CH 4

迴歸分析與預測

Modern Engineering Statistics

工 程 統 計

4-1以最小平方法求線性迴歸

◎線性迴歸（Linear Regression）常用以表示變數Y與X的直線關係，Y= a + bX。

◎線性迴歸之建立：

(1)根據蒐的樣本資料(X , Y)。

(2)利用最小平方（Least Square）法找出式中最配適的斜率（b）及截距（a）。

圖4-1中的直線乃依25個觀測值所建立的迴歸直線，可用以預測X在特定下之Y的反應值。

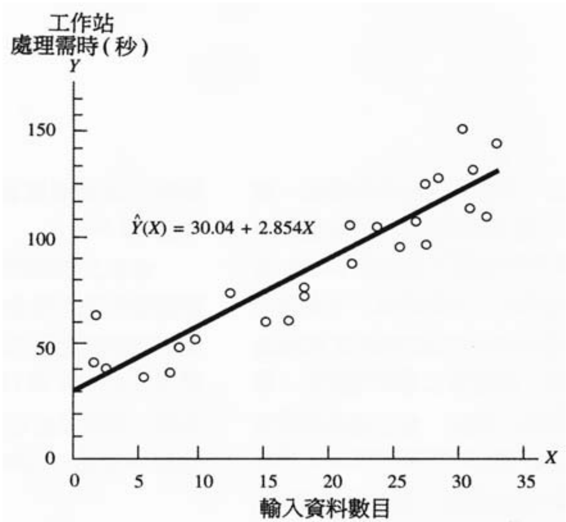


圖 4-1　工作站處理資料需時之散佈圖與迴歸直線

4-1-1最小平方法

最小平方法（Method of Least Squares）乃為求得一最佳配適直線的客觀且有效方法，最佳配適係指利用的配適直線與資料點（X_i,Y_i）之總離差平方和為最小。

$$\hat{Y}(X_i) = a + bX_i$$

$$\sum_{i=1}^{n}\left[Y_i - \hat{Y}(X_i)\right]^2 = \sum_{i=1}^{n}\left[Y_i - (a + bX_i)\right]^2$$

可對a與b取偏微分，然後令微分方程式等於0，得以下二式：

$$\sum_{i=1}^{n} Y_i = na + b\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2$$

可解得係數a與b：

$$b=\frac{n\sum XY-\sum X\sum Y}{n\sum X^2-(\sum X)^2}$$

$$a=\frac{1}{n}\left(\sum Y-b\sum X\right)$$

因 $\overline{X}=(\sum X)/n$ 及 $(\sum Y)/n$ ，故亦即

$$a=\overline{Y}-b\overline{X}$$

表 4-1 利用觀測資料來估計直線迴歸係數

觀測值 i	資料輸入數 X	工作站處理需時（秒）Y	XY	X^2
1	2	66	132	14
2	19	77	1,463	361
3	6	37	222	36
4	23	106	2,438	529
5	10	55	550	100
6	23	89	2,047	529
7	9	52	468	81
8	30	128	3,840	900
9	18	63	1,134	324
10	25	104	2,600	625
11	19	76	1,444	361
12	2	44	88	4
13	27	97	2,619	729
14	28	109	3,052	784
15	8	40	320	64
16	29	124	3,596	841
17	29	98	2,842	841
18	16	63	1,008	256
19	33	131	4,323	1,089
20	3	41	123	9
21	34	111	3,774	1,156
22	32	151	4,832	1,024
23	13	76	988	169
24	33	114	3,762	1,089
25	35	143	5,005	1,225
	506	2,195	52,670	13,130
	$=\sum X$	$=\sum Y$	$=\sum XY$	$=\sum X^2$

$\overline{X}=\frac{506}{25}=20.24$ $\overline{Y}=\frac{2{,}195}{25}=87.80$ $s_X=10.97$ $s_Y=33.99$

