《統計學》

試題評析	這次高考統計的統計學考的題目少只有四大題，但題目有些小細節小變化要注意的，第一二大題分別為抽樣分配和ANOVA，算是基本題型，一定要把握且細心檢查；第三題為混合型變數變換，同學要注意小心；第四題雖然考到數統範圍，算是統計學課內要會的題目且充分、MLE與N-P Lemma考的都是基本題型，這份考卷要上榜的同學一定要考到95分。
考點命中	第一題：1.《高點・高上統計學講義》第四回，5-2常態母體下之抽樣分配，蘇建郎老師編撰，頁24。 2.《高點・高上統計學講義》第二回，3-2生成函數，蘇建郎老師編撰，頁17。 第二題：《高點・高上統計學講義》第七回，8-4二因子變異數分析，蘇建郎老師編撰，頁13-15。 第三題：《高點・高上統計學講義》第二回，3-2期望值與變異數mgf變數變換生成函數，蘇建郎老師編撰，頁9,13,18-21。 第四題：1.《高點・高上統計學講義》第六回，6-1點估計，蘇建郎老師編撰，頁6,28。 2.《高點・高上統計學講義》第六回，7-4 MPT UMPT LRT，蘇建郎老師編撰，頁35-36。

一、(一)令 $Z_1,Z_2,\ldots,Z_n$ 為標準常態分配的隨機樣本，說明隨機變數W的機率分配（不需證明）。（9分）

1. $W=Z_1^2+Z_2^2+Z_3^2+Z_4^2+Z_5^2$
2. $W=Z_5/\sqrt{X/5}$，其中 $X=Z_1^2+Z_2^2+Z_3^2+Z_4^2+Z_5^2$
3. $W=\dfrac{Y/2}{X/5}$，其中 $X=Z_8^2+Z_9^2$，$X=Z_1^2+Z_2^2+Z_3^2+Z_4^2+Z_5^2$

(二)令Z＝Y＋X，Y為卡方分配自由度 v_1 分配、Z為卡方分配自由度 v_3 分配，且Y與X相互獨立，以動差生成函數（moment generating function）推論X的機率分配。（9分）

答：

(一)

1.卡方分配自由度 5。

2. T 分配自由度 5。

註：題目應該想出 T 分配自由度 5 這個答案，但分子的 Z_5 應該改為 Z_6 比較好。

3. F 分配 2 個參數的自由度分別為 2 與 5。

(二)前提已知之 Y, Z 之分配，故有其動差生成函數如下：

$$Y \sim \chi^2(v_1) \iff M_Y(t)=(1-2t)^{-\frac{v_1}{2}},\ t<\frac{1}{2}$$

$$Z \sim \chi^2(v_3) \iff M_Z(t)=(1-2t)^{-\frac{v_3}{2}},\ t<\frac{1}{2}$$

$$(1-2t)^{-\frac{v_3}{2}} = M_Z(t) = M_{X+Y}(t)$$
$$= E(e^{t(X+Y)})$$
$$= E(e^{tX})E(e^{tY}) \ (\because X \perp Y)$$
$$= E(e^{tX})(1-2t)^{-\frac{v_1}{2}},\ t < \frac{1}{2}$$
$$\Rightarrow E(e^{tX}) = (1-2t)^{-\frac{v_3-v_1}{2}},\ t < \frac{1}{2}$$
$\Rightarrow\ X \sim \chi^2(v_3 - v_1)$，即 X 服從卡方分配自由度 $v_3 - v_1$

二、某製造公司設計一項試驗，以決定生產所需原料是以人工或自動方式裝填是否存在差異，以及兩台機器所生產之瑕疵品數目是否會影響瑕疵品數量。下表為生產的瑕疵品數量，已計算得知總平方和SST＝151.5。在顯著水準 α ＝5%下（右尾：$F_{0.05}(1,4)$＝7.71、$F_{0.05}(1,5)$＝6.61、$F_{0.05}(1,6)$＝5.99）：

	裝填方式	
	人工	自動
機器1	30	30
	34	26
機器2	20	24
	22	28

(一)檢定機器、裝填方式及他們的交互作用是否存在顯著效果？（15分）
(二)設若在此試驗設計中，兩台機器是設定為集區變數（Block），則裝填方式是否仍存在顯著效果？（10分）

