第八章　線性迴歸（考古題）

2006 年 4 月 29 日 最後修改

8.1（94-逢甲-國貿）

6. 福特汽車公司想要研究旗下各車型轎車重量與汽油的效能關係。令 X 為汽車的重量(單位:仟磅)，Y 為每公升汽油車子所能行進的車程(單位:公里)。現有八種款式汽車接受測試，所得資料如下。

X	21	24	23	21	22	18	20	26
Y	35	27	31	38	36	40	37	28

Run SPSS software 得到以下結果:

		未標準化係數		標準化係數	t	顯著性
模式		B 之估計值	標準誤	Beta 分配		
1	(常數)	72.776	6.896		10.553	.000
	X	-1.772	.313	-.918	-5.654	.001

a 依變數: y

變異分析表(ANOVA)

模式	平方和SS	自由度df	均方和MS	F	顯著性
迴歸	134.717	1	134.717	31.970	.001
誤差	25.283	6	4.214		
總和	160.000	7			

a 預測變數:(常數), X
b 依變數: y

(a) 請寫出 b_0、b_1 和樣本迴歸模式；(2%)
(b) 請解釋 b_1 的意義；(2%)
(c) 求出 r^2；(2%)
(d) 求出 σ_y^2 的估計值；(2%)
(e) 令 $\alpha = 0.05$，檢定 $\beta_1 = 0$ 是否成立，並解釋你所得的結論。(2%)

【解】

(a)

$$y = 72.776 - 1.772x$$

(b)

β 表示 x 變動一單位會導致 y 變動 $\hat{\beta} = -1.772$ 單位，

即每增加 1,000 磅重量，汽車每公升汽油行駛里程會減少 1.772 公里。

(c)

$$R^2 = \frac{SSR}{SST} = \frac{134.717}{160} = 0.842$$

(d)

$$E\left(\sigma_y^2\right) = MSE = 25.283$$

(e)

$$p = 0.001 < \alpha = 0.05 \quad \Rightarrow \quad \text{拒絕虛無假設} \quad \Rightarrow \quad \beta \neq 0$$

8.2（94-逢甲-保險）

6. 一保險公司要瞭解汽車保險簽單部門加班時間(小時)與新保單件數之關係，以作為費用預算控制之用，經蒐集資料如下：

新保單件數 X	699	504	808	1015	999	1106	323	713	979	542
加班時間 Y	4.2	1.7	4.8	5.9	6	7	1.1	4	5.6	2.9

由此資料得

$$\sum x_i = 7688,\ \bar{x} = 768.8,\ \sum y_i = 43.2,\ \bar{y} = 4.32$$
$$\sum x_i y_i = 37657.0,\ \sum x_i^2 = 6511846.0,\ \sum y_i^2 = 220.36$$

(a) （ 5分）求最小平方線。
(b) （10分）完成ANOVA表。
(c) （ 5分）求新保單件數為500時，平均加班時間的95%信賴區間。

【解】

x	y	x^2	xy	y^2
699	4.2	488601	2935.8	17.64
504	1.7	254016	856.8	2.89
808	4.8	652864	3878.4	23.04
1015	5.9	1030225	5988.5	34.81
999	6	998001	5994	36
1106	7	1223236	7742	49
323	1.1	104329	355.3	1.21
713	4	508369	2852	16
979	5.6	958441	5482.4	31.36
542	2.9	293764	1571.8	8.41
7688	43.2	6511846	37657	220.36

(a)

$$\begin{cases} n\hat{\alpha} + \Sigma x \hat{\beta} = \Sigma y \\ \Sigma x \hat{\alpha} + \Sigma x^2 \hat{\beta} = \Sigma xy \end{cases}$$

$$n = 10, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 10 \times 6,511,846 - (7,688)^2 = 6,013,116$$

$$\hat{\alpha}=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{n\Sigma x^2-\Sigma x\Sigma x}$$

$$=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{|\Sigma|}=\frac{43.2\times 6,511,846-7,688\times 37,657}{6,013,116}=-1.3629$$

$$\hat{\beta}=\frac{n\Sigma xy-\Sigma x\Sigma y}{n\Sigma x^2-\Sigma x\Sigma x}=\frac{n\Sigma xy-\Sigma x\Sigma y}{|\Sigma|}=\frac{10\times 37,657-7,688\times 43.2}{6,013,116}=0.0074$$

$$\hat{\alpha}=\overline{y}-\hat{\beta}\overline{x}=\frac{\Sigma y}{n}-\hat{\beta}\frac{\Sigma x}{n}=-1.3629$$

迴歸線： $y=-1.3629+0.0074x$

(b)

$$\Sigma\hat{y}^2=\Sigma\left(\hat{\alpha}+\hat{\beta}x\right)^2=n\hat{\alpha}^2+2\hat{\alpha}\hat{\beta}\Sigma x+\hat{\beta}^2\Sigma x^2=219.48$$

$$SST=\Sigma y^2-\frac{(\Sigma y)^2}{n}=33.74,\qquad SSE=\Sigma y^2-\Sigma\hat{y}^2=0.88$$

$$SSR=SST-SSE=32.86$$

變異來源	平方和	自由度	均方	F
迴歸變異	32.86	1	32.86	298.638
隨機變異	0.88	8	0.11	
總和	33.74	9		

(c)

（群體 y 值估計）

$x_g=500$

$$y'_g=\hat{\alpha}+\hat{\beta}x_g=-1.3629+0.0074\times 500=2.333$$

$$s^2_{y'_g}=V\left(y'_g\right)=\left(\frac{1}{n}+\frac{n\left(x_d-\overline{x}\right)^2}{n\Sigma x^2-\Sigma x\Sigma x}\right)MSE=0.156^2$$

