# 《迴歸分析》

**試題評析**

本年度考題難易適中，只要同學觀念清楚，靈活應用，應可以拿到80分。

一、現有一實驗，取得有單一解釋變數與應變數之一組獨立數據$(x_i,y_i)$, $i=1,...,n$，如下：

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | −3 | −2 | −1 | −1 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 3 |
| $y_i$ | 2 | 6 | 14 | 9 | 15 | 14 | 13 | 17 | 12 | 16 | 16 | 13 |

擬以簡單線性迴歸模型$y_i=\beta_0+\beta_1 x_i+\in_i, i=1,...,n$，來描述上述應變數與解釋變數間之迴歸關係，其中$\beta_0,\beta_1$均為未知，並假設$\in_i, i=1,...,n$為i.i.d. $N(0,\sigma^2), \sigma^2>0$亦為未知。已知$\sum_{i=1}^{12} y_i=147, \sum_{i=1}^{12} y_i^2=2021, \sum_{i=1}^{12} x_i y_i=58$。

(一)試寫出估計未知參數向量$\theta=(\beta_0,\beta_1,\sigma^2)^T$之概似函數（likelihood function）。（10分）

(二)試求$\theta$之最大概似估計量（maximum likelihood estimator），及$\beta=(\beta_0,\beta_1)^T$最小平方估計量（least squares estimator）。（10分）

(三)試給出此迴歸分析之變異數分析表，及其$R^2$值。（10分）

(四)試求$\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1)^T$之95%聯合信賴域（joint confidence region），並說明$x$變數對解釋$y$之變異是否有幫助？（10分）

（$F_{0.025,1,9}=7.21, F_{0.025,1,10}=6.94, F_{0.025,2,10}=5.48, F_{0.05,1,10}=4.96, F_{0.05,1,9}=5.12, F_{0.05,2,10}=4.10,$
$z_{0.95}=1.65, z_{0.975}=1.96$）

**答**：

(一)Likelihood function：

$$L(\beta_0,\beta_1,\sigma^2)=(2\pi\sigma^2)^{-\frac{n}{2}}\cdot e^{\frac{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}}$$

(二) $$\ln L=-\frac{n}{2}(\ln 2\pi+\ln\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2$$

$$\begin{cases}\frac{\partial \ln L}{\partial \beta_0}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)=0\\ \frac{\partial \ln L}{\partial \beta_1}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)(x_i)=0\\ \frac{\partial \ln L}{\partial \sigma^2}=-\frac{n}{2}\cdot\frac{1}{\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2=0\end{cases}$$

高上高普特考



高上高普特考 www.get.com.tw/goldensun 台北市開封街一段 2 號 8 樓 02-23318268
【中壢】中壢市中山路 100 號 14 樓 03-4256899 【台中】台中市東區復興路四段 231-3 號 1 樓・04-22298699
【台南】台南市中西區中山路 147 號 3 樓之 1・06-2235868 【高雄】高雄市新興區中山一路 308 號 8 樓・07-2358996
【另有淡水・三峽・林口・羅東・中壢・逢甲・東海・中技】

$$\begin{cases} n\ \ \beta_0+\sum_{i=1}^{n}x_iB_1=\sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_i\beta_0+\sum_{i=1}^{n}x_i^2\beta_1=\sum_{i=1}^{n}x_iy_i \\ \sigma^2=\dfrac{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2}{n} \end{cases}$$

$$\therefore \hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{x},\ \hat{\beta}_1=\frac{\sum(x_i-\bar{x})(Y_i-\bar{Y})}{\sum(x_i-\bar{x})^2}$$

