

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。

(三)本科目除專門名詞、數理公式或表格外，應使用本國文字作答。

參考之查表值：F分布 $\alpha=0.05$ 臨界值 $F_{0.05}(df1, df2)$ 。

| | | df1 | |
|-----|----|--------|--------|
| | | 1 | 2 |
| df2 | 17 | 4.4513 | 3.5915 |
| | 43 | 4.0670 | 3.2145 |
| | 44 | 4.0617 | 3.2093 |
| | 87 | 3.9506 | 3.1013 |
| | 88 | 3.9493 | 3.1001 |

一、(一)何謂多重共線性 (multicollinearity)？多重共線性對估計結果有何影響？如何偵測複迴歸模型中存在多重共線性？請詳述所需要的判斷準則。(12分)

(二)一位分析師進行迴歸分析資料並配適複迴歸模型如(1)。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_7 X_{7i} + \varepsilon_i, \quad i = 1, \dots, 100 \quad (1)$$

所獲得初步結果如圖 1。請用圖 1 部分統計電腦套裝軟體輸出結果，說明這位分析師所配適的模型是否合適？如果模型(1)不合適，請說明原因並提供所有可以解決問題的方法。(10分)

圖1

| Analysis of Variance | | | | | |
|----------------------|----|--------------------|----------------|---------|---------|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 7 | 3019.404 | 431.343 | 78.26 | <0.0001 |
| Error | 92 | 507.065 | 5.512 | | |
| Corrected Total | 99 | 3526.470 | | | |
| Parameter Estimates | | | | | |
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
| Intercept | 1 | -0.087 | 0.249 | -0.35 | 0.727 |
| X1 | 1 | 4.874 | 3.855 | 1.26 | 0.209 |
| X2 | 1 | 3.790 | 5.895 | 0.64 | 0.522 |
| X3 | 1 | 2.972 | 5.855 | 0.51 | 0.613 |
| X4 | 1 | 4.964 | 5.723 | 0.87 | 0.388 |
| X5 | 1 | 0.930 | 6.714 | 0.14 | 0.890 |
| X6 | 1 | -12.532 | 6.136 | -2.04 | 0.044 |
| X7 | 1 | 13.154 | 4.479 | 2.94 | 0.004 |

二、醫院分析師希望研究患者滿意度 (Y) 與患者年齡 (X_1 , 以年為單位), 疾病嚴重程度指數 (X_2) 以及焦慮指數 (X_3) 之間的關係。分析師隨機選擇了 46 名患者並收集了數據。請使用圖 2 部分統計電腦套裝軟體輸出結果來回答以下問題。

圖 2

Dependent Variable: Y
Adjusted R-Square Selection Method

| Number in Model | Adjusted R-Square | R-Square | SSE | Variables in Model |
|-----------------|-------------------|----------|-----------|--------------------|
| 2 | 0.6610 | 0.6761 | 4330.4997 | x1 x3 |
| 3 | 0.6595 | 0.6822 | 4248.8407 | x1 x2 x3 |
| 2 | 0.6389 | 0.6550 | 4613.0002 | x1 x2 |
| 1 | 0.6103 | 0.6190 | 5093.9155 | x1 |
| 2 | 0.4437 | 0.4685 | 7106.3941 | x2 x3 |
| 1 | 0.4022 | 0.4155 | 7814.3912 | x3 |
| 1 | 0.3491 | 0.3635 | 8509.0444 | x2 |

SS Total=13369

- (一)請計算額外平方和 (extra sum of squares) $SSR(X_2 | X_1, X_3) = ?$ (4 分)
- (二)假設這位分析師採用模型是 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_3 X_{3i} + \varepsilon_i$ (2)
該分析師想知道在模型(2)之下, 增加疾病嚴重程度指數 (X_2) 此額外變數, 解釋其在顯著水準 $\alpha=5\%$ 下是否有顯著的貢獻, 並敘述對立假設、檢定統計量之值、決策法則和結論。(8 分)
- (三)假設這位分析師採用模型是 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ (3)
請檢定疾病嚴重程度指數 (X_2) 和焦慮指數 (X_3) 兩個解釋變數是否可以從模型(3)中刪除, 也就是在已經有患者年齡 (X_1) 解釋變數之下, 解釋變數 X_2 和 X_3 可否從模型中移除? 請在顯著水準 $\alpha=5\%$ 檢定, 並協助敘述對立假設、檢定統計量之值、決策法則和結論。在本小題的檢定問題中, 請試述需要作何假設, 才能執行這些統計檢定。(10 分)

三、一位分析師考慮對三組數據配適一個簡單迴歸模型 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ，其中 β_0 、 β_1 為參數， ε 為隨機誤差，且假設其為具均數 0，標準差 σ 之常態分配。

(一)配適模型後，三組數據之殘差分析圖分別為 3(a)、3(b)、3(c)，請分別說明配適迴歸模型是否恰當？若模型不合適或偏離模型假設時，請指出不恰當之處並請提出修正的方法。(21 分)

(二)在何種情況下，需要採用加權最小平方法 (Weighted least squares) 估計未知的參數？請協助提供散佈圖和殘差圖說明。(7 分)

圖3(a) 標準化殘差時間序列圖 (Standardized residuals vs. time)

