統　計

《統計學》

試題評析

綜觀今年高考試題，較往年不同的是數理統計方面的試題比例增加，點估計、區間估計與假設檢定的題型皆有，且需化至最簡狀態。

且往年較少考之MVUE的問題及UMP檢定的問題皆出現，難度明顯提高。唯一與過去相同的是必有一題實驗設計題型。分數方面，預料好壞差距勢將擴大，程度好的考生應可考到80分以上，中上程度者應有40~60分的實力。

一、令X1, X2, …, Xn為一組來自Uniform[0,3θ]母體之隨機樣本。

(一)試問Method of Moments，求出θ之估計量$\hat{\theta}_1$。（4分）

(二)試問Method of Maximum Likeihood，求出θ之估計量$\hat{\theta}_2$。（4分）

(三)試問$\hat{\theta}_1$及$\hat{\theta}_2$是否各具有不偏性？請說明理由。（12分）

答：$X_1,\cdots,X_n \xrightarrow{\text{i.i.d}} U(0,3\theta)$

(一)$\because m_1 = E(X) = \int_0^{3\theta} x\left(\frac{1}{3\theta}\right)dx = \frac{3}{2}\theta$

$\therefore \theta = \frac{2}{3}m_1$

則q的動差估計量$\hat{\theta}_1 = \frac{2}{3}m_1' = \frac{2}{3}\overline{X}$

(二)$L(\theta; x_1,\cdots,x_n) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n}\frac{1}{3\theta} I_{(0,3\theta)}(x_i)$

$= \frac{1}{3^n\theta^n} I_{\left(\frac{y_n}{3},\infty\right)}(\theta)$

其中$y_n = \max(x_i)$

$\because L(\theta; x_1,\cdots,x_n)$在$\left(\frac{y_n}{3},\infty\right)$下為$q$的遞減函數

$\therefore \hat{\theta}_2 = \frac{Y_n}{3}$為$q$的最大概似估計量.

(三)$\because E(\hat{\theta}_1) = E\left(\frac{2}{3}\overline{X}\right) = \frac{2}{3}E(\overline{X}) = \frac{2}{3}E(X) = \theta$

$\therefore \hat{\theta}_1$為$q$的不偏估計量.

又Y_n的p.d.f為

$f_n(y) = \frac{n!}{(n-1)!}\left(\frac{y}{3\theta}\right)^{n-1}\left(\frac{1}{3\theta}\right) = n\frac{y^{n-1}}{3^n\theta^n}, 0<y<3\theta$

$\therefore E(\hat{\theta}_2) = \frac{1}{3}E(Y_n) = \frac{1}{3}\int_0^{3\theta} y\left(n\frac{y^{n-1}}{3^n\theta^n}\right)dy$

$= \frac{1}{3}\left(\frac{n}{n+1}\right)(3\theta) = \frac{n}{n+1}\theta$

二、令一組隨機樣本X_1, X_2, X_3, X_4, X_5來自具有指數分配之母體，其機率密度函數（p.d.f.）為$f(x)=(1/\theta)e^{\frac{-x}{\theta}}$, x＞0；θ＞0。

(一)試用動差母函數（moment-generating functions），求出 $\dfrac{2\left(\sum_{i=1}^{5} X_i\right)}{\theta}$ 之機率分配。（12分）

(二)試用小題(一)之結果，推導出一個估計θ之95%信賴區間。（參考附表）（4分）

答：$X_1, \cdots, X_5 \xrightarrow{\text{i.i.d}} f(x) = (1/\theta) e^{\frac{-x}{\theta}}, \theta > 0, x > 0$

(一) $M_{\frac{2\left(\sum_{I=1}^{5} X_i\right)}{\theta}}(t) = E\left(e^{t\frac{2\left(\sum_{i=1}^{5} X_i\right)}{\theta}}\right) = \prod_{i=1}^{5} E\left(e^{\left(\frac{2}{\theta}X_i\right)t}\right)$

$$= \prod_{i=1}^{5} M_{X_i}\left(\frac{2}{q}t\right) = \left(\frac{\frac{1}{q}}{\frac{1}{q} - \frac{2}{q}t}\right)^5 = \left(\frac{1}{1-2t}\right)^5$$

∴根據動差母函數的唯一性可知, $\dfrac{2\left(\sum_{i=1}^{5} X_i\right)}{\theta} \to \chi^2(10)$

(二)以 $\dfrac{2\left(\sum_{i=1}^{5} X_i\right)}{\theta}$ 為樞紐量,則利用樞紐量法

$\because \chi^2_{0.025}(10) = 20.4831, \chi^2_{0.975}(10) = 3.24697$

為q的95%的信賴區間.