(三) $\sum x_i=0,\ \sum x_i^2=30$

$$S_{XX}=\sum x_i^2-\frac{(\sum x_i)^2}{n}=30-\frac{0^2}{12}=30$$

$$S_{XY}=\sum x_iy_i-\frac{(\sum x_i)(\sum y_i)}{n}=58-0=58$$

$$SSTO=S_{YY}=\sum y_i^2-\frac{(\sum y_i)^2}{n}=2021-\frac{(147)^2}{12}=220.25$$

$$SSR=\frac{S_{XY}^2}{S_{XX}}=\frac{58^2}{30}=112.1333$$

$$SSE=SSTO-SSR=108.1167$$

(1) ANOVA table

| Source | SS | df | MS | F – value |
|---|---|---|---|---|
| Regression | 112.1333 | 1 | 112.1333 | F = 10.37 |
| Error | 108.1167 | 10 | 10.8117 | |
| Total | 220.25 | 11 | | |

(2) $R^2=\dfrac{SSR}{SSTO}=\dfrac{112.1333}{220.25}=0.5091$

(四) $\hat{\beta}_1=\dfrac{S_{XY}}{S_{XX}}=\dfrac{58}{30}=1.9333$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=\frac{147}{12}-1.9333\times 0=12.25$$

(1) $\beta_0$ 之 Bonferroni C.I.

$$=\hat{\beta}_0\pm t_{\frac{0.05}{2\times 2}}(12-2)\cdot\sqrt{\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}\right)\cdot MSE}$$

$$=12.25\pm\sqrt{6.94}\cdot\sqrt{\left(\frac{1}{12}+\frac{0^2}{30}\right)\times 10.8117}$$

$$=12.25\pm 2.5=(9.75,14.75)$$

(2) $\beta_1$ 之 95% Bonferroni C.I.

$$=\hat{\beta}_1\pm t_{\frac{0.05}{2\times 2}}(12-2)\cdot\sqrt{\frac{MSE}{S_{XX}}}$$



高上高普特考 www.get.com.tw/goldensun 台北市開封街一段 2 號 8 樓 02-23318268
【中壢】中壢市中山路 100 號 14 樓 03-4256899 【台中】台中市東區復興路四段 231-3 號 1 樓・04-22298699
【台南】台南市中西區中山路 147 號 3 樓之 1・06-2235868 【高雄】高雄市新興區中山一路 308 號 8 樓・07-2358996
【另有淡水・三峽・林口・羅東・中壢・逢甲・東海・中技】

$$=1.9333\pm\sqrt{6.94}\cdot\sqrt{\frac{10.8117}{30}}$$

$$=1.9333\pm1.5815=(0.3518,3.5148)$$

(3) $\because \beta_1$ C.I.不包含 0

$\therefore$ reject $H_0:\beta_1=0$ ，表示 x 變數對解釋 Y 之變異有幫助

【高分閱讀】
1.(一)(二)推導MLE詳參秦大成老師講義第一回P.6。
2.(三)ANOVA table 詳參秦大成老師講義第一回P.31。
3.(四)聯合C.I.詳參秦大成老師講義第二回P.127。

二、某一工程師擬探討某機器之有效使用年限($y$)與使用頻率($x_1$)及其品牌類別間之關係。使用頻率為每週多少小時，品牌類別共有三種，擬建立 $y$ 與二解釋變數間之線性迴歸模型。

(一)由於品牌類別屬於離散型數據，在進行迴歸分析前，三種品牌可以二指標變數（indicator variables）（$x_2,x_3$）表之，如第一種品牌之$(x_2,x_3)=(0,0)$，試寫出其完整定義（$x_2,x_3$）方式。（8分）

(二)根據 $x=(x_1,x_2,x_3)^T$ 以及線性模型 $y(x)=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\in$，試寫出：（12分）

1. 可比較品牌對有效使用年限是否有差異之假設檢定，i.e. $H_0$ 與 $H_1$ 以 $\beta_i$, $i=0,...,3$ 表示為何？
2. 可比較兩兩品牌間差異之假設檢定之 $H_0$ 與 $H_1$。
3. 模型中不考慮 $x_1$ 與 $x_2,x_3$ 交互作用，其模型有何特性？

