林陳 Chap 15 單迴歸 習題

A.一般練習題

14.1 圖書館想知道每天使用圖書館的人數（百人）（ X ）與借出的書本數（百本）（ Y ）之間的關係。已知上個月圖書館共開放 25 天，且得下列資料： $\Sigma X=200$ ， $\Sigma Y=300$ ， $\Sigma X^2=1{,}660$ ， $\Sigma Y^2=3{,}696$ ， $\Sigma XY=2{,}436$ 。

①試求迴歸直線 $\hat{Y}=\hat{\alpha}+\hat{\beta}X$ 。

②試檢定「使用圖書館的人愈多，借出的書也愈多」的假設（ $\alpha=0.05$ ）。

③若某一天有 300 人使用該圖書館，預測當天借出書本數的 95%信賴區間。

解

① $\overline{X}=8$ ， $\overline{Y}=12$ ， $\hat{\beta}=\dfrac{\sum XY-n\overline{X}\overline{Y}}{\sum X^2-n\overline{X}^2}=\dfrac{2436-25(8)(12)}{1660-25(8)^2}=\dfrac{36}{60}=0.6$

$\hat{\alpha}=\overline{Y}-\hat{\beta}\overline{X}=12-0.6(8)=7.2$，迴歸直線為： $\hat{Y}=7.2+0.6X$

②因 $SSE=\sum y^2-\hat{\beta}\sum xy=(\sum Y^2-n\overline{Y}^2)-\hat{\beta}(\sum XY-n\overline{X}\overline{Y})$

$=[3{,}696-25(144)]-0.6[2{,}436-25(8)(12)]=74.4$

$\Rightarrow S_{\hat{\beta}}^2=\dfrac{SSE/(n-2)}{\sum X^2-n\overline{X}^2}=\dfrac{74.4/23}{1{,}660-25(64)}=0.0539$

欲檢定： $H_0:\beta\le 0,\quad H_1:\beta>0$ ，因 $t=\dfrac{0.6-0}{\sqrt{0.0539}}=2.58>t_{23,0.05}=1.714$

故知使用圖書館的人愈多，借出的書也愈多。

③

$$S_{e|x=3}=\sqrt{\left(\frac{SSE}{n-2}\right)\left[1+\frac{1}{n}+\frac{(3-\overline{X})^2}{\sum X^2-n\overline{X}^2}\right]}=\sqrt{\left(\frac{74.4}{23}\right)\left[1+\frac{1}{25}+\frac{(3-8)^2}{1660-25(64)}\right]}=2.17$$

$Y\,|\,X=3$ 的95%信賴區間

$$7.2+0.6\times 3\pm t_{23,0.025}\times 2.17=9\pm 4.49=(4.51,13.49)$$

預測當天借出的書本數約在450本到1,350本之間。

14.2 一唱片公司欲知打歌費用（10 萬元）（ X ）與唱片銷售量（千張）（ Y ）之間的關係，乃從其所發行的唱片中隨機抽選了 10 張，得如下的資料： $\Sigma X = 28$ ， $\Sigma X^2 = 303.4$ ， $\Sigma Y = 75$ ， $\Sigma Y^2 = 598.5$ ， $\Sigma XY = 237$ 。

①試求迴歸直線 $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ 。

②是否打歌費花得愈多，唱片的銷售量就愈高（ $\alpha = 0.05$ ）？

③試分別求 α 與 β 的 95%信賴區間。

④試求判定係數。

解

①

$\overline{X} = 2.8$ ， $\overline{Y} = 7.5$ 。

$$\hat{\beta} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{237 - 10(2.8)(7.5)}{303.4 - 10(2.8)^2} = \frac{27}{225} = 0.12$$

$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = 7.5 - 0.12(2.8) = 7.164$ ，迴歸直線： $\hat{Y} = 7.164 + 0.12X$

②欲檢定 $H_0: \beta \le 0, \quad H_1: \beta > 0$

$$SSE = \sum y^2 - \hat{\beta}\sum xy = (\sum Y^2 - n\overline{Y}^2) - \hat{\beta}(\sum XY - n\overline{X}\overline{Y})$$

$$= [598.5 - 10(7.5)^2] - 0.12[237 - 10(2.8)(7.5)] = 36 - 3.24 = 32.76$$

$$S_{\hat{\beta}} = \sqrt{\frac{SSE/(n-2)}{\sum X^2 - n\overline{X}^2}} = 0.135 \text{ ，} \quad t = \frac{0.12 - 0}{0.135} = 0.89 < t_{8,0.05} = 1.86 \text{ 。}$$

t值不顯著，在5%的信心水準下， β 值不顯著。

③ $S_{\hat{\alpha}} = \sqrt{\dfrac{\sum X^2}{n\sum (X - \overline{X})^2} \cdot \dfrac{SSE}{n-2}} = 0.743$

