

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、某市政府推動智慧便民服務計畫，希望了解各行政區的線上服務使用情形、案件處理效率、服務據點配置與行政區類型，是否會影響民眾對公共服務的滿意度。變數說明如下：

變數	說明
Y	民眾公共服務滿意度，0 至 100 分
X_1	每千人每月線上申辦件數
X_2	平均案件處理天數
X_3	每萬人服務據點數
D_1	一般市區：是為 1，否為 0
D_2	核心商業區：是為 1，否為 0

其中，偏遠地區作為基準組，因此偏遠地區之 $D_1 = 0$ 、 $D_2 = 0$ 。

(一)市政府先以每千人每月線上申辦件數 X_1 作為自變數，解釋民眾公共服務滿意度 Y 。建立簡單線性迴歸模型：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i, \quad i = 1, 2, \dots, 10$$

其中，誤差項 ε_i 相互獨立，且服從常態分配：

$$\varepsilon_i \sim N(0, \sigma^2)$$

已知 10 個行政區的摘要統計量如下：

$$\begin{aligned} n &= 10, & \sum_{i=1}^{10} X_{1i} &= 850, & \sum_{i=1}^{10} Y_i &= 740, \\ \sum_{i=1}^{10} X_{1i}^2 &= 80500, & \sum_{i=1}^{10} Y_i^2 &= 56410, & \sum_{i=1}^{10} X_{1i} Y_i &= 66200 \end{aligned}$$

請以最小平方法 (least squares method) 求出 β_0 與 β_1 的估計值，寫出配適後的簡單線性迴歸模型，並計算判定係數 R^2 ，說明其意義。(15 分)

(二)延續第(一)題的簡單線性迴歸模型，請建立變異數分析表，並在顯著水準 $\alpha = 0.05$ 下檢定迴歸模型是否顯著。已知臨界值為： $F_{0.05}(1,8) = 5.32$ (12 分)

(三)市政府進一步同時考慮線上申辦件數、平均案件處理天數與服務據點數對滿意度的影響，建立複迴歸模型：

$$Y_i = \beta_0 + \beta_1 W_{1i} + \beta_2 W_{2i} + \beta_3 W_{3i} + \varepsilon_i$$

其中：

$$W_1 = \frac{X_1 - 80}{20}, \quad W_2 = \frac{X_2 - 6}{2}, \quad W_3 = \frac{X_3 - 10}{4}$$

令設計矩陣 (design matrix) 為 X ，反應變數向量 (response vector) 為 y ，複迴歸模型可表示為：

$$y = X\beta + \varepsilon$$

且已知：

$$(X'X)^{-1} = \begin{bmatrix} 0.125 & 0 & 0 & 0 \\ 0 & 0.125 & 0 & 0 \\ 0 & 0 & 0.125 & 0 \\ 0 & 0 & 0 & 0.125 \end{bmatrix}, \quad X'y = \begin{bmatrix} 560 \\ 40 \\ -32 \\ 24 \end{bmatrix}$$

請利用最小平方法估計 β ，寫出複迴歸估計模型，並預測某行政區在 $X_1 = 100$ 、 $X_2 = 5$ 、 $X_3 = 14$ 時的公共服務滿意度。(13 分)

(四)市政府為了解不同類型行政區的公共服務滿意度差異，將行政區分為三類：偏遠地區、一般市區與核心商業區。其中以「偏遠地區」作為基準組，並設定兩個屬性自變數如下：

$$D_1 = \begin{cases} 1, & \text{一般市區} \\ 0, & \text{非一般市區} \end{cases} \quad D_2 = \begin{cases} 1, & \text{核心商業區} \\ 0, & \text{非核心商業區} \end{cases}$$

根據資料，估計得到下列迴歸式：

$$\hat{Y} = 46.80 + 0.32X_1 - 1.45X_2 + 5.20D_1 + 10.40D_2$$

請解釋 5.20 與 10.40 兩個迴歸係數的意義。在控制線上申辦件數與平均案件處理天數後，核心商業區的預測滿意度比一般市區高多少分？(10 分)

二、某政府機關希望了解員工使用生成式 AI 協助公務處理的情形。研究團隊依不同部門進行分層抽樣，調查員工每週使用生成式 AI 協助公務的時數。（每小題 10 分，共 50 分）

(一)研究團隊完成比例配置抽樣後，得到下列樣本統計量：

部門	母體人數	樣本數	樣本平均數	樣本變異數
行政部門	1,200	80	7.5	36
技術部門	1,800	120	9.0	81
服務部門	900	60	6.0	25
稽核部門	600	40	8.0	64
合計	4,500	300		

請估計全機關員工每週使用生成式 AI 協助公務的平均時數，並建立全機關平均使用時數的 95% 信賴區間。

(二)研究團隊在正式抽樣前進行前測，得到各部門員工使用生成式 AI 時數的標準差如下：

部門	母體人數	前測標準差
行政部門	1,200	6
技術部門	1,800	9
服務部門	900	5
稽核部門	600	8

若總樣本數為 300，請使用 Neyman allocation 計算各部門建議樣本數，並說明配置結果的意義。

(三)正式調查後，實際回收情況如下：

部門	原抽樣本數	有效回覆數	回覆者平均使用時數
行政部門	80	72	7.4
技術部門	120	96	9.1
服務部門	60	54	6.2
稽核部門	40	28	7.9

若原本採比例配置，各層基本權重皆為 15，請計算各部門非回覆調整後權重，並估計非回覆調整後全機關員工每週平均使用生成式 AI 協助公務的時數，同時說明若不處理非回覆可能造成的問題。

(四)研究團隊發現，不同部門完成一份有效問卷的成本不同，資料如下：

部門	母體人數	前測標準差	單位調查成本
行政部門	1,200	6	1
技術部門	1,800	9	2
服務部門	900	5	1
稽核部門	600	8	3

若總調查成本限制為 500，請使用成本限制下的最適配置計算各部門建議樣本數，並說明此配置與 Neyman allocation 的差異。

(五)若政府機關希望估計全機關員工每週使用生成式 AI 協助公務的平均時數，並希望在 95% 信心水準下估計誤差不超過 0.5 小時。根據前測資料，估計全體標準差為 7.5，母體總人數為 4,500。請分別計算不考慮有限母體修正與考慮有限母體修正時所需樣本數；若預估問卷回覆率為 80%，請計算原始應抽樣本數。