《迴歸分析》

試題評析

第一題：與總複習第5頁＜例題1＞完全相同，百分之百命中！

第二題：考3個自變數複迴歸，雖然計算式稍爲龐大，但計算數值簡易，應可在時間內解完。第(4)小題須使用矩陣，若同學熟悉總複習第44~45頁二例題，應可得分。

第三題：考迴歸報表紙，爲送分題。

總　評：本年度考題除第二題第(4)小題計算較麻煩外，其餘只要觀念清楚，都不難拿到分數，因此，中等程度同學都可以及格，若能考到80分以上，上榜應不成問題。

一、假設截距項為零之簡單線性迴歸模式如下：

$$Y_i = \beta x_i + \varepsilon_i \ ,\ i=1,\ldots,n,$$

其中 ε_i 為 $i.i.d.\, N\left(0,\sigma^2\right)$，$\beta$ 和 σ^2 為未知參數。

(一)求 β 之最小平方估計量（least squares estimator）。（8分）

(二)求 β 之最大概似估計量（maximum likelihood estimator）。（8分）

(三)試證明上述 β 之估計量均為不偏估計量。（5分）

(四)求 σ^2 之最大概似估計量（maximum likelihood estimator）。（8分）

(五)藉由 σ^2 之最大概似估計量找出 σ^2 之不偏估計量。（10分）

答：

(一) 令 $Q=\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2=\sum_{i=1}^{n}\left(Y_i-\hat{\beta}x_i\right)^2$

$$\frac{dQ}{d\hat{\beta}}=-2\sum_{i=1}^{n}\left[\left(Y_i-\hat{\beta}x_i\right)\cdot x_i\right]=0$$

則 $\hat{\beta}=\dfrac{\sum_{i=1}^{n}x_iY_i}{\sum_{i=1}^{n}x_i^2}$

(二)&(四)同解

$$L\left(\beta,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}e^{-\frac{\sum_{i=1}^{n}(y_i-\beta x_i)^2}{2\sigma^2}}$$

$$\ln L=-\frac{n}{2}\left(\ln 2\pi+\ln\sigma^2\right)-\frac{\sum_{i=1}^{n}\left(y_i-\beta x_i\right)^2}{2\sigma^2}$$

$$\begin{cases}\dfrac{\partial \ln L}{\partial\beta}=\dfrac{2\sum_{i=1}^{n}\left[\left(y_i-\beta x_i\right)\cdot x_i\right]}{2\sigma^2}=0\\[2ex]\dfrac{\partial \ln L}{\partial\sigma^2}=-\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2}+\dfrac{\sum_{i=1}^{n}\left(y_i-\beta x_i\right)^2}{2\sigma^4}=0\end{cases}$$

則 $\beta=\dfrac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}$ ， $\sigma^2=\dfrac{\sum_{i=1}^{n}(y_i-\beta x_i)^2}{n}$

滿足 ① $\left.\dfrac{\partial^2\ln L}{\partial\beta^2}\right|_{\beta=\frac{\sum x_iy_i}{\sum x_i^2},\ \sigma^2=\frac{\sum(y_i-\beta x_i)^2}{n}}<0$

② $\left|\begin{matrix}\dfrac{\partial^2\ln L}{\partial\beta^2} & \dfrac{\partial^2\ln L}{\partial\sigma^2\partial\beta}\\ \dfrac{\partial^2\ln L}{\partial\beta\,\partial\sigma^2} & \dfrac{\partial^2\ln L}{\partial\left(\sigma^2\right)^2}\end{matrix}\right|_{\beta=\frac{\sum x_iy_i}{\sum x_i^2},\ \sigma^2=\frac{\sum(y_i-\beta x_i)^2}{n}}>0$