α的信賴區間 $= \hat{\alpha} \pm t_{8,0.025} S_{\hat{\alpha}} = 7.164 \pm 1.713 = (5.451, 8.877)$

β的信賴區間 $= \hat{\beta} \pm t_{8,0.025} S_{\hat{\beta}} = 0.12 \pm 0.311 = (-0.191, 0.431)$

④ $\hat{\beta} = r_{XY} \dfrac{S_Y}{S_X} \quad \Rightarrow r_{XY} = \hat{\beta}\dfrac{S_X}{S_Y} = 0.12\sqrt{\dfrac{225/9}{36/9}} = 0.3$

判定係數： $r^2 = r_{XY}^2 = 0.3^2 = 0.09$

14.3 一家庭問題研究機構想知道是否夫妻所受的教育愈高，愈不願生孩子。現隨機抽樣了 8 對夫妻，計算夫妻所受教育的總年數（ X ）與孩子數（ Y ），得結果如下：

X	19	17	21	18	15	12	14	20
Y	1	3	1	1	2	3	2	1

①試求迴歸直線 $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ 。

②證明殘差值相加會等於零。

③試求迴歸方程式的變異數的 95%信賴區間。

④試檢定是否夫妻所受的教育愈高。愈不願生孩子（$\alpha = 0.05$）？

解

① $\overline{X} = 17$，$\overline{Y} = 1.75$，$\sum X^2 = 2{,}380$，$\sum Y^2 = 30$，$\sum XY = 223$。

$$\hat{\beta} = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{223 - 8(17)(1.75)}{2{,}380 - 8(17)^2} = -0.221$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = 1.75 - (-0.221)(17) = 5.507$$

迴歸直線為：$\hat{Y} = 5.507 - 0.221X$

②

X	19	17	21	18	15	12	14	20
Y	1	3	1	1	2	3	2	1
$\hat{Y}$	1.308	1.75	0.866	1.529	2.192	2.855	2.413	1.087
$e = Y - \hat{Y}$	-0.308	1.25	0.134	-0.529	-0.192	0.145	-0.413	-0.087

$\Rightarrow \sum e = 0$，殘差項之和為零。

③ $SSE = \sum (Y - \hat{Y})^2 = \sum e^2 = 2.19$

變異數的95%信賴區間

$$= \left(\frac{SSE}{\chi^2_{6,0.025}}, \frac{SSE}{\chi^2_{6,0.975}}\right) = (0.152, 1.77)$$

④欲檢定：$H_0: \beta \geq 0, \quad H_1: \beta < 0$

$$S_{\hat{\beta}} = \sqrt{\frac{SSE\ /(n-2)}{\sum X^2 - n\overline{X}^2}} = \sqrt{\frac{2.19/6}{2380 - 8(17)^2}} = 0.073$$

$$t = \frac{-0.221 - 0}{0.073} = -3.027 < -t_{3,0.05} = -1.943$$

拒絕虛無假設，故知夫妻所受教育愈高愈不願生孩子。

14.4 若要分析豬肉的需求與其價格的關係時：

①請問如何利用迴歸分析來分析之？

②承①，若以相關分析來分析，則有何異同？

解

①設 Y：豬肉需求量，X 為豬肉價格

建立迴歸模型如下：

$$Y = \alpha + \beta X + \varepsilon$$

搜集 (Y, X) 資料，利用OLS估計迴歸模型，可估計豬肉價格影響豬肉需求量的大小。

②若以相關分析分析豬肉需求量與豬肉價格間之關係，則利用 (Y, X) 資料計算相關係數的正負及大小，可知兩者的關係程度。但由相關分析無法得知豬肉價格對豬肉需求量的影響力大小。（（參閱課本570頁））

