《統計學概要》

一、令 S 表示一隨機實驗之樣本空間，設 $S=\bigcup_{i=1}^{5}A_i$，其中 $A_i\cap A_j=\phi$，$i\neq j$。A 為一事件，且 $P(A_j)=\frac{j}{15}$，$P(A|A_j)=\frac{5-j}{15}$，$j=1,\ldots,5$。

(一)試求 $P(A_j|A)$，$j=1,\ldots,5$。(15分)

(二)1. 在題(一)中你使用了那個定理？(4分)

2. 請敘述此定理的內容。(6分)

試題評析	本題是考機率論中貝氏定理，屬於基本考題，要拿到分數並不難。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，頁53。

答：

(一) $$\mathrm{P(A_j|A)}=\frac{\mathrm{P(A_j\cap A)}}{\mathrm{P(A)}}=\frac{\mathrm{P(A_j)P(A|A_j)}}{\sum_{i=1}^{5}\mathrm{P(A_i)P(A|A_i)}}$$

$$=\frac{\frac{j}{15}\frac{5-j}{15}}{\frac{1}{15}\frac{4}{15}\times\frac{2}{15}\frac{3}{15}\times\frac{3}{15}\frac{2}{15}\times\frac{4}{15}\frac{1}{15}\times\frac{5}{15}\frac{0}{15}}=\frac{45}{4}\frac{j}{15}\frac{5-j}{15}=\frac{j(5-j)}{20},\ j=1,2,3,4,5$$

(二) 1.貝氏定理

2. 若 $B_1,B_2,B_3,\cdots,B_k$ 把樣本空間分割且 $P(B_i)>0,i=1,2,\cdots,k$，又 $A\in\Im$ 且 $P(A)>0$，則

$$P(B_j|A)=\frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{k}P(B_i)P(A|B_i)}$$，此稱貝氏定理。

二、令 $\{X_i\}_1^n$ 為一組由卜瓦松 $P(\lambda)$ 母體所抽出之隨機樣本，令 $\bar{X}_n=\frac{\sum_{i=1}^{n}X_i}{n}$ 表示樣本平均數：

(一)當樣本大小 $n\geq 30$ 且固定，試寫出 $\bar{X}_n$ 之漸近分配，需說明理由及註明分配名稱與參數。(10分)

(二)設參數 λ 未知，試求 λ 之最大概似估計式，記為 $\hat{\lambda}$。(10分)

試題評析	本題涉及到兩個主題，一個是中央極限定理，另外一個是最大概似估計量，都屬基本題型，要拿到滿分不難。
考點命中	《高點・高上統計學講義》第三回，趙治勳編撰，頁7。 《高點・高上統計學講義》第三回，趙治勳編撰，頁34。

答：

(一) 由中央極限定理得知，$\bar{X}_n \underset{byC.L.T.}{\sim} N(\lambda,\frac{\lambda}{n})$

(二) $L(\lambda)=\prod_{i=1}^{n} f(x_i\mid\lambda)=\dfrac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}$

$$\ln L(\lambda)=-n\lambda+\sum_{i=1}^{n}x_i\ln\lambda-\ln(\prod_{i=1}^{n}x_i!)$$

令 $\dfrac{d\ln L(\lambda)}{d\lambda}=-n+\dfrac{\sum_{i=1}^{n}x_i}{\lambda}=0 \quad \Rightarrow \hat{\lambda}_{MLE}=\dfrac{\sum_{i=1}^{n}x_i}{n}=\bar{x}$

因此，$\bar{X}$ 為參數 λ 之最大概似估計。

三、給下列 $\{(x_i,y_i)\}_1^7$ 成對資料：

x_i	1	2	3	4	5	6	7
y_i	3	5	8	10	10	12	15

設此一資料來自一簡單線性迴歸模型：$Y_i=\beta_0+\beta_1x_i+\varepsilon_i$，$\varepsilon_i\overset{\text{iid}}{\sim}N(0,\sigma^2)$，$i=1,2,\ldots,7$。

