《統計學》

試題評析	今年考題比往年簡單，除了第一題變異數及共變異數之計算較複雜外，另外三題包含了求估計式方法及MP、UMP test典型題目以及多元迴歸分析簡易題型，考生計算能力若不差，應該可拿 80分以上，但對於程度好的同學應該可拿 90 分以上甚至滿分。
考點命中	第一題：《統計學(概要)經典試題精解》，高點文化出版，程大器編撰，頁5-32例35類似。 《高點・高上統計學總複習講義》第一回，程大器編撰，頁64例61類似。 第二題：《統計學(概要)經典試題精解》，高點文化出版，程大器編撰，頁6-46例43。 《高點・高上統計學總複習講義》第一回，程大器編撰，頁81範題77。 第三題：《高點・高上統計學總複習講義》第一回，程大器編撰，頁119範題104。 《統計學(概要)最新版》，高點文化出版，程大器編撰，頁10-52例題33。 第四題：《統計學(概要)最新版》，高點文化出版，程大器編撰，頁12-87習題8。

本試題可能使用之參考值如下：

$Z_{0.10}=1.28$ ，$Z_{0.05}=1.645$ ，$Z_{0.025}=1.96$ ，$F_{0.05}(1,40)=4.08$ ，$F_{0.05}(4,40)=2.61$ ，$F_{0.05}(5,40)=2.45$ ，$F_{0.05}(6,40)=2.34$ ，$F_{0.10}(1,40)=2.84$ ，$F_{0.10}(4,40)=2.09$ ，$F_{0.10}(5,40)=2.00$ ，$F_{0.10}(6,40)=1.93$ 。

一、設 $X_1,X_2,\ldots,X_n$ 為一組抽自平均數為 μ，標準差為 σ 的常態分配之隨機樣本，又 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 為樣本平均數，令隨機變數 $Y_i=X_i-\bar{X}$ ，$i=1,2,\ldots,n$ ，則：

(一)試求 Y_i 的變異數 $V(Y_i)=$？（8分）

(二)試求 Y_1 ，Y_2 的共變異數（Covariance）$Cov(Y_1,Y_2)=$？（8分）

(三)若使 $K\cdot(Y_1+Y_2)^2$ 為 σ^2 之不偏估計量，則常數 $K=$？（8分）

答：

(一)因 $X_i\overset{iid}{\sim}N(\mu,\sigma^2)$ ，$i=1,2,\ldots,n$ ，所以 $E(X_i)=\mu, Var(X_i)=\sigma^2$ ，又

$$Cov(X_i,\bar{X})=Cov\left(X_i,\frac{1}{n}(X_1+X_2+\ldots+X_i+\ldots+X_n)\right)$$

$$=\frac{1}{n}Cov(X_i,X_i)\text{（因 }X_i\text{ 皆獨立，}Cov(X_i,X_j)=0\text{ ，}i\neq j\text{）}$$

$$=\frac{1}{n}Var(X_i)=\frac{\sigma^2}{n}$$

$$\therefore Var(Y_i)=Var(X_i-\bar{X})=Var(X_i)+Var(\bar{X})-2Cov(X_i,\bar{X})=\sigma^2+\frac{\sigma^2}{n}-2\cdot\frac{\sigma^2}{n}=\sigma^2-\frac{\sigma^2}{n}=\sigma^2\left(1-\frac{1}{n}\right)$$

(二) $Cov(Y_1,Y_2)=Cov(X_1-\bar{X},X_2-\bar{X})=Cov(X_1,X_2)-Cov(X_1,\bar{X})-Cov(\bar{X},X_2)+Cov(\bar{X},\bar{X})$

$$=0-\frac{\sigma^2}{n}-\frac{\sigma^2}{n}+Var(\bar{X})=-\frac{\sigma^2}{n}-\frac{\sigma^2}{n}+\frac{\sigma^2}{n}=-\frac{\sigma^2}{n}$$

(三) $\because E(Y_i^2)=Var(Y_i)+[E(Y_i)]^2=\sigma^2\left(1-\frac{1}{n}\right)+(0)^2=\sigma^2\left(1-\frac{1}{n}\right)$

