國立清華大學命題紙

99學年度 統計學研究所 碩士班入學考試　　　*請在【答案卷卡】作答

科目 統計學 科目代碼 0103 共 4 頁第 1 頁

****共31個填充題，【答案卷卡】中必須自行清楚標明每格之編號: (1), (2)...(31)****

**** (1) 至(10) 每格3%; (11) 至(24) 每格4%; (25) 至(31) 配分如格中所列****

****每格只須最後答案或式子; 不須導證過程****

參考值:卡方分布上臨界點(upper critical point)

$\chi^2_{1,5\%}=3.84$，$\chi^2_{2,5\%}=5.99$，$\chi^2_{3,5\%}=7.815$，$\chi^2_{4,5\%}=9.488$，$\chi^2_{5,5\%}=11.071$，$\chi^2_{6,5\%}=12.592$

標準常態分布的上臨界點 $Z_{0.95}=1.64$; $Z_{0.975}=1.96$

題中所有檢定顯著水準均假設為5%

1. **(30%,每格3%)** 觀察100株植物, 對任一株看其基因類型爲 aa, Aa, AA 中那一型, 現觀察之100株植物中各有 aa, Aa, AA 基因類型之數目爲 $(X_1, X_2, X_3)=(10, 60, 30)$株，令三種類型在所有此種植物中其機率各爲 $P(aa)=\pi_1$, $P(Aa)=\pi_2$, $P(AA)=\pi_3$.

(a) Hardy-Weinberg 平衡模式假設 $\pi_1=p^2, \pi_2=2p(1-p), \pi_3=(1-p)^2$ 其中 p 爲未知參數; 則此模型之概似函數 (likelihood function) $L(p)$ 爲何?___(1)___; 此組數據 p 之最大概似估計量及此組數據之估計值爲何? ___(2)___, p 之最大概似估計值之變異數如何得到? ___(3) 簡述方法即可, 不必計算___

(b) 在此 Hardy-Weinberg 平衡模式中, 100株植物三種基因形式類型 aa, Aa, AA 數目之期望值各爲 ___(4)___.如果以 Hardy-Weinberg 平衡模式來配適 (fit) 以上資料, Pearson 卡方適合度檢定 (goodness of fit) 大約爲何值___(5)___ 檢定結論爲___(6) 不可只寫棄卻與否, 必須有完整敘述___

(c) 假若今觀察之植物株數可任意增多L 倍成爲 N = 100×L 株, 即 N/100 = L 並假設觀察之100×L 株植物中各有 aa, Aa, AA 基因類型之數目爲 $(Y_1, Y_2, Y_3)=(10\times L, 60\times L, 30\times L)$株, 則對 Hardy-Weinberg 平衡模式之 Pearson 卡方適合度檢定 (goodness of fit) 改變成何值? ___(7)___

(d) 假若現懷疑 Hardy-Weinberg 平衡模式並不適合，而認爲三個機率除總和爲1外沒有任何限制(即一般模式);現檢定虛無假設 H_0: $\pi_1=p^2, \pi_2=2p(1-p), \pi_3=(1-p)^2$ 及 與對立假設 H_1: 三個機率除總合爲1外沒有任何限制, 則在對立假設下在所有此種植物中其三種類型機率最大概似估計值各爲___(8)___ 寫下可能比檢定(likelihood ratio test) ___(9)___(列出式子即可，但需說明P-value在大樣本假設下如何得到或臨界值如何求得) 。

(e) 假若想知道此種植物中基因形式類型爲 aa 之機率在一般模式下 95%信賴區間，利用一般模式下 X_1 之大樣本分布找出95%之信賴區間___(10) 列出式子即可, 不必計算___ 。

2. **(20%,每格 4%)** 過去生物學家認爲某種蜘蛛雌性的蜘蛛其體長與雄性差不多，現有人懷疑雌性的蜘蛛其體長應較雄性爲長，爲了驗證其觀點，現自兩族群(母體)各抓了雄性蜘蛛 20 隻 (樣本平均爲 $\overline{X}$)，雌性蜘蛛 12 隻(樣本平均爲 $\overline{Y}$)，其 EXCEL 輸出資料(顯著水準爲 5%)如下表內所示:

	A	B	C
1	F 檢定：兩個常態母體變異數的檢定		
2			
3		雄性	雌性
4	平均數	5.87	8.2125
5	變異數	0.588	0.918693182
6	觀察值個數	20	12
7	自由度	19	11
8	F	0.640039582	
9	P(F<=f) 單尾	0.189474685	
10	臨界值：單尾	0.42731152	

	A	B	C	D	E	F
1	t 檢定：兩個母體平均數差的檢定			t 檢定：兩個母體平均數差的檢定		
2	假設變異數相等			假設變異數不相等		
3		雄性	雌性		雄性	雌性
4	平均數	5.87	8.2125	平均數	5.87	8.2125
5	變異數	0.588	0.918693182	變異數	0.588	0.918693182
6	觀察值個數	20	12	觀察值個數	20	12
7	Pooled 變異數	0.709254167		假設的均數差	0	
8	假設的均數差	0		自由度	19	
9	自由度	30		t 統計	-7.196365366	
10	t 統計	-7.617444231		P(T<=t) 單尾	3.8899E-07	
11	P(T<=t) 單尾	8.51432E-09		臨界值：單尾	1.729131327	
12	臨界值：單尾	1.697260359		P(T<=t) 雙尾	7.77981E-07	
13	P(T<=t) 雙尾	1.70286E-08		臨界值：雙尾	2.093024705	
14	臨界值：雙尾	2.042270353				