$\therefore$ MLE $\hat\beta=\dfrac{\sum_{i=1}^{n}x_iY_i}{\sum_{i=1}^{n}x_i^2}$

$$\hat\sigma^2=\frac{\sum_{i=1}^{n}\left(Y_i-\hat\beta x_i\right)^2}{n}$$

(三) $E\hat\beta=E\left[\dfrac{\sum_{i=1}^{n}x_iY_i}{\sum_{i=1}^{n}x_i^2}\right]=\dfrac{\sum_{i=1}^{n}x_iEY_i}{\sum_{i=1}^{n}x_i^2}=\dfrac{\sum_{i=1}^{n}(x_i\cdot\beta x_i)}{\sum_{i=1}^{n}x_i^2}=\dfrac{\beta\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}x_i^2}=\beta$ （unbiased）

(四)已在(二)解過

(五)

$$\begin{cases}\varepsilon_i\sim N\left(0,\sigma^2\right)\\ \hat\beta\sim N\left(\beta\ ,\ \dfrac{\sigma^2}{\sum_{i=1}^{n}x_i^2}\right)\end{cases}$$

則

$$\begin{cases} \dfrac{\sum_{i=1}^{n}\varepsilon_i^2}{\sigma^2} \sim \chi^2(n) \\ \dfrac{\left(\hat{\beta}-\beta\right)^2}{\sigma^2 \Big/ \sum_{i=1}^{n} x_i^2} \sim \chi^2(1) \end{cases}$$

$$\Rightarrow \frac{n\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(Y_i-\hat{\beta}x_i\right)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left[\left(\beta x_i+\varepsilon_i\right)-\hat{\beta}x_i\right]^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\varepsilon_i^2-\left(\hat{\beta}-\beta\right)^2\sum x_i^2}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sigma^2} - \frac{\left(\hat{\beta}-\beta\right)^2}{\sigma^2 \Big/ \sum x_i^2}$$

$$\sim \chi^2(n)-\chi^2(1)=\chi^2(n-1)$$

則 $E\left[\dfrac{n\hat{\sigma}^2}{\sigma^2}\right]=n-1$

$$\Leftrightarrow E\left[\frac{n}{n-1}\hat{\sigma}^2\right]=\sigma^2$$

$$\therefore \frac{n}{n-1}\hat{\sigma}^2 = \frac{n}{n-1}\cdot\frac{\sum_{i=1}^{n}\left(Y_i-\hat{\beta}x_i\right)^2}{n}$$

$$= \frac{\sum_{i=1}^{n}\left(Y_i-\hat{\beta}x_i\right)^2}{n-1}$$ 爲 σ^2 之不偏估計量

二、設現有一應變數 Y 及三自變數 x_1, x_2, x_3，其關係可由下列之迴歸模式描述，

$$Y_i=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon_i,\ i=1,\cdots,7,$$

其中 ε_i 為 $i.i.d.\ N\left(0,\sigma^2\right)$。七組資料如下：

y	x_1	x_2	x_3
1	5	–1	–3
0	0	1	–2
0	–3	1	–1
1	–4	0	0
2	–3	–1	1
3	0	–1	2
3	5	1	3

（一）求出估計之迴歸式（亦即，$\hat{y}=b_0+b_1x_1+b_2x_2+b_3x_3$）。（10分）

（二）估計 $E\left(Y \mid x_1=-3, x_2=-1, x_3=1\right)$。（5分）

（三）當 $x_1=-3, x_2=-1, x_3=1$，預測 Y。（5分）

（四）請檢測 x_2 是否提供足夠預測 Y 之資訊。$\alpha=0.05$，$t_{0.025,3}=3.182$。（請務必將完整之檢定寫出，含 H_0, H_1，檢定量，以及拒絕區域等）（10分）

答：

(一) $\sum x_1=0$，$\sum x_1^2=84$，$\sum x_2=0$，$\sum x_2^2=6$，$\sum x_3=0$，$\sum x_3^2=28$，