答：
(一)

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y})^2 = 151.5$$

SS 機器 $= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{i..} - \bar{Y})^2 = 84.5$

SS 填裝方式 $= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{.j.} - \bar{Y})^2 = 0.5$

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{ij})^2 = 8+8+2+8 = 26$$

SSI=SST- SS 機器- SS 填裝方式-SSE=40.5

ANOVA 表

Source	*SS*	*df*	*MS*	*F*
機器	84.5	1	84.5	13
填裝方式	0.5	1	0.5	0.8
交互	40.5	1	40.5	6.23
誤差	26	4	6.5	
總合	151.5	7		

① H_0 : 機器和填裝方式無交互作用。

H_1 : 機器和填裝方式有交互作用。

②由 ANOVA 表可知檢定統計量 $F_I^* = 6.23$ 。

③ $\alpha = 0.05$, $C = \{F^* \mid F^* > F_{0.05}(1,4) = 7.71\}, F^* \notin C$ 檢定統計量不落入拒絕域。

④故在顯著水準 $\alpha = 0.05$ 情況下，根據樣本資料顯示，機器和填裝方式無交互作用。

(二)

① H_0 : 不同的填裝方式對瑕疵品數量無差異。

H_1 : 不同的填裝方式對瑕疵品數量有差異。

②由 ANOVA 表可知檢定統計量 $F^* = 0.8$ 。

③ $\alpha = 0.05$, $C = \{F^* \mid F^* > F_{0.05}(1,4) = 7.71\}, F^* \notin C$ 檢定統計量不落入拒絕域。

④故在顯著水準 $\alpha = 0.05$ 情況下，根據樣本資料顯示，不同的填裝方式對瑕疵品數量無差異。

三、一隨機變數Y=-2logX，其中隨機變數X具有混合型的機率密度函數如下：

f(x)=0.8　　0<x=<1，f(x)=0.2　　x=1

(一)求隨機變數Y的機率密度函數，並計算Y的中位數。(10分)

(二)求隨機變數Y的動差生成函數，並計算Y的平均數。(15分)

答：

(一)

① 離散部份 $P(X=1)P(Y=0)=0.2$

② 連續部份

$$Y=-2\ln X \;\Leftrightarrow\; x=e^{-\frac{y}{2}},\quad J=\left|\frac{dx}{dy}\right|=\left|-\frac{1}{2}e^{-\frac{y}{2}}\right|=\frac{1}{2}e^{-\frac{y}{2}} \qquad \{x:0<x<1\} \;\Leftrightarrow\; \{y:0<y<\infty\}$$

$$f_Y(y)=f_X(x=e^{-\frac{y}{2}})\,|J|=0.8\times\frac{1}{2}e^{-\frac{y}{2}}=0.4e^{-\frac{y}{2}},\; y>0$$

故隨機變數 Y 機率密度函數為 $f_Y(y)=\begin{cases}0.2 & ,\; y=0\\ 0.4e^{-\frac{y}{2}} & ,\; y>0\\ 0 & ,o.w.\end{cases}$

中位數 M ， $0.5=0.2+\int_0^M 0.4e^{-\frac{y}{2}}dy \;\Rightarrow\; \frac{3}{8}=\int_0^M \frac{1}{2}e^{-\frac{y}{2}}dy=1-e^{-\frac{M}{2}}$

$$\Rightarrow\; e^{-\frac{M}{2}}=\frac{5}{8} \;\Rightarrow\; M=-2\ln\left(\frac{5}{8}\right)=0.94$$

(二)

$$M_Y(t)=E(e^{tY})=0.2\times e^{t\times 0}+\int_0^\infty e^{ty}\,0.4e^{-\frac{y}{2}}dy$$

$$=0.2+0.8\int_0^\infty e^{ty}\frac{1}{2}e^{-\frac{y}{2}}dy$$

$$=0.2+\frac{0.8}{1-2t},\; t<\frac{1}{2}$$

$$M_Y'(t)=-\frac{0.8}{(1-2t)^2}\times(-2)=\frac{1.6}{(1-2t)^2}\text{ ， }E(Y)=M_Y'(t=0)=1.6$$