4-1-2 最小平方法的合理性與意義

第i個觀測值與對應直線點之垂直距離（離差）為Y_i - $Y(X_i)$。例如第22個觀測點（32,151），位於迴歸直線上方之垂直離差為

$$Y_{22}=Y(32)=151\text{-}121.4=29.6(\text{秒})$$

每個垂直離差表示以迴歸直線式所得預測值之誤差（Error）值。

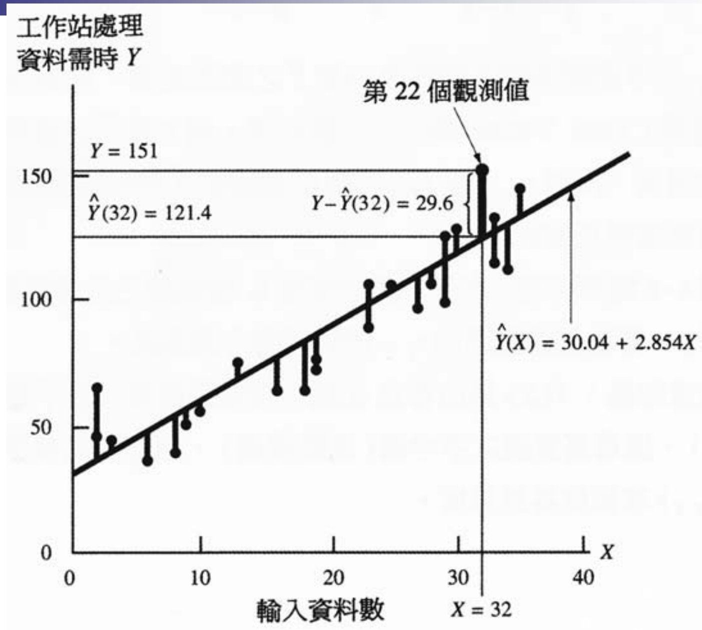


圖 4-2 觀測值與迴歸直線之離差

4-1-3 測度迴歸的變異

迴歸變異量分析主要是用以估量散佈之各觀測點與迴歸直線間的離散程度，此種測度稱為估計的標準誤（Standard Error of the Estimate）

$$\frac{\sum\left[Y-\hat{Y}(X)\right]^2}{n}$$

$$S_{Y\cdot X}=\sqrt{\frac{\sum\left[Y-\hat{Y}(X)\right]^2}{n-2}}$$

表 4-2 估計的標準誤之計算

觀測值 i	資料輸入數 X	工作站處理需時 Y	$Y(X)=30.04+2.854X$	$Y-\hat{Y}(X)$	$[Y-\hat{Y}(X)]^2$
1	2	66	35.7	30.3	918.09
2	19	77	84.3	−7.3	53.29
3	6	37	47.2	−10.2	104.04
4	23	106	95.7	10.3	106.09
5	10	55	58.6	−3.6	12.96
6	23	89	95.7	−6.7	44.89
7	9	52	55.7	−3.7	13.69
8	30	128	115.7	12.3	151.29
9	18	63	81.4	−18.4	338.56
10	25	104	101.4	2.6	6.76
11	19	76	84.3	−8.3	68.89
12	2	44	35.7	8.3	68.89
13	27	97	107.1	−10.1	102.01
14	28	109	110.0	−1.0	1.00
15	8	40	52.9	−12.9	166.41
16	29	124	112.8	11.2	125.44
17	29	98	112.8	−14.8	219.04
18	16	63	75.7	−12.7	161.29
19	33	131	124.2	6.8	46.24
20	3	41	38.6	2.4	5.76
21	34	111	127.1	16.1	259.21
22	32	151	121.4	29.6	876.16
23	13	76	67.1	8.9	79.21
24	33	114	124.2	−10.2	104.04
25	35	143	129.9	13.1	171.61
				−0.2	4,204.86