(一)試說明：

1. β_1 之含意。（5分）
2. 誤差 ε_i 被假設為常態之合理性。（5分）

(二)1. 以最小平方估計法（LSE）計算此資料樣本迴歸線 $\hat{Y}_i=b_0+b_1x_i$。（10分）

2. 以 $\alpha\in(0,1)$ 為顯著水準，如何檢定此迴歸線是否顯著？（10分）

試題評析	本題涉及為簡迴歸範圍，在考古題中也常有，要拿到滿分不難。
考點命中	《迴歸分析熱門題庫》，趙治勳編著，高點・高上出版，頁2-10、頁2-21。

答：

(一)1. β_1 表自變數X每增加一個單位時，對於應變數Y平均值之影響。

2. ε_i 假設為常態分配是為了得到模型中未知參數估計量之抽樣分配，有利於進行假設檢定與信賴區間之推導。這個假設在實務上也是合理的，因為誤差項中所含之變數就是除了自變數X外也會對應變數Y產生影響之其他變數，但是這些變數通常影響不大或無法控制或缺乏規則性的，它們都應該時隨機地出現，可見常態假設是合理的。

(二)1. $n=7$，$\sum X_i=28$，$\sum X_i^2=140$，$\sum Y_i=63$，$\sum Y_i^2=667$，$\sum X_iY_i=304$，

$SS_X=28$，$SS_Y=100$，$SS_{XY}=52$

$\hat{\beta}_1=\dfrac{SS_{XY}}{SS_X}=1.8571$，$\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}=1.5716$

$\therefore \hat{y}=1.5716+1.8571x$

2. $SSR=\hat{\beta}_1^2SS_X=96.567$，$SST=SS_Y=100$

ANOVA TABLE

source	SS	d.f.	MS	F
Reg	96.567	1	96.567	$F^*=140.6452$
Error	3.433	5	0.6866	
Total	100	6		

H_0：模型是不適當的 vs H_1：模型是適當的

T.S.： $F=\dfrac{MSR}{MSE}\sim F_{(1,5)}$

R.R.：Reject H_0 at α if $F^*>F_{(1,5)\alpha}$

四、設X_1，X_2，…，X_9，為一組來自常態$N(\mu_1,5^2)$母體之隨機樣本，Y_1，Y_2，…，Y_{16}為另一組來自常態$N(\mu_2,6^2)$母體之隨機樣本，X_i與Y_j之間獨立。設：

$$\bar{X}=\frac{\sum_{i=1}^{9}X_i}{9}\text{，}\bar{Y}=\frac{\sum_{j=1}^{16}Y_j}{16}$$

(一)試寫出$\bar{X}-\bar{Y}$之抽樣分配。（10分）

(二)若由此兩個母體分別抽出之特定樣本得$\bar{x}=64$，$\bar{y}=59$，利用題(一)之結果：

1. 求$\mu_1-\mu_2$之90%信賴區間。（8分）
2. 並解釋其意義。（2分）

（已知$Z\sim N(0,1)$，$P(|Z|<1.96)=0.95$，$P(|Z|<1.645)=0.9$）

(三)依據題(二)之結果，可否作$H_0:\mu_1=\mu_2$ vs. $H_1:\mu_1\neq\mu_2$之檢定？需說明理由。（5分）

試題評析	本題涉及兩獨立母體平均數之信賴區間及假設檢定，在考古題中也常有，要拿到滿分不難。
考點命中	《高點．高上統計學講義》第三回，趙治勳編撰，頁54。 《高點．高上統計學講義》第四回，趙治勳編撰，頁17。

答：

(一) $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\dfrac{5^2}{9}+\dfrac{6^2}{16}=\dfrac{181}{36})$

(二)1. 樞紐量：$\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{181}{36}}}\sim N(0,1)$

機率區間：$P(-z_{0.05}\le\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{181}{36}}}\le z_{0.05})=0.9$

結論：$\mu_1-\mu_2$之90%信賴區間為

$$((\bar{X}-\bar{Y})\pm z_{0.05}\sqrt{\frac{181}{36}})=((64-59)\pm1.645\sqrt{\frac{181}{36}})=(1.3115,8.6885)$$

2. 我們有90%信賴區間$(1.3115,8.6885)$會包含$\mu_1-\mu_2$。

(三)可以。由於信賴水準90%下之信賴區間$(1.3115,8.6885)$未含0，故我們有足夠證據去推論兩母體平均數之差異是顯著的。