**答**：

(一)令 $x_2=\begin{cases}1,\text{第二品牌}\\0,\text{o.w.}\end{cases}$，$x_3=\begin{cases}1,\text{第三品牌}\\0,\text{o.w.}\end{cases}$

where $(x_2,x_3)=(0,0)$ 表第一種品牌
(1,0)表第二種品牌
(0,1)表第三種品牌

(二) $E(Y\mid x_2=0,x_3=0)=\beta_0+\beta_1x_1$

$E(Y\mid x_2=1,x_3=0)=\beta_0+\beta_1x_1+\beta_2x_2$

$E(Y\mid x_2=1,x_3=1)=\beta_0+\beta_1x_1+\beta_3x_3$

1. $\begin{cases}H_0:\beta_2=\beta_3=0\\H_1:\text{不全爲}0\end{cases}$

$\Leftrightarrow\begin{cases}H_0:\text{品牌對有效使用年限無顯著差異}\\H_1:\text{品牌對有效使用年限有顯著差異}\end{cases}$

2.① $\begin{cases}H_0:\beta_2=0\\H_1:\beta_2\neq0\end{cases}$

$\Leftrightarrow\begin{cases}H_0:\text{第一種品牌與第二種品牌對有效使用年限無顯著差異}\\H_1:\text{第一種品牌與第二種品牌對有效使用年限有顯著差異}\end{cases}$



高上高普特考 www.get.com.tw/goldensun 台北市開封街一段 2 號 8 樓 02-23318268
【中壢】中壢市中山路 100 號 14 樓 03-4256899 【台中】台中市東區復興路四段 231-3 號 1 樓・04-22298699
【台南】台南市中西區中山路 147 號 3 樓之 1・06-2235868 【高雄】高雄市新興區中山一路 308 號 8 樓・07-2358996
【另有淡水・三峽・林口・羅東・中壢・逢甲・東海・中技】

② $\begin{cases} H_0 : \beta_3 = 0 \\ H_1 : \beta_3 \neq 0 \end{cases}$

$\Leftrightarrow \begin{cases} H_0$：第一種品牌與第三種品牌對有效使用年限無顯著差異 $\\ H_1$：第一種品牌與第三種品牌對有效使用年限有顯著差異 $\end{cases}$

3.第一種品牌與第二種品牌迴歸模型爲平行線，截距差爲 $\beta_2$

第一種品牌與第三種品牌迴歸模型爲平行線，截距差爲 $\beta_3$

【高分閱讀】
1. 定義啞變數：詳參秦大成老師講義第三回 P.51。
2. (1)(2)兩小題：啞變數解 ANOVA，詳參秦大成老師講義第三回 P.51 敘述及例題 8。
3. 第3小題：啞變數與交互作用模型分析詳參秦大成老師講義第三回P.42上課補充內容。

三、在一項有關體脂肪之研究中，收集到11位健康女姓之資料。擬探討應變數體脂肪($y$)，與解釋變數手臂三項肌皮摺厚度($x_1$)、臀圍($x_2$)及臂中圍($x_3$)之相關性。根據這11筆數據，在 $y$ 服從常態分布假設下，擬健立一 $y$ 與 $x_i$ 間之線性模型

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \in_i$

首先得到以下 $x_1, x_2, x_3$ 之相關矩陣（correlation matrix）

$$R = \begin{pmatrix} 1.00 & 0.94 & 0.35 \\ 0.94 & 1.00 & 0.03 \\ 0.35 & 0.03 & 1.00 \end{pmatrix}$$

(一)試求 $R^{-1}$，並求各解釋變數之變異膨脹因子（variance inflation factors, VIF）之值。（10分）

(二)說明 $x_i$ 間是否有共線性（multicollinearity）？若有共線性，對建立上述模型有何影響？（10分）

(三)完成下列變異數分析表（ANOVA table）。（10分）

| 來源（Source） | 平方和(SS) | 自由度（df） | 均方（MS） | F |
|---|---|---|---|---|
| 迴歸（Regression） | 202.59 | | | |
| $x_1$ | 183.92 | | | |
| $x_2 \mid x_1$ | 9.42 | | | |
| $x_3 \mid x_1, x_2$ | 9.25 | | | |
| 誤差（Error） | 28.79 | | | |
| 總和（Total） | 231.38 | | | |

