《統計學概要》

一、設X為一間斷隨機變數，其隨機變量為$x_1, x_2, ..., x_{2n+1}$，對應x_i的每一數值有唯一的機率與之對應，設機率值為$p(x_i) = f_i$, $i = 1,2,...,2n+1$。若$x_1 < x_2 < x_3 < ... < x_{2n} < x_{2n+1}$，且$f_1 = f_{2n+1} < f_2 = f_{2n} < f_3 = f_{2n-1} < ... < f_{n-1} = f_{n+3} < f_n = f_{n+2} < f_{n+1}$，請只用$x_i$及$f_i$回答下列問題：（每小題5分，共20分）

(一)眾數（Mode）

(二)全距（Range）

(三)平均數（Mean）

(四)變異數（Variance）

試題評析	本題主要是要看懂題目給的資訊：變量個數奇數及對稱鐘形分配。
考點命中	《高上統計學講義第一回》，趙治勳編撰，第四章第二節。

答：

(一) 由於$r.v.X$之變量中x_{n+1}發生機率f_{n+1}最大, 故x_{n+1}為$r.v.X$之眾數。

(二) 由於$x_1 < x_2 < \cdots < x_{2n+1}$, 故$x_{2n+1} - x_1$為$r.v.X$之全距。

(三) 由於$f_1 = f_{2n+1} < f_2 = f_{2n} < \cdots < f_n = f_{n+2} < f_{n-1}$可得知$r.v.X$為對稱鐘形分配且由於$2n+1$為奇數可得知$r.v.X$之變量個數為奇數，

故 $E(X) = \sum_{i=1}^{2n+1} x_i f_i = x_{n+1} = \text{mod}\,e$　　$V(X) = E(X^2)\text{-}[E(X)]^2 = \sum_{i=1}^{2n+1} x_i^2 f_i - (x_{n+1})^2$

二、一電視趣味節目準備了4個綠色和4個藍色的球放在一個箱子。

(一)主持人請你伸手至箱子裡連續抽取8個球，如果你前4次都拿到綠色的球，並且後4次都拿到藍色的球，那你將獲得一份14天歐洲旅遊的招待。主持人告訴你，你可挑選不置回抽取（without replacement）或者置回抽取（with replacement）中任一個方法來連續抽取此8個球。何種方法獲得此份14天歐洲旅遊招待的機會較高？（10分）

(二)在你第一次嘗試抽取此8個球，結果失敗了。主持人又給你另一次機會，這一次他請你伸手至箱子裡連續抽取8個球，如果你最後的4次都是拿到藍色的球（前面4次拿到什麼顏色的球並無任何關係），那你將獲得此份14天歐洲旅遊的招待。主持人告訴你，你可挑選不置回抽取（without replacement）或者置回抽取（with replacement）中任一個方法來連續抽取此8個球。何種方法獲得此份14天歐洲旅遊招待的機會較高？（10分）

試題評析	本題是考抽後放回及抽後不放回下機率之計算與比較。
考點命中	《高上統計學講義第一回》，趙治勳編撰，第三章第一節。

答：

(一) 不置回抽取(抽後不放回)

$P(\text{獲得旅遊招待}) = \frac{4}{8} \times \frac{3}{7} \times \frac{2}{6} \times \frac{1}{5} \times 1 \times 1 \times 1 \times 1 = 0.0143$

置回抽取(抽後放回)

P(獲得旅遊招待)=$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=0.00391$

故不置回抽取(抽後不放回)下，獲得旅遊招待之機率較高。

(二) 不置回抽取(抽後不放回)

P(獲得旅遊招待)=$\frac{4}{8}\times\frac{3}{7}\times\frac{2}{6}\times\frac{1}{5}\times 1\times 1\times 1\times 1=0.0143$ (與(一)相同遊戲條件)

置回抽取(抽後放回)

P(獲得旅遊招待)=$1\times 1\times 1\times 1\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=0.0625$

故置回抽取(抽後放回)下，獲得旅遊招待之機率較高。

三、(一)何謂抽樣分配（Sampling Distribution）？（5分）

(二)試討論樣本平均數 $\overline{x}$ 分配的形狀。

1.若母體分配為常態。（5分）

2.若母體分配不為常態。（5分）

(三)請寫出簡單隨機抽樣樣本大小為n之樣本比例 $\hat{p}$ 的抽樣分配。（10分）

試題評析	本題是考樣本平均數及樣本成功比例之抽樣分配，考生要注意討論過程中有那些必要條件，一定要清楚列明。
考點命中	《高上統計學講義第三回》，趙治勳編撰，第七章第一節。

答：

(一)統計量所服從之機率分配稱爲抽樣分配，其中統計量爲樣本之函數且不含母體未知參數。

(二) 1.母體：$X\sim N(\mu,\sigma^2)$ 假設隨機樣本

樣本：$X_1,X_2,\cdots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$

$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$ (真實分配)

2.母體：$X\sim(\mu,\sigma^2)$ 假設隨機樣本

樣本：$X_1,X_2,\cdots,X_n \overset{iid}{\sim}(\mu,\sigma^2)$

當n夠大時

$\overline{X}\underset{byC.L.T.}{\sim}N(\mu,\frac{\sigma^2}{n})$ (逼近分配)