14.5 某五星級大飯店的住屋率（%）（ X ）與每天每間客房的成本（元）（ Y ）如下：

X	100	75	65	55	50
Y	2,000	2,500	2,800	3,200	4,000

①試求迴歸直線 $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ 。

②試檢定此迴歸直線的斜率是否為零（ $\alpha = 0.05$ ）？

③若將 Y 的單位換成千元，試求此新的迴歸直線，並檢定其斜率是否為零（ $\alpha = 0.05$ ）？

④若 Y 的單位仍為元，但將 X 以小數點表示（如 75%表為 0.75），求此新的迴歸直線，並檢定其斜率是否為零（ $\alpha = 0.05$ ）？

⑤比較以上 4 小題的答案，若將自變數或依變數的單位改變，是否會影響斜率的檢定？

解

① $\overline{X} = 69$ ， $\overline{Y} = 2{,}900$ ， $\sum X^2 = 25{,}375$ ， $\sum Y^2 = 44{,}330{,}000$ ，

$\sum XY = 945{,}500$

$$\hat{\beta} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{945500 - 5(69)(2900)}{25375 - 5(69)^2} = \frac{-55{,}000}{1{,}570} = -35.0318$$

$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = 2900 - (-35.0318)(69) = 5{,}317.1942$ 。

迴歸直線為 $\hat{Y} = 5{,}317.1942 - 35.0318X$

②欲檢定： $H_0\!:\beta = 0, \quad H_1\!:\beta \neq 0$

計算得

$SSE = \sum (Y - \hat{Y})^2 = \sum y^2 - \hat{\beta}^2 \sum x^2 = 2{,}280{,}000 - 1{,}926{,}746.3 = 353{,}253.7$ ，

$$S_{\hat{\beta}} = \sqrt{\frac{SSE/(n-2)}{\sum X^2 - n\overline{X}^2}} = \sqrt{\frac{353253.7/3}{1{,}570}} = 8.660$$

因 $t=\dfrac{-35.0318-0}{8.660}=-4.045<-t_{3,0.025}=-3.182$，故拒絕虛無假設，在 5%的信心水準下，β值不為零，住屋率與客房成本有顯著相關。

③ $Y^*=Y/1{,}000 \quad \Rightarrow \overline{Y}^*=\overline{Y}/1{,}000,\quad \sum XY^*=\sum XY/1{,}000$。

$\Rightarrow \hat{\beta}_1=\hat{\beta}/1{,}000=-0.03503$，$\hat{\alpha}_1=\hat{\alpha}/1{,}000=5.3172$

迴歸直線為：

$\hat{Y}^*=5.3172-0.03503X$

$SSE_1=SSE/1{,}000{,}000,\quad S_{\hat{\beta}_1}=S_{\hat{\beta}}/1{,}000=0.00866$

$t=\dfrac{-0.03053-0}{0.00866}=-4.045<-t_{3,0.025}=-3.182$，故拒絕虛無假設，斜率不為零。

④ $X^*=X/100 \quad \Rightarrow \overline{X}^*=\overline{X}/100,\quad \sum(X^*)^2=\sum X^2/10{,}000,\quad \sum X^*Y=\sum XY/100$

$\Rightarrow \hat{\beta}_2=100\hat{\beta}=3{,}503.18,\quad \hat{\alpha}_2=\hat{\alpha}=5{,}317.1942$。

迴歸直線為：

$\hat{Y}=5{,}317.1942-3{,}503.18X^*$

$SSE_2=SSE,\quad S_{\hat{\beta}_2}=100S_{\hat{\beta}}=866$

因 $t=\dfrac{-3503.18-0}{866}=-4.045<-t_{3,0.025}=-3.182$，故拒絕虛無假設，斜率不為零。

⑤當自變數或依變數的單位改變時，斜率的檢定統計量之值不會變，因為其分子與分母的變化會相互抵消掉，故不影響斜率的檢定。

14.8 在簡單迴歸分析中請說明「樣本判定係數」之意義？若為 0 則其意義為何？其與相關係數、迴歸係數（β）間有何關係？

解

樣本判定係數是指迴歸方程式解釋的變異佔總變異的比例，亦即代表迴歸方程式的釋能力。若判定係數為 0，表示迴歸方程式無解釋能力。在簡單迴歸方程式中，判定係數恰好為 X 與 Y 相關係數的平方，且與估計的迴歸係數 $\hat{\beta}$ 間有如下之關係：