$\dfrac{y'_g-y_g}{s_{y_g}}$ 爲 $df=n-2=8$ 的 t 分配，$1-\alpha=95\%$ ，$t^*=2.3060$

$$CI_{y_g}=\left\{\left|y_g-y'_g\right|\le t^*\times s_{y'_g}\right\}=\left\{1.974\le y_g\le 2.692\right\}$$

8.3（94-逢甲-工工）

5. n = 5 的(x, y)數據如下：

i	1	2	3	4	5
x_i	1	2	3	4	5
y_i	1	2	2	3	4

(a) (8 分) 試以最小平方法配合迴歸直線 $E(Y)=\beta_0+\beta_1 x$。

(b) (7 分) $\alpha=0.05$ 檢定 β_1 是否為 0。

【解】

x	y	x²	xy	y²
1	1	1	1	1
2	2	4	4	4
3	2	9	6	4
4	3	16	12	9
5	4	25	20	16
15	12	55	43	34

(a)

$$\begin{cases} n\hat{\alpha} + \Sigma x\hat{\beta} = \Sigma y \\ \Sigma x\hat{\alpha} + \Sigma x^2\hat{\beta} = \Sigma xy \end{cases}$$

$n = 5, \quad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 5 \times 55 - (15)^2 = 50$

$$\hat{\alpha} = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{n\Sigma x^2 - \Sigma x \Sigma x} = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{|\Sigma|} = \frac{12 \times 55 - 15 \times 43}{50} = 0.3$$

$$\hat{\beta} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - \Sigma x \Sigma x} = \frac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = \frac{5 \times 43 - 15 \times 12}{50} = 0.7$$

迴歸線： $y = 0.3 + 0.7x$

(b)

$$s_{\hat{\beta}}^2 = MSE \times \frac{n}{|\Sigma|} = 0.100^2$$

(1) $H_0 : \beta = 0$

(2)檢定統計量 $\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}$ 為 $df = n - 2 = 3$ 的 t 分配

(3)雙尾檢定、$df = 3$ 的 t 分配、$\alpha = 0.05$，拒絕區域 $R = \{|t| > 3.1824\}$

(4)樣本檢定統計量 $\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} = \frac{0.7}{0.1} = 7 \in R$，拒絕虛無假設 H_0

(5)統計上，應視 $\beta \neq 0$。

8.4（94-淡江-國貿）

5) Suppose that the height x (inches) and weight y (pounds) of a women basketball player has a relation with equation of the form $y = b_0 + b_1 x$.

(a) What are the least squares estimates $\hat{b}_0$ *and* $\hat{b}_1$ of b_0 *and* b_1, respectively, if we have n players' data; $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. (6%)

(b) The following data were the height and weight of the starting line-ups of a team:

height x	68	64	62	65	66
weight y	132	108	102	115	128

(8%)

Compute the estimated regression equation.

(c) If a player's height is 63 inches, what would you estimated her weight to be ? (4%)

【解】

(a)

$$y=\alpha+\beta x$$

$$y_i=\hat{\alpha}+\hat{\beta}x_i+\varepsilon_i$$

$$E=\Sigma\varepsilon_i^2=\Sigma\left(\hat{\alpha}+\hat{\beta}x_i-y_i\right)^2=n\hat{\alpha}^2+2\hat{\alpha}\hat{\beta}\Sigma x+\hat{\beta}^2\Sigma x^2-2\hat{\alpha}\Sigma y-2\hat{\beta}\Sigma xy+\Sigma y^2$$

$$\begin{cases}\dfrac{\partial}{\partial\hat{\alpha}}E=2n\hat{\alpha}+2\Sigma x\hat{\beta}-2\Sigma y=0\\ \dfrac{\partial}{\partial\hat{\beta}}E=2\Sigma x\hat{\alpha}+2\Sigma x^2\hat{\beta}-2\Sigma xy=0\end{cases}\Rightarrow\begin{cases}\hat{\alpha}=\dfrac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{n\Sigma x^2-\Sigma x\Sigma x}\\ \hat{\beta}=\dfrac{n\Sigma xy-\Sigma x\Sigma y}{n\Sigma x^2-\Sigma x\Sigma x}\end{cases}$$

(b)

x	y	x^2	xy	y^2
68	132	4624	8976	17424
64	108	4096	6912	11664
62	102	3844	6324	10404
65	115	4225	7475	13225
66	128	4356	8448	16384
325	585	21145	38135	69101

$$n=5,\qquad |\Sigma|=n\Sigma x^2-\Sigma x\Sigma x=5\times 21{,}145-(325)^2=100$$

$$\hat{\alpha}=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{|\Sigma|}=\frac{585\times 21{,}145-325\times 38{,}135}{100}=-240.5$$

$$\hat{\beta}=\frac{n\Sigma xy-\Sigma x\Sigma y}{|\Sigma|}=\frac{5\times 38{,}135-325\times 585}{100}=5.5$$

迴歸線： $y=-240.5+5.5x$

(c)

（個體 y 值估計）

$$x_d=63$$

$$y_d'=\hat{\alpha}+\hat{\beta}x_d=-240.5+5.5\times 63=106$$

8.5（94-雲科大-資管）

21. The equation of regression line for the paired data below is $\hat{y}=3x$. Find SSE.

x	2	4	5	6
y	7	11	13	20

(A) 14.25 (B) 10.00 (C) 88.75 (D) 78.75 (E) 16.78

【解】

x	y	x^2	y^2
2	7	4	49
4	11	16	121
5	13	25	169
6	20	36	400
17	51	81	739

$\Sigma x^2 = 81, \quad \Sigma y^2 = 739, \quad \alpha = 0, \quad \beta = 3$

$\Sigma \hat{y}^2 = \Sigma\left(\hat{\alpha} + \hat{\beta}x\right)^2 = n\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}\Sigma x + \hat{\beta}^2\Sigma x^2 = 9\times 81 = 729$