$\sum y=10$，$\sum x_1x_2=0$，$\sum x_1x_3=0$，$\sum x_2x_3=0$，

$\sum x_1y=10$，$\sum x_2y=-3$，$\sum x_3y=14$

$$\begin{cases} nb_0+\sum x_1b_1+\sum x_2b_2+\sum x_3b_3=\sum y \\ \sum x_1b_0+\sum x_1^2b_1+\sum x_1x_2b_2+\sum x_1x_3b_3=\sum x_1y \\ \sum x_2b_0+\sum x_1x_2b_1+\sum x_2^2b_2+\sum x_2x_3b_3=\sum x_2y \\ \sum x_3b_0+\sum x_1x_3b_1+\sum x_2x_3b_2+\sum x_3^2b_3=\sum x_3y \end{cases}$$

$$\begin{cases} 7b_0+0b_1+0b_2+0b_3=10 \\ 0b_0+84b_1+0b_2+0b_3=10 \\ 0b_0+0b_1+6b_2+0b_3=-3 \\ 0b_0+0b_1+0b_2+28b_3=14 \end{cases}$$

則 $b_0=1.4286$，$b_1=0.1190$，$b_2=-0.5$，$b_3=0.5$

$\therefore\ \hat{Y}=1.4286+0.1190x_1-0.5x_2+0.5x_3$

【另解】

$$Y=\begin{bmatrix}1\\0\\0\\1\\2\\3\\3\end{bmatrix}，\quad X=\begin{bmatrix}1&5&-1&-3\\1&0&1&-2\\1&-3&1&-1\\1&-4&0&0\\1&-3&-1&1\\1&0&-1&2\\1&5&1&3\end{bmatrix}$$

$$X^TX=\begin{bmatrix}7&0&0&0\\0&84&0&0\\0&0&6&0\\0&0&0&28\end{bmatrix}$$

$$\left(X^TX\right)^{-1}=\frac{\operatorname{adj}\left(X^TX\right)}{\det\left(X^TX\right)}=\frac{\begin{bmatrix}14112&0&0&0\\0&1176&0&0\\0&0&16464&0\\0&0&0&3528\end{bmatrix}}{98784}=\begin{bmatrix}\frac{1}{7}&0&0&0\\0&\frac{1}{84}&0&0\\0&0&\frac{1}{6}&0\\0&0&0&\frac{1}{28}\end{bmatrix}$$

$$X^TY=\begin{bmatrix}10\\10\\-3\\14\end{bmatrix}$$

$$\hat{\beta}=\left(X^TX\right)^{-1}X^TY=\begin{bmatrix}1.4286\\0.1190\\-0.5\\0.5\end{bmatrix}$$

$\therefore\ \hat{Y}=1.4286+0.1190x_1-0.5x_2+0.5x_3$

(二) $\hat{Y}_{x_1=-3,\,x_2=-1,\,x_3=1}$

$=1.4286+0.1190(-3)-0.5(-1)+0.5\times1=2.0716$

$\therefore\ E\left(Y\mid x_1=-3,x_2=-1,x_3=1\right)$ 之估計值為2.0716

(三) Y 之預測值

$=E\left(Y\mid x_1=-3,x_2=-1,x_3=1\right)$ 之估計值

$=2.0716$

(四) $SSE=Y^TY-\hat{\beta}^TX^TY=24-23.976=0.024$

$MSE=\dfrac{SSE}{n-k}=\dfrac{0.024}{7-4}=0.008$

① Hypothesis

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

② critical region

$$C=\left\{T\,\middle|\,|T|>t_{\frac{0.05}{2}}(3)=3.182\right\}$$

③ statistic

$$\left| T \right| = \left| \frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} \right| = \left| \frac{-0.5}{\sqrt{\frac{1}{6} \times 0.008}} \right| = 13.69 \quad \in C$$

$\therefore$ reject H_0，x_2足夠提供Y之資訊。

三、台灣房屋仲介聯盟欲調查房屋仲介業者之佣金（commissions）和其取得執照之年限以及業者性別（gender）的關係。此調查案之應變數為佣金（單位：萬元），自變數為執照取得之總月份數（month）和性別（1 = 女性，0= 男性）。

