《統計學》

試題評析	今年考題難度適中，但是題目之小題多且計算流程繁瑣，非常考驗考生之耐力與作答速度。本卷基本分60分，數理程度較佳的考生若能夠應付第一大題的話，本卷應該可拿到90分以上。
高分命中	第一題：(一)《高點統計學上課講義第二回》，趙治勳編撰，頁14，7.5節。 (二)《高點統計學上課講義第一回》，趙治勳編撰，頁65，4.4節。 (三)《高點統計學上課講義第一回》，趙治勳編撰，頁58。 (五)《高點統計學上課講義第二回》，趙治勳編撰，頁14<註>。 第二題：(一)《高點統計學上課講義第二回》，趙治勳編撰，頁19，例1。 (二)《高點統計學上課講義第一回》，趙治勳編撰，頁24，例1。 (三)至(五)《高點統計學上課講義第二回》，趙治勳編撰，頁26，例1。 (六)(七)《高點統計學上課講義第二回》，趙治勳編撰，頁46。 (八)(九)《高點統計學上課講義第二回》，趙治勳編撰，頁76。 第三題：《高點統計學上課講義》，第二回，趙治勳編撰，頁37及65。

一、若$X_1, X_2, X_3,\ldots, X_n$係隨機抽自同質分配（Uniform Distribution）母體之隨機樣本（Random Sample），其機率密度函數（Probability Density Function）為

$$f_X(x)=1 \qquad \forall x\in[\theta-1/2,\ \ \theta+1/2]$$
$$=0 \qquad \text{其他}$$

且定義$Y_1=\min\{X_1, X_2, X_3,\ldots, X_n\}$與$Y_n=\max\{X_1, X_2, X_3,\ldots, X_n\}$，則

(一)試求隨機變數Y_1與Y_n之聯合機率密度函數（Joint Probability Density Function）。（5分）

(二)試求隨機變數$R=Y_n-Y_1$與$M=(Y_1+Y_n)/2$之聯合機率密度函數。（6分）

(三)試分別求隨機變數R與M之邊際機率密度函數（Marginal Probability Density Function）。（10分）

(四)試問與MR是否為θ之輔助統計式（Ancillary Statistic）？簡要說明為什麼？（4分）

(五)試分別求隨機變數R與M之期望值（Expected Value）與變異數（Variance）。（15分）

答：

$$F_X(x)=P(X\le x)=\int_{\theta-\frac{1}{2}}^{x}1dt=x-\theta+\frac{1}{2},\ \ \theta-\frac{1}{2}<x<\theta+\frac{1}{2}$$

(一) $$f_{Y_1Y_n}(y_1,y_n)=\frac{n!}{0!(n-2)!0!}[(y_n-\theta+\frac{1}{2})-(y_1-\theta+\frac{1}{2})]^{n-2}\times1\times1$$
$$=n(n-1)(y_n-y_1)^{n-2},\ \ \theta-\frac{1}{2}<y_1\le y_n<\theta+\frac{1}{2}$$

(二) $$\begin{matrix}R=Y_n-Y_1\\ M=\dfrac{Y_1+Y_n}{2}\end{matrix}\ \Rightarrow\ \begin{matrix}Y_1=\dfrac{2M-R}{2}\\ Y_n=\dfrac{2M+R}{2}\end{matrix}$$

$$J=\begin{vmatrix}1 & -\frac{1}{2}\\ 1 & \frac{1}{2}\end{vmatrix}=(1)(\frac{1}{2})-(-\frac{1}{2})(1)=1$$

$$f_{RM}(r,m)=n(n-1)(\frac{2m+r}{2}-\frac{2m-r}{2})^{n-2}\,|1|=n(n-1)r^{n-2}$$

$$,\theta-\frac{1}{2}<\frac{2m-r}{2}\le\frac{2m+r}{2}<\theta+\frac{1}{2}$$

(三) $$f_M(m)=\begin{cases}\int_0^{2m-2\theta+1}n(n-1)r^{n-2}dr=n(2m-2\theta+1)^{n-1},\theta-\frac{1}{2}<m<\theta\\ \int_0^{2\theta-2m+1}n(n-1)r^{n-2}dr=n(2\theta-2m+1)^{n-1},\theta<m<\theta+\frac{1}{2}\\ 0 \qquad ,o.w.\end{cases}$$

$$f_R(r)=\int_{\theta-\frac{1}{2}+\frac{r}{2}}^{\theta+\frac{1}{2}-\frac{r}{2}}n(n-1)r^{n-2}dm=n(n-1)r^{n-2}[(\theta+\frac{1}{2}-\frac{r}{2})-(\theta-\frac{1}{2}+\frac{r}{2})]$$

