香港統計學會考試

統計學高級證書，2004

試卷 I：統計理論
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 爲底的對數。
以其他數爲底的對數會明確標示，例如 $\log_{10}$.

注意：$\binom{n}{r}$ 等價於 ${}^{n}C_{r}$

本試卷共有 9 頁，**採用單面印刷**

本頁爲第一頁

題目 1 從第二頁開始

試卷共有八個題目

1. 某大城市的成年人人口中，60%贊成設立一個新的消閒中心，30%反對，10%則沒有意見。從該人口中以隨機抽樣方式抽選了 4 個成年人，並取得他們對新消閒中心的意見。

(i) 找出以下情況的概率：

(a) 這 4 個人的想法一致;

(b) 4 人中沒有一個反對設立該中心;

(c) 所抽選樣本包含了 3 種不同意見(即贊成、反對、沒有意見);

(d) 若已知道 4 人中無人持反對意見，則 4 人均爲贊成消閒中心的人士。

(10 分)

(ii) 請說明樣本中贊成消閒中心的人數的期望數目及方差。

(3 分)

(iii) 在該城市，四分之一的成年人被分類爲 “年青人士” (年齡 <30)而四分之三則爲 “較年長人士” (年齡爲 30 或以上)。你已知道 12%的年青人士反對設立該消閒中心。請推論較年長人士中反對設立該消閒中心的比例。 (3 分)

(iv) 若已知道樣本中包含一個年青人士及三個較年長人士，請

找出樣本中只有一位人士反對設立消閒中心的概率。

(4 分)

2. (i) 隨機變量 X 服從泊松(Poisson)分佈，其概率質量函數爲

$$f(x) = e^{-\lambda}\frac{\lambda^{x}}{x!}, \qquad x = 0,1,2,\ldots$$

(a) 請找出 X 的矩母函數。

(b) 由此或其他方法證明 X 的均值及方差均爲 λ 。

請解釋泊松近似二項分佈，及指出適用於哪一種情況。

(8 分)

(ii) 某公務員計算支付給失業成年人士的每星期社會保障金。這金額是按申領人士的情況而厘定，而計算可能會有錯誤。經過一段長時間後，發現錯誤計算的概率爲 0.0075。請找出在一個 200 人的樣本中，包含(a)一個錯誤計算的概率;及(b) 4 個錯誤計算的精確概率(請顯示小數點後 4 個位)。

(5 分)

(iii) 請使用普阿泊松近似二項分佈的方法，再計算(ii)部。在每項計算中，請找出使用松方法計算的百分誤差(請顯示 2 個有效數字)。並簡要地解釋你所得的結果。

(7 分)

3. 機器把蘋果汁分配入紙盒內。每盒蘋果汁的名義容量爲一公升(1000 毫升)。而紙盒內的蘋果汁的實質體積，可視爲獨立地服從正態分佈，其均值定爲 1010 毫升，標準差則爲 8 毫升。

(i) 請找出在長期生產情况下，實際載有較名義容量爲少的紙盒蘋果汁的比例。 (4 分)

(ii) 紙盒裝的蘋果汁通常以 6 盒爲一包的方式發售。請列出每包 6 盒果汁的總容量的分佈。請找出一包 6 盒的蘋果汁的總容量少於 6 公升的概率，並解釋這概率低於(i)部答案的原因。 (6 分)

(iii) 現已配備一部新的及更準確的機器，每盒分配出來的果汁的體積仍然按正態分佈，但標準差則較小，只有 4 毫升。若不增加目前實際容量低於名義容量的紙盒的比例，請問目前分配入每一紙盒的果汁的體積的均值，可以減少多少?假設該部新機器的額外成本爲 200 英鎊，而蘋果汁的成本爲每公升 1 英鎊，請問該部較準確的機器須注滿多少盒蘋果汁，才能證明其較高的成本是合理的。 (10 分)

4. (i) 請說明何謂正態近似二項分佈，及在甚麼條件下有效。一種地對空飛彈在 50 次發射中，有 30 次成功擊中目標。請找出

某一次飛彈擊中目標的概率 p 的 95%的近似置信區間。

(ii) 地對空飛彈通常是以一對方式獨立地向目標發射。只須其中一枚擊中目標，便可摧毀目標。(可假設兩枚飛彈成功擊中目標是各自獨立的事件。)請以 p 找出當一對飛彈向某目標發射，該目標被摧毀的概率，並提供其點估計數。

(iii) 請把(i)部內 p 的置信區間適當地變換，找出當發射一對飛彈時目標被摧毀的概率的 95%的置信區間。

(iv) 國防專家稱敵方飛行目標在剛進入國家上空範圍前可以被偵察到。國防部的策略是當偵察到某一目標時，數對地對空飛彈(假設是 n 對)要同時發射。假設你在(ii)部的估計是準確的，請找出能夠把不成功摧毀目標的概率減小於 0.0005 的 n 的最小值。

5. (i) 某連續隨機變量 T 的概率密度函數 (pdf) f(t) 爲:

