102 年特種考試地方政府公務人員 衛生行政試題

等別：三等考試
類科：衛生行政
科目：生物統計學

一、為瞭解市售高麗菜農藥 Acephate 殘留狀況，定期在批發市場隨機抽取高麗菜樣本檢驗。

㈠除了平均數與標準差以外，請指出其他可以描述農藥殘留的集中趨勢與分散程度的量數各兩種，並說明每一種量數之計算方法或公式。
(8 分)

㈡請提出兩種適用來呈現這種類型資料的圖示法。(4 分)

㈢根據過去的經驗，市售高麗菜 Acephate 殘留量為常態分配，其平均值(μ)為 0.75 ppm，標準差(σ)為 0.25 ppm。請問理論上有多少%的市售高麗菜，其 Acephate 殘留量低於容許標準 1 ppm？(5 分)

㈣某日抽取一個 9 顆高麗菜的樣本，其平均 Acephate 殘留量低於容許標準 1 ppm 的機會為多少%？(5 分)

㈤該日抽取之 9 顆高麗菜，若其 Acephate 平均殘留量為 0.8 ppm，請計算 Acephate 平均殘留量的 95%信賴區間，並據以判斷這一批高麗菜的 Acephate 是否低於容許殘留量(1 ppm)。(8 分)

【擬答】

㈠

1. 集中趨勢除了平均數外，尚可使用中位數 Md 以及眾數 Mo 來衡量。
   中位數是將資料由小而大排列後，最中央的數，即有一半的資料小於等於中位數，有另一半的資料大於等於中位數。
   次數出現最多的稱眾數。
2. 分散程度的量數除了標準差外，亦可使用四分位距或變異係數
   四分位距(IQR)= $Q_3 - Q_1$
   其中 Q_1 是第 1 四分位數，Q_3 是第 3 四分位數
   變異係數是衡量相對離散量數，藉以標準差除以平均數，得到標準差佔平均數的比例。
   $$C.V = \frac{\sigma}{\mu} * 100\%$$

㈡屬量的資料可以利用莖葉圖、盒鬚圖來呈現資料分布的情況莖葉圖是一個與直方圖相類似的特殊工具，但與直方圖不同處是莖葉圖保留原始資料的資訊。
盒鬚圖是第 1 四分位數與第 3 四分位數中間的區域繪成方框，此圖中盒子包含資料中間 50%部分。並直接連接最小值與最大值，即鬚的部分，便可繪成盒鬚圖。

㈢假設 X_i :市售高麗菜 Acephate 殘留量

$$X_i \sim N(\mu = 0.75, \sigma^2 = 0.25^2)$$

$$P(X<1)=P(Z<\frac{1-0.75}{0.25})$$
$$=P(Z<1)=0.8413=84.13\%$$

㈣ $P(\overline{X}<1)=P(<Z<\frac{1-0.75}{0.25/\sqrt{9}})$
$$=P(Z<3)=0.9987=99.87\%$$

㈤Acephate 平均殘留量之 95%信賴區間為

$$\bar{x}\pm Z_{0.025}\cdot\frac{\sigma}{\sqrt{n}}$$
$$\Rightarrow 0.8\pm1.96\cdot\frac{0.25}{\sqrt{9}}$$
$$\Rightarrow[0.6367,0.9633]$$

信賴區間不包含 1，所以有顯著證據 Acephate 平均殘留量小於 1ppm，但因為此問題為單尾的問題，故在此考慮單尾信賴區間

$$\bar{x}+Z_{0.05}\cdot\frac{\sigma}{\sqrt{n}}$$
$$\Rightarrow 0.8+1.645\cdot\frac{0.25}{\sqrt{9}}$$
$$\Rightarrow[0,0.9371]$$

單尾信賴區間仍不包含 1
故有顯著證據 Acephate 平均殘留量小於 1ppm

二、為瞭解某一新型流感疫苗的效果，找到 100 名符合條件的自願者施打疫苗，另找 100 名條件相同者當作對照組。以下是其後流感罹患狀況：

	疫苗組	對照組
n	100	100
罹患流感人數	5	15

㈠請計算兩組流感的發生率及 95%信賴區間。(10 分)
㈡請計算兩組罹患流感發生率的差異及其 95%信賴區間。(5 分)
㈢請問注射流感疫苗是否會顯著降低罹患流感的風險 $(\alpha=0.05)$？請列出統計假設及相關步驟。(10 分)