$SSE = \Sigma y^2 - \Sigma\hat{y}^2 = 739 - 729 = 10$

8.6（94-雲科大-財金）

7. A simple regression produces the regression equation $\hat{Y} = 5X + 7$.

[a] If we add 2 to all the X values in the data (and keep the Y values the same as the original), what will be the new regression equation be? ___________. (3 points)

[b] If we add 2 to all the Y values in the data (and keep the X values the same as the original), what will be the new regression equation be? ___________. (3 points)

[c] If we multiply all the X values in the data by 2 (and keep the Y values the same as the original), what will be the new regression equation be? ___________. (2 points)

[d] If we multiply all the Y values in the data by 2 (and keep the X values the same as the original), what will be the new regression equation be? ___________. (2 points)

【解】

(a)

$y = 7 + 5x,\ X = x + 2,\ Y = y \quad\Rightarrow\quad Y = 7 + 5\times\left(X - 2\right) \quad\Rightarrow\quad Y = -3 + 5X$

(b)

$y = 7 + 5x,\ X = x,\ Y = y + 2 \quad\Rightarrow\quad Y - 2 = 7 + 5X \quad\Rightarrow\quad Y = 9 + 5X$

(c)

$y = 7 + 5x,\ X = 2x,\ Y = y \quad\Rightarrow\quad Y = 7 + 5\times\dfrac{X}{2} \quad\Rightarrow\quad Y = 7 + 2.5X$

(d)

$y = 7 + 5x,\ X = x,\ Y = 2y \quad\Rightarrow\quad \dfrac{Y}{2} = 7 + 5X \quad\Rightarrow\quad Y = 14 + 10X$

8.7（94-淡江-企管）

4. In a study to determine how the skill in doing a complex assembly job is influenced by the amount of training, 15 new recruits were given varying amounts of training ranging between 3 and 12 hours. After the training, their times to perform the job were recorded. After denoting x = duration of training (in hours) and y = time to do the job (in minutes), the following summary statistics were calculated

$\overline{x}=7.2$ $\overline{y}=45.6$ $\sum(x-\overline{x})^2=33.6$ $\sum(x-\overline{x})(y-\overline{y})=-57.2$ $\sum(y-\overline{y})^2=160.2$

(1) Determine the equation of the best fitting straight line. (6 分)

(2) Do the data substantiate the claim that the job time decreases with more hours of training? Use $\alpha=0.05$. (7 分)

(3) Estimate the mean job time for 9 hours of training and construct a 95% confidence interval. (7 分)

(4) Find the predicted y for x = 35 hours and comment on the result. (5 分)

【解】

$n=15$

$\Sigma x=n\overline{x}=15\times 7.2=108,\qquad \Sigma y=n\overline{y}=15\times 45.6=684$

$\Sigma x^2=\Sigma\left(x-\overline{x}\right)^2+n\overline{x}^2=811.2,\qquad \Sigma y^2=\Sigma\left(y-\overline{y}\right)^2+n\overline{y}^2=31{,}350.6$

$\Sigma xy=\Sigma\left(x-\overline{x}\right)\left(y-\overline{y}\right)+n\overline{x}\,\overline{y}=4{,}867.6$

(1)

$n=15,\qquad |\Sigma|=n\Sigma x^2-\Sigma x\Sigma x=504$

$\hat{\alpha}=\dfrac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{|\Sigma|}=57.86\qquad \hat{\beta}=\dfrac{n\Sigma xy-\Sigma x\Sigma y}{|\Sigma|}=-1.70$

迴歸線： $y=57.86-1.70x$

(2)

$s_{\hat{\beta}}^2=MSE\times\dfrac{n}{|\Sigma|}=0.379^2$

(1) $H_0:\beta=0$

(2)檢定統計量 $\dfrac{\hat{\beta}-\beta}{s_{\hat{\beta}}}$ 為 $df=n-2=13$ 的 t 分配

(3)雙尾檢定、$df=13$ 的 t 分配、$\alpha=0.05$，拒絕區域 $R=\{|t|>2.1604\}$

(4)樣本檢定統計量 $\dfrac{\hat{\beta}-\beta}{s_{\hat{\beta}}}=\dfrac{-1.70}{0.379}=-4.489\in R$，拒絕虛無假設 H_0

(5)統計上，應視 $\beta\neq 0$。

即 $\beta<0$，顯示作業時間會隨訓練時間的增加而減少。

(3)

（群體 y 值估計）

$x_g = 9$

$y'_g = \hat{\alpha} + \hat{\beta} x_g = 57.86 - 1.70 \times 9 = 43.54$

$s^2_{y'_g} = V\left(y'_g\right) = \left(\dfrac{1}{n} + \dfrac{n\left(x_d - \overline{x}\right)^2}{n\Sigma x^2 - \Sigma x \Sigma x}\right) MSE = 0.888^2$

$\dfrac{y'_g - y_g}{s_{y_g}}$ 爲 $df = n - 2 = 13$ 的 t 分配，$1 - \alpha = 95\%$ ，$t^* = 2.1604$

$CI_{y_g} = \left\{ \left| y_g - y'_g \right| \le t^* \times s_{y'_g} \right\} = \left\{ 40.618 \le y_d \le 44.454 \right\}$

(4)

$x_g = 35$

$y'_g = \hat{\alpha} + \hat{\beta} x_g = 57.86 - 1.70 \times 35 = -1.726$

這是不合理的結果，因爲完成一見工作的時間不能爲負。

8.8（94-政大-國貿）

9. The following data were obtained by measuring cement bonding strength y, in response to setting times x, for 10 different locations.

