《統計學》

試題評析	今年試題多是數統範圍，經建考生作答時一定非常吃力。但題目並沒有出現繁瑣的數學運算，有準備過數統範圍的考生，應該可以拿到75分以上；如果放棄數統範圍，本卷成績可能只有30~40分。
高分命中	第一題：《高點統計學上課講義第二回》，趙治勳編撰，頁46、頁76。 第二題：《高點迴歸分析熱門題庫書籍》，趙治勳編著，頁1-81例9。 第三題：《高點統計學上課講義第二回》，趙治勳編撰，第8.5節。 第四題：《高點迴歸分析熱門題庫書籍》，趙治勳編著，頁1-56。 《高點統計學上課講義第二回》，趙治勳編撰，頁24、頁21〈註3〉。

※本試題可能使用之標準常態之如下：

$z_{0.025}=1.96$，$z_{0.05}=1.645$，$z_{0.1}=1.28$

$t_{0.05}(15)=1.753$，$t_{0.025}(15)=2.131$，$t_{0.05}(16)=1.746$，$t_{0.025}(16)=2.120$

一、在農業試驗之某種果樹可能生出既大又酸的水果，也可能生出雖小又甜的水果。若此種果樹長出既大又酸的水果之機率為 p_1 及長出雖小又甜的水果之機率為 p_2。令 X 為此試驗種出 $n=30$ 棵果樹中長出既大又酸水果之棵數。但已知由 X 估計 p_1 的 $100(1-\alpha)\%$ 信賴區間為 $(0.4,0.7)$，試求：

（一）p_2 的 $100(1-\alpha)\%$ 信賴區間。（15分）

（二）當顯著水準為 $\alpha\in(0,1)$，但未知時，請檢定 H_0：$p_2=0.4$ vs. H_1：$p_2\neq 0.4$。（10分）

答：本題需要假設隨機樣本

(一)由題意，$(\hat{p}_1\pm z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{30}})=(0.4,0.7)$

可得知 $\hat{p}_1=0.55$，$z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{30}}=0.15$

又因爲 $\hat{p}_2=1-\hat{p}_1=0.45$ 及 $z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{30}}=z_{\frac{\alpha}{2}}\sqrt{\dfrac{(1-\hat{p}_1)\hat{p}_1}{30}}=0.15$

故 p_2 之 $100(1-\alpha)\%$ 信賴區間爲 $(\hat{p}_2\pm z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{30}})=(0.3,0.6)$

(二)信賴區間檢定法

由於 p_2 之 $100(1-\alpha)\%$ 信賴區間爲 $(0.3,0.6)$ 有包含 H_0 爲真下之 $p_2=0.4$，故在顯著水準爲 α 下，我們沒有足夠證據去推論長出雖小又甜的水果之機率不爲0.4。

二、中部地區平均每公畝生產A水果5單位（每單位為100公斤）。近年因氣候異常，令人質疑每公畝平均產量是否仍為5單位。

（一）若隨機樣本 $X_1,X_2,\ldots,X_n$ 具有常態分配 $N(\mu,1)$，則請推導顯著水準為 α 之概似比（Likelihood Ratio）檢定。（15分）

（二）若顯著水準 $\alpha=0.05$、樣本數 $n=25$ 且樣本平均 $\bar{x}=4$，則每公畝平均產量是否仍為5單位？（10分）

答：

(一) $X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu,1)$

$H_0: \mu = 5$ vs $H_1: \mu \neq 5 (=\mu_1)$

$$令\ \lambda = \frac{L(\mu=5)}{L(\mu=\mu_1)} = \frac{(\frac{1}{\sqrt{2\pi}})^n e^{-\frac{\sum(x_i-5)^2}{2}}}{(\frac{1}{\sqrt{2\pi}})^n e^{-\frac{\sum(x_i-\mu_1)^2}{2}}} = e^{-\frac{2(\mu_1-5)\sum x_i + n(5^2-\mu_1^2)}{2}} \le k$$