其中 $E(Y_i)=E(X_i-\bar{X})=E(X_i)-E(\bar{X})=\mu-\mu=0$

又 $Cov(Y_1,Y_2)=E(Y_1Y_2)-E(Y_1)\cdot E(Y_2)$

$$\Rightarrow E(Y_1Y_2)=Cov(Y_1,Y_2)+E(Y_1)E(Y_2)=-\frac{\sigma^2}{n}+0\times 0=-\frac{\sigma^2}{n}$$

$$\therefore E[k(Y_1+Y_2)^2]=\sigma^2 \Rightarrow k\{E(Y_1^2+Y_2^2+2Y_1Y_2)\}=\sigma^2 \Rightarrow k\{E(Y_1^2)+E(Y_2^2)+2E(Y_1Y_2)\}=\sigma^2$$

$$\Rightarrow k\left\{\sigma^2\left(1-\frac{1}{n}\right)+\sigma^2\left(1-\frac{1}{n}\right)+2\left(-\frac{\sigma^2}{n}\right)\right\}=\sigma^2 \Rightarrow k\left\{\sigma^2-\frac{\sigma^2}{n}+\sigma^2-\frac{\sigma^2}{n}-\frac{2\sigma^2}{n}\right\}=\sigma^2$$

$$\Rightarrow k\left\{2\sigma^2-\frac{4\sigma^2}{n}\right\}=\sigma^2 \Rightarrow k\left\{2-\frac{4}{n}\right\}=1 \Rightarrow k=\frac{n}{2n-4}=\frac{n}{2(n-2)}$$

二、已知隨機變數X的機率密度函數為 $f(x)=\begin{cases}(\theta+1)x^{\theta} & ,0<x<1\\ 0 & ,\text{其他}\end{cases}$，其中未知參數 $\theta>-1$。又令 $X_1,X_2,\ldots,X_n$ 為抽自X之一組大小為n的隨機樣本，則：

(一)試以動差法（method of moments）求 θ 之點估計量。（12分）

(二)試以最大概似法（method of maximum likelihood）求 θ 之點估計量。（12分）

答：

(一) $\because E(X)=\int_0^1 x(\theta+1)x^{\theta}dx=\int_0^1(\theta+1)x^{\theta+1}dx=\left(\frac{\theta+1}{\theta+2}x^{\theta+2}\right)_0^1=\frac{\theta+1}{\theta+2}$

故知 $\mu_1'=E(X)=\frac{\theta+1}{\theta+2}$，又 $m_1'=\frac{1}{n}\sum_{i=1}^{k}x_i=\bar{x}$

令 $\mu_1'=m_1' \Rightarrow \frac{\theta+1}{\theta+2}=\bar{x}$，$\Rightarrow \theta\bar{x}+2\bar{x}=\theta+1 \Rightarrow \theta=\frac{1-2\bar{x}}{\bar{x}-1}$

亦即 θ 之動差估計式為 $\hat{\theta}=\frac{1-2\bar{X}}{\bar{X}-1}$

(二)因概似函數為

$$L(\theta)=f(x_1;\theta)\cdot f(x_2;\theta)\cdots f(x_n;\theta)$$

$$=(\theta+1)x_1^{\theta}\cdot(\theta+1)x_2^{\theta}\cdots(\theta+1)x_n^{\theta}=(\theta+1)^n\prod_{i=1}^{n}x_i^{\theta}$$

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

又 $\frac{d\ln L(\theta)}{d\theta}=\frac{n}{\theta+1}+\sum_{i=1}^{n}\ln x_i$

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

$$\Rightarrow \frac{n}{\theta+1}=-\sum_{i=1}^{n}\ln x_i \Rightarrow \theta+1=-\frac{n}{\sum_{i=1}^{n}\ln x_i} \Rightarrow \theta=\frac{-n}{\sum_{i=1}^{n}\ln x_i}-1$$

又 $\frac{d^2\ell nL(\theta)}{d\theta^2}=\frac{-n}{(\theta+1)^2}$，故 $\frac{d^2\ell nL(\theta)}{d\theta^2}<0$

故知 θ 之最大可能估計式為 $\hat{\theta}=\frac{-n}{\sum_{i=1}^{n}\ln X_i}-1$

三、設 $X_1,X_2,\ldots,X_n$ 為一組抽自平均數為 μ，變異數為 $\sigma^2=100$ 的常態母體之隨機樣本，μ 為未知參數。

(一)試求（導出）檢定問題 $H_0：\mu=80$ vs. $H_1：\mu=86$ 的最強力檢定（most powerful test）之最佳拒絕域R為何？（8分）

(二)對於(一)之檢定問題，試求滿足 $P\left((X_1,X_2,\ldots,X_n)\in R\mid H_0\right)=P(\bar{X}\ge k\mid H_0)=0.05$ 及 $P\left((X_1,X_2,\ldots,X_n)\in R\mid H_1\right)=P(\bar{X}\ge k\mid H_1)=0.95$ 要求下之樣本大小 $n=$？及最佳拒絕域R之臨界值 $k=$？（10分）

