111 年公務人員高等考試三級考試試題

類　　科：統計
科　　目：統計學

吳迪老師

一、隨機變數 X 服從指數分配，其機率密度函數為 $f(x)=\begin{cases}e^{-x}, & \text{if } x>0\\ 0, & \text{if } x\le 0\end{cases}$。（已知自然對數值：ln(0.05)≒-3, ln(0.25)≒-1.386, ln(0.5)≒-0.693, ln(0.75)≒-0.288, ln(0.95)≒-0.051）（每小題 10 分，共 20 分）

㈠ 請計算其「中位數」及「期望值」。

㈡ 請計算其四分位距（interquartile range）。

1.《考題難易》：★★
2.《解題關鍵》：積分求中位數及四分位距,小心計算應可拿分
3.《命中特區》：P3-2~P3-8

【擬答】：

㈠

1. $\int_o^{Me} e^{-x}dx=\frac{1}{2}$

$\Rightarrow -e^{-x}\Big|_o^{Me}=1-e^{-Me}=\frac{1}{2}$

$\Rightarrow e^{-Me}=2^{-1}\Rightarrow Me=\ell_n 2=0.693$

2. $\int_o^{Q_1} e^{-x}dx=\frac{1}{4}$

$\Rightarrow -e^{-x}\Big|_o^{Q_1}=\frac{1}{4}\Rightarrow Q_1=-\ell_n 0.75=0.288$

$\int_o^{Q_3} e^{-x}dx=\frac{3}{4}$

$\Rightarrow -e^{-x}\Big|_o^{Q_3}=\frac{3}{4}\Rightarrow Q_3=-\ell_n 0.25=1.386$

$\therefore IQR=Q_3-Q_1=1.386-0.288=1.098$

二、研究者自某未知分配之母體 A，收集了樣本 $X_1 \dots X_n$；自另一母體 B，收集樣本 $Y_1 \dots Y_m$

㈠ 請描述中央極限定理，及使其成立所需要的假設。(14 分）

㈡ 若想比較母體 A 與 B 的中心點是否有差異，如何應用中央極限定理及其他機率性質？（6 分）

1.《考題難易》：★★
2.《解題關鍵》：考中央極限定理及其應用
3.《命中特區》：P6-15

【擬答】：

㈠

1. 中央極限定理(Central Limit Theorem)

若母體平均數為μ，變異數為σ^2。從母體中抽出無限多組樣本數為 n 的隨機樣本。若 n 趨近於無限大($n\ge 30$)，不論母體為何種分配，則樣本平均數$\bar{X}$之抽樣分配會趨近於常態分配

2. 中央極限定理所要的假設

⑴中央極限定理適用於大樣本($n\ge 30$)。若樣本數小於 30，則不適合用中央極限定理。

⑵若母體不為常態分配(即為偏態)，則樣本數 n 愈大，則$\overline{X}$之抽樣分配圖形愈接近對稱。若樣本數 n 趨近於無限大，則$\overline{X}$之抽樣分配會趨近於常態分配。

⑶若母體不為常態分配，但$(n \geq 30)$，則可利用常態分配查表求機率的近似值。

參考：林惠玲，陳正倉著統計學方法與應用三版(上，下)；2008 年 9 月；雙葉書廊；P407～P409

㈡若$n \geq 30, m \geq 30$，由中央極限定理

$\overline{X} \sim N(\mu_A, \frac{{\sigma_A}^2}{n})$

$\overline{Y} \sim N(\mu_B, \frac{{\sigma_B}^2}{m})$

但${\sigma_A}^2, {\sigma_B}^2$未知，可由 Slutzky Theorem

得樞紐量$Z = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{{S_A}^2}{n} + \frac{{S_B}^2}{m}}} \sim N(0,1)$

求得$\mu_A - \mu_B$的信賴區間

若信賴區間包含 0，即μ_A, μ_B無顯著差異

則 A 與 B 的中心點無顯著差異

三、A、B、C 為三種硬幣，其正面的概率分別為 1/3, 1/2, 2/3。今隨機選了一個，拋 8 次，得到 5 次正面 3 次反面。(每小題 10 分，共 20 分)