Location #	x	y	Location #	x	y
1	5	21	6	30	49
2	10	28	7	35	60
3	15	34	8	40	52
4	20	39	9	45	58
5	25	47	10	50	66

$n = 10, \quad \sum\left(x - \overline{x}\right)\left(y - \overline{y}\right) = 1925$

$\sum x = 275, \quad \sum\left(x - \overline{x}\right)^2 = 2062.5$

$\sum y = 454, \quad \sum\left(y - \overline{y}\right)^2 = 1924.4$

a. Plot a rough scatter diagram and fit a regression line. (3%)

b. What amount of variation of y is accounted for by variation in x. (3%)

c. Now suppose that there is an extra observation from the 11th location with $x_{11} = 27.5$, and $y_{11} = 45.4$. Without recalculating the slope parameter estimate $\hat{\beta}$, show that $\hat{\beta}$ is unchanged when this observation is added to the sample. Does the standard error of $\hat{\beta}$ change? Does this extra observation contribute any new information about the relationship between x and y. (5%)

【解】

(a)

$n = 15 \qquad \Sigma x = 275 \qquad \Sigma y = 454$

$$\Sigma x^2 = \Sigma\left(x-\bar{x}\right)^2 + \frac{\left(\Sigma x\right)^2}{n} = 9,625, \qquad \Sigma y^2 = \Sigma\left(y-\bar{y}\right)^2 + \frac{\left(\Sigma y\right)^2}{n} = 22,536.1$$

$$\Sigma xy = \Sigma\left(x-\bar{x}\right)\left(y-\bar{y}\right) + \frac{\Sigma x \Sigma y}{n} = 14,410$$

$$\left|\Sigma\right| = n\Sigma x^2 - \Sigma x \Sigma x = 20,625$$

$$\hat{\alpha} = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{\left|\Sigma\right|} = 1,796.67 \qquad \hat{\beta} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\left|\Sigma\right|} = 127.83$$

迴歸線： $y = 1,796.67 + 127.83x$

(b)

x 變動一單位會導致 y 變動 $\hat{\beta} = 127.83$ 單位。

(c)

（群體 y 值估計）

$x_g = 27.5$

$y'_g = \hat{\alpha} + \hat{\beta} x_g = 1,796.67 + 127.83 \times 27.5 = 45.4$

$\left(27.5, 45.4\right)$ 在迴歸線上，因此多加入此觀察值不會更動迴歸線係數。

此觀察值會降低 $\hat{\beta}$ 之標準差。

同時，迴歸線的判定係數 R^2 也會增加。

8.9（94-政大-財管）

7. (25 points) Use the classical Ordinary Least Squares (OLS) to estimate the following Linear Regression Model: $Y_i = A + BX_i + E_i$
where Y is the annual Food Expenditure(in \$1000), X is the annual Income(in \$1000), E is the random error term.

Data

	1	2	3	4	5	6	7	8	9	10
Food Expenditure	5.2	5.1	8.1	7.8	5.8	5.1	5.2	4.8	7.9	6.4
Income	28	26	59	44	30	40	28	20	42	47

【解】

x	y	x^2	xy	y^2
28	5.2	784	145.6	27.04
26	5.1	676	132.6	26.01
59	8.1	3481	477.9	65.61
44	7.8	1936	343.2	60.84
30	5.8	900	174	33.64
40	5.1	1600	204	26.01
28	5.2	784	145.6	27.04
20	4.8	400	96	23.04
42	7.9	1764	331.8	62.41
47	6.4	2209	300.8	40.96
364	61.4	14534	2351.5	392.6

$n = 10, \quad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 12,844$

$\hat{\alpha} = \dfrac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{|\Sigma|} = 2.837 \qquad \hat{\beta} = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = 0.091$

迴歸線： $y = 2.837 + 0.0091x$

$\Sigma \hat{y}^2 = \Sigma\left(\hat{\alpha} + \hat{\beta} x\right)^2 = n\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}\Sigma x + \hat{\beta}^2 \Sigma x^2 = 387.57$

$SST = \Sigma y^2 - \dfrac{(\Sigma y)^2}{n} = 15.60, \qquad SSE = \Sigma y^2 - \Sigma \hat{y}^2 = 5.03$

$SSR = SST - SSE = 10.57$

變異來源	平方和	自由度	均方	F
迴歸變異	10.57	1	10.57	16.819
隨機變異	5.03	8	0.63	
總和	15.60	9		

8.10（94-台大-商學）

4. 某公司為決定其新產品訂價與銷售量的關係，試賣期間獲得以下資料：

訂　價 (X)	11	6	8	7	4	5	3	5
銷售量 (Y)	9	8	7	9	11	10	15	11

試求算：

(1) 最小平方估計迴歸式 ＝________；

(2) 樣本相關係數 ＝________；

(3) 當新產品訂價為 9 時，產品銷售量期望值 95%信賴區間 ＝________。

【解】

x	y	x^2	xy	y^2
11	9	121	99	81
6	8	36	48	64
8	7	64	56	49
7	9	49	63	81
4	11	16	44	121
5	10	25	50	100
3	15	9	45	225
5	11	25	55	121
49	80	345	460	842

(1)

$n = 8, \quad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 359$

$$\hat{\alpha}=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{|\Sigma|}=14.095 \qquad \hat{\beta}=\frac{n\Sigma xy-\Sigma x\Sigma y}{|\Sigma|}=-0.669$$