Regression Analysis

R^2	0.642			
Adjusted R^2	0.600	n = 20	k = 2	
Std. Error	3.219	Dep. Var.	Commissions	

ANOVA table

Source	SS	df	MS	F	*p*-value
Regression	315.9291	2	157.9645	15.25	.0002
Residual	176.1284	17	10.3605		
Total	492.0575	19			

Regression output

Variables	Coefficients	Std. error	*t*(df = 17)	*p*-value
Intercept	15.7625	3.0782	5.121	.0001
Months	0.4415	0.0839	5.263	.0001
Gender	3.8598	1.4724	2.621	.0179

（一）寫出估計之迴歸式。（5分）

（二）請檢測佣金和取得執照之年限以及性別間之關係是否均為線性。（請務必將完整之檢定寫出，含H_0, H_1，檢定量，以及拒絕區域等）（5分）

（三）請估計取得執照之總月份數為30個月之女性房屋仲介業者的平均佣金。（5分）

（四）請問女性仲介業者的佣金高過或低於男性仲介業者的佣金？高（低）過多少？（8分）

（五）請檢測性別之自變數是否應納入分析內。α = 0.05（請務必將完整之檢定寫出，含H_0, H_1，檢定量，以及拒絕區域等）（8分）

答：

(一) $\hat{Y} = 15.7625 + 0.4415x_1 + 3.8598x_2$

(二) 本題：檢定關係是否"均"爲線性？

須採用$\beta - t$係數各別檢定，解答與(五)重覆，

$\therefore$ 本題若將"均"字拿掉，採ANOVA F-test 較合理。

設 model：$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

where Y：佣金，x_1：取得執照年限，x_2：性別

(1) ① Hypothesis

$$\begin{cases} H_0: \text{佣金與取得執照年限關係不爲線性} \\ H_1: \text{佣金與取得執照年限關係爲線性} \end{cases} \Leftrightarrow \begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

② critical region

$$C = \left\{ t \,\middle|\, |t| > t_{\frac{0.05}{2}}(17) = 2.1098 \right\}$$

③ statistic

$|t| = 5.263 \in C$ （ $p\text{-value} = 0.0001 < \alpha = 0.05 \Rightarrow$ 統計值$|t| \in C$ ）

$\therefore$ reject H_0，佣金與取得執照年限關係爲線性。

(2) ① Hypothesis

$$\begin{cases} H_0: \text{佣金與性別關係不爲線性} \\ H_1: \text{佣金與性別關係爲線性} \end{cases} \Leftrightarrow \begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases}$$

② critical region

$$C = \left\{ t \,\middle|\, |t| > t_{\frac{0.05}{2}}(17) = 2.1098 \right\}$$

③ statistic

$|t| = 2.621 \in C$ （ $p\text{-value} = 0.0179 < \alpha = 0.05 \Rightarrow$ 統計值$|t| \in C$ ）

$\therefore$ reject H_0，佣金與性別關係爲線性。

$\therefore$ 佣金與年限，性別關係均爲線性

(三) $\hat{Y}_{x_1=30\,,\,x_2=1} = 15.7625 + (0.4415 \times 30) + (3.8598 \times 1) = 32.8673$

(四) (1) $\because x_2$ 係數估計值爲 $3.8598 > 0$

$\therefore$ 女性高過男性

(2) 當 x_1 固定下（男女性執照取得之總月份數相同之下）

女性平均佣金高於男性3.8598萬元。

(五) ① Hypothesis

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

② critical region

$$C = \left\{ t \,\middle|\, |t| > t_{\frac{0.05}{2}}(17) = 2.1098 \right\}$$

③ statistic

$|t| = 2.621 \in C$ （ $p\text{-value} = 0.0179 < \alpha = 0.05 \Rightarrow$ 統計值$|t| \in C$ ）

$\therefore$ reject H_0，性別之自變數應納入分析內。