三、設隨機變數X為某部機器在一天中所發生故障之次數，且X具有Poisson分配，其平均值為λ。已知這部機器每天之維修費為 $C=3X^2$。令$X_1, X_2, \cdots, X_n$為隨機選出之n天中，各天所發生故障之次數。

(一)試求E(C)（即C之期望值）。（5分）

(二)試求E(C)之MVUE（Minimum Variance Unbiased Estimator）。（15分）

答：$X_1, \cdots, X_n \xrightarrow{\text{i.i.d}} Po(\lambda)$

(一) $E(C) = E(3X^2) = 3E(X^2) = 3\left\{(E(X))^2 + Var(X)\right\}$

$= 3(\lambda^2 + \lambda).$

(二) $\because f(x;\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \left(e^{-\lambda}\right)\left(\dfrac{1}{x!}\right)\exp\{(\ln\lambda)x\}$

則可令 $c(l) = e^{-l}, h(x) = \dfrac{1}{x!}, w(l) = \ln l, t(x) = x$

$\because W(\lambda) = \{\ln\lambda | \lambda > 0\} = R$ 至少包含1個1維度的矩形

$\therefore \sum_{i=1}^{n} X_i$ 是 l 的完備充份統計量.

又 $E\left(3\left(\overline{X}^2 + \overline{X}\right)\right) = 3\left\{E\left(\overline{X}\right)^2 + E\left(\overline{X}\right)\right\}$

$$= 3\left(l^2 + \frac{l}{n} + l\right)$$

$\therefore E\left\{3\left(\overline{X}^2 + \left(1 - \frac{1}{n}\right)\overline{X}\right)\right\} = 3(\lambda^2 + \lambda)$

$\because 3\left(\overline{X}^2 + \left(1 - \frac{1}{n}\right)\overline{X}\right)$ 是完備充份統計量的函數，且其為 $3(\lambda^2 + \lambda)$ 的不偏估計。

則根據Rao-Blackwell-Lehmann-Scheffe定理

$3\left(\overline{X}^2 + \left(1 - \frac{1}{n}\right)\overline{X}\right)$ 是 $3(\lambda^2 + \lambda)$ 的MVUE.

四、令$X_1, X_2, \cdots, X_{10}$為一組來自常態母體，其平均數為μ（已知），變異數為σ^2（未知）之隨機樣本。

(一)在$\alpha=0.05$時，試求H_0：$\sigma^2=\sigma_0^2$對H_1：$\sigma^2=\sigma_1^2$（其中$\sigma_1^2>\sigma_0^2$）之MPT（Most Powerful Test）拒絕域。（參考附表）（15分）

(二)試問在小題(一)中之MPT（Most Powerful Test），是否對於H_1：$\sigma^2=\sigma_0^2$為UMPT（Uniformly Most Powerful Test）？請說明理由。（5分）

答：$X_1,\cdots,X_{10}\xrightarrow{\text{i.i.d}}N(\mu,\sigma^2)$, μ已知, σ^2未知

(一)$H_0:\sigma^2=\sigma_0^2, H_1:\sigma^2=\sigma_1^2\ (\sigma_1^2>\sigma_0^2)$

$$f(x_1,\cdots,x_{10};\mu,\sigma^2)=\prod_{i=1}^{10}f(x_i;\mu,\sigma^2)$$

$$=\frac{1}{(2\pi\sigma^2)^5}\exp\left\{-\frac{\sum_{i=1}^{10}(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$\therefore\frac{f(x_1,\cdots,x_{10};\mu,\sigma_1^2)}{f(x_1,\cdots,x_{10};\mu,\sigma_0^2)}$$

$$=\left(\frac{\sigma_0^2}{\sigma_1^2}\right)^5\exp\left\{-\left(\frac{1}{2\sigma_1^2}-\frac{1}{2\sigma_0^2}\right)\sum_{i=1}^{10}(x_i-\mu)^2\right\}>k$$

$$\Leftrightarrow\sum_{i=1}^{10}(x_i-\mu)^2>c$$

則根據Neyman-Pearson引理可知,

對於檢定問題$H_0:\sigma^2=\sigma_0^2, H_1:\sigma^2=\sigma_1^2\ (\sigma_1^2>\sigma_0^2)$

其MP-test為$C=\left\{\sum_{i=1}^{10}(x_i-\mu)^2>c\right\}$

其中常數c由$\alpha=0.05=P\left\{\sum_{i=1}^{10}(x_i-\mu)^2>c\,\middle|\,\sigma^2=\sigma_0^2\right\}$

又$\dfrac{\sum_{i=1}^{10}(x_i-\mu)^2}{\sigma_0^2}\xrightarrow{H_0}\chi^2(10)$

則可知, $c=\sigma_0^2\chi_{0.05}^2(10)=\sigma_0^2(18.3070)$

即$C=\left\{\sum_{i=1}^{10}(x_i-\mu)^2>\sigma_0^2(18.3070)\right\}$為$\alpha=0.05$下的MP-test.