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0; \quad \lambda > 0$$

(a) 畫出 f(t)的圖形。

(b) 證明其累積分佈函數爲:

$$F(t) = 1 - e^{-\lambda t}, \quad t > 0; \quad \lambda > 0$$

(c) 推論出:

$$P(a < T \le b) = e^{-\lambda a} - e^{-\lambda b}, \quad 0 < a < b$$

(7 分)

(ii) 某建築公司的會計經理假設一份貨單收到付款，所需的時間 T 是一個隨機變量，其概率密度函數 pdf f(t)如上述而其中的 λ 是未知值。從一組隨機抽選的 100 份貨單中，他發現 50 份在一星期內收到付款，35 份在第二個星期內收到付款，而 15 份在兩個星期後收到付款。請清楚解釋爲何這些數據的似然可以表達如下:

$$L(\lambda) = k(1-e^{-\lambda})^{50}(e^{-\lambda}-e^{-2\lambda})^{35}(e^{-2\lambda})^{15} \quad ,$$

其中 k 爲常數。 從而請證明

$$\log L(\lambda) = \log(k) + 85\log(1-e^{-\lambda}) - 65\lambda$$

並請推論 λ 的極大似然估計值近似爲 0.836。

(iii) 假設 λ =0.836，請分別計算第一個星期內、第二個星期內和第二個星期之後收到付款的貨單的期望數目。從而請簡要地評論使用這 pdf f(t)模型配合有關數據有多好。 (5 分)

6. (i) 隨機變量 X 的概率密度函數 f(x)爲：

$$f(x) = \frac{k}{x^{k+1}}, \qquad x \geq 1, \quad k > 0$$

請畫出 f(x)的曲綫圖，並找出其累積分佈函數 F(x)。 (5 分)

(ii) 請找出 X 的中位數、下四分位數及上四分位數，並推論 X 的

半內四分位數間距。 (5 分)

(iii) 假設 k>2，請找出 X 的期望和方差。X 會大於它的期望的概率是多少? (7分)

(iv) 在烏托邦這國家，收入是以 10,000 英鎊爲單位，其分佈正如 X，而其中 k=3。請找出(a)收入的中位數，(b)收入的均值，(c)收入大過 100,000 英鎊的比例。 (3 分)

7. 在一木匠的工場裏，木材的下腳料的長度 X 服從一個連續均勻分佈，其概率密度函數(pdf)如下：

$$f(x) = \frac{1}{\theta}, \quad 0 \le x \le \theta,$$

其中 θ (>0)爲一未知參數。

(i) 請找出 X 的均值和方差。 (5 分)

(ii) 該木匠抽選了一個下腳料的隨機樣本，長度爲 X_1，X_2 ，…，X_n。請解釋爲何：

P(樣本中最長下腳料的長度≦x)= $\left[\frac{x}{\theta}\right]^n$, 0≦x≦ θ ,並推論樣本的極大值的 pdf，設其爲 $X_{(n)}$ ，並證明

$$E(X_{(n)}) = \frac{n\theta}{n+1}$$

及

$$Var(X_{(n)}) = \frac{n\theta^2}{(n+1)^2(n+2)},$$

請列出一個以 θ 爲無偏估計量的 $X_{(n)}$ 的多重倍數，並找出其方差。 (11 分)

(iii) 請證明 $\frac{2}{n}\sum_{i=1}^{n} X_i$ 是 θ 的矩估計量，並找出這估計量的方差。

(4 分)

8. 在一個辦公室效率研究中，某公司就十項常要處理的辦公室的工作，設立了基準時間 x_1，$x_2 \ldots$，x_{10}。一位新聘用的見習生處理這十項工作所需時間 y_1，$y_2 \ldots$，y_{10}，作爲工作表現的根據。以分鐘計算的時間(xi，yi)，i=1, 2, …, 10，列於下表。

工作項目	A	B	C	D	E	F	G	H	I	J
基準時間 x	5	5	10	10	10	15	15	20	20	40
見習生所用時間 y	8	12	16	16	21	18	20	25	31	53

附註：

$\sum x_i = 150,\quad \sum x_i^2 = 3200,_{i}\quad \sum y_i = 220,\quad \sum y_i^2 = 6280\quad \sum x_i\, y_i = 4440$

(i) 根據上述數據繪製一個散點圖，並簡要地說明在這情況下，使用簡單綫性回歸模型分析是否適合。 (5 分)

(ii) 在清楚說明你的假設後，請用最小平方法爲這些數據配上一個簡單綫性回歸模型，並計算(a)其剩餘均方(該回歸模型內的隨機項的方差 σ^2 的無偏估計值)及(b)其可決系數 R^2。

(7 分)

(iii) 若其斜率估值的方差爲 $\frac{\sigma^2}{\sum (x_i - \bar{x})^2}$，其中 $\bar{x}$ 代表樣本中基準時間的均值，以 5%顯著水平測試這回歸模型中的斜率是 1 的零假設。 (4 分)

(iv) 該公司的經理不滿意你所得到的回歸關係式。他認爲當 x=0 時則 y=0。請說明綫性回歸模型中那一個形式已包含了這特殊情況。找出計算其斜率參數估值的公式，並按上列數據估計這參數。