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

$\alpha = P(\bar{X} \le c_1 or \bar{X} \ge c_2 \mid \mu = 5)$ 其中 $\bar{X} \overset{H_0 is true}{\sim} N(5, \frac{1}{n})$

$\Rightarrow c_1 = 5 - z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}}$，$c_2 = 5 + z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}}$

$\therefore C = \{\bar{x} \mid \bar{x} \le 5 - z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}} or \bar{x} \ge 5 + z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}}\}$ 爲在檢定大小 α 下 $H_0: \mu = 5$ vs $H_1: \mu \neq 5$ 之LRT

(二) R.R.：$C = \{\bar{x} \mid \bar{x} \le 5 - z_{0.025} \frac{1}{\sqrt{25}} = 4.608 or \bar{x} \ge 5 + z_{0.025} \frac{1}{\sqrt{25}} = 5.392\}$

$\because \bar{x} = 4$ $\qquad \therefore don't reject H_0$

我們沒有足夠證據去推論平均產量不爲5單位。

三、若 $X_1, X_2, \ldots, X_n$ 為抽自 $N(\mu, \sigma^2)$ 之隨機樣本，且 X_{n+1}, X_{n+2} 為獨立並具有相同分配。盼以 $X_1, X_2, \ldots, X_n$ 來預測 $Y = X_{n+1} + X_{n+2}$，則：

(一)請推導 Y 的 $100(1-\alpha)\%$ 信賴區間。（15分）

(二)若已知 $n=16$，$s^2 = \frac{1}{15}\sum_{i=1}^{16}(X_i - \bar{x})^2 = 1.44$ 及 $\bar{x} = \frac{1}{16}\sum_{i=1}^{16} X_i = 2$，試求 Y 的95%信賴區間。（10分）

答：

(一) $X_1, X_2, \cdots, X_n, X_{n+1}, X_{n+2} \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\sum_{i=1}^{n} X_i - \frac{n}{2}Y \sim N(0, (n + \frac{n^2}{2})\sigma^2)$$

$$\frac{(\sum_{i=1}^{n} X_i - \frac{n}{2}Y) - (0)}{\sqrt{(n + \frac{n^2}{2})\sigma^2}} \sim N(0,1)$$

得知，$\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

樞紐量：$\dfrac{\dfrac{(\sum_{i=1}^{n} X_i - \frac{n}{2}Y) - (0)}{\sqrt{(n+\frac{n^2}{2})\sigma^2}}}{\sqrt{\dfrac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \dfrac{\sum_{i=1}^{n} X_i - \frac{n}{2}Y}{\sqrt{(n+\frac{n^2}{2})S^2}} \sim t_{(n-1)}$

機率區間：$P(-t_{\frac{\alpha}{2}(n-1)} \le \dfrac{\sum_{i=1}^{n} X_i - \frac{n}{2}Y}{\sqrt{(n+\frac{n^2}{2})S^2}} \le t_{\frac{\alpha}{2}(n-1)}) = 1-\alpha$

信賴區間：

$$P(\frac{2[\sum_{i=1}^{n} X_i - t_{\frac{\alpha}{2}(n-1)}\sqrt{(n+\frac{n^2}{2})S^2}\,]}{n} \le Y \le \frac{2[\sum_{i=1}^{n} X_i + t_{\frac{\alpha}{2}(n-1)}\sqrt{(n+\frac{n^2}{2})S^2}\,]}{n}) = 1-\alpha$$

結論：Y 之$100(1-\alpha)\%$ 信賴區間為

$$(\frac{2[\sum_{i=1}^{n} X_i - t_{\frac{\alpha}{2}(n-1)}\sqrt{(n+\frac{n^2}{2})S^2}\,]}{n}, \frac{2[\sum_{i=1}^{n} X_i + t_{\frac{\alpha}{2}(n-1)}\sqrt{(n+\frac{n^2}{2})S^2}\,]}{n})$$

(二) $(\dfrac{2[32-2.131\sqrt{(16+\frac{16^2}{2})1.44}\,]}{16}, \dfrac{2[32+2.131\sqrt{(16+\frac{16^2}{2})1.44}\,]}{16})$