(三)若檢定問題為 $H_0：\mu=80$ vs. $H_1：\mu>80$，且樣本大小為 $n=25$，試求在 $\alpha=0.05$ 下，此檢定問題之齊一最強力檢定（uniformly most powerful test）之最佳拒絕域R為何？（8分）

答：

(一)因概似函數爲

$$L(\mu)=\prod_{i=1}^{n}f(x_i;\mu)=\left(\frac{1}{\sqrt{2\pi}\cdot 10}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\times 100}\right\}$$

所以

$$\frac{L(80)}{L(86)}=\frac{\left(\frac{1}{\sqrt{2\pi}\cdot 10}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-80)^2}{2\times 100}\right\}}{\left(\frac{1}{\sqrt{2\pi}\cdot 10}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-86)^2}{2\times 100}\right\}}=\exp\left\{-\frac{\sum_{i=1}^{n}(X_i-80)^2-\sum_{i=1}^{n}(X_i-86)^2}{200}\right\}$$

$$=\exp\left\{-\frac{\sum_{i=1}^{n}X_1^2-160\sum_{i=1}^{n}X_i+6400-\sum_{i=1}^{n}X_1^2+172\sum_{i=1}^{n}X_i-7396}{200}\right\}=\exp\left\{-\frac{12\sum_{i=1}^{n}X_i-996}{200}\right\}$$

令 $\frac{L(80)}{L(86)}\le k\Rightarrow \exp\left\{-\frac{12\sum_{i=1}^{n}X_i-996}{200}\right\}\le k\Rightarrow\left(-\frac{12}{200}\right)\sum_{i=1}^{n}X_i+\frac{996}{200}\le \ln k\Rightarrow\left(-\frac{12}{200}\right)\sum_{i=1}^{n}X_i\le \ell nk-\frac{996}{200}$

$$\Rightarrow\sum_{i=1}^{n}X_i\ge\left(-\frac{200}{12}\right)\left(\ell nk-\frac{996}{200}\right)\Rightarrow\bar{X}\ge\left(-\frac{200}{12n}\right)\left(\ell nk-\frac{996}{200}\right)=k'$$

故知 $\frac{L(80)}{L(86)}\le k\overset{\text{相當於}}{\Rightarrow}\bar{X}\ge k'$，由Neyman-Pearson定理知，此檢定之最佳危險域R爲

$R=\{\bar{X}\mid\bar{X}\ge k'\}$，其中 k' 由 α 決定

(二)因 $P(\bar{X}\ge k\mid\mu=80)=0.05\Rightarrow P\left(Z>\frac{k-80}{10/\sqrt{n}}\right)=0.05$

$\therefore\frac{k-80}{10/\sqrt{n}}=1.645\Rightarrow k=80+1.645\frac{10}{\sqrt{n}}\cdots\cdots$①

又 $P(\bar{X}\ge k\mid\mu=86)=0.95\Rightarrow P\left(Z\ge\frac{k-86}{10/\sqrt{n}}\right)=0.95$

$\therefore\frac{k-86}{10/\sqrt{n}}=-1.645\Rightarrow k=86-1.645\frac{10}{\sqrt{n}}\cdots\cdots$②

由①，②可知

$$80+1.645\frac{10}{\sqrt{n}}=86-1.645\frac{10}{\sqrt{n}} \Rightarrow n=\frac{(1.645+1.645)^2\cdot(10)^2}{(86-80)^2}=30.067$$，

取 $n=31$，又將 $n=31$ 代入①，可知最佳危險域

$$k=80+1.645\frac{10}{\sqrt{31}} \quad 82.955$$

(三)令 μ_1 爲大於 80 之任一數，故知

$$\frac{L(80)}{L(\mu_1)}=\frac{\left(\frac{1}{\sqrt{2\pi}\cdot 10}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-80)^2}{2\times 100}\right\}}{\left(\frac{1}{\sqrt{2\pi}\cdot 10}\right)^n \exp\left\{-\frac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{2\times 100}\right\}}=\exp\left\{-\frac{\sum_{i=1}^{n}(X_i-80)^2-\sum_{i=1}^{n}(X_i-\mu_1)^2}{200}\right\}$$

$$=\exp\left\{-\frac{\sum_{i=1}^{n}X_1^2-160\sum_{i=1}^{n}X_i+6400-\sum_{i=1}^{n}X_1^2+2\mu_1\sum_{i=1}^{n}X_i-\mu_1^2}{200}\right\}=\exp\left\{-\frac{(2\mu_1-160)\sum_{i=1}^{n}X_i-(\mu_1^2-6400)}{200}\right\}$$