(二)∵ 當$\sigma_1^2>\sigma_0^2$,則MP-test型態皆為$C=\left\{\sum_{i=1}^{10}(x_i-\mu)^2>c\right\}$

且常數c的選取僅與σ_0^2有關.

即對與所有$H_0:\sigma^2=\sigma_0^2, H_1:\sigma^2=\sigma_1^2\ (\sigma_1^2>\sigma_0^2)$

$C=\left\{\sum_{i=1}^{10}(x_i-\mu)^2>\sigma_0^2(18.3070)\right\}$皆為$\alpha=0.05$下的MP-test.

$\therefore C=\left\{\sum_{i=1}^{10}(x_i-\mu)^2>\sigma_0^2(18.3070)\right\}$為$\alpha=0.05$下,

對於檢定問題$H_0:\sigma^2=\sigma_0^2, H_1:\sigma^2>\sigma_0^2$的UMP-test.

五、某位統計學教授想要研究學生熟悉統計軟體所需要之時間，是否會受到統計軟體種類與學生性別之影響？下列為二因子完全隨機設計且重複實驗兩次所得之資料（小時）：

	性別	
軟體	男性	女性
1	28,20	26,34
2	26,28	32,22
3	18,24	23,31

試以顯著水準$\alpha=0.05$檢定各小題：（參考附表）

(一)軟體種類與學生性別是否對熟悉軟體所需要之時間有交互作用？（12分）

(二)軟體種類對熟悉軟體所需要之時間是否有影響？（6分）

(三)學生性別對熟悉軟體所需要之時間是否有影響？（6分）

答：

軟體	性別	
	男性	女性
1	28,20	26,34
2	26,28	32,22
3	18,24	23,31

$$SST=\sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{2}Y_{ijk}^2-\frac{1}{12}Y_{\bullet\bullet\bullet}^2=262$$

$$SSR=\sum_{i=1}^{3}\frac{Y_{i\bullet\bullet}^2}{4}-\frac{1}{12}Y_{\bullet\bullet\bullet}^2=24$$

$$SSC=\sum_{j=1}^{2}\frac{Y_{\bullet j\bullet}^2}{6}-\frac{1}{12}Y_{\bullet\bullet\bullet}^2=48$$

$$SSCe=\sum_{i=1}^{3}\sum_{j=1}^{2}\frac{Y_{ij\bullet}^2}{2}-\frac{1}{12}Y_{\bullet\bullet\bullet}^2=96$$

SSI=*SSCe*-*SSR*-*SSC*=24

SSE=*SST*-*SSR*-*SSC*-*SSI*=166

ANOVA Table

Source	*S.S*	*d.f*	*M.S*	*F*
列處理	24	2	12	0.4337
行處理	48	1	48	1.7350
交互作用	24	2	12	0.4337
誤差	166	6	27.6667	
總和	262	11		

(一) H_0 **：軟體種類與性別間無交互作用**

H_1**:軟體種類與性別間交互作用存在**

檢定統計量: $F=\frac{MSI}{MSE}\xrightarrow{H_0}F(2,6)$

檢定規則:Reject H_0 if $F>F_{0.05}(2,6)$

$\because F=0.4337<F_{0.05}(2,6),\therefore$ Do not reject H_0 at $\alpha=0.05$.

表示無足夠證據顯示軟體種類與性別間交互作用存在.

(二) H_0 **：軟體種類與熟悉時間無影響**.

H_1**:軟體種類與熟悉時間有影響**.

檢定統計量: $F=\frac{MSR}{MSE}\xrightarrow{H_0}F(2,6)$

檢定規則:Reject H_0 if $F>F_{0.05}(2,6)$

$\because F=0.4337<F_{0.05}(2,6),\therefore$ Do not reject H_0 at $\alpha=0.05$.

表示無足夠證據顯示軟體種類與熟悉時間有影響.

(三) H_0 **：學生性別與熟悉時間無影響**.

H_1**:學生性別與熟悉時間有影響**.

檢定統計量: $F=\frac{MSC}{MSE}\xrightarrow{H_0}F(1,6)$

檢定規則:Reject H_0 if $F>F_{0.05}(1,6)$

$\because F=1.7350<F_{0.05}(1,6),\therefore$ Do not reject H_0 at $\alpha=0.05$.

表示無足夠證據顯示學生性別與熟悉時間有影響.