《統計學》

試題評析	今年度試題以統計學之基礎觀念爲主，沒有太難的數學運算，但是非常考驗考生在各單元中計算之熟悉度及解題之速度。第二大題與100年地特完全相同。本卷基本分70分。
考點命中	第一題：1.《高點統計學講義第一回》，趙治勳老師編撰，頁96，指數分配c.d.f.。 2.《高點統計學講義第二回》，趙治勳老師編撰，頁16，順序統計量抽樣分配。 3.《高點統計學講義第二回》，趙治勳老師編撰，頁29，不偏性。 第二題：《高點迴歸分析熱門題庫書籍》，趙治勳老師編著，頁1-92。 第三題：《高點統計學講義第二回》，趙治勳老師編撰，頁101。 第四題：1.《高點統計學講義第二回》，趙治勳老師編撰，頁5，抽樣分配之類題。 2.《高點統計學講義第二回》，趙治勳老師編撰，頁29，不偏性。 第五題：《高點統計學講義第二回》，趙治勳老師編撰，頁15，抽樣分配之類題。 第六題：1.《高點統計學講義第二回》，趙治勳老師編撰，頁90。 2.《高點統計學講義第二回》，趙治勳老師編撰，頁40。 第七題：《高點統計學講義第二回》，趙治勳老師編撰，頁83，題型二。

參考值：$t_{20,0.05}=1.7247$，$t_{22,0.05}=1.7171$，$t_{20,0.025}=2.086$，$t_{22,0.025}=2.0739$，
$\chi^2_{2,0.05}=5.9915$，$\chi^2_{3,0.05}=7.8147$，$\chi^2_{5,0.05}=11.0705$，$F_{2,10,0.05}=4.1028$，
$F_{3,12,0.05}=3.4903$，$F_{3,40,0.05}=2.8387$，$F_{3,45,0.05}=2.8115$，$F_{3,40,0.025}=3.4633$，
$F_{3,45,0.025}=3.4224$

一、從參數為λ的指數分配中，隨機抽出一組樣本（大小是5個），說明此樣本中位數並不是母體中位數的不偏估計量。（13分）

答：

母體：$X \sim Exp(\lambda)$

樣本：$X_1, X_2, \cdots, X_5 \overset{iid}{\sim} Exp(\lambda)$

令 $F_X(x)=P(X\le x)=1-e^{-\lambda x}=0.5 \Rightarrow x=-\dfrac{\ln(0.5)}{\lambda}$

故母體中位數 $Md=-\dfrac{\ln(0.5)}{\lambda}$

設 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(5)}$ 爲樣本之順序統計量

樣本中位數爲 $X_{(3)}$

$$g(x_{(3)})=\frac{5!}{2!2!}(1-e^{-\lambda x_{(3)}})^2[1-(1-e^{-\lambda x_{(3)}})]^2\lambda e^{-\lambda x_{(3)}}=30\lambda(1-e^{-\lambda x_{(3)}})^2(e^{-3\lambda x_{(3)}}),0<x_{(3)}$$

$$E(X_{(3)})=\int_0^\infty x_{(3)}g(x_{(3)})dx_{(3)}=\int_0^\infty x_{(3)}30\lambda(1-e^{-\lambda x_{(3)}})^2(e^{-3\lambda x_{(3)}})dx_{(3)}=\frac{47}{60\lambda}$$

$\because E(X_{(3)})=\dfrac{47}{60\lambda}\ne Md=-\dfrac{\ln(0.5)}{\lambda}$

$\therefore$ 樣本中位數不爲母體中位數之不偏估計量

二、若指數分配其參數為λ，當檢定H_0：$\lambda=\lambda_0$ vs. H_a：$\lambda>\lambda_0$，導出概似比檢定（likelihood ratio test），並以$\bar{x}$表示之。（13分）

答：

注意：題意是要利用GLR尋找拒絕域，而非Neyman-Pearson Lemma

假設 母體：$X \sim Exp(\lambda)$

樣本：$X_1, X_2, \cdots, X_n \overset{iid}{\sim} Exp(\lambda)$

令$\Theta = \{\lambda \mid 0<\lambda\}$

$\Theta_0 = \{\lambda \mid 0<\lambda=\lambda_0\}$

在Θ下，未知參數λ之MLE，$\hat{\lambda}=\dfrac{1}{\overline{X}}$

在Θ_0下，未知參數λ取值爲λ_0，$\lambda=\lambda_0$

$$令\lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\lambda_0^n e^{-\lambda_0\sum x_i}}{\frac{1}{\bar{x}^n} e^{-\frac{1}{\bar{x}}\sum x_i}} = \bar{x}^n \lambda_0^n e^{-\lambda_0\sum x_i + n} \le k$$

