{\beta}$  的估計方差。 (6)

- (iv) 在 PS 考試中得 80 分的學生，寫出他們在 SM 考試中平均得分的 95% 置信區間推算。評論你的答案。 (5)

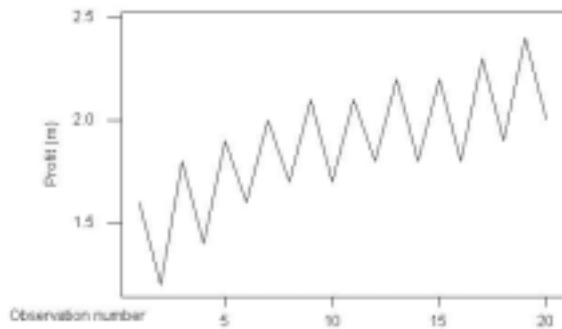
4. (i) 請描述時間數列可分解的 4 個部份。以這些部份，定義時間數列的加法模型，並示範乘法模型如何能轉換成加法模型。 (4)

- (ii) 一間公司在過去十年的半年利潤（以百萬英磅計）展示於下表。這

20個數字已經扣除通脹因素以表示今天的價值。

年份	二月至 七月	八月至 一月	MA1	MA2
1	1.6	1.2		1.45
2	1.8	1.4	1.55	1.625
3	1.9	1.6	1.7	1.775
4	2.0	1.7	1.825	1.875
5	2.1	1.7	1.9	1.9
6	2.1	1.8	1.925	1.975
7	2.2	1.8	2	2
8	2.2	1.8	2	2.025
9	2.3	1.9	2.075	2.125
10	2.4	2.0	2.175	

- (a) 下圖展示有關數據，請描述它的主要特性。 (3)
- (b) 假定是一個加法模型及上表展示了半年移動平均數值。根據這些數值，識別出所採用移動平均的格式。評論所採用的移動平均的合適性。建議其他可用的移動平均，並與現在所採用的作比較。 (4)
- (c) 根據上表所採用的移動平均，得出調整趨勢後的數列，從而估計季節性的組成部份。並且計算不規則組成部份及繪製其時間數列圖。評論模型的合適性。 (9)



5. (i) 從一大批大量生產的裝飾物中，簡單隨機抽選 20 件作為一個樣本。樣本中的每一件裝飾物會被檢查是否符合品質要求。如果樣本包含 0 或 1 件有缺點的裝飾物，該批產品將會被接受。

在一簡單隨機樣本中，每一件裝飾物有相同的概率被抽選為樣本。如果一批產品包含  $p$  比例的裝飾物有缺點，以  $p=0.01, 0.05, 0.1$ ，計算接受它的準確概率。 (4)

- (ii) 以上方案現作出修改，如果樣本包含多過 2 件有缺點的裝飾物，該批產品仍被拒絕接受。但如果一個有 20 件裝飾物的樣本只包含 2 件有缺點的產品，另外抽選一個容量為 20 件產品的隨機樣本，可假設它與第一個樣本完全獨立。如果第二個樣本沒有包含有缺點的裝飾物，該批產品將被接受，否則它將被拒絕接受。

在兩個樣本中，列出有缺點裝飾物的可能數目，而使該批產品被接受。以  $p=0.01, 0.05, 0.1$ ，因此計算接受該批產品的概率。 (7)

- (iii) 被拒絕接受的每批產品會百分百被檢查，全部有缺點的裝飾物會被拿走。如果一批產品的容量是 1 000，以每個給出的  $p$  值，在(i)和(ii)的兩個抽樣方案中，計算產品預期被檢查的總數。簡要評論你的結果。 (9)

6. (i) 你對簡單隨機抽樣這術語的理解是什麼？描述在那些情況下，它可能不是一個適當的抽樣程序，或它與某些其他抽樣方法結合是合適的。 (4)

- (ii) 選擇三種不同類型的非抽樣誤差，並簡要描述在統計調查中甚麼情況下會發生這些誤差。 (7)

(iii) 略述電話調查的主要缺點。 (3)

(iv) 一項研究公眾支持政府衛生政策新建議的民意調查已完成。被調查的 1 015 人中，853 人表示支持。一年後，一項研究它是否仍然得到高支持的類似民意調查得出 1 005 人中有 780 人表示支持。確定是否有任何證據顯示市民支持政策的比例減少了。  
(6)