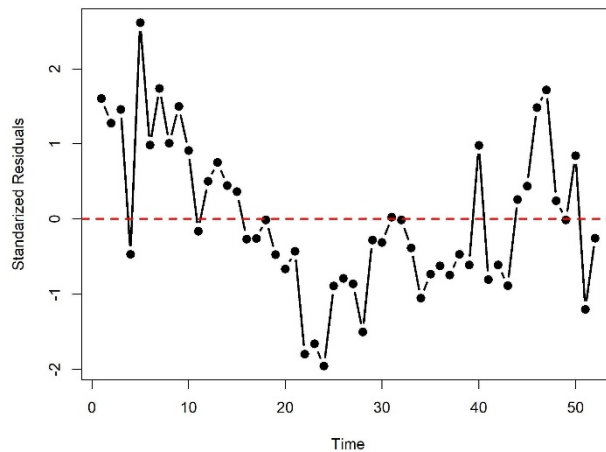


圖3(b) 標準化殘差對預測值圖 (Standardized residuals vs. \hat{y}_i)

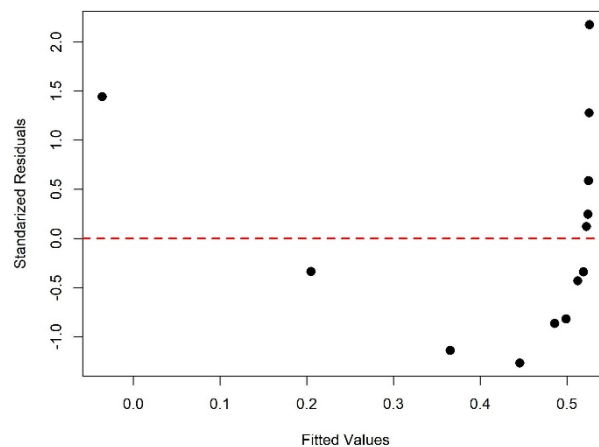
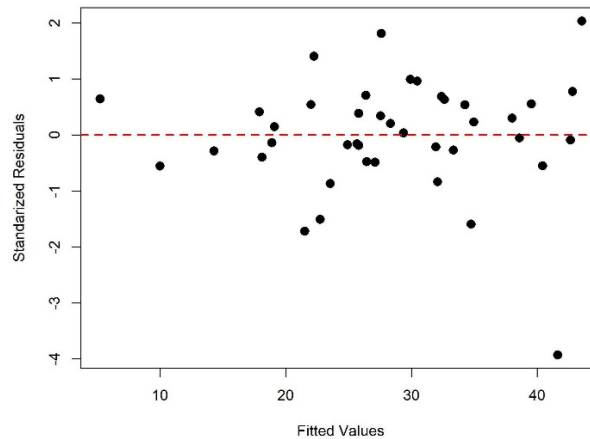


圖 3(c) 標準化殘差對預測值圖 (Standardized residuals vs. \hat{y}_i)



四、一位資料分析師受託分析一組數據，想要了解一個特定基因，稱之 GT 基因，是否有影響老鼠斷奶時的重量。該分析師預計配適模型 1 和模型 2。

Y=斷奶時的重量 (公克為單位)

X1=年齡 (以日為單位)

X2=品種 (品種 A=1, B=0)

X3=GT 基因 (有此基因=1, 無此基因=0)

X4=性別 (公老鼠=1, 母老鼠=0)

模型 1: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$,

模型 2: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{1i} X_{2i} + \varepsilon_i$,

請使用圖 4 和圖 5 中部分統計電腦套裝軟體輸出變異數分析 (ANOVA, Analysis of Variance) 回答下列問題：

- (一)請計算模型 1 和模型 2 的調整的複判定係數 R^2 (the adjusted R-squared)。試述其意義，並判斷何種模型為佳。(8 分)
- (二)在顯著水準 5% 下，請檢定「GT 基因」在模型 1 中是否影響老鼠的重量？(4 分)
- (三)請解釋在考慮模型 1 下，請說明如何檢定老鼠的性別之兩條迴歸線是相同的迴歸線。並請列出虛無假設、對立假設、檢定統計量及決策法則。(4 分)
- (四)在顯著水準 5% 下，請檢定 $X_{1i} X_{2i}$ 相乘項在模型 2 中是否對解釋反應變數 Y 有顯著貢獻？請試述虛無假設、檢定統計量之值、決策法則和結論，以及所需要之假設。請解釋 $X_{1i} X_{2i}$ 該項在迴歸模型的意義。(12 分)

圖 4 模型 1 的變異數分析

| Analysis of Variance | | | | | |
|----------------------|----|----------------|-------------|---------|--------|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 4 | 2067.1901 | 516.798 | 74.40 | <.0001 |
| Error | 88 | 611.2830 | 6.946 | | |
| Corrected Total | 92 | 2678.4731 | | | |

| Parameter Estimates | | | | | |
|---------------------|----|--------------------|----------------|---------|---------|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
| Intercept | 1 | -3.5571 | 1.5904 | -2.24 | 0.0278 |
| X1 | 1 | 0.7971 | 0.0540 | 14.75 | <.0001 |
| X2 | 1 | 3.1411 | 0.5614 | 5.59 | <.0001 |
| X3 | 1 | -2.1286 | 0.5958 | -3.57 | 0.0006 |
| X4 | 1 | -3.1229 | 0.6234 | -5.01 | <.0001 |

圖 5 模型 2 的變異數分析

| Analysis of Variance | | | | | |
|----------------------|----|----------------|-------------|---------|--------|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 5 | 2161.1224 | 432.224 | 72.68 | <.0001 |
| Error | 87 | 517.3508 | 5.947 | | |
| Corrected Total | 92 | 2678.4731 | | | |