

113年特種考試地方政府公務人員及 離島地區公務人員考試試題

考試別：地方政府公務人員考試

等別：三等考試

類科：統計

科目：抽樣方法與迴歸分析

考試時間：2小時

座號：_____

※注意：(一)可以使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、有母體如下：

i	1	2	3	4
y_i	10	30	10	50

考慮以下設計於該母體中選擇兩個樣本：

1. 將母體分為 (1, 2) 及 (3, 4) 兩組。
2. 以簡單隨機抽樣選擇一個母體單元 i ，觀察得 y_i 。
3. 若 $y_i > 20$ ，則以 i 所在該組之另一單元為第二個樣本單元，反之若 $y_i \leq 20$ ，則由另一組隨機選擇第二個樣本單元。例如若第一個被選到的單元是第 1 個單元，則下一個單元則從第 3 跟第 4 個單元中隨機選擇，若第一個被選擇到的是第 4 個單元，則第二個被選單元必須是第 3 個單元。

請回答下列問題：(每小題 5 分，共 20 分)

- (一) 請問單元 2 之包含機率 (inclusion probability)，亦即在本設計下選擇之樣本組合中包含單元 2 之機率。
- (二) 若以觀察值樣本平均 \bar{y} 為母體平均 μ 之估計量，請問該估計量在本設計及本母體下是否不偏？
- (三) 請計算 \bar{y} 之均方誤 (mean squared error)。
- (四) 請計算 \bar{y} 之變異數。

二、人行道垃圾筒的設置可以方便行人處理隨身垃圾，但也會因為附近住家丟棄居家垃圾而造成髒亂，某里長為了解里民是否贊成增設人行道垃圾筒，對里民進行一項問卷調查。該里家戶共 950 戶，居住狀況可以分為獨棟別墅 10 戶，每棟 4~5 戶之舊式公寓 50 棟共 240 戶，以及集合式大樓 4 座共 700 戶三類。抽樣設計及觀察之樣本資料彙整如下：

1. 獨棟別墅之 10 戶全查。贊成比例 0.2。
2. 以簡單隨機抽樣取出不放回選擇 5 棟舊式公寓，再調查所選公寓中之全體住戶。各樣本公寓戶數及贊成之比例如下：

公寓編號	公寓戶數	贊成比例
15	4	1
27	4	0.5
35	5	1
39	5	0.6
47	5	0

3. 每座集合式大樓以簡單隨機抽樣取出不放回各選擇 20 戶。各大樓戶數及贊成之樣本比例如下：

大樓編號	戶數	贊成比例
1	75	0.2
2	400	0.3
3	100	0.2
4	125	0.4

請回答下列問題：

- (一) 請問舊式公寓住戶之抽樣設計為何？並請以不偏估計推估舊式公寓住家贊成之比例，並推估該不偏估計量之變異數估計。(12分)
- (二) 請問集合式大樓住戶之抽樣設計為何？請以不偏估計推估集合式大樓住家贊成之比例，並推估該不偏估計量之變異數估計。(12分)
- (三) 請估計本里住家贊成之比例以及該估計量在 95% 信心水準下之最大誤差 ($z_{0.025}=1.96$)。(6分)

三、教師收集 30 位學生考試成績 (Y) 與讀書時間 (X) 資料，用以了解兩者之間的關係，得到以下的數據：

$$\sum_{i=1}^{30} x_i = 396, \sum_{i=1}^{30} y_i = 1,644.302, \sum_{i=1}^{30} x_i^2 = 5,852, \sum_{i=1}^{30} y_i^2 = 94,202.131,$$
$$\sum_{i=1}^{30} x_i * y_i = 23,255.832$$

(一)以 Y 做為應變數 (dependent variable)， X 做為自變數 (independent variable)，假設簡單線性迴歸模型為 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ，其中 ε_i 為相互獨立且具常態分配 $N(0, \sigma^2)$ 的隨機誤差，試以最小平方法 (least squares method) 求出 β_0 及 β_1 的估計值。(8 分)

(二)若以 $x_i^* = \frac{x_i - \bar{x}}{s_x}$ ， $y_i^* = \frac{y_i - \bar{y}}{s_y}$ 分別做為新的自變數及新的應變數 (其中

\bar{x} ， \bar{y} 為原本資料的樣本平均數， s_x ， s_y 為原本資料的樣本標準差)，建立新的迴歸模型 $Y_i^* = \beta_0^* + \beta_1^* X_i^* + \varepsilon_i^*$ ，試以最小平方法求出 β_0^* 及 β_1^* 的估計值。(8 分)

(三)若學生的居住地區分為北、中、南三個地區，今定義三個虛擬變數 (dummy variable) D_1, D_2, D_3 ，其中 $D_1=1$ 代表居住北部， $D_1=0$ 代表其他； $D_2=1$ 代表居住中部， $D_2=0$ 代表其他； $D_3=1$ 代表居住南部， $D_3=0$ 代表其他。如果以 Y 做為應變數， X, D_1, D_2, D_3 做為自變數建立複迴歸模型，請問有何問題？(4 分)

(四)如果以 Y 做為應變數， X, D_1, D_2 做為自變數建立的複迴歸方程式為 $\hat{Y} = 20.03 + 2.02X + 1.03D_1 + 3.12D_2$ ，試求以 Y 做為應變數， X, D_2, D_3 做為自變數建立的複迴歸方程式為何？(5 分)

四、利用 20 個樣本，計算複迴歸模型 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ 參數估計如下：

	估計值 (estimate)	標準誤 (standard error)
β_0	20.03	10.15
β_1	2.02	1.05
β_2	-1.03	2.01
β_3	3.12	1.56

並計算出複判定係數 (coefficient of multiple determination) 為 $R^2=0.8$ 。

(一) 計算並解釋調整複判定係數 (adjusted coefficient of multiple determination)。其與複判定係數的差異為何？(8 分)

(二) 在顯著水準 $\alpha=0.05$ 下，試檢定 $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ vs. $H_1: \beta_i$ 不全為 0, $i=1, 2, 3$ 。(10 分)

(三) 在顯著水準 $\alpha=0.05$ 下，試檢定 $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ 。(7 分)

參考之查表值：

$$F_{0.05}(3, 16)=3.239, \quad F_{0.05}(3, 18)=3.160, \quad F_{0.05}(3, 20)=3.098$$

$$t_{0.05}(16)=1.746, \quad t_{0.05}(18)=1.734, \quad t_{0.05}(20)=1.725$$

$$t_{0.025}(16)=2.120, \quad t_{0.025}(18)=2.101, \quad t_{0.025}(20)=2.086$$