7. 一保險公司支付由某一類政策引起的索賠總額 (£) 被認為跟隨一個以  $\mu$  為均值的指數分佈。

(i) 就一個有  $n$  宗索賠  $x_1, x_2, \dots, x_n$  的隨機樣本，以  $\mu$  之方式，寫下它的似然函數，和從而找出  $\mu$  的極大似然估計。 (5)

(ii) 在一年來，收到有 96 宗索賠。按類別分組，保險公司支付的總額 ( $x$ ，以£為單位) 給出如下。索賠樣本的均值容量為£2 989.8。基於指數模型，計算預期索賠數量欄內的缺省項目，和進行一個合適統計檢驗去確定一個指數模型對這些數據是否適合。

區間	索賠數量	預期索賠數量
$0 \leq x < 250$	11	7.70
$250 \leq x < 500$	16	7.08
$500 \leq x < 1\,000$	16	
$1\,000 \leq x < 1\,500$	10	10.58
$1\,500 \leq x < 2\,000$	10	8.95
$2\,000 \leq x < 3\,000$	11	
$3\,000 \leq x < 4\,000$	7	10.01
$4\,000 \leq x < 6\,000$	5	12.29
$6\,000 \leq x < 8\,000$	4	
$8\,000 \leq x < 10\,000$	2	
$10\,000 \leq x < 20\,000$	2	3.27
$20\,000 \leq x < 40\,000$	1	0.12
$40\,000 \leq x < 60\,000$	1	0.00

(10)

(iii) 用這合適的指數模型，估計支付一宗索賠總額超過£20 000 的概率。評論你的答案與數據的關係。 (3)

- (iv) 用你給(i)和(ii)的答案，對此類政策索賠款項的改良模型，討論它的特徵。 (2)

8. 一麵粉製造商完成一個對某烘焙過程中生產線變化的研究。鬆脆的餡餅方塊放置在一個可橫放 7 個方塊和直放 10 個方塊的盤子內烘焙。

對剛出烤爐的完成產品，搜集它的容量數據。量度每個方塊的寬度、長度和高度(mm)；又基於這些尺寸，計算它的大概體積(mm<sup>3</sup>)。所得的數據總結如下。

儘可能以資訊性的方法，包括適當的圖，但避免正式統計資料（例如顯著檢驗或置信區間），表述你對烘焙過程的結論。你應考慮的問題，諸如是否有可查覺的效果是因鬆脆的餡餅方塊放在盤子上的位置而產生；方塊的長闊高和任何對於使用那些可放置餡餅方塊的影響的可能關係。指出任何關於可能有用原始數據的附加資料。 (20)

#### 以寬位置（即橫放在盤子寬度的位置）分類的摘要數據

長度變數，N = 10

寬位置	均值(長)	標準誤差(長)	中位數(長)	最小(長)	最大(長)
1	86.4	1.42984	86.5	85	89
2	87.3	1.25167	87.0	86	90
3	87.5	1.43372	88.0	84	89
4	86.7	1.70294	86.5	84	90
5	86.2	1.81353	86.0	83	89
6	84.7	1.63639	84.5	83	88
7	84.3	1.49443	84.0	82	87

寬度變數，N = 10

寬位置	均值(寬)	標準誤差(寬)	中位數(寬)	最小(寬)	最大(寬)
1	77.1	1.91195	77.5	72	79
2	78.1	1.44914	77.5	77	81
3	77.0	2.78887	76.5	74	83
4	76.1	1.66333	76.0	73	79
5	76.8	0.91894	77.0	76	79
6	78.3	1.41814	78.0	77	81
7	79.1	1.66333	79.5	76	81

高度變數，N = 10

寬位置	均值(高)	標準誤差(高)	中位數(高)	最小(高)	最大(高)
1	27.8	2.44040	27.5	24	31
2	27.9	2.84605	28.0	24	32
3	29.3	2.75076	28.0	27	34
4	29.2	2.44040	30.0	25	32
5	28.4	2.01108	29.0	24	31
6	30.6	2.50333	30.0	27	36
7	31.5	3.89444	31.5	27	38

體積變數，N = 10

寬位置	均值(體積)	標準誤差(體積)	中位數(體積)	最小(體積)	最大(體積)
1	185044	15396.0	183344	159120	207669
2	190217	19868.5	187910	167475	216832
3	196930	13628.0	193604	178524	221408
4	192254	11153.5	194940	174174	204972
5	187874	11941.2	189535	162336	200970
6	202796	15521.1	201192	181305	238392
7	209629	22467.7	205096	183222	242609