【擬答】

㈠設疫苗組流感發生率為 p_1，對照組流感發生率為 p_2

$$\hat{p}_1=\frac{5}{100}=0.05$$

疫苗組流感發生率 p_1 之 95%信賴區間為

$$\hat{p}_1 \pm Z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}}$$

$$\Rightarrow 0.05 \pm 1.96 \cdot \sqrt{\frac{0.05\times 0.95}{100}}$$

$$\Rightarrow [0.0073, 0.0927]$$

$$\hat{p}_2 = \frac{15}{100} = 0.15$$

對照組流感發生率 p_2 之 95%信賴區間為

$$\hat{p}_2 \pm Z_{0.025}\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

$$\Rightarrow 0.15 \pm 1.96 \cdot \sqrt{\frac{0.15\times 0.85}{100}}$$

$$\Rightarrow [0.08001, 0.21999]$$

㈡兩組罹患流感發生率差 $p_2 - p_1$ 的 95%信賴區間為

$$(\hat{p}_2 - \hat{p}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{m} + \frac{\hat{p}_1(1-\hat{p}_1)}{n}}$$

$$\Rightarrow (0.15-0.05) \pm 1.96\sqrt{\left(\frac{0.15\times 0.85}{100} + \frac{0.05\times 0.95}{100}\right)}$$

$$\Rightarrow [0.018, 0.182]$$

㈢ $\hat{p} = \dfrac{\sum x_i + \sum y_i}{n+m} = \dfrac{5+15}{100+100} = 0.1$

$H_0: p_1 = p_2 \quad H_1: p_1 < p_2$

$\alpha = 0.05$

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{0.05-0.15}{\sqrt{0.1\times 0.9\left(\frac{1}{100}+\frac{1}{100}\right)}} = -2.357 \in C$$

$C = \{Z^* < Z_{0.05} = -1.645\}$

拒絕 H_0 ，有顯著的證據說註冊流感疫苗顯著降低罹患流感的風險

三、某檢驗師想測試食用油高溫加熱時間（X）及酸價（Y）的關係，測試結果如下：

	加熱時間（小時）	酸價
	1.0	0.1
	3.0	0.2
	6.0	0.8
	9.0	1.2
	12.0	1.5
	16.0	1.8
	20.0	2.0
	24.0	2.2
	36.0	3.0
	48.0	4.0
平均數	17.5	1.68
標準差	15.1	1.21

$$\sum X = 175, \sum Y = 16.8, \sum XY = 455.9, \sum X^2 = 5103, \sum Y^2 = 41.5$$

㈠請完成 X 與 Y 的迴歸分析摘要表及該模型的決定係數（R2）（不需含 p-value）。（10 分）

㈡請完成 X 與 Y 的迴歸方程式，並說明兩個係數的意義。（10 分）

㈢根據這個迴歸方程式，若加熱時間為 30 小時，預期其酸價應該為多少？（3 分）

㈣請列出簡單線性迴歸分析之統計前提、虛無假設及對立假設。（12 分）

【擬答】

㈠ $$\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{455.9 - \frac{175 \times 16.8}{10}}{5103 - \frac{(175)^2}{10}} = \frac{161.9}{2040.5} = 0.0793$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 1.68 - 0.0793 \cdot 17.5 = 0.29225$$

$$SSR = \hat{\beta}_1^2 \cdot SS_X = 0.0793^2 \times 2040.5 = 12.8317$$

$$SSTO = SS_Y = 41.5 - \frac{16.8^2}{10} = 13.276$$

$$SSE = SSTO - SSR = 0.4443$$

AVOVA 表

	SS	df	MS	F
Regression	12.8317	1	12.8317	231.2
Error	0.4443	8	0.0555	

Total	13.276	9

判定係數 $R^2 = \dfrac{SSR}{SSTO} \times 100\% = \dfrac{12.8317}{13.276} = 96.65\%$

$$S(\hat{\beta}_1) = \sqrt{\frac{MSE}{SS_X}} = \sqrt{\frac{0.0555}{2040.5}} = 0.0052$$

$$S(\hat{\beta}_0) = \sqrt{MSE(\frac{1}{n} + \frac{\overline{X}^2}{SS_X})} = \sqrt{0.0555 \cdot (\frac{1}{10} + \frac{17.5^2}{2040.5})} = 0.1178$$

所以係數估計表為

數	數	準誤	值
數	9225	178	81
熱時間 X	793	052	.25

㈡迴歸方程式為 $\hat{y} = 0.29225 + 0.0793 \cdot X$

$\hat{\beta}_1 = 0.0793$，代表每增加一小時加熱時間，則酸價增加 0.0793 單位

$\hat{\beta}_0 = 0.29225$，代表不加熱時，基本的平均酸價為 0.29225

㈢ $\hat{Y}\Big|_{X=30} = 0.29225 + 0.0793 \times 30 = 2.67125$

㈣簡單線性迴歸模型可表示為：

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

前提假設即對 ε_i 作常態、同質、獨立的假設，在此我們將 X 視為常數，所以 Y 便可看作常態的線性組合，為常態分配。

若要檢定迴歸模型是否顯著，虛無假設與對立假設可設定為

$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$

四、請說明統計檢力（power）對假設檢定的意義。透過那些方法，可以提高檢力？（10 分）

【擬答】

檢定力(power)是在 H_1 假設為真的情況下，正確拒絕 H_0 的機率，可視為檢定的強度。

若要求較大的檢定力，就是要較小的 β，因為 $power = 1 - \beta$。

而 α 與 β 互相影響，所以當要有較小的 β，其代價必然是使得 α 上升。

但若要兩者同時降低，可行的辦法便是增加更多的樣本數。