《統計學概要》

試題評析	這次的地特四等經建行政的統計學很簡單，四題都是非常基本的題型，但第四大題在資料的前處理要花蠻多時間，所以前三題請以最快的速度完成，可以看的出來用性質的地方就用，把剩下的1小時留給第四大題與檢查，若要上榜，這份考卷應該要考到100分，不容有錯。
考點命中	1.《高點・高上統計學講義》第二回，蘇建郎編撰，頁1、13，3-1隨機變數。 2.《高點・高上統計學講義》第一回，蘇建郎編撰，頁3、7，1-1緒論。 3.《高點・高上統計學講義》第二回，蘇建郎編撰，頁27，3-4多變數隨機變數。 第二回，蘇建郎編撰，頁35，3-5條件分配。 4.《高點・高上統計學講義》第七回，蘇建郎編撰，頁18-19，10-1適合度檢定。 第五回，蘇建郎編撰，頁32，7-2檢定統計量法決策準則。

一、投擲兩個公正六面骰子一次，設隨機變數X為出現點數2的個數。請回答下列問題：

(一)寫出題目所述的實驗（experiment）的樣本空間（sample space）。（5分）

(二)求隨機變數X的動差母函數（moment generating function）。（6分）

(三)求機率P(X≥1)。（3分）

答：

(一)樣本空間 $S=\begin{Bmatrix}(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\\(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\\(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\\(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\\(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\\(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\end{Bmatrix}$ or $S=\{(x,y)\,|\,1\le x\le 6, 1\le y\le 6,\ x和y為整數\}$

(二) $f_X(x)=\begin{cases}\dfrac{25}{36}, & x=0\\[2mm] \dfrac{10}{36}, & x=1\\[2mm] \dfrac{1}{36}, & x=2\\[2mm] 0, & o.w.\end{cases}$ $\quad M_X(t)=E(e^{tX})=\sum_{x=0}^{2}e^{tx}\cdot f_X(x)=e^0\times\dfrac{25}{36}+e^t\times\dfrac{10}{36}+e^{2t}\times\dfrac{1}{36}$

$$=\frac{25+10e^t+e^{2t}}{36}\ ,\ t\in\mathbb{R}$$

(三) $P(X\ge 1)=P(X=1)+P(X=2)=\dfrac{11}{36}$

【版權所有，重製必究！】

二、請回答下列問題：

(一)寫出樣本平均數與樣本變異數的計算公式。（8分）

(二)寫出母體（population）平均數與母體變異數的定義及計算公式。（8分）

(三)說明什麼狀況下，會以樣本平均數來替代（估計）母體平均數。（5分）

答：

(一)假設資料$(x_1,x_2,......,x_N)$之母體個數為N，樣本個數為n

樣本平均數 $\overline{x}=\dfrac{\sum_{i=1}^{n}x_i}{n}=\dfrac{x_1+x_2+......+x_n}{n}$

樣本變異數 $s^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$ ；計算公式$s^2=\dfrac{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}{n-1}$

(二)母體平均數為$\mu=\dfrac{\sum_{i=1}^{N}x_i}{N}=\dfrac{x_1+x_2+......+x_N}{N}$

母體變異數為 $\sigma^2=\dfrac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$ ；計算公式 $\sigma^2=\dfrac{\sum_{i=1}^{N}x_i^2-N\mu^2}{N}=\dfrac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2$

(三)在估計母體變異數時，若母體平均數未知時，可以用樣本平均數來替代(估計)母體平均數。

三、已知隨機變數X與Y的聯合機率密度函數為$f(x,y)=\dfrac{1}{c}$，$0 \le x \le y \le 2$：其中c為常數。

請回答下列問題：

(一)求常數c，以滿足題目所述的f(x,y)構成一機率密度函數。（6分）

(二)分別求X與Y的邊際密度函數，$f_X(x)$與$f_Y(y)$。（6分）

(三)求給定X＝x下，Y的條件機率密度函數，$f_{Y|x}(y|x)$。（5分）

(四)求給定X＝x下，Y的條件平均數與條件變異數，$\mu_{Y|x}$和$\sigma^2_{Y|x}$。（6分）

(五)根據題(三)之結果，回答給定X＝x下，Y的條件分配是何種分配？（回答分配名稱及其參數。）（6分）

答：

(一)$\int\int_A \dfrac{1}{c}\ dxdy$，$A=\{(x,y)\,|\,0\le x\le y\le 2\}$ 且面積為2。

$\Rightarrow\ c=\dfrac{1}{2}$

(二)

$$f_{X,Y}(x,y)=\frac{1}{2}\ ,\ 0\le x\le y\le 2$$

$$f_X(x)=\int_x^2\frac{1}{2}\ dy=\frac{2-x}{2}\ \ 0\le x\le 2$$

【版權所有，重製必究！】

$$f_Y(y)=\int_0^y\frac{1}{2}\ dx=\frac{y}{2}\ ,\ 0\le y\le 2$$

(三) $f_{Y|X}(y|x)=\dfrac{f_{X,Y}(x,y)}{f_X(x)}=\dfrac{\frac{1}{2}}{\frac{2-x}{2}}=\dfrac{1}{2-x}$,$0\le x\le y\le 2$

