《統計學》

試題評析	今年試題難度中等，除了第三大題有一些數學推導外，其他考題在講義中都有一模一樣的例題。預期平均分數為80分，程度較好之考生應該可以得到90分以上。
考點命中	第一題：《高點・高上統計學講義》第二回，趙治勳編撰，第7.3節。 第二題：《高點・高上統計學講義》第一回，趙治勳編撰，第五章。 第三題：《高點・高上統計學講義》第二回，趙治勳編撰，第十章。 第四題：《高點・高上統計學講義》第二回，趙治勳編撰，第十章例題完全命中！ 第五題：《高點・高上統計學講義》第二回，趙治勳編撰，第十章。 第六題：《高點・高上統計學講義》第三回，趙治勳編撰，第十三章。

一、某國營事業單位今年準備以招聘考試招收 202 人，其中 177 人為正取人員，25 人為備取人員，今年報考人數共計 2500 人，考試滿分為 400 分，考試後得知應考人成績 X 服從平均數為 μ，標準差為 σ 的常態分配，即 $X\sim N(\mu,\sigma)$，且。$\mu=256$ 分，成績在 360 分以上者共有 47 人，請依據上面之訊息，試求：（每小題 8 分，共 24 分）

(一) 這 2500 位應考人成績分配之標準差為 $\sigma=$？

(二) 此項考試正取人員的最低錄取分數為何？

(三) 若某甲參加了這個國營事業單位的考試，而他考試的分數為 327 分，則他是否被錄取為正取？還是備取，抑或是落榜了？

答：

$X\sim N(\mu=256,\sigma^2)$

(一) $P(X>360)=P(Z>\dfrac{360-256}{\sigma})=\dfrac{47}{2500}=0.0188$

$\Rightarrow \dfrac{360-256}{\sigma}=z_{0.0188}=2.08 \quad \Rightarrow \sigma=50$

(二) 令a為最低錄取分數

$P(X>a)=P(Z>\dfrac{a-256}{50})=\dfrac{177}{2500}=0.0708$

$\Rightarrow \dfrac{a-256}{50}=z_{0.0708}=1.47 \quad \Rightarrow a=329.5$

(三) $P(X>b)=P(Z>\dfrac{b-256}{50})=\dfrac{202}{2500}=0.0808$

$\Rightarrow \dfrac{b-256}{50}=z_{0.0808}=1.4 \quad \Rightarrow b=326$

由上述得知，326分以下為落榜，(326,329..5)為備取，329.5以上為正取
故甲考327分為備取。

二、設二維隨機變量（X, Y）的聯合機率密度函數為

$$f(x,y)=\begin{cases}\dfrac{6}{(x+y+1)^4} & ，x>0,y>0\\ 0 & ，其他\end{cases}$$

試求：（每小題 6 分，共 12 分）

(一) 當 $Y=y$ 時，X 的條件機率密度函數 $f_{X|Y}(x|y)$ 為何？

(二) $P(0 \le X \le 1 | Y=1) =$ ？

答：

(一) $f_{X|Y}(x|y) = \dfrac{\dfrac{6}{(x+y+1)^4}}{\dfrac{2}{(y+1)^3}} = 3\dfrac{(y+1)^3}{(x+y+1)^4}, 0<x<\infty, 0<y<\infty$

其中 $f_Y(y) = \int_0^\infty \dfrac{6}{(x+y+1)^4}dx = \dfrac{2}{(y+1)^3}, 0<y<\infty$

(二) $P(0 \le X \le 1 | Y=1) = \int_0^1 3\dfrac{2^3}{(x+2)^4}dx = 0.7037$

三、設由兩個常態母體 $N(\mu_1, \sigma_1 = 1)$ 及 $N(\mu_2, \sigma_2 = 2)$ 中分別抽出大小為 n_1 及 n_2 的兩組獨立隨機樣本，且令 $\bar{X}_1$ 及 $\bar{X}_2$ 分別為此兩組隨機樣本的樣本平均數，則：（每小題 8 分，共 16 分）

(一) 試以 $\bar{X}_1$ ，$\bar{X}_2$ 估計 $\mu_1 - \mu_2$ 的 95%的信賴區間。

(二) 若為使(一)中所求的 95%信賴區間長度為最短，且 $n_1 + n_2 = 100$ 時，試問 n_1 與 n_2 應為多少？為多少？

答：

(一) 母體: $X_1 \sim N(\mu_1, \sigma_1^2 = 1) \perp X_2 \sim N(\mu_2, \sigma_2^2 = 4)$