$$r^2=\hat{\beta}^2\cdot\frac{S_X^2}{S_Y^2}$$

B. 應用題

14.9 下表為證券交易所統計之金融保險類股 2014 年第三季每股淨值與 2014 年 9 月 30 日的股市收盤價：

公司	每股淨值 X	股市收盤價 Y	公司	每股淨值 X	股市收盤價 Y
1	14.79	18.65	16	16.06	16.10
2	22.77	33.65	17	17.09	16.95
3	12.37	10.30	18	13.61	10.00
4	16.15	26.75	19	18.16	17.85
5	15.69	11.20	20	14.98	17.75
6	16.95	25.10	21	36.26	46.70
7	19.44	22.85	22	31.47	49.55
8	11.57	8.98	23	11.07	9.34
9	15.04	9.05	24	14.77	18.45
10	12.64	10.60	25	16.32	15.00
11	12.36	10.25	26	20.31	24.95
12	12.61	9.76	27	13.15	14.25
13	15.92	14.50	28	10.38	9.23
14	23.02	23.95	29	11.03	8.60
15	16.09	16.30	30	13.15	13.05

資料來源：證券交易所，2014 年 12 月。

①試求迴歸估計式 $\hat{Y}=\hat{\alpha}+\hat{\beta}X$ ，並檢定此迴歸模型是否具解釋能力？（ $\alpha=0.05$ ）

②試分別計算 α 與 β 的 95%信賴區間。

解

①利用最小平方法或 Excel 可得下列迴歸分析結果：

	自由度	SS	MS	F	顯著值
迴歸	1	2728.84	2728.84	195.18	<0.0001
殘差	28	391.48	13.98		
總和	29	3120.32			

	係數	標準誤	t 統計	P-值
截距	-9.97523	2.11484	-4.72	<0.0001
每股淨值 X	1.69403	0.12126	13.97	<0.0001

R 平方	調整的 R 平方	觀察值個數
0.8745	0.8701	30

迴歸估計式為：$\hat{Y} = -9.97523 + 1.69403X$ 。

因F檢定統計量的值小於顯著值$\alpha = 0.05$ ，故知此迴歸模型具解釋能力。

②截距α 的95%信賴區間：

$\hat{\alpha} \pm t_{n-2,\alpha/2} S_{\hat{\alpha}} = -9.97523 \pm (2.048 \times 2.11484) = (-14.3064, -5.6440)$

斜率β 的95%信賴區間：

$\hat{\beta} \pm t_{n-2,\alpha/2} S_{\hat{\beta}} = 1.69403 \pm (2.048 \times 0.12126) = (1.4457, 1.9424)$

14.11 某人以臺北市政府主計處 2004 至 2012 年家庭收支概況調查為主要資料來源，研究每戶每年可支配所得和每戶家庭消費支出的關係，資料如下表所示（單位：元）：

年度	每戶可支配所得	每戶家庭消費支出
2004	1,225,096	925,845
2005	1,236,014	937,499
2006	1,262,406	945,344
2007	1,287,803	963,713
2008	1,271,060	961,630
2009	1,246,310	952,723
2010	1,298,640	988,691
2011	1,251,519	996,646
2012	1,278,278	973,747

資料來源：臺北市家庭收隻概況調查報告，台北市政府主計處，2013 年。

①請畫出每戶每年可支配所得和每戶家庭消費支出散佈圖，並說明兩者之關係。

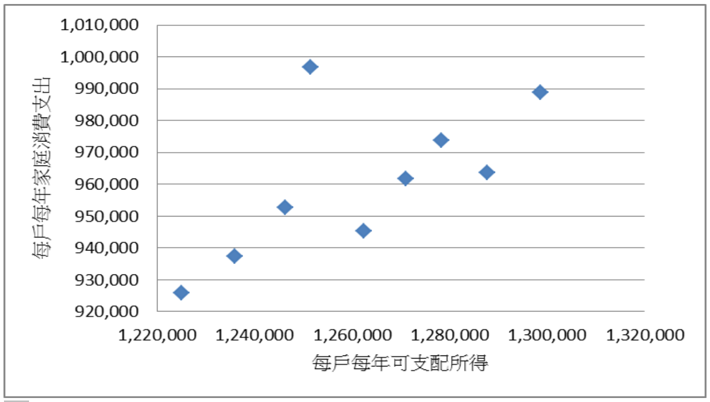
②利用最小平方法求出估計的迴歸方程式$\hat{Y}$ 。

③解釋$\hat{\beta}$之含義為何？及此一含義在 X 的範圍為何時才有意義？

④承③，若假設 2013 年台北市每戶每年平均可支配所得為 1,250,000 元，試問根據②所估計的迴歸方程式，2013 年每戶家庭消費支出為多少？此一結果是否有意義？