迴歸線： $y=14.095-0.669x$

(2)

$$r=\frac{\Sigma(x-\overline{x})(y-\overline{y})}{\sqrt{\Sigma(x-\overline{x})^2}\sqrt{\Sigma(y-\overline{y})^2}}=\frac{\Sigma xy-\Sigma x\Sigma y/n}{\sqrt{\Sigma x^2-\Sigma x\Sigma x/n}\sqrt{\Sigma y^2-\Sigma y\Sigma y/n}}$$

$$=\frac{460-49\times 80/8}{\sqrt{345-49\times 49/8}\sqrt{842-80\times 80/8}}=-0.6910$$

(3)

$$\Sigma\hat{y}^2=\Sigma\left(\hat{\alpha}+\hat{\beta}x\right)^2=n\hat{\alpha}^2+2\hat{\alpha}\hat{\beta}\Sigma x+\hat{\beta}^2\Sigma x^2=820.06$$

$$SSE=\Sigma y^2-\Sigma\hat{y}^2=842-820.06=21.94$$

$$MSE=\frac{SSE}{n-2}=\frac{21.94}{6}=3.66$$

（群體 y 值估計）

$$x_g=9$$

$$y'_g=\hat{\alpha}+\hat{\beta}x_g=14.095-0.669\times 9=8.078$$

$$s^2_{y'_g}=V\left(y'_g\right)=\left(\frac{1}{n}+\frac{n\left(x_d-\overline{x}\right)^2}{n\Sigma x^2-\Sigma x\Sigma x}\right)MSE=1.063^2$$

$\dfrac{y'_g-y_g}{s_{y_g}}$ 爲 $df=n-2=6$ 的 t 分配，$1-\alpha=95\%$ ，$t^*=2.4480$

$$CI_{y_g}=\left\{\left|y_g-y'_g\right|\le t^*\times s_{y'_g}\right\}=\left\{5.476\le y_g\le 10.680\right\}$$

8.11（94-台大-商學）

6. 下表爲甲地近 8 年來的鐵路貨運量，試用最小平方法配合直線長期趨勢方程式 = ________________，並預測第 9 年的貨運量 = _____ 。

年別	1	2	3	4	5	6	7	8
貨運量 Y (億延噸公里)	24.72	24.84	24.86	25.08	25.42	25.20	25.26	25.22

【解】

x	y	x^2	xy	y^2
1	24.72	1	24.72	611.0784
2	24.84	4	49.68	617.0256
3	24.86	9	74.58	618.0196
4	25.08	16	100.32	629.0064
5	25.42	25	127.1	646.1764
6	25.2	36	151.2	635.04
7	25.26	49	176.82	638.0676
8	25.22	64	201.76	636.0484
36	200.6	204	906.18	5030.4624

(1)

$n=8, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 336$

$\hat{\alpha} = \dfrac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{|\Sigma|} = 24.702 \qquad \hat{\beta} = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = 0.083$

迴歸線： $y = 24.702 + 0.083x$

(2)

$x_g = 9$

$y'_g = \hat{\alpha} + \hat{\beta} x_g = 24.702 + 0.083 \times 9 = 25.448$

這個題目作一下 y 數值轉換會比較好計算（數值較小）:

令 $Y = y - 25$ ，其中 y 爲原變數，Y 爲新權宜變數。

若求得之迴歸線爲

$Y = \hat{\alpha} + \hat{\beta} x$

則原迴歸線應爲

$y - 25 = \hat{\alpha} + \hat{\beta} x \quad \Rightarrow \quad y = (\hat{\alpha} + 25) + \hat{\beta} x$

x	y	x^2	xy	y^2
1	-0.28	1	-0.28	0.0784
2	-0.16	4	-0.32	0.0256
3	-0.14	9	-0.42	0.0196
4	0.08	16	0.32	0.0064
5	0.42	25	2.1	0.1764
6	0.2	36	1.2	0.04
7	0.26	49	1.82	0.0676
8	0.22	64	1.76	0.0484
36	0.6	204	6.18	0.4624

$n=8, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 0.3333$

$\hat{\alpha} = \dfrac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{|\Sigma|} = -0.298 \qquad \hat{\beta} = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = 0.083$

迴歸線： $y = -0.298 + 0.083x$

真正迴歸線應爲： $y = (-0.298 + 25) + 0.083x = 24.702 + 0.083x$

（群體 y 值估計）

$x_g = 9$

$y'_g = \hat{\alpha} + \hat{\beta} x_g = 24.702 - 0.083 \times 9 = 25.448$

8.12（93-雲科大-企管）

二、群生生技公司為瞭解廣告的促銷效果，隨機抽出 8 家分店比較廣告前銷售金額 x_i 與廣告後銷售金額 y_i，$i = 1,2,3,\cdots,8$，其資料如下：

項目	1	2	3	4	5	6	7	8
廣告前銷售金額 x_i	3.2	4.6	3.6	5.3	6.2	3.2	3.6	4.5
廣告後銷售金額 y_i	2.9	4.7	3.2	5.0	5.7	3.3	3.4	4.3

試問：

(1) a. 廣告前後銷售金額之樣本相關係數為何。

b. 由樣本相關係數是否可以接受廣告前後銷售金額有線性關係。(5%)

(2) （x_i，y_i）$i = 1,2,3,\cdots,8$，是否可以被接受滿足線性迴歸模型，亦即寫出檢定 $H_0 : \beta_1 = 0$ 之 ANOVA 表。(10%)