$$s_{Y\cdot X}=\sqrt{\frac{4{,}204.86}{25-2}}=13.52$$

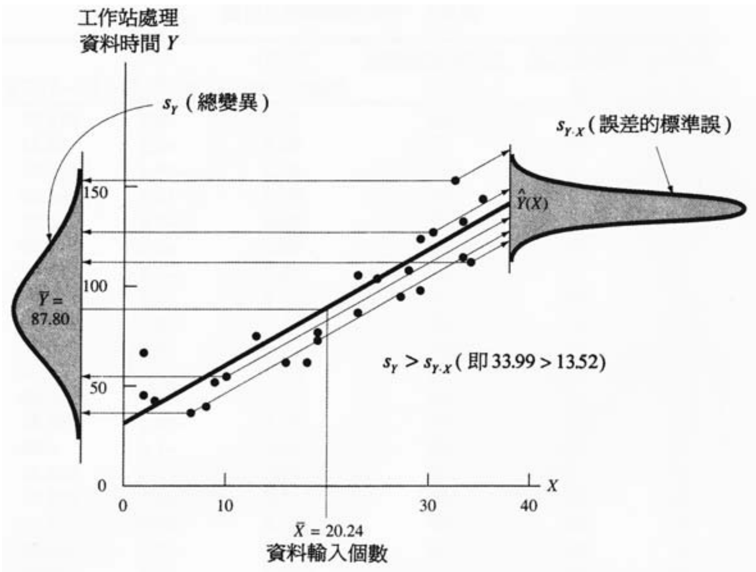


圖 4-3 迴歸式之總變異與估計的標準誤之示意圖

4-2 相關與迴歸分析

◎相關係數

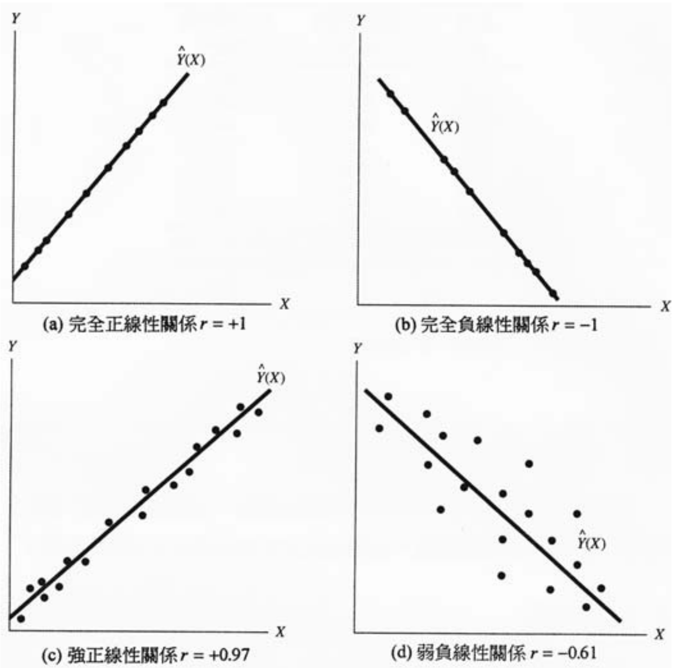
相關係數（Correlation Coefficient）r表示X與Y間直線關係的強度，r的正負符號表示相關的方向 而r值必介於+1與-1之間。

◎相關係數的計算

相關係數可用以測度每一個觀測點與其樣本平均值$\overline{X}$及 $\overline{Y}$之交叉的差異。

$$S_X=\sqrt{\frac{\sum\left(X-\overline{X}\right)^2}{n-1}}=\sqrt{\frac{\sum X^2-\left(\sum X\right)^2/n}{n-1}}$$

$$S_Y=\sqrt{\frac{\sum\left(Y-\overline{Y}\right)^2}{n-1}}=\sqrt{\frac{\sum Y^2-\left(\sum Y\right)^2/n}{n-1}}$$



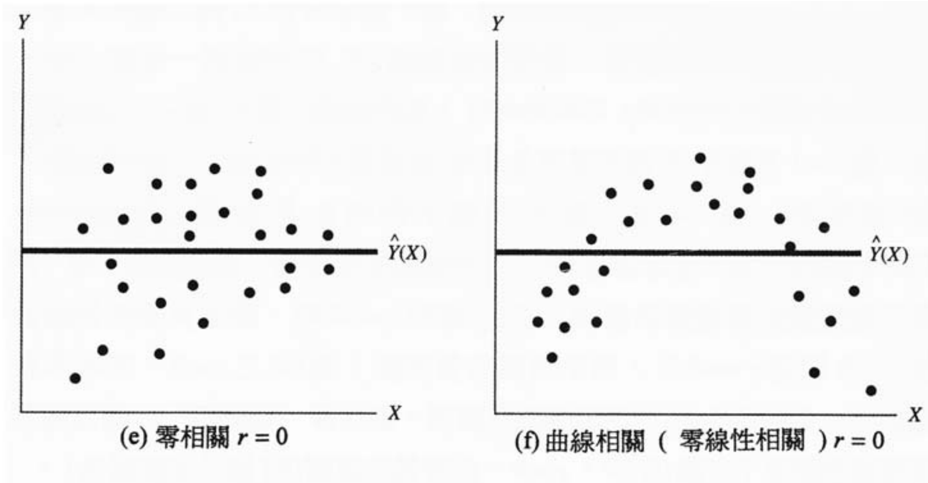


圖 4-6　散佈圖與各種相關強度的對應關係

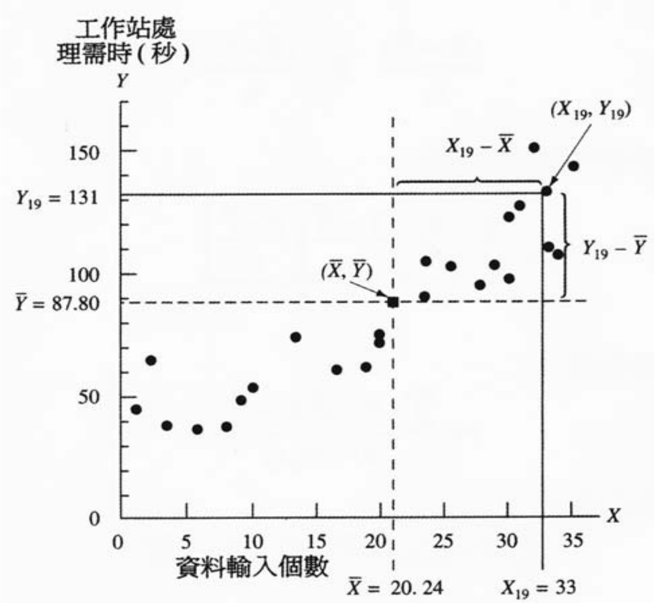


圖 4-7　工作站資料處理需時之散佈圖

將觀測點與 $\overline{X}$ 與 $\overline{Y}$ 差距予以標準為1而期望值為0，其計算式如下：

$$\frac{\left(X-\overline{X}\right)}{S_X} \text{及} \frac{\left(Y-\overline{Y}\right)}{S_Y}$$