四、設 $X_1,X_2,\ldots,X_n$ 為一組隨機樣本服從母體X具機率密度函數 $f(x;\theta)=\theta_e^{-\theta x}$ ，$x>0$。

(一)證明 $T=\sum_{i=1}^{n}X_i$ 是參數 θ 的一個充分統計量。（10分）

(二)試求參數 θ 的最大概似估計量MLE（Maximum Likelihood Estimator）。（10分）

(三)若要檢定 H_0：$\theta=1$ 對應 H_1：$\theta=2$，依據Neyman-Pearson Lemma，試求檢定統計量及其顯著水準 α。（12分）

答：

(一) $X_i \overset{iid}{\sim} f_X(x;\theta)=\theta e^{-\theta x},\ x>0\ \ ,\theta\in\Omega=(0,\infty)$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x};\theta)=\prod_{i=1}^{n}f_{X_i}(x_i;\theta)=\prod_{i=1}^{n}[\theta e^{-\theta x_i}]=\theta^n e^{-\theta\sum x_i}$$

其中 $g(t;\theta)=\theta^n e^{-\theta\sum x_i}$ ， $h(\underset{\sim}{x})=1$

由分解定理可知 $T(\underset{\sim}{X})=\sum_{i=1}^{n}X_i$ 參數 θ 之充分統計量

(二)指數分配 $X_i \overset{iid}{\sim} f_X(x;\theta)=\theta e^{-\theta x},\ x>0\ \ ,\theta\in\Omega=(0,\infty)$

① $L(\theta)=\prod_{i=1}^{n}f_{X_i}(x_i;\theta)=\prod_{i=1}^{n}[\theta e^{-\theta x_i}]=\theta^n e^{-\theta\sum x_i}$

② $\ln L(\theta)=n\ln\theta-\theta\sum x_i$

③ $\frac{\partial}{\partial\theta}\ln L(\theta)=\frac{n}{\theta}-\sum x_i=0\ \Rightarrow\ \frac{n}{\theta}=\sum x_i\ \Rightarrow\ \theta=\frac{n}{\sum x_i}=\frac{1}{\bar{x}}$

④ $\frac{\partial^2}{\partial\theta^2}\ln L(\theta)=-\frac{n}{\theta^2}<0$, $\forall\theta>0\ \Rightarrow\ \hat{\theta}_{MLE}=\frac{1}{\bar{X}}$

(三) $X_i \overset{iid}{\sim} f_X(x;\theta)=\theta e^{-\theta x},\ x>0\qquad L(\theta)=\prod_{i=1}^{n}f_{X_i}(x_i;\theta)=\prod_{i=1}^{n}[\theta e^{-\theta x_i}]=\theta^n e^{-\theta\sum x_i}$

Neyman-Pearson Lemma 建構檢定統計量

$$\lambda(\underset{\sim}{x})=\frac{L(\theta=2)}{L(\theta=1)}=\frac{2^n e^{-2\sum x_i}}{1^n e^{-\sum x_i}}=2^n e^{-\sum x_i}>k$$

$$\Leftrightarrow\ -\sum x_i>\ln(k/2^n)\Leftrightarrow\ \sum x_i<-\ln(k/2^n)=c$$

由 N-P Lemma 可知 $\phi(\underset{\sim}{x})=\begin{cases}1\ ,\ \sum x_i<c\\ 0\ ,\ \sum x_i\ge c\end{cases}$ 為檢定 $H_0:\theta=1\ v.s.\ H_1:\theta=2$ 之 MPT。

其中檢定統計量 $\sum X_i\sim gamma(\alpha=n,\lambda=\theta)$。

顯著水準 $\alpha=P_{\theta=1}(\sum X_i<c)=P_{\theta=1}(2\sum X_i<2c)\ \Rightarrow\ \chi^2_{1-\alpha}(2n)=2c\ \Rightarrow\ c=\frac{\chi^2_{1-\alpha}(2n)}{2}$

H_0 為真下 $2\sum X_i\sim\chi^2(2n)$。