樣本: $X_{11}, X_{12}, \cdots, X_{1n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2 = 1)$

$X_{21}, X_{22}, \cdots, X_{2n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2 = 4)$

點估計: $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \dfrac{1}{n_1} + \dfrac{4}{n_2})$

樞紐量: $\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{1}{n_1} + \dfrac{4}{n_2}}} \sim N(0,1)$

機率區間: $P(-z_{0.025} \le \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{1}{n_1} + \dfrac{4}{n_2}}} \le z_{0.025}) = 0.95$

信賴區間: $P((\bar{X}_1 - \bar{X}_2) - z_{0.025}\sqrt{\dfrac{1}{n_1} + \dfrac{4}{n_2}} \le \mu_1 - \mu_2 \le (\bar{X}_1 - \bar{X}_2) + z_{0.025}\sqrt{\dfrac{1}{n_1} + \dfrac{4}{n_2}}) = 0.95$

結論: $\mu_1 - \mu_2$ 之95%信賴區間為 $((\bar{X}_1 - \bar{X}_2) \mp 1.96\sqrt{\dfrac{1}{n_1} + \dfrac{4}{n_2}})$

(二) 信賴區間長度為 $2 \times 1.96\sqrt{\dfrac{1}{n_1} + \dfrac{4}{n_2}}$

$\min\limits_{n_1, n_2}(\dfrac{1}{n_1} + \dfrac{4}{n_2})$ s.t. $n_1 + n_2 = 100$

設 $L=\frac{1}{n_1}+\frac{4}{n_2}+\lambda(100-n_1-n_2)$

令 $\frac{\partial L}{\partial n_1}=-\frac{1}{n_1^2}-\lambda=0 \quad \Rightarrow \lambda=-\frac{1}{n_1^2}$ -------(1)

$\frac{\partial L}{\partial n_2}=-\frac{4}{n_2^2}-\lambda=0 \quad \Rightarrow \lambda=-\frac{4}{n_4^2}$ -------(2)

$\frac{\partial L}{\partial \lambda}=100-n_1-n_2=0$ --------------------(3)

以上(1)(2)(3)可得 $n_1=33.3333$

取 $n_1=33$ 及 $n_2=67$

四、設隨機變數 X 服從在 $[\theta_1,\theta_2]$ 上之均等分配（uniform distribution），其中 θ_1,θ_2 為未知參數，即隨機變數 X 的機率密度函數為 $f(x)=\begin{cases}\frac{1}{\theta_2-\theta_1}，\forall x\in[\theta_1,\theta_2]\\ 0，其他\end{cases}$

又令 $X_1,X_2,\ldots,X_n$ 為抽自 X 之一組大小於 n 之隨機樣本，則：（每小題 8 分，共 24 分）

(一) 隨機變數 X 之期望值 $E(X)=$？變異數 $V(X)=$？

(二) 試以動差估計法（the method of moments estimation）求 θ_1,θ_2 之點估量。

(三) 試以最大概似估計法（the method of maximum likelihood estimation）求 θ_1,θ_2 之點估計量。

答：

母體: $X\sim Unifrom(\theta_1,\theta_2)$

樣本: $X_1,X_2,\mathrm{L}\ ,X_n \overset{iid}{\sim} Unifrom(\theta_1,\theta_2)$

(一) $E(X)=\frac{\theta_1+\theta_2}{2}$ 與 $V(X)=\frac{(\theta_2-\theta_1)^2}{12}$

(二) 令 $E(X)=\frac{\theta_1+\theta_2}{2}=\bar{X}$ ---------------------------------(1)

$$E(X^2)=\frac{(\theta_2-\theta_1)^2}{12}+(\frac{\theta_1+\theta_2}{2})^2=\frac{\sum_{i=1}^{n}X_i^2}{n} \text{ ----------(2)}$$

以上(1)(2)解聯立可得，$\hat{\theta}_1=\bar{X}-\sqrt{\frac{n-1}{n}3S^2}$ 與 $\hat{\theta}_2=\bar{X}+\sqrt{\frac{n-1}{n}3S^2}$