則此兩者乘積和即為相關係數：

$$r=\frac{\sum\left[\frac{X-\overline{X}}{S_X}\right]\left[\frac{Y-\overline{Y}}{S_Y}\right]}{n-1}$$

上式可加以轉換為：

$$r=\frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{(n-1)_{S_X S_Y}}$$

其中分子部份可寫成：

$$\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)=\sum XY-n\overline{X}\,\overline{Y}$$

最後相關係數的計算式為：

$$r=\frac{\sum XY-n\overline{X}\,\overline{Y}}{(n-1)_{S_X S_Y}}$$

或是寫成：

$$r=\frac{\sum XY-n\overline{X}\,\overline{Y}}{\sqrt{\left(\sum X^2-n\overline{X}^2\right)\left(\sum Y^2-n\overline{Y}^2\right)}}$$

4-3 複迴歸分析

有二個自變數的複迴歸式可表示成

$$\mathrm{Y}=\mathrm{a}+\mathrm{b}_1\mathrm{X}_1+\mathrm{b}_2\mathrm{X}_2$$

我們採用最小平方法來估計複迴歸式的係數，使得垂直離差的 $\mathrm{Y}-\hat{\mathrm{Y}}$ 平方和最小，即

$$\sum(\mathrm{Y}-\hat{\mathrm{Y}})^2=\sum\left[\mathrm{Y}-(a+b_1X_1+b_2X_2)\right]^2$$

欲求得各係數，分別以上式對a ,b及b取偏微分並令其值為零

$$\sum Y=na+b_1\sum X_1+b_2\sum X_2$$

$$\sum X_1Y=a\sum X_1+b_1\sum X_1^{\ 2}+b_2\sum X_1X_2$$

$$\sum X_2Y=a\sum X_2+b_1\sum X_1X_2+b_2\sum X_2^{\ 2}$$

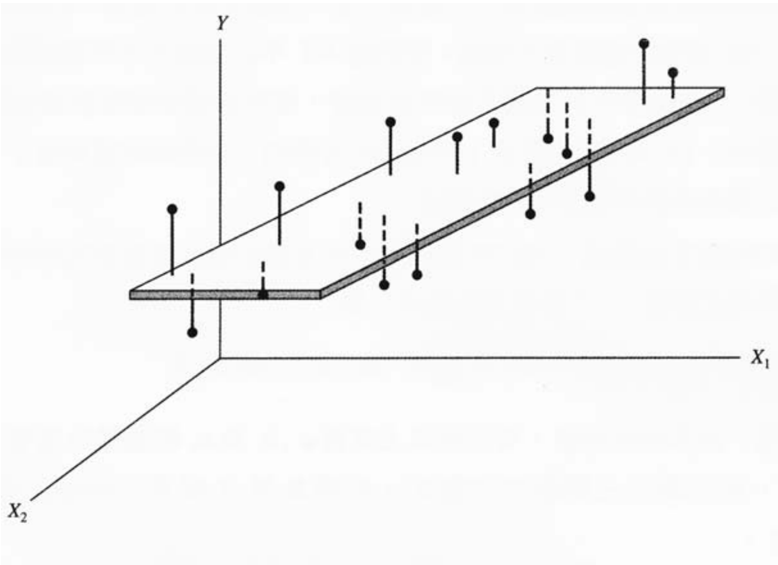


圖 4-8 散佈圖與由 3 個變數構成之三維迴歸平面

表 4-4 燃油里程、辛烷含量及車速之試驗資料

試驗	燃油里程 (mpg) Y	辛烷量 X_1	平均車速 (mph) X_2
1	24.8	88	52
2	30.6	93	60
3	31.1	91	58
4	28.2	90	52
5	31.6	90	55
6	29.9	89	46
7	31.5	92	58

表 4-4 （續）

試驗	燃油里程 (mpg) Y	辛烷量 X_1	平均車速 (mph) X_2
8	27.2	87	46
9	33.3	94	55
10	32.6	95	62
11	30.6	88	47
12	28.1	89	58
13	25.2	90	63
14	35.0	93	54
15	29.2	91	53
16	31.9	92	52
17	27.7	89	52
18	31.7	94	53
19	34.2	93	54
20	30.1	91	58