(3) 倘若(3)為線性迴歸可以接受，若再增加一筆資料恰好為（$\bar{x}$，$\bar{y}$），變成 9 筆，試求：(10%)

a. 廣告前後銷後金額相關係數。

b. 廣告後銷售金額對廣告前銷售金額之線性迴歸式。

c. 請寫出迴歸 ANOVA 表。

【解】

x	y	x^2	y^2	xy
3.2	2.9	10.24	8.41	9.28
4.6	4.7	21.16	22.09	21.62
3.6	3.2	12.96	10.24	11.52
5.3	5	28.09	25	26.5
6.2	5.7	38.44	32.49	35.34
3.2	3.3	10.24	10.89	10.56
3.6	3.4	12.96	11.56	12.24
4.5	4.3	20.25	18.49	19.35
34.2	32.5	154.34	139.17	146.41

(1)

$$r = \frac{\Sigma xy - \Sigma x \Sigma y / n}{\sqrt{\Sigma x^2 - \Sigma x \Sigma x / n}\sqrt{\Sigma y^2 - \Sigma y \Sigma y / n}} = 0.9806$$

由相關係數得知 x、y 接近完全相關（$|r| = 1$），兩者自有線性關係。

(2)

$n = 8, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 65.08$

$$\hat{\alpha}=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{|\Sigma|}=0.136 \qquad \hat{\beta}=\frac{n\Sigma xy-\Sigma x\Sigma y}{|\Sigma|}=0.919$$

迴歸線： $y=0.136+0.919x$

$$\Sigma\hat{y}^2=\Sigma\left(\hat{\alpha}+\hat{\beta}x\right)^2=n\hat{\alpha}^2+2\hat{\alpha}\hat{\beta}\Sigma x+\hat{\beta}^2\Sigma x^2=138.90$$

$$SST=\Sigma y^2-\frac{(\Sigma y)^2}{n}=7.14, \qquad SSE=\Sigma y^2-\Sigma\hat{y}^2=0.27$$

$$SSR=SST-SSE=6.86$$

變異來源	平方和	自由度	均方	F
迴歸變異	6.86	1	6.86	149.870
隨機變異	0.27	6	0.05	
總和	7.14	7		

假設檢定：

(1) $H_0:\alpha=\beta=0$

(2)檢定統計量 $F=\frac{MSR}{MSE}$ 爲 $df=(1,n-2)=(1,6)$ 的 F 分配

(3)右檢定、 $df=(1,6)$ 的 F 分配、 $\alpha=0.05$ ，拒絕區域 $R=\{F>5.9874\}$

(4)樣本檢定統計量 $F=\frac{MSR}{MSE}=149.87\in R$ ，拒絕虛無假設 H_0

(5)迴歸係數不全爲零。（該迴歸線可以接受。）

(3)

$$r=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

由以上公式得知，新增的觀察值$(\bar{x},\bar{y})$ ，不會變動

$\Sigma(x-\bar{x})(y-\bar{y})$ 、 $\Sigma(x-\bar{x})^2$ 、 $\Sigma(y-\bar{y})^2$

等值，因此相關係數不會改變。

新增的觀察值$(\bar{x},\bar{y})$在迴歸線上，因此迴歸線也會改變。

$$SST=\Sigma(y-\bar{y})^2, \qquad SSE=\Sigma(y-\hat{y})^2$$

由以上公式得知，新增的觀察值$(\bar{x},\bar{y})$ ，不會改變 SST、SSE、SSR 等值，

但是樣本數變成$n=9$ ，新迴歸 ANOVA 表如下：

變異來源	平方和	自由度	均方	F
迴歸變異	6.86	1	6.86	174.848
隨機變異	0.27	7	0.04	
總和	7.14	8		

8.13（93-逢甲-經濟）

四. (25 分) 假設牛奶的攝取量(MILK, 加侖)與身高(HEIGHT, 公尺)的關係為 HEIGHT=$\beta_1+\beta_2$(MILK)。依據 190 個研究樣本，SAS 軟體所得出的迴歸結果如下：

Dependent Variable: Consumption

Analysis of Variance （2 分）

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.2376	0.2376	0.436	0.5100
Error	189	103.0196	(a)_______		
C Total	190	103.2572			

Root MSE	0.7300	R-square	(b)_______
Dep Mean	0.3513		

Parameter Estimates （2 分）

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob> |T|
INTERCEPT	1	0.3440	0.0528	(c)_____	0.0001
MILK	1	0.0054	(d)_____	1.96	0.0514

Covariance of Estimates（6 分）

COVB	INTERCEPT	X
INTERCEPT	(e)_____	(f) ________
X	-0.6510	(g) ________

依據以上資料，請回答下列問題：

(1) 請填空 (a) 到 (g)，需示出計算過程。

(2) (5 分) 在無表可查臨界值的情況下，在 5%顯著水準下是否仍可以對 β_1=0 和 β_2=0 的虛無假設做出拒絕或接受的決定？若您的答案是"yes"，請說明原因。

(3) (10 分)從迴歸結果中，在 10%顯著水準下，對牛奶的攝取量是否影響身高請用臨界值法作檢定。需寫出虛無假設、對立假設和檢定過程。若無法拒絕虛無假設，那麼其所代表的意義為何？請說明

【解】

(1)

$$MSE=\frac{SSE}{n-2}=\frac{103.0196}{189}=0.5451 \quad \Leftarrow (a)$$

$$R^2=\frac{SSR}{SST}=\frac{0.2376}{103.2572}=0.0023 \quad \Leftarrow (b)$$

$$t=\frac{\hat{\alpha}}{s_{\hat{\alpha}}}=\frac{\hat{\beta}}{s_{\hat{\beta}}}$$

$$t=\frac{\hat{\alpha}}{s_{\hat{\alpha}}}=\frac{0.3440}{0.0528}=6.515 \quad \Leftarrow (c)$$

$$s_{\hat{\beta}}=\hat{\beta}\times t=0.0054\times 1.96=0.0106 \quad \Leftarrow (d)$$