(四)在 $\alpha = 0.05$ 水準下檢定 $x_2, x_3$ 對解釋體脂肪多寡是否有幫助？（10分）

即檢定 $H_0 : \beta_2 = \beta_3 = 0$ v. s. $\beta_2 \neq 0$ or $\beta_3 \neq 0$。

（$F_{0.025,2,6} = 7.26, F_{0.025,2,7} = 6.54, F_{0.05,2,6} = 5.14, F_{0.05,2,7} = 4.74, t_{0.025,2} = 4.30,$
$t_{0.025,7} = 2.365, t_{0.05,2} = 2.92, t_{0.05,7} = 1.895$）



高上高普特考 www.get.com.tw/goldensun 台北市開封街一段 2 號 8 樓 02-23318268
【中壢】中壢市中山路 100 號 14 樓 03-4256899 【台中】台中市東區復興路四段 231-3 號 1 樓・04-22298699
【台南】台南市中西區中山路 147 號 3 樓之 1・06-2235868 【高雄】高雄市新興區中山一路 308 號 8 樓・07-2358996
【另有淡水・三峽・林口・羅東・中壢・逢甲・東海・中技】

**答**：

(一)

(1) $A_{11}=\begin{vmatrix}1 & 0.03\\ 0.02 & 1\end{vmatrix}=0.9994,\ A_{12}=-\begin{vmatrix}0.94 & 0.03\\ 0.35 & 1\end{vmatrix}=-0.9295$

$A_{13}=\begin{vmatrix}0.94 & 1\\ 0.35 & 0.03\end{vmatrix}=-0.3218,\ A_{21}=-\begin{vmatrix}0.94 & 0.35\\ 0.03 & 1\end{vmatrix}=-0.9295$

$A_{22}=\begin{vmatrix}1 & 0.35\\ 0.35 & 1\end{vmatrix}=0.8775,\ A_{23}=-\begin{vmatrix}1 & 0.94\\ 0.35 & 0.03\end{vmatrix}=0.299$

$A_{31}=\begin{vmatrix}0.94 & 0.35\\ 1 & 0.03\end{vmatrix}=-0.3218,\ A_{32}=-\begin{vmatrix}1 & 0.35\\ 0.94 & 0.03\end{vmatrix}=0.299$

$A_{33}=\begin{vmatrix}1 & 0.94\\ 0.94 & 1\end{vmatrix}=0.1164$

$$\text{adj}(A)=\begin{bmatrix}0.9994 & -0.9295 & -0.3218\\ -0.9295 & 0.8775 & 0.299\\ -0.3218 & 0.299 & 0.1164\end{bmatrix}^{T}$$

$$=\begin{bmatrix}0.9994 & -0.9295 & -0.3218\\ -0.9295 & 0.8775 & 0.299\\ -0.3218 & 0.299 & 0.1164\end{bmatrix}$$

$$\det(A)=1+0.00987+0.00987-0.1225-0.0009-0.8836$$
$$=0.01274$$

$$A^{-1}=\frac{\text{adj}(A)}{\det A}=\begin{bmatrix}78.4458 & -92.959 & -25.259\\ -92.959 & 68.8776 & 23.4694\\ -25.259 & 23.4694 & 9.1366\end{bmatrix}$$

(2) $C=(X^{T}X)^{-1}=\begin{bmatrix}78.4458 & -92.959 & -25.259\\ -92.959 & 68.8776 & 23.4694\\ -25.259 & 23.4694 & 9.1366\end{bmatrix}$

① $X_1=\beta_0'+\beta_2'X_2+\beta_3'X_3$

$VIF_1=C_{11}=78.4458$

② $X_2=\beta_0''+\beta_1''X_1+\beta_3''X_3$

$VIF_2=C_{22}=68.8776$

③ $X_3=\beta_0'''+\beta_1'''X_1+\beta_2'''X_2$

$VIF_3=C_{33}=9.1366$

(二)