(四)由(三)可知 $Y|X=x\sim U(x,2)$

故 $\mu_{Y|x}=\dfrac{x+2}{2}$ ，$0\le x\le 2$ 與 $\sigma^2_{Y|x}=\dfrac{(2-x)^2}{12}$ ，$0\le x\le 2$

(五)給定 $X=x$ 下，Y 之條件分配是均勻分配（i.e. $Y|X=x\sim U(x,2)$）

高點高上

四、假設有組50筆的資料如下所示：

70 40 63 58 77 53 69 65 67 41 63 62 42 63 57 59 77 59 67 57
69 62 47 63 42 56 69 66 70 55 54 66 67 60 77 49 74 75 77 76
52 61 42 44 68 60 53 49 58 67。（不提供統計分配的百分位數值，依觀念可判斷是否棄卻假設。）（$z_{0.1}=-1.282$, $z_{0.2}=-0.8146$, $z_{0.25}=-0.6745$。）

請回答下列問題：

(一)以卡方適合度檢定對此資料的分配檢定是否為常態分配，顯著水準為0.05。（需包括虛無假設與對立假設、檢定統計量、計算過程、棄卻域（rejection region）與結論。）（分四組檢定即可）（20分）

(二)檢定此組資料的平均數是60，顯著水準為0.1。（需包括虛無假設與對立假設、檢定統計量、計算過程，棄卻域與結論。）（16分）

本試題可能使用之查表值如下：

1. $\chi^2_{n,a}$（具有自由度n之卡方分配右尾之機率為α的χ^2值）：

$\chi^2_{1,0.05}=3.841$， $\chi^2_{2,0.05}=5.991$， $\chi^2_{3,0.05}=7.815$， $\chi^2_{1,0.025}=5.024$

$\chi^2_{2,0.025}=7.378$， $\chi^2_{3,0.025}=9.348$ 。

2. $t_a(n)$（具有自由度n之t分配右尾之機率為α的t值）：

$t_{0.1}(49)=1.299$, $t_{0.1}(50)=1.299$, $t_{0.05}(49)=1.677$, $t_{0.05}(50)=1.676$。

答：

(一)找出50筆資料之最小值為40 與 最大值為77

並計算樣本平均數 $\bar{x}=\dfrac{\sum_{i=1}^{50}x_i}{50}=60.74$ 與 樣本標準差 $s=\sqrt{\dfrac{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}{n-1}}=10.4174$

分四組檢定定組界：
$$\bar{x}+z_{0.25}s=60.74-0.6745\times 10.4174=53.7135$$
$$\bar{x}+z_{0.75}s=60.74+0.6745\times 10.4174=67.7665$$

組界	40至53.71	53.71至60.74	60.74至67.77	67.77至77
觀察 o_i	12	11	14	13
理論 e_i	12.5	12.5	12.5	12.5

【版權所有，重製必究！】

其中理論次數每組都是 $50\times\dfrac{1}{4}=12.5$ ，觀察次數 o_i 將原始資料編入適合的組裡

(1) H_0 : X 服從常態分配

H_1 : X 不服從常態分配，$a=0.05$

(2) $c^2=\sum_{i=1}^{4}\frac{\left(|o_i-e_i|-0.5\right)^2}{e_i}=\frac{[\left(|12-12.5|-0.5\right)^2+\cdots+\left(|13-12.5|-0.5\right)^2]}{12.5}=\frac{2}{12.5}=0.16$

(3) $a=0.05$ ，拒絕域 $C=\left\{c^2\,|\,c^2>c^2_{0.05}\left(4-1-m\right)=c^2_{0.05}\left(1\right)=3.841\right\}$

$c^2\notin C$ ，*don't reject* H_0

(4)故在顯著水準 $a=0.05$ 的情況下，根據樣本資料顯示，我們沒有證據說資料不服從常態分配，即此為常態分配。

(二)

(1) $\begin{cases}H_0:\mu=60\\H_1:\mu\neq 60\end{cases}$ $\alpha=0.1$, $n=50$, $\bar{x}=60.74$, $s=10.4174$

(2) $t^*=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}=\frac{60.74-60}{10.4174/\sqrt{6}}=0.174$

(3) $\alpha=0.1$, $C=\{t^*:|t^*|>t_{\alpha/2}(n-1)=t_{0.05}(49)=1.676\}$

$t^*=0.174\notin C$ ， *don't reject* H_0

(4)故在顯著水準 $\alpha=0.01$ 的情況下，根據樣本資料顯示，我們沒有證據說 $\mu\neq 60$ 。

高點・高上

【版權所有，重製必究！】