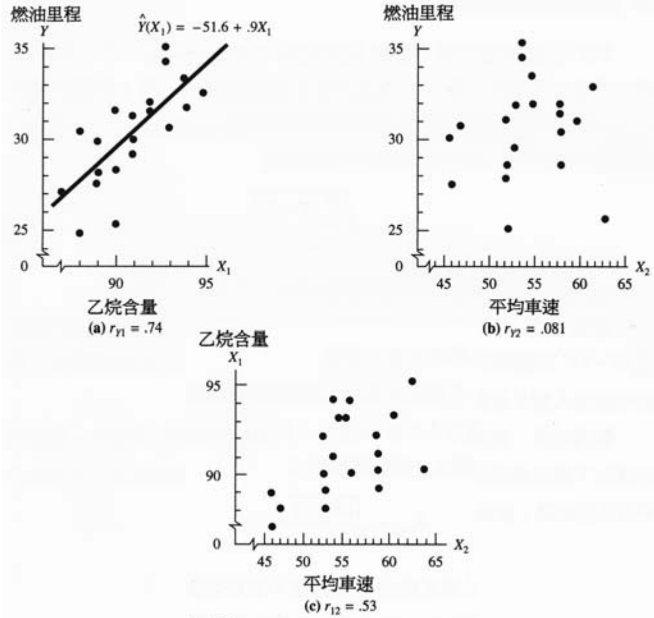


圖 4-9 不同的 2 個變數組的散佈圖

4-3-1 殘差與估計標準誤差

其迴歸式的估計標準誤差為：

$$S_{Y\cdot 12} = \sqrt{\frac{\sum\left(Y - \hat{Y}\right)^2}{n-3}}$$

垂直離差，$Y - \hat{Y}$，又稱為殘差(Residuals)。殘差常用於衡量 $\sum\left(Y - \hat{Y}\right)^2$ 於迴歸式中為無法解釋的變異的部份。

表 4-6 燃油里程之真實值與預測值及殘差的計算表

真實值 Y	預測值 $\hat{Y}$	殘 差 $Y-\hat{Y}$	離差平方 $(Y-\hat{Y})^2$
24.8	27.34	−2.54	6.4516
30.6	31.26	−0.66	0.4356
31.1	29.39	1.71	2.9241
28.2	29.70	−1.50	2.2500
31.6	28.96	2.64	6.9696
29.9	30.00	−0.10	0.0100
31.5	30.57	0.93	0.8649
27.2	27.65	−0.45	0.2025
33.3	33.67	−0.37	0.1369
32.6	33.12	−0.52	0.2704
30.6	28.58	2.02	4.0804
28.1	27.04	1.06	1.1236
25.2	26.98	−1.78	3.1684
35.0	32.74	2.26	5.1076
29.2	30.63	−1.43	2.0449
31.9	32.06	−0.16	0.0256
27.7	28.52	−0.82	0.6724
31.7	34.17	−2.47	6.1009
34.2	32.74	1.46	2.1316
30.1	29.39	0.71	0.5041
604.5	604.51	−0.01	45.4751