令 $\frac{L(80)}{L(\mu_1)}\le k \Rightarrow \exp\left\{-\frac{(2\mu_1-160)\sum_{i=1}^{n}X_i-(\mu_1^2-6400)}{200}\right\}\le k \Rightarrow \left(-\frac{2\mu_1-160}{200}\right)\sum_{i=1}^{n}X_i+\frac{\mu_1^2-6400}{200}\le \ell nk$

$$\Rightarrow \left(-\frac{2\mu_1-160}{200}\right)\sum_{i=1}^{n}X_i \le \ell nk-\left(\frac{\mu_1^2-6400}{200}\right)$$

$$\Rightarrow \sum_{i=1}^{n}X_i \ge \left(-\frac{200}{2\mu_1-160}\right)\left(\ell nk-\frac{\mu_1^2-6400}{200}\right)$$ （$\because 2\mu_1-160>0$）

$$\Rightarrow \bar{X}\ge -\frac{200}{n(2\mu_1-160)}\left(\ell nk-\frac{\mu_1^2-6400}{200}\right)=k'$$

故知 $\frac{L(80)}{L(\mu_1)}\le k \overset{\text{相當於}}{\Rightarrow} \bar{X}\ge k'$，

又 μ_1 爲大於 80 之任一數，故知檢定 H_0：$\mu=80$ ，H_1：$\mu>80$ 之均匀

最佳危險域爲 $R=\{\bar{X}\,|\,\bar{X}\ge k'\}$，因 $\alpha=0.05$ ，$n=25$，故

$$P(\bar{X}\ge k'\,|\,\mu=80)=0.05 \Rightarrow P\left(Z\ge \frac{k-80}{10/\sqrt{n}}\right)=0.05$$

$$\therefore \frac{k'-80}{10/\sqrt{25}}=1.645 \Rightarrow k'=80+1.645\frac{10}{\sqrt{25}}=83.29$$

亦即 $R=\{\bar{X}\,|\,\bar{X}\ge 83.29\}$

四、設因變數Y與自變數 X_1,X_2,X_3,X_4,X_5 做複迴歸，模式為：

$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5\cdots(1)$

蒐集樣本資料後，經資料分析，得到下列部分變異數分析（ANOVA）表：

變異來源	平方和 SS	自由度 DF	均方和 MS	F
迴歸	①	②	④	⑥
殘差	200	③	⑤	
總變異	250	45		

(一)請完成上面的變異數分析表，即在試卷上填答①～⑥之值。(6分)

(二)試求複判定係數(coefficient of multiple determination) $R^2=?$(6分)

(三)試問在 $\alpha=0.05$ 下，檢定 $H_0:\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=0$ 是否顯著？(7分)

(四)若Y對自變數 X_2,X_3,X_4,X_5 做複迴歸，得殘差平方和為 $SSE=210$，試問對模式(1)中檢定 $H_0:\beta_1=0$ 是否顯著？(取 $\alpha=0.05$)(7分)

答：

(一)變異數分析表為

變異來源	平方和 SS	自由度 DF	均方和 MS	F
迴歸	(50)	(5)	(10)	(2)
殘差	200	(40)	(5)	
總變異	250	45		

(二) $R^2=\dfrac{SSR}{SSTO}=\dfrac{50}{250}=0.2$

(三)1. $H_0:\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=0$

2. $H_1:\beta_i$ 不全為 0，$i=1,2,\ldots,5$

3. $\alpha=0.05$

4. $C=\{F\,|\,F>F_{0.05}(5,40)=2.45\}$

5.計算：$\because F=\dfrac{MSR}{MSE}=2\notin C$

6.結論：不拒絕 H_0

(四)1. $H_0:\beta_1=0$

2. $H_1:\beta_1\neq 0$

3. $\alpha=0.05$

4. $C=\{F\,|\,F>F_{0.05}(1,40)=4.08\}$

5.計算：因 $SSE(X_2,X_3,X_4,X_5)=210$

$SSE(X_1,X_2,X_3,X_4,X_5)=200$

$$F=\frac{[SSE(X_2,X_3,X_4,X_5)-SSE(X_1,X_2,X_3,X_4,X_5)]\div[(n-5)-(n-6)]}{SSE(X_1,X_2,X_3,X_4,X_5)\div(n-6)}=\frac{(210-200)\div 1}{200\div 40}=2\notin C$$

6.結論：不拒絕 H_0