

等 別：三等考試

類 科：統計

科 目：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。

一、假定 $y$ 表示公司產品一年銷售量。與 $y$ 相關的變數（如廣告費用、人事成本等）有 $x_1, x_2, \dots, x_p$ ，我們考慮線性迴歸（Regression）模型：

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$$

其中 $\varepsilon$ 服從常態分配。公司一資深經理認為，變數 $x_2$ 為 $x_1$ 的兩倍效力而 $x_3$ 又為 $x_2$ 的1/3倍效力。請問當我們有 $n$ 個觀察值 $(y_i, x_{i1}, \dots, x_{ip}), i = 1, \dots, n$ 時，您要如何驗證這個經理的說法是否正確（請說明您的假設）。（25分）

二、假定我們有一線性迴歸模型：

$$y_i = x_i' \beta + \varepsilon_i, i = 1, \dots, n$$

其中 $\varepsilon_1, \dots, \varepsilon_n$ 為iid隨機變數其平均數為0及變異數為 $\sigma^2$ 。令 $\hat{y}_i$ 為 $y_i$ 之預測值。

(一)請問若 $\varepsilon_1, \dots, \varepsilon_n$ 不具有常態分配，那 $\sum_{i=1}^n (y_i - \hat{y}_i) = 0$ 是否仍是對的。請驗證。（13分）

(二)我們假設 $\varepsilon_1, \dots, \varepsilon_n$ 為iid  $N(0,1)$ 的隨機變數。令 $\bar{\varepsilon}$ 為 $\varepsilon_1, \dots, \varepsilon_n$ 的平均，請找出 $\bar{\varepsilon}$ 的 $100(1-\alpha)\%$ 信賴區間。（12分）

三、我們考慮一個工業產品某一品質變數 $y$ ，它滿足下面迴歸模型：

$$y_i = x_i' \beta + \varepsilon_i, i = 1, \dots, n$$

其中 $x_i$ 包含截距項的 $(p+1)$ 向量且 $\varepsilon_1, \dots, \varepsilon_n$ 為iid  $N(0, \sigma^2)$ 的變數。若有一向量 $x_a$ 想知道這時對應變數 $y_a$ 的表現。我們重複觀察 $m$ 次產生下面樣本模型：

$$y_{aj} = x_a' \beta + \varepsilon_{aj}, j = 1, \dots, m$$

其中 $\varepsilon_{aj}$ 's與 $\varepsilon_i$ 's獨立且具相同分配，令 $y_a = \frac{1}{m} \sum_{j=1}^m y_{aj}$ 。

(一)找出 $y_a$ 的 $100(1-\alpha)\%$ 信賴區間。（13分）

(二)令 $E(y_a | x_a) = x_a' \beta$ ，考慮假設 $H_0: E(y_a | x_a) = 3$  vs.  $H_1: E(y_a | x_a) > 3$ ，請導出顯著水準為 $\alpha$ 的檢定。（12分）

(請接背面)

等 別：三等考試  
類 科：統計  
科 目：迴歸分析

四、令 $y$ 表示嬰兒身高。一奶粉公司考慮如何選取對身高有益的元素（變數）。現有 $x_1, x_2, x_3, x_4$ 四種變數供選取。若 $A, B$ 為 $x_1, x_2, x_3, x_4$ 的部分集合。我們令 $F(A|B) = \frac{MSR(A|B)}{MSE(A, B)}$ ，

表示當 $B$ 集合變數已在模型內時 $A$ 集合變數的部分 $F$ （partial  $F$ ）檢定統計量。再令 $P(A|B)$ 表示 $F(A|B)$ 觀察值的 $P$ 值。我們現在有下面 $P$ 值：

$$P(x_1) = 0.03, \quad P(x_2) = 0.02, \quad P(x_3) = 0.01, \quad P(x_4) = 0.02$$

$$P(x_i|x_1) = 0.04, \quad i = 2, 3, \quad P(x_4|x_1) = 0.01$$

$$P(x_i|x_2) = 0.02, \quad i = 1, 3, \quad P(x_4|x_2) = 0.01$$

$$P(x_1|x_3) = 0.04, \quad P(x_2|x_3) = 0.05, \quad P(x_4|x_3) = 0.035$$

$$P(x_i|x_4) = 0.05, \quad i = 1, 2, 3, \quad P(x_i|x_1, x_2) = 0.06, \quad i = 3, 4$$

$$P(x_i|x_1, x_3) = 0.07, \quad i = 2, 4, \quad P(x_i|x_1, x_4) = 0.07, \quad i = 2, 3$$

$$P(x_i|x_2, x_3) = 0.06, \quad i = 1, 4, \quad P(x_i|x_2, x_4) = 0.06, \quad i = 1, 3$$

$$P(x_i|x_3, x_4) = 0.06, \quad i = 1, 2$$

我們考慮 Forward selection 選取變數。在下面問題請說明理由及每階段選取結果。

(一)當  $\alpha = 0.05$  時，請找出選取的模型。（15分）

(二)當  $\alpha = 0.03$  時，請找出選取的模型。（10分）