(1) $X_1$ 與 $X_2$ 高度共線性

$X_2$ 與 $X_3$ 低度共線性

高點·高上高普特考



高上高普特考 www.get.com.tw/goldensun 台北市開封街一段 2 號 8 樓 02-23318268
【中壢】中壢市中山路 100 號 14 樓 03-4256899 【台中】台中市東區復興路四段 231-3 號 1 樓・04-22298699
【台南】台南市中西區中山路 147 號 3 樓之 1・06-2235868 【高雄】高雄市新興區中山一路 308 號 8 樓・07-2358996
【另有淡水・三峽・林口・羅東・中壢・逢甲・東海・中技】

$X_1$與$X_3$中度共線性

(2)當模型存在共線性時

①全部係數檢定 F－test 與個別係數檢定 t－test 產生矛盾。

②複判定係數$R^2$很高，但是許多迴歸係數的t檢定卻都傾向不顯著異於0。

③複判定係數$R^2$很高，但許多偏判定係數卻很低。

④理論上很重要的自變數，它們的t檢定結果卻不顯著。

⑤將自變數變少，發現迴歸係數估計值變動很大。

(三) $SSE(x_1) = SSTO - SSR(x_1) = 231.38 - 183.92 = 47.46$

$SSR(x_2 \mid x_1) = SSE(x_1) - SSE(x_1x_2) = 47.46 - SSE(x_1x_2) = 9.42$

$\Rightarrow SSE(x_1x_2) = 38.04$

| 來源 | SS | df | MS | F |
|---|---|---|---|---|
| 迴歸(Regression) | 202.59 | 3 | 67.53 | 16.4191 |
| $x_1$ | 183.92 | 1 | 183.92 | $F = \frac{183.92}{\frac{47.46}{(11-2)}} = 34.877$ |
| $x_2 \mid x_1$ | 9.42 | 1 | 9.42 | $F = \frac{9.42}{\frac{38.04}{(11-3)}} = 1.981$ |
| $x_3 \mid x_1x_2$ | 9.25 | 1 | 9.25 | 2.249 |
| 誤差(error) | 28.79 | 7 | 4.1129 | |
| 總和(Total) | 231.38 | 10 | | |

(四) $SSR(x_2x_3 \mid x_1) = SSR(x_1x_2x_3) - SSR(x_1) = 202.59 - 183.92 = 18.67$

① $\begin{cases} H_0 : \beta_2 = \beta_3 = 0 \\ H_1 : \beta_2 \neq 0 \text{或} \beta_3 \neq 0 \end{cases}$

② $C = \{F \mid F > F_{0.05}(2, 11-4) = 4.74\}$

③ $F = \frac{SSR(x_2x_3 \mid x_1)/2}{SSE(x_1x_2x_3)/7} = \frac{18.67/2}{4.1129} = 2.27 \notin C$

$\therefore$ Not reject $H_0, x_2 \cdot x_3$對解釋體脂肪無幫助

【高分閱讀】

1. 求逆矩陣：詳參秦大成老師講義第二回 P.5 例題 7。
2. 共線性與 VIF：詳參秦大成老師講義第三回 P. 73-74。
3. 偏F值與(四)偏F檢定：詳參秦大成老師講義第二回P. 74-76。

高點高上高普特考



高上高普特考 www.get.com.tw/goldensun 台北市開封街一段 2 號 8 樓 02-23318268
【中壢】中壢市中山路 100 號 14 樓 03-4256899 【台中】台中市東區復興路四段 231-3 號 1 樓・04-22298699
【台南】台南市中西區中山路 147 號 3 樓之 1・06-2235868 【高雄】高雄市新興區中山一路 308 號 8 樓・07-2358996
【另有淡水・三峽・林口・羅東・中壢・逢甲・東海・中技】