其中 $S^2=\frac{\sum(X_i-\bar{X})^2}{n-1}$

(三) 令 $a=\theta_1$ 與 $b=\theta_2$

$$L(a,b)=f(x_1,x_2,\cdots,x_n;a,b)=\frac{1}{(b-a)^n}\prod_{i=1}^{n}I_{(a,b)}(x_i)$$

$$=\frac{1}{(b-a)^n}I_{(-\infty,x_{(1)})}(a)I_{(x_{(n)},\infty)}(b)$$

其中 $I_{(-\infty,x_{(1)})}(a)=\begin{cases}1, -\infty<a<x_{(1)}\\0, o.w.\end{cases}$, $I_{(x_{(n)},\infty)}(b)=\begin{cases}1, x_{(n)}<b<\infty\\0, o.w.\end{cases}$

因此, 母體未知參數a,b之MLE為 $\hat{a}_{MLE}=X_{(1)}$, $\hat{b}_{MLE}=X_{(n)}$

五、某汽車電瓶公司宣稱其所製造的小汽車電瓶之平均壽命為 3 年，標準差為 1 年。中華民國消費者文教基金會為了檢驗該公司的宣稱是否屬實，在市場上隨機抽取五個該公司所生產的小汽車電瓶做測試，結果得到電瓶壽命資料（單位：年）為 1.9, 2.4, 4.2, 3.5, 3.0。假設該公司所生產的小汽車電瓶壽命具常態分配 $N(\mu,\sigma)$，試根據上述資料，建立該公司所生產的小汽車電瓶壽命標準差 σ 之 95%信賴區間，並請根據此信賴區間的結果，判定是否有足夠的證據在顯著水準 $\alpha=0.05$ 下，判定是否相信該公司對電瓶壽命標準差的宣稱。（10 分）

答：令X表壽命(年)

母體: $X\sim N(\mu,\sigma^2)$

樣本: $X_1,X_2,\cdots,X_5\overset{iid}{\sim}N(\mu,\sigma^2)$

點估計: $S^2=\dfrac{\sum_{i=1}^{5}(X_i-\bar{X})^2}{5-1}$ (註：未知題意是否要把 μ 視為已知)

樞紐量: $\dfrac{(5-1)S^2}{\sigma^2}\sim\chi^2_{(5-1=4)}$

機率區間: $P(\chi^2_{(4)0.975}\le\dfrac{(5-1)S^2}{\sigma^2}\le\chi^2_{(4)0.025})=0.95$

信賴區間: $P(\sqrt{\dfrac{(5-1)S^2}{\chi^2_{(4)0.025}}}\le\sigma\le\sqrt{\dfrac{(5-1)S^2}{\chi^2_{(4)0.975}}})=0.95$

結論: σ 之95%信賴區間為

$$(\sqrt{\frac{(5-1)S^2}{\chi^2_{(4)0.025}}},\sqrt{\frac{(5-1)S^2}{\chi^2_{(4)0.975}}})=(\sqrt{\frac{(5-1)0.815}{11.14}},\sqrt{\frac{(5-1)0.815}{0.48}})=(0.5409,2.6061)$$

$H_0:\sigma=1$ vs $H_1:\sigma\neq1$

由於 σ 之95%信賴區間包含1，因此在顯著水準0.05下不拒絕 $H_0:\sigma=1$，我們沒有足夠證據去推論該公司對壽命標準差之宣稱不正確。

六、為瞭解台灣彩卷公司 4 星彩中獎號碼是否為隨機產生，記錄最近 30 期中獎號碼，得到下列資料：

數字	0	1	2	3	4	5	6	7	8	9
出現次數	11	12	10	9	9	10	15	14	17	13

試以 $\alpha=0.05$ 之顯著水準，檢定最近 30 期台灣彩卷公司 4 星彩中獎號碼之數字出現的次數分配可否合於均等分配（Uniform distribution）（即檢定 0,1,…,9 等 10 個數字被搖出之機率是否相等）？

(一) 可使用那一統計方法？（2 分）

(二) 如何利用(一)之方法進行檢定（請寫出完整的檢定步驟）？又檢定的結論為何？（12 分）

答：

(一) 卡方適合度檢定

(二)

	0	1	2	3	4	5	6	7	8	9
Oi	11	12	10	9	9	10	15	14	17	13
Ei	12	12	12	12	12	12	12	12	12	12

H_0：資料服從均勻分配 vs H_1：not H_0

T.S.: $\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i} \overset{d}{\sim} \chi^2_{(10-1=9)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{(9)0.05} = 16.92$

$\because \chi^{2*} = 5.5$ $\therefore$ don't reject H_0

我們沒有足夠證據去推論資料不服從均勻分配。