$$\Rightarrow (\lambda_0\bar{x})^n e^{-n(\lambda_0\bar{x})} \le k' \overset{令y=\lambda_0\bar{x}}{\Rightarrow} y^n e^{-ny} \le k''$$

$$\Rightarrow \bar{x} \le k'''$$

$$\alpha = P(\overline{X} \le c \mid \lambda=\lambda_0) = P(\chi^2_{(2n)} \le 2n\lambda_0 c \mid \lambda=\lambda_0) \text{ 其中} 2\lambda_0 n\overline{X} \overset{H_0 isture}{\sim} \chi^2_{(2n)}$$

$$\Rightarrow 2n\lambda_0 c = \chi^2_{1-\alpha(2n)} \Rightarrow c = \frac{\chi^2_{1-\alpha(2n)}}{2n\lambda_0}$$

$$\therefore C = \left\{\bar{x} \,\middle|\, \bar{x} \le \frac{\chi^2_{1-\alpha(2n)}}{2n\lambda_0}\right\}$$ 爲顯著水準α下$H_0:\lambda\le\lambda_0$ vs. $H_1:\lambda>\lambda_0$之LRT

三、王研究員想研究3種廠牌咖啡的口味是否有顯著差異，找了6位品評員做測試，每位都隨機次序品嚐這3種廠牌的咖啡，評分結果如下：

品評員	1	2	3	4	5	6
廠牌A	8	7	4	5	7	6
廠牌B	5	3	4	4	3	4
廠牌C	4	6	3	2	5	1

若對廠牌A、B及C評分，是服從常態分配，且變異數都相等。在$\alpha=0.05$，試問廠牌A、B及C評分是否有相同的平均數？（15分）

答：

假設模型：$y_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}$,$\varepsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2)$

ANOVA TABLE				
Source	SS	d.f.	MS	F
廠牌	25.333	2	12.6665	$F_1^*=7.0369$
品評員	13.167	5	2.6334	$F_2^*=1.463$
Error	18	10	1.8	
Total	56.5	17		

$H_0:\mu_1=\mu_2=\mu_3$ vs H_1:至少一個$\mu_i\neq\mu_j$

T.S.: $F_1=\dfrac{MSA}{MSE}\sim F_{(2,10)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $F_1^*>F_{0.05(2,10)}=4.1028$

$\because F_1^*=7.0369$ $\quad\therefore reject H_0$

我們有足夠證據去推論廠牌A,B,C之平均評分不盡相同

四、從一母體為{1, 2, 3, 4}，隨機抽出n＝2個為樣本。每抽一個元素為1、3或4的機率各為0.2，抽中2的機率為0.4，採用取出放回的抽樣法。設$\hat{\tau}=\frac{1}{n}\sum_{i=1}^{n}y_i/\delta_i$，其中$y_i$為第i次抽中的元素，$\delta_i$為抽中的元素$y_i$的機率。詳細說明$E(\hat{\tau})=\tau$（$\tau$為母體元素總和）是否成立？（15分）

答：

$\tau=1+2+3+4=10$

$Y_1(\delta_1)$	$Y_2(\delta_2)$	$\hat{\tau}=\frac{1}{2}\sum_{i=1}^{2}\frac{Y_i}{\delta_i}$	機率
1(0.2)	1(0.2)	5	0.04
2(0.4)	2(0.4)	5	0.16
3(0.2)	3(0.2)	15	0.04
4(0.2)	4(0.2)	20	0.04
1(0.2)	2(0.4)	5	0.16
1(0.2)	3(0.2)	10	0.08
1(0.2)	4(0.2)	12.5	0.08
2(0.4)	3(0.2)	10	0.16
2(0.4)	4(0.2)	12.5	0.16
3(0.2)	4(0.2)	17.5	0.08

$\hat{\tau}$	5	10	12.5	15	17.5	20
機率	0.36	0.24	0.24	0.04	0.08	0.04

$E(\hat{\tau})=5(0.36)+10(0.24)+\cdots+20(0.04)=10=\tau$

五、若X_1、X_2、…、X_5是服從 $N(0,\sigma^2)$ 且互相獨立。求k值，使得 $\dfrac{k(X_1-X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$ 為一個t分配，並求此t分配的自由度為何？（9分）

答：

$$X_1-X_2\sim N(0,2\sigma^2)\Rightarrow \frac{X_1-X_2}{\sqrt{2}\sigma}\sim N(0,1)$$

$$\frac{X_i^2}{\sigma^2}\overset{iid}{\sim}\chi^2_{(1)},i=3,4,5$$

$$\frac{X_3^2}{\sigma^2}+\frac{X_4^2}{\sigma^2}+\frac{X_5^2}{\sigma^2}=\frac{X_3^2+X_4^2+X_5^2}{\sigma^2}\sim\chi^2_{(3)}$$

$$\frac{\dfrac{X_1-X_2}{\sqrt{2}\sigma}}{\sqrt{\dfrac{\dfrac{X_3^2+X_4^2+X_5^2}{\sigma^2}}{3}}}=\frac{\sqrt{3}}{\sqrt{2}}\frac{X_1-X_2}{\sqrt{X_3^2+X_4^2+X_5^2}}\sim t_{(3)}$$

