104 年公務人員特種考試關務人員考試試題

考試別：關務人員考試
等別：四等考試
類科：關務統計
科目：統計學概要

一、某大學生去年有 1,000 位畢業同學找到了第一份工作，根據最近的調查，這 1,000 位畢業生的月薪平均 25,600 元，月薪的標準差為 2,200 元。試問：

㈠如果月薪的分布為一鐘形分布，有多少畢業生每月的薪資是介於 19,000 元及 32,200 元間？(10 分)

㈡如果月薪的分布不為鐘形分布，有多少畢業生每月的薪資介於 19,000 元及 32,200 元間？(10 分)

㈢請說明及比較㈠與㈡兩小題的回答。(5 分)

【擬答】：

㈠

$$X \sim N(\mu = 25600, \sigma^2 = 2200^2)$$

$$\Rightarrow P(19000 \le X \le 32200)$$

$$\Rightarrow P(\frac{19000 - 25600}{2200} \le Z \le \frac{32200 - 25600}{2200})$$

$$\Rightarrow P(-3 \le Z \le 3) = 0.9974$$

$$\Rightarrow 1000 \times 0.9974 = 997.4$$

$\therefore$ 約有997個畢業生每月薪資介於19000元~32200元之間

㈡

資料不為鐘形分布,利用chebyshev's Inequality

$$\Rightarrow P(\ \left|x - \mu\right| \le k\sigma\ ) \ge 1 - \frac{1}{k^2}$$

$$\Rightarrow P(19000 \le X \le 32200) = P(\ \left|X - 25600\right| \le 3 \times 2200)$$

$$\ge 1 - \frac{1}{3^2} = \frac{8}{9}\text{(下界)}$$

$$\Rightarrow 1000 \times \frac{8}{9} = 888.89$$

$\therefore$ 至少有888個畢業生每月薪資介於19000元 − 32200元之間

㈢㈠母體為鐘形分配,則利用標準常態分配查表求機率的準確值。
㈡母體非鐘形分配,則利用 chebyshev's Inequality 求機率的下界值。
二者的意義不相同。

二、有一機械零件製造公司員工生產零件數量為常態分配。員工 A 每天平均生產 75 個零件，標準差為 20 個;員工 B 每天平均生產 65 個零件，標準差為 21 個。試問:

㈠某一單日，員工 A 生產的零件高於員工 B 的機率為何？(10 分)

㈡在某一週內(平均工作 5 日)，員工 A 每日生產的平均零件數高於員工 B 的機率為何？(10 分)

㈢請說明造成㈠與㈡兩小題所得機率值有差異的原因為何？(5 分)

【擬答】：

㈠設 X 為員工 A 每天生產零件個數

$\Rightarrow X \sim N(\mu = 75, \sigma^2 = 400)$

設Y為員工B每天生產零件個數

$\Rightarrow Y \sim N(\mu = 65, \sigma^2 = 441)$

若X, Y為獨立的隨機變數

$\Rightarrow X - Y \sim N(\mu = 10, \sigma^2 = 841)$

$\Rightarrow P(X > Y) = P(X - Y > 0) = P(Z > \frac{0-10}{\sqrt{841}})$

$P(Z > -0.34) = 0.6331$

㈡

$\overline{X} \sim N(75, \frac{400}{5}), \overline{Y} \sim N(65, \frac{441}{5})$

若X, Y為獨立隨機變數

$\Rightarrow \overline{X} - \overline{Y} \sim N(10, \frac{841}{5})$

$\Rightarrow P(\overline{X} > \overline{Y}) = P(\overline{X} - \overline{Y} > 0) = P(Z > \frac{0-10}{\sqrt{\frac{841}{5}}})$

$= P(Z > -0.77) = 0.7794$

㈢㈠為單日生產零件個數

㈡為 5 日平均生產零件個數

二者抽樣分配標準差不同, 即 n 愈大, 樣本平均數抽樣分配標準差愈小, 所以二者所得機率值不同。

三、自 A 及 B 兩個國家進口大量的水果中抽驗是否有農藥殘留超標，結果發現來自 A 國家的 100 件水果抽樣中測出有 67 件有農藥殘留超標，來自 B 國家的 150 件水果抽樣中測出有 93 年有農藥殘留超標。由於進口農藥殘留超標的水果對人民的健康有很大的影響，政府想瞭解來自 A 國家的水果有農藥殘留超標的比例是否大於來自 B 國家的水果？假設顯著水準為 10%，試回答下列問題:

㈠請設立假說。(5 分)

㈡請以信賴區間法驗證假說。(10 分)

㈢請以 P 值法驗證假說。(10 分)

【擬答】:

㈠設 P_A 為 A 國家農藥殘留超標比例。

P_B 為 B 國家農藥殘留超標比例。

$$\Rightarrow \begin{cases} H_0 : P_A \le P_B \\ H_1 : P_A > P_B \end{cases}$$

㈡

$\widehat{P_A} = \frac{67}{100} = 0.67, \widehat{P_B} = \frac{93}{150} = 0.62$

$\widehat{P} = \frac{67+93}{100+150} = 0.64$

因為 $n_1 = 100 \ge 30, n_2 = 150 \ge 30$

$$\therefore \hat{P}_A - \hat{P}_B \sim N(P_A - P_B, \frac{\hat{P}(1-\hat{P})}{n_A} + \frac{\hat{P}(1-\hat{P})}{n_B})$$

$$\Rightarrow Z = \frac{(\hat{P}_A - \hat{P}_B) - (P_A - P_B)}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n_A} + \frac{\hat{P}(1-\hat{P})}{n_B}}} \sim N(0.1)$$

$$\Rightarrow P(Z \le Z_{0,10}) = 0.9$$

$$\Rightarrow P(\frac{(\hat{P}_A - \hat{P}_B) - (P_A - P_B)}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n_A} + \frac{\hat{P}(1-\hat{P})}{n_B}}} \le Z_{0.10}) = 0.9$$