表 4-7　加入第 3 個自計數 - 載重的原始資料

試驗	燃油里程 (mpg) Y	乙烷含量 X_1	平均車速 (mph) X_2	載重 (磅) X_3
1	24.8	88	52	646
2	30.6	93	60	465
3	31.1	91	58	359
4	28.2	90	52	665
5	31.6	90	55	214
6	29.9	89	46	606
7	31.5	92	58	458
8	27.2	87	46	557
9	33.3	94	55	605
10	32.6	95	62	407
11	30.6	88	47	259
12	28.1	89	58	423
13	25.2	90	63	596
14	35.0	93	54	286
15	29.2	91	53	612
16	31.9	92	52	399
17	27.7	89	52	444
18	31.7	94	53	697
19	34.2	93	54	376
20	30.1	91	58	363


```
                         GASOLINE MILEAGE DATA
Model: MODEL1
Dep Variable: MILEAGE

                         Analysis of Variance

                         Sum of          Mean
   Source      DF       Squares        Square      F Value       Prob>F

   Model        3     134.59603      44.86534       83.070       0.0001
   Error       16       8.64147       0.54009
   C Total     19     143.23750

      Root MSE  Sy.123 0.73491     R-Square       0.9397
      Dep Mean      30.22500  Ȳ   Adj R-Sq       0.9284
      C.V.           2.43146

                         Parameter Estimates

                   Parameter         Standard       T for H0:
Variable   DF       Estimate            Error     Parameter=0    Prob > |T|

INTERCEP    1     -59.004933  a    7.09881835         -8.312        0.0001
OCTANE      1       1.209731  b1   0.08812242         13.728        0.0001
SPEED       1      -0.296972  b2   0.04181721         -7.102        0.0001
LOAD        1      -0.009833  b3   0.00119102         -8.256        0.0001

                       Predict     Std Err                  Std Err    Student
  Obs     MILEAGE        Value     Predict     Residual    Residual   Residual
              Y            Ŷ
    1     24.8000      25.6566       0.344      -0.8566       0.649     -1.319
    2     30.6000      31.1093       0.263      -0.5093       0.686     -0.742
    3     31.1000      30.3261       0.248       0.7739       0.692      1.118
    4     28.2000      27.8892       0.288       0.3108       0.676      0.460
    5     31.6000      31.4331       0.356       0.1669       0.643      0.260
    6     29.9000      29.0415       0.357       0.8585       0.642      1.337
    7     31.5000      30.5623       0.208       0.9377       0.705      1.330
    8     27.2000      27.1038       0.381       0.0962       0.629      0.153
    9     33.3000      32.4272       0.340       0.8728       0.652      1.339
   10     32.6000      33.5051       0.369      -0.9051       0.635     -1.424
   11     30.6000      30.9469       0.431      -0.3469       0.595     -0.583
   12     28.1000      27.2773       0.327       0.8227       0.658      1.250
   13     25.2000      25.3010       0.483      -0.1010       0.554     -0.182
   14     35.0000      34.6512       0.341       0.3488       0.651      0.536
   15     29.2000      29.3231       0.236      -0.1231       0.696     -0.177
   16     31.9000      32.9243       0.257      -1.0243       0.688     -1.488
   17     27.7000      28.8526       0.223      -1.1526       0.700     -1.646
   18     31.7000      32.1165       0.425      -0.4165       0.600     -0.695
   19     34.2000      33.7663       0.280       0.4337       0.679      0.638
   20     30.1000      30.2867       0.245      -0.1867       0.693     -0.270

Sum of Residuals                -2.59348E-13
Sum of Squared Residuals              8.6415
Predicted Resid SS (Press)           12.7560
```

圖 4-10　統計軟體 SAS 的計算結果 (總變數 $m = 4$)

表 4-8　車輛停等數與燃油里程之試驗資料

試驗	燃油里程 (mpg) Y	車輛停等數 / 哩 X_3	試驗	燃油里程 (mpg) Y	車輛停等數 / 哩 X_3
1	24.8	1.2	11	30.6	0.2
2	30.6	0.7	12	28.1	0.7
3	31.1	0.3	13	25.2	0.6
4	28.2	0.6	14	35.0	0.3
5	31.6	1.2	15	29.2	0.8
6	29.9	0.1	16	31.9	0.7
7	31.5	0.2	17	27.7	1.4
8	27.2	0.1	18	31.7	1.1
9	33.3	0.3	19	34.2	0.2
10	32.6	0.4	20	30.1	0.1

表 4-9　罐重減輕的相關資料表

溫度	接觸時間	每 1 M 化學劑濃度下減輕的罐重 (毫克)				
X_2 , °C	X_3 , 小時	0.115 M	0.057 M	0.028 M	0.01 M	0.0025 M
25	1	3.85	1.89	0.5	–	–
	3	13.75	7.1	3.0	–	–
	6	34.6	15.4	6.4	–	–
	12	88.4	36.8	15.9	–	–
45	1	19.2	11.7	4.4	2.6	1.5
	3	76.9	39.7	16.9	10.8	3.2
	6	152.2	77.1	36.4	14.6	5.2
	12	220.2	100.2	49.3	19.3	5.1
65	1	101.5	41.0	15.4	7.1	2.9
	3	139.8	113.0	44.0	9.4	4.0
	6	228.0	100.0	54.7	17.0	4.6
	12	276.7	100.0	54.1	26.0	4.7
78	1	112.6	49.7	27.5	–	–
	3	211.1	86.9	43.9	–	–
	6	265.9	112.9	48.0	–	–
	12	280.0	122.0	65.8	–	–

例 **13-11** 已知某河川之水位 H(m) 及流量 Q(cms) 之資料如下：

水位，H(m)	0.00	0.49	0.78	0.88	1.03	1.20	1.28	1.34	1.36	1.51
流量，Q(cms)	9.8	15.3	23.1	29.8	31.8	40.2	46.2	49.8	54.5	61.0

H	1.55	1.76	1.86	1.92	21.0
Q	71.0	82.5	93.2	105.1	113.4

試以 $Q = aH^n + b$ 型式，作迴歸相關分析。

解：已知 H = 0 時，Q = 9.8 cms，故先假設 b = 9.8 cms，

$\therefore \quad Q - 9.8 = aH^n$

取對數 $\log(Q-9.8) = \log a + n \log H$

將 log(Q − 9.8) 及 log H 資料求出如下表：

號數	Q−9.8	log(Q−9.8)	H	log H
1	6.1	0.7853	0.49	−0.3098
2	13.3	1.1239	0.78	−0.1079
3	20.0	1.3010	0.88	−0.0555
4	22.0	1.3424	1.03	0.0128
5	30.4	1.4829	1.20	0.0792
6	36.4	1.5611	1.28	0.1072
7	40.0	1.6021	1.34	0.1271
8	44.7	1.6503	1.36	0.1335
9	51.2	1.7093	1.51	0.1790
10	61.2	1.7868	1.55	0.1903
11	72.7	1.8615	1.76	0.2455
12	83.4	1.9212	1.86	0.2695
13	95.3	1.9791	1.92	0.2833
14	103.6	2.0154	2.10	0.3222
		Σ 22.1223		Σ 1.4764