$s_{\hat{\alpha}}^2 = 0.0528^2 = 0.002788 \quad \Leftarrow (e)$

$s_{\hat{\alpha},\hat{\beta}}^2 = s_{\hat{\beta},\hat{\alpha}}^2 = -6510 \quad \Leftarrow (f)$

$s_{\hat{\beta}}^2 = (d)^2 = 0.0106^2 = 0.000112 \quad \Leftarrow (g)$

(2)

變異數分析表中得知

$$p = P\left(F > F^*\right) = 0.5100 > \alpha = 0.05$$

因此會拒絕虛無假設。

(3)

(1) $H_0 : \beta = 0$

(2)檢定統計量 $\dfrac{\hat{\beta}-\beta}{s_{\hat{\beta}}}$ 爲 $df = n-2 = 189$ 的 t 分配（大樣本，可查 z 表）

(3)雙尾檢定、 $df = 189$ 的 t 分配、 $\alpha = 0.10$ ，拒絕區域 $R = \{|t| > 1.645\}$

(4)樣本檢定統計量 $\dfrac{\hat{\beta}-\beta}{s_{\hat{\beta}}} = 1.96 \in R$ ，拒絕虛無假設 H_0

(5)統計上，應視 $\beta \neq 0$ 。

即 $\beta > 0$ ，顯示牛奶的攝取量對身高有正面影響。

8.14（93-淡江-國貿）

六、(20%)某研究單位欲從6個家庭所得與儲蓄額的資料，研究家庭所得與儲蓄額之關係，其資料如下：

家庭所得(萬元)	6	2	5	4	6	7
儲蓄額 (千元)	8	1	2	3	4	6

進行簡單線性迴歸分析時，以儲蓄額爲反應變數Y，以家庭所得爲解釋變數X。

(1)求樣本迴歸線。

(2)求判定係數R^2，並解釋之。

(3)求家庭所得每增加一萬元，其平均儲蓄額增加或減少多少？

(4)張三家庭所得爲 5.6 萬元，預測其儲蓄額爲多少？並求 90%的信賴區間。

【解】

x	y	x²	y²	xy
6	8	36	64	48
2	1	4	1	2
5	2	25	4	10
4	3	16	9	12
6	4	36	16	24
7	6	49	36	42
30	24	166	130	138

(1)

$$n = 6, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 96$$

$$\hat{\alpha}=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{|\Sigma|}=-1.625 \qquad \hat{\beta}=\frac{n\Sigma xy-\Sigma x\Sigma y}{|\Sigma|}=1.125$$

迴歸線： $y=-1.625+1.125x$

(2)

$$\Sigma\hat{y}^2=\Sigma\left(\hat{\alpha}+\hat{\beta}x\right)^2=n\hat{\alpha}^2+2\hat{\alpha}\hat{\beta}\Sigma x+\hat{\beta}^2\Sigma x^2=116.25$$

$$SST=\Sigma y^2-\frac{\left(\Sigma y\right)^2}{n}=34, \qquad SSR=\Sigma\hat{y}^2-\frac{\left(\Sigma y\right)^2}{n}=20.25$$

$$R^2=\frac{SSR}{SST}=\frac{20.25}{34}=0.5956$$

(3)

x 增加一單位，y 會增加 $\beta=1.125$ 單位。

故所得增加一萬元，儲蓄會增加 1125 元。

(4)

$$SSE=\Sigma y^2-\Sigma\hat{y}^2=138-116.25=13.75$$

$$MSE=\frac{SSE}{n-2}=\frac{13.75}{4}=3.44$$

（個體 y 值估計）

$$x_d=5.6$$

$$y_d'=\hat{\alpha}+\hat{\beta}x_d=-1.625+1.125\times 5.6=4.675$$

$$s_{y_d'}^2=V\left(y_d'\right)=\left(1+\frac{1}{n}+\frac{n\left(x_d-\overline{x}\right)^2}{n\Sigma x^2-\Sigma x\Sigma x}\right)MSE=2.002^2$$

$\dfrac{y_d'-y_d}{s_{y_d}}$ 為 $df=n-2=4$ 的 t 分配，$1-\alpha=95\%$ ，$t^*=2.7764$

$$CI_{y_d}=\left\{\left|y_d-y_d'\right|\le t^*\times s_{y_d'}\right\}=\left\{-0.938\le y_d\le 10.288\right\}$$

8.15（93-政大-國貿）

5. Based on the following 10 pairs of data,

X	6	3	4	-3	2	-4	0	-1	-5	-2
Y	10	7	8	1	6	2	5	3	2	1

and assuming the following linear regression model,

$$Y = \alpha + \beta X + U,$$

where U has zero mean and variance equal to σ^2.

answer the following questions.

(a) Compute the least-squares point estimates of α and β, and the coefficient of determination. (6%)

(b) Calculate variance estimates for $\widehat{\alpha}$ and $\widehat{\beta}$, and the covariance estimate between $\widehat{\alpha}$ and $\widehat{\beta}$. (6%)

(c) With 10% significance level, test the hypothesis: $H_0 : \sigma^2 = 1$ using the following critical values: (5%)

$Z_{.90} = 1.28$, $Z_{.95} = 1.65$; $t_{0.90,\,8} = 1.40$, $t_{0.95,\,8} = 1.86$; $t_{0.90,9} = 1.38$, $t_{0.95,\,9} = 1.83$; $t_{0.90,\,10} = 1.37$, $t_{0.95,\,10} = 1.81$; $\chi^2_{0.9,\,8} = 13.36$, $\chi^2_{0.95,\,8} = 15.51$; $\chi^2_{0.9,\,9} = 14.68$, $\chi^2_{0.95,\,9} = 16.92$; $\chi^2_{0.9,\,10} = 15.99$, $\chi^2_{0.95,\,10} = 18.31$.