$\therefore P_A - P_B$信賴度90%之單尾信賴區間為

$$(\hat{P}_A - \hat{P}_B - Z_{0.10}\sqrt{\frac{\hat{P}(1-\hat{P})}{n_A} + \frac{\hat{P}(1-\hat{P})}{n_B}}\ ,\ \infty)$$

$$\Rightarrow (0.67 - 0.62 - 1.28\sqrt{\frac{0.64 \times 0.36}{100} + \frac{0.64 \times 0.36}{150}}\ ,\ \infty)$$

$$\Rightarrow (-0.0293\ ,\ \infty)$$

因為此信賴區間包含0 $\Rightarrow$ notReH_0

結論:沒有證據顯示 A 國家的水果有農藥殘留超標的比例大於來自 B 國家的水果。

㈢

$$\begin{cases} H_0 : P_A \le P_B \\ H_1 : P_A > P_B \end{cases}$$

$$Z = \frac{(\hat{P}_A - \hat{P}_B) - (P_A - P_B)}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n_A} + \frac{\hat{P}(1-\hat{P})}{n_B}}} = \frac{(0.67 - 0.62) - 0}{\sqrt{\frac{0.64 \times 0.36}{100} + \frac{0.64 \times 0.36}{150}}}$$

=0.81

$\Rightarrow$ P-Value=P(Z>0.81)=0.209

因為$p - value = 0.209 > 0.1 = \alpha \Rightarrow$ notReH_0

結論:沒有證據顯示 A 國家的水果有農藥殘留超標的比例大於來自 B 國家的水果。

四、我們想知道學生考試準備時間與考試答對題數是否有線性關係，因此抽出 10 位學生為樣本，配適一條最小平方迴歸線，下列為所計算出的一些數值：

$$\sum (X - \bar{X})^2 = 32,\ \bar{X} = 3,\ \sum (Y - \bar{Y})^2 = 26,\ \bar{Y} = 4,\ \sum (X - \bar{X})(Y - \bar{Y}) = 28$$

㈠請問簡單最小平方迴歸線為何？(10 分)

㈡迴歸的變異數為何？(5 分)

㈢當 $X_p = 2.5$小時時，請問考試答對平均題目數的 95%信賴區間為何？(10 分)

【擬答】：

(一)

$$SS_X = \sum (X - \overline{X})^2 = 32, \overline{X} = 3$$

$$SS_Y = \sum (Y - \overline{Y})^2 = 26, \overline{Y} = 4$$

$$SS_{XY} = \sum (X - \overline{X})(Y - \overline{Y}) = 28, n = 10$$

設樣本最小平方迴歸線為

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X$$

1. $\hat{\beta} = \dfrac{SS_{XY}}{SS_X} = \dfrac{28}{32} = 0.875$

2.

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = 4 - 0.875 \times 3 = 1.375$$

$$\Rightarrow \hat{Y} = 1.375 + 0.875X$$

(二)迴歸方程式的變異數σ^2的估計式為

$$MSE = \frac{\sum_{\lambda=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2}{n-2} = \frac{SSE}{n-2} = \frac{SST - SSR}{n-2}$$

$$= \frac{SS_Y - \hat{\beta}^2 SS_X}{n-2} = \frac{26 - 0.875^2 \times 32}{10 - 2}$$

$$= 0.1875$$

(三)

$$X_P = 2.5 \Rightarrow \hat{Y} = 1.375 + 0.875 \times 2.5 = 3.5625$$

以 $\hat{Y}$ 估計 $\mu_{Y|X}$

$$\Rightarrow t = \frac{\hat{Y} - \mu_{Y|X}}{\sqrt{MSE\left[\frac{1}{n} + \frac{(X - \overline{X})^2}{SSx}\right]}} \sim t(n-2)$$

$$\Rightarrow P(-t_{0.025}(8) \le t_{0.025}(8)) = 0.95$$

$$\Rightarrow P(-t_{0.025}(8) \le \frac{\hat{Y} - \mu_{Y|X}}{\sqrt{MSE\left[\frac{1}{n} + \frac{(X - \overline{X})^2}{SSx}\right]}}) \le t_{0.025}(8)) = 0.95$$

$\therefore$ $\mu_{Y|X}$信賴度95%之信賴區間為

$$(\hat{Y} - t_{0.025}(8)\sqrt{MSE\left[\frac{1}{n} + \frac{(X - \overline{X})^2}{SSx}\right]}, \hat{Y} + t_{0.025}(8)\sqrt{MSE\left[\frac{1}{n} + \frac{(X - \overline{X})^2}{SSx}\right]}$$

$$\Rightarrow (3.5625 - 2.306\sqrt{0.1875\left[\frac{1}{10} + \frac{(2.5 - 3)^2}{32}\right]}), 3.5625 + 2.306\sqrt{0.1875\left[\frac{1}{10} + \frac{(2.5 - 3)^2}{32}\right]}$$

$$\Rightarrow (3.2346 \ , \ 3.8904)$$