$$=n(n-1)r^{n-2}(1-r),\ 0<r<1 \qquad \therefore R\sim Beta(n-1,2)$$

(四) R 爲 θ 之輔助統計量，因爲 $R=Y_n-Y_1$ 與 θ 無關且其分配也與 θ 無關。

M 不爲 θ 之輔助統計量，因爲雖然 $M=\frac{Y_1+Y_n}{2}$ 與 θ 無關但其分配與 θ 無關的。

(五) $R=Y_n-Y_1\sim Beta(n-1,2)$

$$E(R)=\frac{n-1}{n+1},\ V(R)=\frac{2(n-1)}{(n+2)(n+1)^2}$$

又 $Y_1-\theta+\frac{1}{2}\sim Beta(1,n)$, $Y_n-\theta+\frac{1}{2}\sim Beta(n,1)$

$$E(M)=E(\frac{Y_1+Y_n}{2})=\frac{E(Y_1)+E(Y_n)}{2}=\frac{\frac{1}{n+1}+\frac{n}{n+1}}{2}=\frac{1}{2}$$

$$V(M)=V(\frac{Y_1+Y_n}{2})=\frac{V(Y_1)+V(Y_n)+2Cov(Y_1,Y_n)}{4}$$

$$=\frac{\frac{n}{(n+2)(n+1)^2}+\frac{n}{(n+2)(n+1)^2}+2\frac{1}{(n+2)(n+1)^2}}{4}=\frac{1}{2(n+2)(n+1)}$$

其中 $V(R)=V(Y_n-Y_1)=V(Y_1)+V(Y_n)-2Cov(Y_1,Y_n)=\frac{2(n-1)}{(n+2)(n+1)^2}$

$$\Rightarrow -2Cov(Y_1,Y_n)=\frac{2(n-1)}{(n+2)(n+1)^2}-V(Y_1)-V(Y_n)$$

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

$$\Rightarrow Cov(Y_1,Y_n)=\frac{1}{(n+2)(n+1)^2}$$

二、若考慮ABC地區適婚男性未婚之比率p，遂於該區之適婚男性抽出（$X_1, X_2, X_3,\ldots, X_n$）為隨機樣本，且（$x_1, x_2, x_3,\ldots, x_n$）為觀測樣本（Observed Sample），則

(一)請用動差法（The Method of Moments）求適婚男性未婚比率p之點估計式（Point Estimator）。（3分）

(二)請用最適法（The Method of Maximum Likelihood Estimation）求適婚男性未婚比率p之點估計式。（7分）

(三)證明該點估計式為不偏估計式（Unbiased Estimator）。（2分）

為瞭解該區適婚男性未婚之詳情，遂於該區之適婚男性隨機抽出300人為樣本，發現適婚男性未婚者81人，則

(四)請列適婚男性未婚比率p之估計式（Estimator）。（1分）

(五)請計算適婚男性未婚比率p之估計值（Estimates）。（2分）

若相關母體為常態，且顯著水準為（The Level of Significance）5%，則

(六)請列適婚男性未婚比率p之信賴區間（Confidence Interval）估計式。（1分）

(七)請計算適婚男性未婚比率p之信賴區間估計值。（4分）

(八)請問適婚男性未婚之比率p是否為0.3？（5分）

(九)請問適婚男性未婚之比率p是否低於0.3？（5分）

答：

令 X 表適婚男性而未婚者

母體: $X \sim Ber(p)$

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

(一) 令 $E(X)=p=\overline{X}$，可得 $\hat{p}_{MME}=\overline{X}$

(二) $L(p)=p^{\sum x_i}(1-p)^{n-\sum x_i}$

$$\ln L(p)=\sum x_i \ln p \ +(n-\sum x_i)\ln(1-p)$$

$$令\ \frac{d\ln L(p)}{dp}=\frac{\sum x_i}{p}-\frac{n-\sum x_i}{1-p}=0$$

$$\hat{p}_{MLE}=\frac{\sum x_i}{n}=\bar{x}$$

$\therefore\ \hat{p}_{MLE}=\overline{X}$ 爲 p 之MLE

(三) $\because E(\hat{p}_{MME})=E(\hat{p}_{MLE})=E(\overline{X})=\dfrac{np}{n}=p$

$\therefore \overline{X}$ 爲 p 之不偏估計式

樣本: $X_1, X_2, \cdots, X_{300} \overset{iid}{\sim} Ber(p)$

(四) $\overline{X}=\dfrac{\sum_{i=1}^{300} X_i}{300}$

(五) $\bar{x} = \dfrac{81}{300} = 0.27$

(六) 點估計: $\hat{p} = \dfrac{\sum X_i}{300} \underset{byC.L.T.}{\sim} N(p, \dfrac{p(1-p)}{300})$