(三)情況一：樣本數 n 夠大

母體：$X\sim Ber(p)$

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

當 n 夠大時

$$\hat{p}=\frac{\sum X_i}{n}\underset{byC.L.T.}{\sim}N(p,\frac{p(1-p)}{n})$$

情況二：樣本數 n 不夠大

母體： $X \sim Ber(p)$

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

$$\sum_{i=1}^{n} X_i \sim Bin(n,p)$$

令 $\mathrm{Y}=\hat{p}=\frac{\sum X_i}{n}$

$$\mathrm{f_Y(y)}=\binom{\mathrm{n}}{\mathrm{ny}}p^{ny}(1-p)^{n-ny}, y=0,\frac{1}{n},\frac{2}{n},\cdots,\frac{n-1}{n},1$$

四、(一)當H_0：此人是有罪的或H_0：此人是無罪的兩種情況之下，分別寫出其型Ⅰ錯誤與型Ⅱ錯誤的事件。（5分）

(二)依照一般的角度，請說明統計假設會傾向採用何者？（5分）

試題評析	本題是考假設檢定針對虛無假設及對立假設的設立準則。
考點命中	《高上統計學講義第四回》，趙治勳編撰，第九章第一節。

答：

(一)當 H_0 ：此人是有罪的情況下

型 I 錯誤：本來此人是有罪，錯誤判決成無罪。

型 II 錯誤：本來此人是無罪，錯誤判決成有罪。

當 H_0 ：此人是無罪的情況下

型 I 錯誤：本來此人是無罪，錯誤判決成有罪。

型 II 錯誤：本來此人是有罪，錯誤判決成無罪。

(二)統計假設比較傾向採用 H_0 ：此人是無罪，因為設立 H_0 ， H_1之準則係把研究者想要找證據拒絕的事件放在 H_0 ，反言之，把研究者想要找證據驗證的事件放在 H_1的。而法庭上基於人人平等原則，在進行審判時應該先假設被告無罪，而利用所提供的證據去拒絕被告無罪之假設，從而找證據證明被告有罪，因此比較傾向於採用 H_0 ：此人是無罪。

又在 H_0 ：此人是無罪之下，犯型 I 錯誤所造成的損失遠大於犯型 II 錯誤，故一個「理想檢定」應該先控制犯型 I 錯誤之最大機率，從而求正確拒絕 H_0 之機率(檢定力 power)之極大。

五、某百貨公司記錄了顧客在該公司消費時所使用的信用卡種類（包括VISA、MasterCard和AE）和每筆信用卡的交易金額。若該公司欲比較三種信用卡的平均每筆交易金額是否有差異，從三種信用卡隨機各抽出30位顧客，做變異分析，以下是部分的ANOVA表格：

註：$F_{0.05}(2, 60)=3.15$, $F_{0.05}(2, 120)=3.07$, $F_{0.05}(3, 60)=2.76$, $F_{0.05}(3, 120)=2.68$

變異來源	平方和（SS）	自由度（df）	平均平方和（MS）	F
因子變異	a	c	e	3.7
隨機變異	b	d	159	
總變異				

(一)請算出上述表格中的a、b、c、d、e。（10分）

(二)若在顯著水準5%之下，此三種信用卡的平均每筆交易金額是否有差異？（5分）

試題評析	變異數分析基礎計算題，獲得滿分不難。
考點命中	《高上統計學講義第四回》，趙治勳編撰，第十章第二節。

答：

(一) (a)1176.6 (b)13833 (c)2 (d)87 (e)588.3

(二) 假設模型：$y_{ij}=\mu+\alpha_i+\varepsilon_{ij}$ ， $\varepsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2)$

$H_0:\mu_1=\mu_2=\mu_3$ vs $H_1:$至少一個$\mu_i\neq\mu_j$

T.S.: $F=\dfrac{MSR}{MSE}\sim F_{(2,87)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(2,87)}=(3.07,3.15)$

$\because F^*$=3.7 $\therefore rejectH_0$

結論：我們有足夠的統計證據去推論三種信用卡的平均每筆交易金額有差異。

六、若想推論母體比率p，希望估計誤差能控制在3%之內，信賴度訂為95%，假設母體數未知，且p根據經驗已知在0.3上下，則樣本數應至少有多少？（10分）

試題評析	利用信賴區間之觀念去得到合理樣本數之計算題。
考點命中	《高上統計學講義第三回》，趙治勳編撰，第十章第七節。

答：

母體：$X\sim Ber(p)$ 假設隨機樣本

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

點估計：$\hat{p}=\dfrac{\sum_{i=1}^{n}X_i}{n}\underset{byC.L.T.}{\sim}N(p,\dfrac{p(1-p)}{n})$

$P(|\hat{p}-p|\leq 0.03)=P(|Z|\leq\dfrac{0.03}{\sqrt{p(1-p)}/\sqrt{n}})=0.95$

$\Rightarrow\dfrac{0.03}{\sqrt{p(1-p)}/\sqrt{n}}=1.96\overset{p=0.3}{\Rightarrow}n=896.37$ ，故n取897個。