

※ 請依序作答，並標明作答之大題及其題號。

一、【共 30 分】

由一組資料：(1,2,3,4,5)中，以歸還方式隨機抽取 2 個數字構成一個樣本，請回答以下問題：

- (1) 計算此組資料的平均 μ 和變異數 σ^2 。(5 分)
- (2) 列出所有可能被抽出的樣本，以 K 表示這些可能的樣本個數。(5 分)
- (3) 以 \bar{x}_i, S_i^2 分別表示每一個可能樣本的樣本平均與樣本變異數， $i=1, \dots, K$ ，計算這些 K 個 \bar{x}_i 和 K 個 S_i^2 。(8 分)

(4) 取 $\bar{\bar{x}} = \frac{1}{K} \sum_{i=1}^K \bar{x}_i$ 和 $\bar{S}_1^2 = \frac{1}{K} \sum_{i=1}^K S_i^2$ 和 $\bar{S}_2^2 = \frac{1}{K} \sum_{i=1}^K \frac{(x_{i1} - x_{i2})^2}{2}$ ，將此三個

數值和 μ 和 σ^2 比較，作出一個有意義的結論。(註： \bar{S}_2^2 的等式右側之 x_{i1} 和 x_{i2} 分別表示第 i 個樣本中的二個資料數值。)(12 分)

二、【共 16 分】

一醫學研究欲探討飲用綠茶是否會降低血中膽固醇的效用，因血中膽固醇因人而異，故隨機收集 n 個高膽固醇患者，在研究開始時測量每個人的(前測)膽固醇值，記為 $Y_{i1}(i=1, \dots, n)$ ，研究過程中指導每個人於三餐飯後飲用 250cc 的綠茶並持續三個月，三個月後再測量每個人的(後測)膽固醇值，記為 $Y_{i2}(i=1, \dots, n)$ 。回答下列問題：

- (1) 由於 Y_{i1}, Y_{i2} (前測、後測膽固醇值)源自同一人，所以 Y_{i1}, Y_{i2} 應為相依而非獨立。根據此醫學研究設計、目的、與資料性質，寫出每個人之前測、後測膽固醇值的線性回歸模式($Y_{ij}, j=1, 2, i=1, \dots, n$ 的 $2n$ 個線性回歸模式)，並說明此模式中所有設定符號的意義以及所需之機率統計的假設。(8 分)
- (2) 由(1)所設定之模式、假設、與符號，寫出以該模式之回歸參數表示此檢定之虛無擬說與所對應之檢定統計量，並說明此檢定統計量是否為成對樣本的 t 檢定統計量。(8 分)

三、【共 24 分】

某研究欲探討 A 與 B 兩個二元變數的關係，收集資料整理成一個 2×2 列聯表如下：

	B=1	B=2	
A=1	Y_{11}	Y_{12}	Y_{1+}
A=2	Y_{21}	Y_{22}	Y_{2+}
	Y_{+1}	Y_{+2}	Y_{++}

在此，A、B 變數各有兩種狀態且均以 1、2 表示之， $Y_{i+} = Y_{i1} + Y_{i2}$ 且 $Y_{+j} = Y_{1j} + Y_{2j}$ ， $i, j=1, 2$ 。令 $p_{ij} = P(A=i, B=j)$ ， $i, j=1, 2$ ，而且 $p_{11} + p_{12} + p_{21} + p_{22} = 1$ 。定義：兩機率之差值為 $D = P(A=B) - P(A \neq B)$ 。

- (1) 若此 2×2 列聯表之資料係自一群體隨機選取 Y_{++} 個個體之樣本並調查紀錄每個個體其 A、B 的狀態整理所得之 4 筆數據： Y_{ij} ， $i, j=1, 2$ 。

根據此數據回答下列問題：

- (i) 推導出 D 之最大概似估計量 (maximum likelihood estimator，記為 \hat{D})。(4 分)
- (ii) 求出 \hat{D} 之變異數，即 $Var(\hat{D})=?$ (4 分)
- (iii) \hat{D} 是否可以用來檢定 A 與 B 兩變數間獨立與否？需有推導驗證之過程。(4 分)

- (2) 若此 2×2 列聯表之四筆數據係來自收集以下兩組獨立樣本：

樣本一：由一具有 A=1 之群體中隨機選取 Y_{1+} 個個體，

樣本二：由一具有 A=2 之群體中隨機選取 Y_{2+} 個個體，

並調查紀錄每個個體其 B 的狀態整理所得之 4 筆數據： Y_{ij} ， $i, j=1, 2$ 。

根據 4 筆數據數據回答下列問題：

- (i) 寫出此 4 筆數據之概似函數 (likelihood function)。(4 分)
- (ii) 根據(i)是否可以求出 \hat{D} 之最大概似估計量？需有推導驗證之過程。(4 分)
- (iii) 欲檢定 A 與 B 兩變數獨立與否，寫出以 p_{ij} 表示所對應的虛無擬說 (null hypothesis) 與對立擬說 (alternative hypothesis)。(4 分)

四、【共 30 分】

考慮隨機變數 Y 來自母群體平均值為 5 之布瓦松分佈, $Y \sim \text{Poisson}(\mu)$

- (1) 計算 $E(Y^2)$ 之值。(3 分)
- (2) 計算 $Y \geq 2$ 之機率。(3 分)
- (3) 若吾人觀察 6 個樣本值分別為 11, 0, 2, 3, 4, 3, 請問樣本變異數為母群體變異數幾倍?(3 分)
- (4) 假設上述樣本代表 6 個人在半年內中耳感染次數, 請問上述樣本變異數過大, 很可能是所觀察樣本違反布瓦松分佈之何種假定?(3 分)
- (5) 寫出上述(3)所得樣本之對數概似函數。(3 分)
- (6) 以最大概似法, 利用上述(3)之樣本估計 μ 之最大概似估計值 $\hat{\mu}$ 。(3 分)
- (7) 寫出估計 μ 過程中之充分統計量 (Sufficient statistics)。(3 分)
- (8) 求取上述(6)估計值 $\hat{\mu}$ 之變異數。(3 分)
- (9) 假若自母群體隨機抽取 50 個獨立樣本, 請依大樣本理論描述抽樣分佈之平均值及變異數。(3 分)
- (10) 假設 50 個樣本平均值 \bar{y} 和上述(3)相同, 利用上述(9)之抽樣分佈求取以此樣本估計 μ 之 95% 信賴區間。(3 分)