㈠ 以似然比（Likelihood ratio）方法估計這是那個硬幣的可能性最大。

㈡ 若 A 硬幣有 3 個，B 硬幣有 2 個，C 硬幣有 1 個；從這 6 個隨機選一個拋 8 次，得 5 次正面 3 次反面。以似然比方法估計這是那種硬幣。

1.《考題難易》：★★★
2.《解題關鍵》：考最大概似估計式
3.《命中特區》：P7-14～P7-17

【擬答】：

㈠

$A：L(\theta) = C_5^8(\frac{1}{3})^5(\frac{2}{3})^3 = 0.0683$

$B：L(\theta) = C_5^8(\frac{1}{2})^5(\frac{1}{2})^3 = 0.21875$

$C：L(\theta) = C_5^8(\frac{2}{3})^5(\frac{1}{3})^3 = 0.2731$

因為 C 的$L(\theta)$最大，所以 C 硬幣的可能性最大

㈡

$A：L(\theta) = \frac{3}{6} \times C_5^8(\frac{1}{3})^5(\frac{2}{3})^3 = 0.03415$

$B：L(\theta) = \frac{2}{6} \times C_5^8(\frac{1}{2})^5(\frac{1}{2})^3 = 0.07292$

$C：L(\theta) = \frac{1}{6} \times C_5^8(\frac{2}{3})^5(\frac{1}{3})^3 = 0.04552$

因為 B 的$L(\theta)$最大，所以 B 硬幣的可能性最大

四、令 *log* 為以 10 為底的對數、ε 為期望值 0 的誤差。 *y* 和 *x* 的關係可以下式 表達：$log(y_i) = 1 + \beta \cdot x_i + \varepsilon$。

請導出 β 的最小平方法估計公式。（20 分）

1.《考題難易》：★★
2.《解題關鍵》：考簡單迴歸分析最小平方法估計公式

3.《命中特區》：P10-20

【擬答】：

$$\log(y_i) = 1 + \beta x_i + \varepsilon$$

$$\Rightarrow \varepsilon = \log(y_i) - 1 - \beta x_i$$

$$\Rightarrow \text{SSE} = \sum_{i=1}^{n} \varepsilon_i{}^2 = \sum_{i=1}^{n} [\log(y_i) - 1 - \beta x_i]^2$$

$$\Rightarrow \frac{\partial SSE}{\partial \beta} = 2\sum_{i=1}^{n} [\log(y_i) - 1 - \beta x_i] \times (-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} [\log(y_i) - 1 - \beta x_i]\, x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i \log(y_i) - \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i{}^2 = 0$$

$$\Rightarrow \beta \sum_{i=1}^{n} x_i{}^2 = \sum_{i=1}^{n} x_i \log(y_i) - \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \beta = \frac{\sum_{i=1}^{n} x_i \log(y_i) - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i{}^2}$$

∴ β的最小平方估計式為 $\frac{\sum_{i=1}^{n} x_i \log(y_i) - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i{}^2}$

五、以下為 A、B 兩組中風病人手術後復發的時間（以年為單位）。

A	4.2,	6.5,	7.9,	13.2,	16
B	0.4,	1.2,	2.9,	5.6,	6.7

以 Wilcoxon rank sum 檢定，在 $\alpha = 0.05$ 顯著水準下檢定 H_1：A 組的術後效果較好。需列計算過程，不能用逼近方法。需算出 p-value。（20 分）

1.《考題難易》：★★★
2.《解題關鍵》：考 Wilcoxon rank sum 檢定,但要求的做法有點變化
3.《命中特區》：P12-25~P12-26

【擬答】：

$\begin{cases} H_0：\text{A 組的術後效果沒有較好} \\ H_1：\text{A 組的術後效果較好} \end{cases}$

A	4.2	6.5	7.9	13.2	16
rank	4	6	8	9	10
B	0.4	1.2	2.9	5.6	6.7
rank	1	2	3	5	7

$n_1 = n_2 = 5$，因為 B 的平均數較小

$\Rightarrow W = 1 + 2 + 3 + 5 + 7 = 18$

$$E(W) = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{5(5 + 5 + 1)}{2} = 27.5$$

$$Var(W) = \frac{n_1 n_2(n_1 + n_2 + 1)}{12} = \frac{5 \times 5(5 + 5 + 1)}{12} = 22.92$$

$$Z = \frac{W - E(W)}{\sqrt{Var(W)}} = \frac{18 - 27.5}{\sqrt{22.92}} = -1.98$$

$P - value = P(Z < -1.98) < 0.05 = \alpha \Rightarrow ReHo$

結論：有證據顯示 A 組的術後效果較好