故$k=\dfrac{\sqrt{3}}{\sqrt{2}}$，自由度爲3

六、有一研究員從4家類似的公司，隨機各抽取11位員工進行考試，所得考試分數結果之ANOVA表如下：

變異來源	平方和	自由度	均方和	F	p－值
公司	1154	(a)		(c)	(d)
誤差			109.475		
總和		(b)			

(一)計算出(a)、(b)、(c)及(d)之數值。（8分）

(二)若其中兩家公司之考試資料分為x與y，且$\sum_{i=1}^{11}x_i=630$，$\sum_{i=1}^{11}y_i=592$，$\sum_{i=1}^{11}(x_i-\bar{x})^2=998.2$，$\sum_{i=1}^{11}(y_i-\bar{y})^2=779.6$。假設$\alpha=0.05$，檢定兩家公司員工之平均考試分數是否相同？假設兩家員工考試資料的變異數相等。（8分）

(三)以上(二)與(一)之結果，是否有矛盾的現象？（4分）

答

(一) (a)4-1=3 (b)44-1=43 (c)384.667/109.475=3.5137

(d) $P(F_{(3,40)}>3.5137)\approx 0.025$

(二)

由題目可得，$\bar{x}=57.2727, \bar{y}=53.8181, s_X^2=\frac{998.2}{11-1}=99.82, s_Y^2=\frac{779.6}{11-1}=77.96$

母體：$X \sim N(\mu_X,\sigma^2) \perp Y \sim N(\mu_Y,\sigma^2)$ 假設(1) $X \perp Y$ (2) $X \sim N, Y \sim N$

樣本：$X_1, X_2, \cdots, X_{11} \overset{iid}{\sim} N(\mu_X,\sigma^2)$，$Y_1, Y_2, \cdots, Y_{11} \overset{iid}{\sim} N(\mu_Y,\sigma^2)$

點估計：$\overline{X}-\overline{Y} \sim N(\mu_X-\mu_Y, \sigma^2(\frac{1}{11}+\frac{1}{11}))$

$H_0: \mu_X=\mu_Y$ vs $H_1: \mu_X \neq \mu_Y$

T.S.：$T=\frac{\overline{X}-\overline{Y}-(0)}{\sqrt{S_p{}^2(\frac{1}{11}+\frac{1}{11})}} \sim t_{(11+11-2=20)}$ 其中 $S_p^2=\frac{(11-1)S_1^2+(11-1)S_2^2}{11+11-2}=88.89$

R.R.：Reject H_0 at $\alpha=0.05$ if $|T^*|>t_{0.025(20)}=2.086$

$\because T^*=0.8593 \qquad \therefore don't\,reject H_0$

我們沒有足夠證據去推論兩家公司員工之平均考試分數不相同

(三)

(一)(二)之結論沒有存在矛盾之現象，在(一)中可推論四家公司員工之平均考試分數不盡相同，「不盡相同」不表示其中兩家公司員工之平均考試分數不可以相同，故(二)中，推論出其中兩家公司員工之平均考試分數是相同的，沒有存在矛盾。

七、有一機率密度函數 $f(y)=1/\theta$，$0 \le y \le \theta$。在顯著水準 $\alpha=0.10$ 之下，欲檢定 H_0：$\theta=2.0$ vs. H_a：$\theta<2.0$，所採用檢定統計量為 $Y_{(8)}=\max\{y_1,y_2,...,y_8\}$，當 $\theta=1.7$ 時，犯型二誤差（type II error）的機率為多少？（15分）

答

母體：$Y \sim Uniform(0,\theta)$

樣本：$Y_1, Y_2, \cdots, Y_8 \overset{iid}{\sim} Uniform(0,\theta)$

令 $Z=\frac{Y_{(8)}}{\theta} \sim Beta(8,1)$

$H_0: \theta=2$ vs $H_1: \theta<2$

$$\alpha=0.1=P(Y_{(8)} \le c \mid \theta=2)=P(Z \le \frac{c}{2})=\int_0^{\frac{c}{2}} 8z^7 dz=(\frac{c}{2})^8 \Rightarrow c=2\sqrt[8]{0.1}=1.5$$

$$\beta=P(Y_{(8)}>1.5 \mid \theta=1.7)=P(Z>0.8824)=\int_{0.8824}^{1} 8z^7 dz=0.6324$$