(a) 假設雄性蜘蛛與雌性蜘蛛族群(母體)體長之分布各爲 $N(\mu_1,\sigma_1^2)$與 $N(\mu_2,\sigma_2^2)$，由報表，可否假設兩族群之變異大致一樣？_____ **(11)** _____

(b) 此時懷疑雌性的蜘蛛其體長應較雄性爲長的人其虛無假設與對立假設各爲何?_____ **(12)** _____

(c) 樣本平均 $\bar{X}-\bar{Y}$ 大約要小過多少_____ **(13)** _____(顯著水準爲 5%)才可下結論說資料足以說明雌性的蜘蛛其體長較雄性爲長。

(d) 針對此數據與報表，檢定的結論爲_____ **(14)** 不可只寫棄卻與否，必須有完整敘述_____

(e) 雄性蜘蛛平均體長 μ_1 之 95%之信賴區間_____ **(15)** 列出式子，並說明式中每一項在此資料下之數值即可_____

參考數值: $(0.588)^{0.5} = 0.767$, $(0.9187)^{0.5} = 0.958$, $(0.7093)^{0.5} = 0.842$, $(1/20+1/12)^{0.5} = 0.365$, $(20)^{0.5} = 4.472$

3. **(16%,每格 4%)** 假若給定 $\lambda_i > 0$，隨機變數 X_i 之條件分布爲 Poisson (λ_i)，pdf 爲 $P(X_i = x \mid \lambda_i) = \lambda_i^x e^{-\lambda_i} / x!,\ x = 0, 1, 2\ldots,\quad i = 1, 2, \ldots, S$ 且 $\lambda_1, \lambda_2, \ldots, \lambda_S$ 爲自 gamma 分布取出之一組 i.i.d. 隨機變數具有 pdf $f(\lambda) = \dfrac{1}{\Gamma(k)\beta^k}\lambda^{k-1}e^{-\lambda/\beta},\quad \lambda > 0,\quad k > 0,\quad \beta > 0$

   (a) 隨機變數 X_i 之非條件(unconditional)分布之 pdf 爲何? 寫出 $P(X_i = x) =$ ____**(16)____

   (b) 當 $k \to 0$，$P(X_i = x \mid X_i > 0)$ 趨近 ____**(17)**____

   (c) $(X_1, X_2, \ldots, X_S \mid \sum_{i=1}^{S} X_i = n)$ 之分布之 pdf 爲何? ____**(18)**____

   (d) 當 $k \to \infty$，$(X_1, X_2, \ldots, X_S \mid \sum_{i=1}^{S} X_i = n)$ 之分布之 pdf 趨近 ____**(19)**____

4. **(20%,每格 4%)**若某產品壽命爲指數分布具 pdf 如下 (其中 θ 爲未知參數)

$$f_\theta(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}},\quad 0 < x < \infty,\quad \theta > 0.$$

   定義 $R(t) = P(X > t)$ 爲任一產品壽命超過時間 t 之機率，t 是一固定正數.

   (a) 現若有 i.i.d. 數據 $(X_1, X_2, \ldots X_n)$，θ 之均勻最小變異不偏估計量(UMVUE, uniformly minimum variance unbiased estimator) 爲 ____**(20)**____；

   (b) $R(t)$ 之一不偏(unbiased) 估計量爲 $U = \begin{cases} 1 & if\ X_1 > t \\ 0 & if\ X_1 \le t \end{cases}$。利用 Rao-Blackwell 定理，則 $R(t)$ 之 UMVUE 爲 ____**(21)** 必須寫出估計量形式____

   (c) 若實驗同時測試 n 產品，但只有得到最先壞掉的 r 個產品壽命 $X_{(1)} < X_{(2)} < \cdots < X_{(r)}$，$(r < n)$ 即因時間關係要停止實驗，因此其它沒壞掉的 $n-r$ 個產品壽命至少爲 $X_{(r)}$；此組數據之聯合 pdf 爲 ____**(22)**____；此組數據下 θ 之 UMVUE 爲 ____**(23)**____ 此組數據下 $R(t)$ 之 UMVUE 爲 ____**(24)**____

5. **(6%,每格 2%)**

   假若丟骰子 n 次(各面出現機率相同)，每次結果視爲獨立

   (a) 令 X_n 爲此 n 次中出現 3 的次數，當 $n \to \infty$

$$\frac{X_n}{n} \xrightarrow{P} \text{\_\_\_\_(25)\_\_\_\_}$$

   (b) 利用中央極限定理，當 $n \to \infty$

$$\frac{X_n - a}{b} \xrightarrow{D} N(0,1)$$

   $a =$ ____**(26)**____，$b =$ ____**(27)**____

6. **(8%, 配分如每格中所列)** 甲、乙二醫院有下列治療記錄

	甲醫院				乙醫院		
	死亡	生存	Row sum		死亡	生存	Row sum
病重者	80	720	800	病重者	60	440	500
病輕者	4	196	200	病輕者	60	1440	1500
Column sum	84	916		Column sum	120	1880	

(a) 甲、乙二醫院之總死亡率何者較高？ **(28) 1%**

(b) 對病重者而言，甲、乙二醫院死亡率何者較高？ **(29) 1%**

(c) 對病輕者而言，甲、乙二醫院死亡率何者較高？ **(30) 1%**

(d) 以上有無矛盾？爲什麼？ **(31)** 簡潔說明理由 **5%**