《統計學》

試題評析	本年度考題對經建類同學而言，第二題(20分)，第四題(15分)可能會覺得稍有難度，其他三題只要計算方面細心一點，相信不難拿到分數，預估中上程度的考生分數大概分佈於50-60分之間，如果能考到70分以上應該勝券在握。
考點命中	第一題：(一)與秦大成《統計學講義》第二回P5例題3相同 (二)與秦大成《迴歸分析》第二回P35例題1相同 第二題：與秦大成《統計學講義》第三回P18例題完全相同 第三題：與秦大成《統計學講義》第六回P42例題1相同 第四題：與秦大成《統計學講義》第二回P89例題2相同 第五題：與秦大成《統計學講義》第四回P102例題5相同

一、令隨機變數X的機率密度函數為$f(x)=a+bx2$，$0\le x\le 1$，且X之期望值為$E(x)=\dfrac{3}{4}$，同時令$Y=-2\log X$，試求：（每小題10分，共20分）

（一）a、b之值。

（二）Y的機率密度函數為何？

答：

(一)$EX=\int_0^1 x(a+bx^2)dx=\frac{a}{2}+\frac{b}{4}=\frac{3}{4}$

$$\int_0^1 a+bx^2dx=a+\frac{b}{3}=1$$

$$\Rightarrow a=0\ ,\ b=3$$

(二)$y=-2\log x\ ,\ y\ge 0$

$$\Rightarrow x=10^{-\frac{y}{2}}\ ,\qquad \left|\frac{dx}{dy}\right|=\frac{\ln 10}{2}10^{-\frac{y}{2}}$$

$$f(y)=f_X\left(10^{-\frac{y}{2}}\right)\cdot\left|\frac{dx}{dy}\right|=3(10^{-\frac{y}{2}})^2\cdot\frac{\ln 10}{2}10^{-\frac{y}{2}}$$

$$=\frac{3\cdot\ln 10}{2}10^{-\frac{3y}{2}}\ ,\ y\ge 0$$

二、某聯歡晚會活動舉辦喝飲料拼字兌獎活動，若集滿「F、U、N、G、O、趣」6個飲料瓶蓋，即可兌換平板電腦一台，假定每瓶飲料皆有此6個字中的一個，且機率相同。請問平均要喝多少瓶才有兌獎機會？（20分）

答：

令X_i：**直到 ith 個字母出現，所需要喝的瓶數**

則 $X_i\sim Geo\left(P_i=\frac{7-i}{6}\right)\ ,\ i=1,2,\dots,6$

$E(\sum_{i=1}^{6} X_i) = \sum_{i=1}^{6}(EX_i) = \sum_{i=1}^{6}\frac{1}{P_i} = \sum_{i=1}^{6}\frac{6}{7-i} = 14.7$

∴平均要喝14.7瓶才有對獎機會

三、為分析員工的工作效率，下表為A、B、C三位員工在不同工作時間下之工作效率的ANOVA表，其平均工作量依序為 $\bar{Y}_{.1} = 30$ ，$\bar{Y}_{.2} = 25$ ，$\bar{Y}_{.3} = 23$ 。

變異來源	平方和	自由度
時間	18	3
員工	104	2
誤差	12	6

(一)試寫出變異數分析所需的假設。(8分)
(二)在顯著水準為5%下，檢定時間之不同是否顯著地影響工作量？(9分)
(三)在顯著水準為5%下，求算各個小母體共同變異數σ 2之信賴區間。(8分)

答：
(一)1.每一母體皆爲常態分配
2.每一母體σ^2都相等(齊一性)
3.母體之間彼此獨立，而且每個樣本都隨機抽得

即 $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

(二)1. $\begin{cases} H_0：時間之不同對工作量會有顯著影響 \\ H_1：時間之不同對工作量不會有顯著影響 \end{cases}$

2.$C = \{F | F > F_{0.05}(3, 6) = 4.7571\}$

3.$F = \frac{SSR/(r-1)}{SSE/(r-1)(c-1)} = \frac{18/3}{12/6} = 3 \notin C$

∴ Do not reject H_0，無充分證據顯示時間之不同對工作量不會有顯著影響

(三) σ^2 之 95% C. I. $= \left(\frac{SSE}{\chi^2_{\frac{0.05}{2}}(r-1)(c-1)}, \frac{SSE}{\chi^2_{1-\frac{0.05}{2}}(r-1)(c-1)}\right)$

$= \left(\frac{12}{14.4494}, \frac{12}{1.2373}\right) = (0.8305, 9.6985)$

四、小江每週路跑2次，平均每次跑步的時間是40分鐘，標準差為5分鐘，且每週平均為1小時20分，標準差為4分鐘，假設在星期一他只跑15分鐘下，求他在星期五會跑步的預期時間為何？(15分)

答：

令X，Y分別表示每周兩次路跑時間

則 $V(X+Y)=V(X)+2cov(X,Y)+V(Y)$

$$=V(X)+2\rho\sigma_X\sigma_Y+V(Y)$$

$$4^2=5^2+2\rho\cdot5\cdot5+5^2 \Rightarrow \rho=-0.68$$

Y對X之最佳線性預測是：$E(Y|X)=\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$

$$\Rightarrow E(Y|X=15)=40+(-0.68)\cdot\frac{5}{5}(15-40)=57$$

∴小江在星期五會跑步的預測時間為57分鐘

五、臺灣地區滿16歲以上的人口約1300萬人，為估計這些人中抽菸的比例p，今以隨機抽樣法抽出樣本數為n的一組樣本，設X表示樣本抽菸的人數，以樣本比例 $\hat{p}=\frac{X}{n}$ 估計p值，則：（每小題10分，共20分）

（一）假設n＝100人，x＝50人，在95%的信心水準之下，試估計誤差界限（margin of error）之值為何？

（二）在（一）中，若希望誤差界限不超出0.02，則n＝100是否太多或太少，需要增加或減少多少？

答：

(一) $n=100\geq30$，採C.L.T.查standard normal表

$$\hat{P}=\frac{X}{n}=\frac{50}{100}=0.5$$

誤差界限 $b=Z_{\frac{0.05}{2}}\cdot\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}=1.96\cdot\sqrt{\frac{0.5\times0.5}{100}}=0.098$

(二) $b=Z_{\frac{0.05}{2}}\cdot\sqrt{\frac{P_0(1-P_0)}{n}} \Rightarrow 0.02=1.96\cdot\sqrt{\frac{0.5(1-0.5)}{n}}$

$\Rightarrow n=2401$ （至少要抽2401個樣本）

∴ $n=100$ 樣本數太少，至少需要增加2301個樣本