樞紐量: $Z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{300}}} \underset{byC.L.T.}{\sim} N(0,1)$

機率區間: $P(-z_{0.025} \le \dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{300}}} \le z_{0.025}) = 0.95$

信賴區間: $P(\hat{p} - z_{0.025}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{300}} \le p \le \hat{p} + z_{0.025}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{300}}) = 0.95$

結論: p 之95%信賴區間

$$(\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{300}}, \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{300}})$$

(七) p 之95%信賴區間為 $(0.27 - 1.96\sqrt{\dfrac{0.27(0.73)}{300}}, 0.27 + 1.96\sqrt{\dfrac{0.27(0.73)}{300}})$

$= (0.2198, 0.3202)$

(八) $H_0: p = 0.3 vs H_1: p \ne 0.3$

T.S.: $Z = \dfrac{\hat{p}-0.3}{\sqrt{\dfrac{0.3(1-0.3)}{300}}} \underset{byC.L.T.}{\sim} N(0,1)$

R.R.: Reject H_0 at $\alpha = 0.05$ if $|Z^*| > z_{0.025} = 1.96$

$\because |Z^*| = |\dfrac{0.27-0.3}{\sqrt{\dfrac{0.3(1-0.3)}{300}}}| = 1.134 \qquad \therefore don't reject H_0$

我們沒有足夠證據去推論適婚男性未婚之比率不為0.3

(九) $H_0: p \ge 0.3 vs H_1: p < 0.3$

T.S.: $Z = \dfrac{\hat{p}-0.3}{\sqrt{\dfrac{0.3(1-0.3)}{300}}} \underset{byC.L.T.}{\sim} N(0,1)$

R.R.: Reject H_0 at $\alpha = 0.05$ if $Z^* < -z_{0.05} = -1.645$

$\because Z^* = \dfrac{0.27-0.3}{\sqrt{\dfrac{0.3(1-0.3)}{300}}} = -1.134 \qquad \therefore don't reject H_0$

我們沒有足夠證據去推論適婚男性未婚之比率低於0.3

三、為研究試驗個體間之抽樣變異（Sampling Variability），遂於參加SAT考試之學生中隨機抽出分

數分別為580、585、590、595與600之5位男學生與分數分別為540、545、550、555與560之5位女學生為樣本1。又於參加SAT考試之學生中隨機抽出分別為530、560、590、620與650之5位男學生與分數分別為490、520、550、580與610之5位女學生為樣本2。若相關母體為常態，且顯著水準（The Level of Significance）為5%，則請以均數差之推論或變異數分析（ANOVA）等統計方法分析前述兩樣本試驗個體間之抽樣變異，但至少須考慮並回答下列問題：

(一)請問各樣本男學生與女學生之平均分數是否不同？（8分）

(二)請問各樣本男學生之平均分數是否比女學生之平均分數高30分但不高於35分？（8分）

(三)試問各樣本男學生與女學生分數之相關係數是否大於0？（5分）

(四)試問各樣本男學生與女學生分數之相關係數是否超過0.8？（5分）

(五)簡要說明你的發現與理由。（4分）

答：

		男學生	女學生
樣本1	樣本數	5	5
	樣本平均數	590	550
	樣本變異數	62.5	62.5
	樣本相關係數(假設成對抽樣)	1	
樣本2	樣本數	5	5
	樣本平均數	590	550
	樣本變異數	2250	2250
	樣本相關係數(假設成對抽樣)	1	

令 X_i, Y_i 分別表樣本 i 之男學生與女學生之SAT分數, $i=1,2$

母體: $X_i \sim N(\mu_{Xi},\sigma_{Xi}^2) \perp Y_i \sim N(\mu_{Yi},\sigma_{Yi}^2)$ 假設 $X_i \perp Y_i$

樣本: $X_{i1},\cdots,X_{i5} \overset{iid}{\sim} N(\mu_{Xi},\sigma_{Xi}^2) \perp Y_{i1},\cdots,Y_{i5} \overset{iid}{\sim} N(\mu_{Yi},\sigma_{Yi}^2)$

(一)由各樣本之樣本變異數得知, $H_0:\sigma_{Xi}^2=\sigma_{Yi}^2$ 均爲不拒絕

$H_0:\mu_{X1}=\mu_{Y1} vs H_1:\mu_{X1}\neq\mu_{Y1}$

T.S.: $T=\dfrac{\overline{X}_1-\overline{Y}_1-(0)}{\sqrt{S_p^2(\frac{1}{5}+\frac{1}{5})}} \sim t_{(8)}$ 其中 $S_p^2=\dfrac{(5-1)S_{X1}^2+(5-1)S_{Y1}^2}{5+5-2}=62.5$