【解】

x	y	x^2	y^2	xy
6	10	36	100	60
3	7	9	49	21
4	8	16	64	32
-3	1	9	1	-3
2	6	4	36	12
-4	2	16	4	-8
0	5	0	25	0
-1	3	1	9	-3
-5	2	25	4	-10
-2	1	4	1	-2
0	45	120	293	99

(a)

$$n = 10, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 1200$$

$$\hat{\alpha} = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{|\Sigma|} = 4.5 \qquad \hat{\beta} = \frac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = 0.825$$

迴歸線： $y = 4.5 + 0.825x$

$$\Sigma\hat{y}^2 = \Sigma\left(\hat{\alpha} + \hat{\beta}x\right)^2 = n\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}\Sigma x + \hat{\beta}^2\Sigma x^2 = 284.18$$

$$SST = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 90.5, \qquad SSR = \Sigma\hat{y}^2 - \frac{(\Sigma y)^2}{n} = 81.68$$

$$R^2 = \frac{SSR}{SST} = \frac{81.68}{90.5} = 0.9025$$

(b)

$$V\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \begin{bmatrix}\sigma^2_{\hat{\alpha}} & \sigma_{\hat{\alpha},\hat{\beta}}\\ \sigma_{\hat{\alpha},\hat{\beta}} & \sigma^2_{\hat{\beta}}\end{bmatrix} = \begin{bmatrix} n & \Sigma x\\ \Sigma x & \Sigma x^2\end{bmatrix}^{-1}\sigma^2 = \frac{1}{|\Sigma|}\begin{bmatrix}\Sigma x^2 & -\Sigma x\\ -\Sigma x & n\end{bmatrix}\sigma^2$$

其中 $|\Sigma| = \begin{vmatrix} n & \Sigma x\\ \Sigma x & \Sigma x^2\end{vmatrix} = n\Sigma x^2 - \Sigma x\Sigma x$

即

$$\sigma^2_{\hat{\alpha}} = \frac{\Sigma x^2}{|\Sigma|}\sigma^2,\qquad \sigma^2_{\hat{\beta}} = \frac{n}{|\Sigma|}\sigma^2,\qquad \sigma^2_{\hat{\alpha},\hat{\beta}} = \frac{-\Sigma x}{|\Sigma|}\sigma^2$$

$$SSE = \Sigma y^2 - \Sigma\hat{y}^2 = 293 - 284.18 = 8.82$$

$$E\left(\sigma^2\right) = MSE = \frac{SSE}{n-2} = \frac{8.82}{8} = 1.10$$

$$E\left(\sigma^2_{\hat{\alpha}}\right) = \frac{\Sigma x^2}{|\Sigma|}E\left(\sigma^2\right) = 0.1102$$

$$E\left(\sigma^2_{\hat{\beta}}\right) = \frac{n}{|\Sigma|}E\left(\sigma^2\right) = 0.0092$$

$$E\left(\sigma^2_{\hat{\alpha},\hat{\beta}}\right) = \frac{-\Sigma x}{|\Sigma|}E\left(\sigma^2\right) = 0$$

(c)

(1) $H_0 : \sigma^2 = 1$

(2)檢定統計量 $\frac{(n-2)s^2}{\sigma^2} = \frac{SSE}{\sigma^2}$ 爲 $df = n-2 = 8$ 的 χ^2 分配

(3)雙尾檢定、$df = 8$ 的 χ^2 分配、$\alpha = 0.1$，拒絕區域 $R = \left\{\chi^2 < 2.73 \text{ 或 } \chi^2 > 15.51\right\}$

(4)樣本檢定統計量 $\frac{SSE}{\sigma^2} = 8.82 \notin R$ ，無法拒絕虛無假設 H_0

(5)統計上，應視 $\sigma^2 = 1$ 。

8.16（93-政大-財管）

5. (10%)

X	213	196	184	202	221	247
Y	76	65	62	68	71	75

If you were to develop a regression line to predict Y by X, what value would the coefficient of determination have?

【解】

x	y	x²	y²	xy
213	76	45369	5776	16188
196	65	38416	4225	12740
184	62	33856	3844	11408
202	68	40804	4624	13736
221	71	48841	5041	15691
247	75	61009	5625	18525
1263	417	268295	29135	88288

$n = 6, \qquad |\Sigma| = n\Sigma x^2 - \Sigma x \Sigma x = 14,601$

$\hat{\alpha} = \dfrac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{|\Sigma|} = 25.43 \qquad \hat{\beta} = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{|\Sigma|} = 0.209$

迴歸線： $y = 25.43 + 0.209x$

$\Sigma \hat{y}^2 = \Sigma\left(\hat{\alpha} + \hat{\beta}x\right)^2 = n\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta}\Sigma x + \hat{\beta}^2 \Sigma x^2 = 29,088.2$

$SST = \Sigma y^2 - \dfrac{\left(\Sigma y\right)^2}{n} = 153.5, \qquad SSR = \Sigma \hat{y}^2 - \dfrac{\left(\Sigma y\right)^2}{n} = 106.67$

$R^2 = \dfrac{SSR}{SST} = \dfrac{106.67}{153.5} = 0.6949$