令 $\log(Q-9.8) = y$， $\bar{y} = \dfrac{22.1223}{14} = 1.5802$

$\log H = x$， $\bar{x} = \dfrac{1.4764}{14} = 0.1055$

再以下表計算 $(y-\bar{y})$，$(x-\bar{x})$，$(y-\bar{y})^2$，$(x-\bar{x})^2$ 及 $(x-\bar{x})(y-\bar{y})$：

號數	$y-\bar{y}$	$x-\bar{x}$	$(y-\bar{y})^2$	$(x-\bar{x})^2$	$(y-\bar{y})(x-\bar{x})$
1	−0.7949	0.4153	0.6319	0.1725	0.3301
2	−0.4563	0.2134	0.2082	0.0455	0.0974
3	−0.2792	0.1610	0.0779	0.0259	0.0449
4	−0.2378	0.0927	0.0565	0.0086	0.0220
5	−0.0973	0.0263	0.0095	0.0007	0.0026
6	0.0191	0.0017	0.0004	0.0000	0.0000
7	0.0219	0.0216	0.0005	0.0005	0.0005
8	0.0701	0.0280	0.0049	0.0008	0.0020
9	0.1291	0.0735	0.0167	0.0054	0.0095
10	0.2066	0.0848	0.0427	0.0072	0.0175
11	0.2813	0.1400	0.0791	0.0196	0.0394
12	0.3410	0.1640	0.1163	0.0269	0.0559
13	0.3989	0.1778	0.1591	0.0316	0.0709
14	0.4352	0.2167	0.1894	0.0469	0.0943
Σ			1.5931	0.3921	0.7870

x_i 之標準偏差：

$$S_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}} = \sqrt{\frac{0.3921}{14-1}} = 0.1737$$

y_i 之標準偏差：

$$S_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{N-1}} = \sqrt{\frac{1.5931}{14-1}} = 0.3501$$

相關係數：

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\Sigma(\bar{y}-\bar{y})^2}} = \frac{0.7870}{\sqrt{0.3921 \times 1.5931}} = 0.996$$

代入式 (13-78)，迴歸方程式為 $y-\bar{y} = r \cdot \dfrac{s_y}{s_x}(x-\bar{x})$

$$\therefore \quad y - 1.5802 = 0.996\,\frac{0.3501}{0.1737}(x-0.1055)$$

或 $y - 1.5802 = 2.0075(x - 0.1055)$

即 $y = 2.0075x + 1.3684$

因 $y = \log(Q-9.8)$， $x = \log H$

$\therefore \quad \log(Q-9.8) = 2.0075 \log H + \log 23.36$

故所求之迴歸方程式為：

$$Q = 23.36H^{2.0075} + 9.8$$

估計之標準誤差：

$$S_{\bar{y}} = S_y\sqrt{1-r^2} = 0.3501\sqrt{1-0.996^2} = 0.0313$$

迴歸方程式之上、下限：

$$y_1 = 23.36H^{2.0075} + 9.8 \pm 0.0313\, z_{\frac{1}{2}\alpha}$$

相關係數顯著性之變方分析：

$$F = \frac{r^2(N-2)}{1-r^2} = \frac{0.996^2(14-2)}{1-0.996^2} = \frac{11.904}{0.00798} = 1491$$