$= (0.1642, 7.8358)$

四、假設 $X_1, X_2, \ldots, X_n$ 為一組隨機樣本具有如下分配：

$f(x,\theta) = \dfrac{1}{\theta^2} x e^{-\frac{x}{\theta}}$ ，$x > 0$ 。

(一)請找出 θ 的Cramer-Rao下限。（9分）

(二)請找出 θ 的最大概似估計量（mle）。（8分）

(三)請由mle之極限分配導出 θ 的 $100(1-\alpha)\%$ 信賴區間。（8分）

答：

$X_1, X_2, \cdots, X_n \overset{iid}{\sim} Gamma(\alpha = 2, \lambda = \dfrac{1}{\theta})$

(一) $\ln f_X(x\,|\,\theta) = -2\ln\theta + \ln x - \dfrac{x}{\theta}$

$\dfrac{d \ln f_X(x\,|\,\theta)}{d\theta} = -\dfrac{2}{\theta} + \dfrac{x}{\theta^2}$

$\dfrac{d^2 \ln f_X(x\,|\,\theta)}{d\theta^2} = \dfrac{2}{\theta^2} - \dfrac{2x}{\theta^3} = \dfrac{2(\theta - x)}{\theta^3}$

$$CRLB(\theta)=\frac{[\tau'(\theta)]^2}{-nE[\frac{\partial^2}{\partial\theta^2}\ln f(X\mid\theta)]}=\frac{1}{-nE(\frac{2(\theta-X)}{\theta^3})}$$

$$=\frac{1}{-n(\frac{2(\theta-2\theta)}{\theta^3})}=\frac{\theta^2}{2n}$$

(二) $L(\theta)=\frac{1}{\theta^{2n}}\prod_{i=1}^{n}x_i e^{-\frac{\sum_{i=1}^{n}x_i}{\theta}}$

$$\ln L(\theta)=-2n\ln\theta+\prod_{i=1}^{n}x_i-\frac{\sum_{i=1}^{n}x_i}{\theta}$$

令 $\frac{d\ln L(\theta)}{d\theta}=\frac{-2n}{\theta}+\frac{\sum_{i=1}^{n}x_i}{\theta^2}=0$

$$\Rightarrow\hat{\theta}_{MLE}=\frac{\sum_{i=1}^{n}x_i}{2n}$$

因此，$\hat{\theta}_{MLE}=\frac{\sum_{i=1}^{n}X_i}{2n}$ 為 θ 之最大概似估計量。

(三) $\hat{\theta}_{MLE}=\frac{\sum_{i=1}^{n}X_i}{2n}\overset{d}{\sim}N(\theta,\frac{\theta^2}{2n})$

$$\frac{\hat{\theta}_{MLE}-\theta}{\frac{\theta}{\sqrt{2n}}}=\sqrt{2n}(\frac{\hat{\theta}_{MLE}}{\theta}-1)\overset{d}{\sim}N(0,1)$$

樞紐量：$\sqrt{2n}(\frac{\hat{\theta}_{MLE}}{\theta}-1)\overset{d}{\sim}N(0,1)$

機率區間：$P(-z_{\frac{\alpha}{2}}\le\sqrt{2n}(\frac{\hat{\theta}_{MLE}}{\theta}-1)\le z_{\frac{\alpha}{2}})=1-\alpha$

信賴區間：$P(\frac{\hat{\theta}_{MLE}}{z_{\frac{\alpha}{2}}\frac{1}{\sqrt{2n}}+1}\le\theta\le\frac{\hat{\theta}_{MLE}}{-z_{\frac{\alpha}{2}}\frac{1}{\sqrt{2n}}+1})=1-\alpha$

結論：θ 之 $100(1-\alpha)\%$ 信賴區間為

$$\left(\frac{\hat{\theta}_{MLE}}{z_{\frac{\alpha}{2}}\frac{1}{\sqrt{2n}}+1},\frac{\hat{\theta}_{MLE}}{-z_{\frac{\alpha}{2}}\frac{1}{\sqrt{2n}}+1}\right)$$