解

①每戶每年可支配所得和每戶每年家庭消費支出散佈圖如下圖，由圖可知，兩者略呈正向線性之關係。



② $\hat{Y} = 175263 + 0.62238X$ 。

③ $\hat{\beta}$ 之含義為：當每戶每年平均可支配所得增加 1 元時，每戶每年家庭消費支出用平均增加約 0.62238 元。因所取得樣本數全距為 [1,225,096, 1,298,640]，此一迴歸方程式 $\hat{Y}$ 在每戶每年平均可支配所得介於 1,225,096 元和 1,298,640 才有意義。

④若 2013 年台北市每戶每年平均可支配所得為 1,250,000 元，則 2013 年每戶每月平均育樂費用為 $\hat{Y} = 175263 + 0.62238 \times 1250000 = 953238$ 元。因為每戶每年可支配所得 125,000 元在樣本範圍內，故此計算結果有意義。

14.13 某人欲知國內研究發展經費（ X ）和本國人專利核准數（ Y ）是否有關，從行政院主計處網站和 2013 年的科學技術統計要覽中得到下表的資料。（單位：件、新台幣百萬元）

年	本國人專利核准數（件）	全國研究展經費（新台幣百萬元）
2003	30,955	242,942
2004	33,517	263,271
2005	42,324	280,980
2006	33,773	307,037
2007	34,068	331,386
2008	32,364	351,405
2009	33,457	367,174
2010	35,056	394,960
2011	36,924	413,293
2012	39,645	431,296

資料來源：行政院主計處網站和科學技術統計要覽，2013 年版。

並利用普通最小平方法求出 Y 對 X 的迴歸直線為：

$$\hat{Y} = 29{,}598 + 0.01658X$$

已知 $n = 10$，$S_X = 64{,}567.979$，$S_Y = 3{,}476.15426$，$\overline{X} = 33{,}8374.4$ 。

①試求 X 與 Y 之相關係數與判定係數。

②試求相關係數的95%信賴區間。

③當 $X = 400{,}000$，Y 平均值為何？並求其95%信賴區間。

④當 $X = 450{,}000$，Y 預期值為何？並求其95%信賴區間。

解

① $r_{XY} = \hat{\beta}\dfrac{S_X}{S_Y} = 0.01658\dfrac{64567.979}{3476.15426} = 0.307966$，$r^2 = r_{XY}^2 = 0.094843$ 。

② $Z_r = 0.5\ln(\dfrac{1+0.307966}{1-0.307966}) = 0.318297$，$Z_\rho$ 的信賴區間

$= 0.318297 \pm Z_{0.025}\sqrt{1/(10-3)} = 0.318297 \pm 0.74 = (-0.42, 1.06)$，故

$\dfrac{1+\rho}{1-\rho}$ 的信賴區間 $= (e^{2\times(-0.42)}, e^{2\times 1.06}) = (0.43, 8.32)$，可得$\rho$的信賴區間=

(-0.40,0.79)。

③ $E(Y|X = 400{,}000) = 29{,}598 + 0.01658(400{,}000) = 36230$，故

$SSE = (n-1)[S_Y^2 - \hat{\beta}^2 S_X^2] = 98438389$，可得

$S_{\hat{Y}/X=250000} = \sqrt{(\dfrac{SSE}{n-2})[\dfrac{1}{n} + \dfrac{(400000-\overline{X})^2}{\sum(X-\overline{X})^2}]} = 1573.502978$ 因此，

$E(Y|X = 800)$ 的信賴區間

$= 36230 \pm Z_{0.025}S_{\hat{Y}/X=400000} = 36230 \pm 3084.065837 = (33145.93, 39314.07)$

④ $E(Y|X = 450000) = 29598 + 0.01658(450000) = 37059$

$$S_{e/X=450000} = \sqrt{(\frac{SSE}{n-2})[1 + \frac{1}{n} + \frac{(450000-\overline{X})^2}{(n-1)S_X^2}]} = 4197.80$$

$Y|X = 450000$ 的信賴區間

$= 37059 \pm Z_{0.025}S_{e/X=450000} = 37059 \pm 8227.689497 = (28831.31, 45286.69)$

。