由表 13-6 得 $F_{0.01(1,\,14-2=12)} = 9.93$

因 $1491 \gg 9.93$ ，故知該流量 —— 水位關係十分顯著。

補充資料

如何找出相關

針對五個選美之評分，找出對美之相同看法：

STEP 1：

量化：第一名給5分，第二名給4分………

表一：為甲與乙兩位評審者針對5個參賽者之評分。

	A	B	C	D	E	
甲	5	4	3	2	1	
乙	5	4	3	2	1	
	25	16	9	4	1	總分：55

表二：

	A	B	C	D	E	
甲	5	4	3	2	1	
乙	1	2	3	4	5	
	5	8	9	8	5	總分：35

STEP 2：

數量化之相關強度：

將每一個參賽者之得分相乘，表一之總分為55，表二之總分為35，進一步發現，任何評分均介於55~35之間。

STEP 3：

(1) 將範圍55~35扣去適當值（平均值45）：移動中心位置得 10 ~ -10

(2) 除去一個適當值（i.e 就大之差距）：統一變化範圍在 1~ -1之間

另例如參賽者為6位會得：

量化相關強度為：91 ~ 56

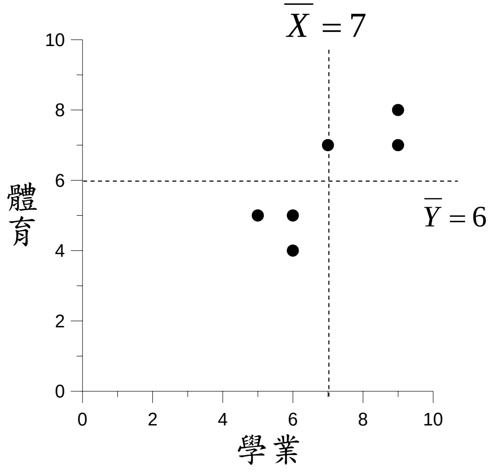
分　成： 17.5 ~ -17.5

再標準化： 1 ~ -1

相關係數

例：6個同學之學業與體育之相關

	學業	體育
A	6	5
B	5	5
C	9	7
D	9	8
E	6	4
F	7	7



STEP 1. 先標準化

移動中心位置：

$$X_i \rightarrow X_i - \overline{X}$$

$$Y_i \rightarrow Y_i - \overline{Y}$$

每一點均以 $\overline{X},\overline{Y}$ 為原點，如位在第一個與第三個象限均表示正相關，反之在第二個與第四個象限則為負相關。

STEP 2. 量化相關強度：

將每一個座標之相關強度計算如下：

$$\left(X_i-\overline{X}\right)\cdot\left(Y_i-\overline{Y}\right)$$

所有資料相關強度之總和為：

$$\sum\left(X_i-\overline{X}\right)\cdot\left(Y_i-\overline{Y}\right)$$

STEP 3. 統一範圍在 1～-1 之間

將 $\sum\left(X_i-\overline{X}\right)\cdot\left(Y_i-\overline{Y}\right)$ 除以最大變化差距為

$$\sqrt{\sum\left(X_i-\overline{X}\right)^2\cdot\left(Y_i-\overline{Y}\right)^2}$$

得

$$r=\frac{\sum\left(X_i-\overline{X}\right)\cdot\left(Y_i-\overline{Y}\right)}{\sqrt{\sum\left(X_i-\overline{X}\right)^2\cdot\left(Y_i-\overline{Y}\right)^2}}$$

[証明]：

$$\sum(a_i)^2\cdot\sum(b_i)\geq\left[\sum(a_i)(b_i)\right]^2$$

$$\sqrt{\sum(a_i)^2\sum(b_i)^2}\geq\sum(a_i)(b_i)$$

令 $a_i=X_i-\overline{X}$ $\quad b_i=Y_i-\overline{Y}$

$$\sqrt{\sum\left(X_i-\overline{X}\right)^2\cdot\left(Y_i-\overline{Y}\right)^2}\geq\sum\left(X_i-\overline{X}\right)\cdot\left(Y_i-\overline{Y}\right)$$

X_i	$X_i-\overline{X}$	$\left(X_i-\overline{X}\right)^2$	Y_i	$Y_i-\overline{Y}_i$	$\left(Y_i-\overline{Y}_i\right)^2$	$\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}_i\right)$
6	-1	1	5	-1	1	1
5	-2	4	5	-1	1	2
9	2	4	7	1	1	2
9	2	4	8	2	4	4
6	-1	1	4	-2	4	2
7	0	0	7	1	1	0
$\overline{X}=7$		$\sum\left(X_i-\overline{X}\right)^2=14$	$\overline{Y}=6$		$\sum\left(Y_i-\overline{Y}_i\right)^2=12$	$\sum\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}_i\right)=11$

$$r=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(X_i-\overline{X})^2\sum(Y_i-\overline{Y})^2}}=\frac{11}{\sqrt{14\times12}}=0.85$$

$$r^2 = \frac{\sum (\hat{Y}_i - \overline{Y})^2}{\sum (Y_i - \overline{Y})^2} = \frac{\text{Sum of squares due to regression}}{\text{Sum of squares corrected for mean}}$$

$$= \frac{\text{Y預測值隨著X值沿直線移動時產生的變異}}{\text{Y觀測值的總變異}}$$

例：

月份	10	11	12	1	2	3	4	5
溫度X	49.4	38.2	27.2	28.6	29.5	46.4	49.7	57.1
瓦斯消耗量Y	520	610	870	850	880	490	450	250

$$r^2 = 0.966 \qquad (r = -0983)$$

瓦斯消耗量大約97%之變異可以用他與溫度的直線關聯來說明