R.R.: Reject H_0 at $\alpha=0.05$ if $|T^*|>t_{0.025(8)}=2.306$

$\because T^*=\dfrac{590-550-(0)}{\sqrt{62.5(\frac{1}{5}+\frac{1}{5})}}=8$ $\quad\therefore reject H_0$

我們有足夠證據去推論樣本1之男學生與女學生之平均分數爲不同

$H_0:\mu_{X2}=\mu_{Y2} vs H_1:\mu_{X2}\neq\mu_{Y2}$

T.S.: $T=\dfrac{\overline{X}_2-\overline{Y}_2-(0)}{\sqrt{S_p^2(\frac{1}{5}+\frac{1}{5})}} \sim t_{(8)}$ 其中 $S_p^2=\dfrac{(5-1)S_{X1}^2+(5-1)S_{Y1}^2}{5+5-2}=2250$

R.R.: Reject H_0 at $\alpha = 0.05$ if $|T^*| > t_{0.025(8)} = 2.306$

$$\because T^* = \frac{590-550-(0)}{\sqrt{2250(\frac{1}{5}+\frac{1}{5})}} = 1.333 \qquad \therefore don't reject H_0$$

我們沒有足夠證據去推論樣本1之男學生與女學生之平均分數為不同

(二) $H_0 : \mu_{X1} - \mu_{Y1} \le 30 \, vs \, H_1 : \mu_{X1} - \mu_{Y1} > 30$

T.S.: $T = \dfrac{\overline{X}_1 - \overline{Y}_1 - (30)}{\sqrt{S_p^2(\frac{1}{5}+\frac{1}{5})}} \sim t_{(8)}$ 其中 $S_p^2 = \dfrac{(5-1)S_{X1}^2 + (5-1)S_{Y1}^2}{5+5-2} = 62.5$

R.R.: Reject H_0 at $\alpha = 0.05$ if $T^* > t_{0.05(8)} = 1.86$

$$\because T^* = \frac{590-550-(30)}{\sqrt{62.5(\frac{1}{5}+\frac{1}{5})}} = 2 \qquad \therefore reject H_0$$

我們有足夠證據去推論樣本1之男學生之平均分數比女學生之平均分數高30分

$H_0 : \mu_{X1} - \mu_{Y1} \le 35 \, vs \, H_1 : \mu_{X1} - \mu_{Y1} > 35$

T.S.: $T = \dfrac{\overline{X}_1 - \overline{Y}_1 - (35)}{\sqrt{S_p^2(\frac{1}{5}+\frac{1}{5})}} \sim t_{(8)}$ 其中 $S_p^2 = \dfrac{(5-1)S_{X1}^2 + (5-1)S_{Y1}^2}{5+5-2} = 62.5$

R.R.: Reject H_0 at $\alpha = 0.05$ if $T^* > t_{0.05(8)} = 1.86$

$$\because T^* = \frac{590-550-(35)}{\sqrt{62.5(\frac{1}{5}+\frac{1}{5})}} = 1 \qquad \therefore don't reject H_0$$

我們沒有足夠證據去推論樣本1之男學生之平均分數比女學生之平均分數高35分

$H_0 : \mu_{X2} - \mu_{Y2} \le 30 \, vs \, H_1 : \mu_{X2} - \mu_{Y2} > 30$

T.S.: $T = \dfrac{\overline{X}_2 - \overline{Y}_2 - (30)}{\sqrt{S_p^2(\frac{1}{5}+\frac{1}{5})}} \sim t_{(8)}$ 其中 $S_p^2 = \dfrac{(5-1)S_{X1}^2 + (5-1)S_{Y1}^2}{5+5-2} = 2250$

R.R.: Reject H_0 at $\alpha = 0.05$ if $T^* > t_{0.05(8)} = 1.86$

$$\because T^* = \frac{590-550-(30)}{\sqrt{2250(\frac{1}{5}+\frac{1}{5})}} = 0.3333 \qquad \therefore don't reject H_0$$

我們沒有足夠證據去推論樣本1之男學生之平均分數比女學生之平均分數高30分,更不用說35分

(三)(四)：

由於各樣本中之樣本相關係數都已經達到正線性相關之最大值1,故針對假設檢定 $H_0 : \rho_{XY} \le 0 \, vs \, H_1 : \rho_{XY} > 0$ 及 $H_0 : \rho_{XY} \le 0.8 \, vs \, H_1 : \rho_{XY} > 0.8$ 必然會得到拒絕 H_0 之決策，因此，我們有足夠證據去推論男學生與女學生分數之相關係數大於0及大於0.8。

(五)

以上(一)(二)兩小題之結論在樣本1與樣本2間完全相異，原因在於共同母體變異數之估計值 S_p^2 在兩組樣本下差

異甚大(即抽樣變異大)所導致的。