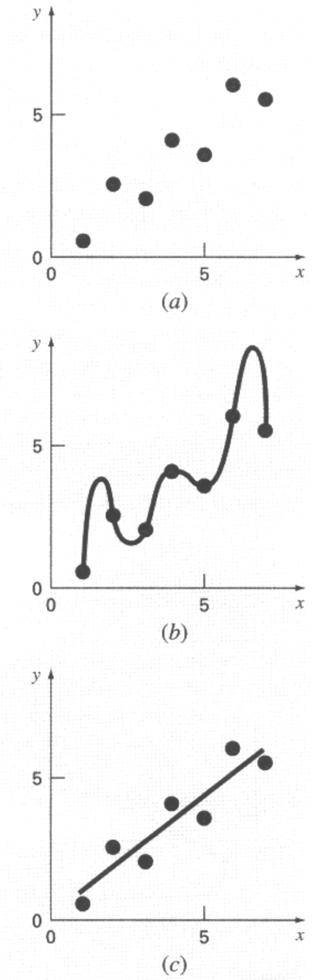
บทที่ 8

การปรับเส้น และการประมาณค่า

ข้อมูลที่ได้จากการทดลอง จะมีการเบี่ยงเบนมากเราอาจจะใช้กราฟเส้นตรงหรือเส้นโค้งใด ๆ มาแทนข้อมูลที่ทดลองได้นี้ ในบทนี้จะศึกษาการคำนวณว่าฟังก์ชันใดเป็นตัวแทนของชุดข้อมูลได้ดีที่สุดและได้ด้วยวิธีใด



รูปที่ 8.1 ข้อมูล a) จากการทดลอง b) จากการ fit ด้วยโพลีโนเมียล c) จากการ fit ด้วยเส้นตรง

8.1 การถดถอยเชิงเส้น (Linear Regression)

สมมุติว่าจากการทดลองได้ข้อมูล 1 ชุด คือ $(x_1,y_1),(x_2,y_2),........,(x_n,y_n)$ วิธีง่ายที่สุดคือ แทนข้อมูลชุดนี้ด้วยกราฟเส้นตรง

$$y = a_0 + a_{1x} + e$$

เมื่อ a_0 = จุดตัดบนแกน y

a_1 = ความชันของกราฟ

e = ความคลาดเคลื่อนจากค่าทดลอง

เส้นตรงที่นำมาแทนข้อมูลจะต้องเป็นเส้นตรงที่ทำให้ผลรวมของความคลาดเคลื่อนยกกำลังสอง มีค่าน้อยที่สุด หรือเขียนเป็นสมการได้ว่า

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i \right)$$

$$S_i = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(y_{i,..}, y_{i,\text{model}} \right)^2$$

$$= \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i \right)^2$$

ต้องการหาค่า a_0 และ a_1

$$\frac{\partial S_r}{\partial a_0} = -2\sum \left(y_i - a_0 - a_1 x_i \right)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum \left| \left(y_i - a_0 - a_1 x_i \right) x_i \right|$$

หาค่า S_r ที่น้อยที่สุด โดย set derivative = 0

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$
$$0 = \sum y_i x_i \sum a_0 x_i \sum a_1 x_i^2$$

เนื่องจาก $\sum a_0 = na_0$ ดังนั้น

$$na_0 + \left(\sum x_i\right) a_1 = \sum y_i$$
$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

สมการทั้งสองนี้ มีชื่อเรียกว่า normal equation ได้

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

และหา a_0 จาก $a_0 = \overline{y} - a_1 \overline{x}$

ความคลาดเคลื่อนจากการประมาณค่า

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

ถ้า S_t = total sum of square

$$S_t = \sum \left(y_i - \bar{y}\right)^2$$

เมื่อ r^2 = coefficient of determination

$$r^2 = \frac{S_t - S_r}{S_t}$$

r = $\sqrt{r^2}$ = correlation coefficient

สำหรับการ fit ที่ดี $S_r = 0$ และ $r = r^2 = 1$ และสำหรับการที่ fit ไม่ดี $r = r^2 = 0$, $S_r = S_t$

สูตรที่ใช้คำนวณ r ของคอมพิวเตอร์ คือ

$$r = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}$$

ตัวอย่างที่ 8.1 จง fit กราฟเส้นตรงกับค่า x และ y ใน ตาราง ซึ่งเป็นค่าที่ได้จากการทดลอง และจงเขียนโปรแกรมเพื่อ fit ข้อมูลนี้

วิธีทำ จะต้องคำนวณค่าเหล่านี้ก่อน

$$n = 7, \sum x_i y_i = 119.5, \sum x_i^{\ 2} = 140$$
$$\sum x_i = 28, \bar{x} = \frac{28}{7} = 4$$
$$\sum y_i = 24, \bar{y} = \frac{24}{7} = 3.428571$$
$$a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$
$$a_0 = 3.428571 - 0.8392857(4) = 0.07142857$$

จะได้กราฟเส้นตรง $y = 0.07142857 + 0.8392857x$

x_i	y_i	$\left(y_i - \overline{y}\right)^2$	$\left(y_i - a_0 - a_1 x_i\right)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
$\sum$	24.0	22.7143	2.9911

โปรแกรมที่ใช้ fit ข้อมูลนี้ คือ

```
*=====================================================
      PROGRAM L_REG
*=====================================================
*     SET THIS BLOCK FOR NEW PROBLEM
*------------------------------------------------------
      PARAMETER (N=7)
      DIMENSION X(100),Y(100)
      DATA (X(I),I=1,N) /1.,2.,3.,4.,5.,6.,7./
      DATA (Y(I),I=1,N) /0.5,2.5,2.,4.,3.5,6.,5.5/
*=====================================================
*     LINEAR REGRESSION PROCESS
*------------------------------------------------------
      SUMX=0.
      SUMY=0.
      SUMX2=0.
      SUMY2=0.
```

```
      SUMXY=0.
      ST=0.
      SR=0.
      DO 10 I=1,N
        SUMX=SUMX+X(I)
        SUMY=SUMY+Y(I)
        SUMX2=SUMX2+X(I)*X(I)
        SUMY2=SUMY2+Y(I)*Y(I)
        SUMXY=SUMXY+X(I)*Y(I)
10       CONTINUE
      DETER=N*SUMX2-SUMX*SUMX
      A0=(SUMY*SUMX2-SUMXY*SUMX)/DETER
      A1=(N*SUMXY-SUMX*SUMY)/DETER
*=====================================================
*        OUTPUT REPORT
*-----------------------------------------------------
      WRITE(*,100)
      WRITE(*,200)
      WRITE(*,300)
      WRITE(*,200)
      DO 20 ITH=1,N
        WRITE(*,400) ITH,X(ITH),Y(ITH)
20       CONTINUE
      WRITE(*,200)
      WRITE(*,800)
      READ(*,*)
      WRITE(*,200)
      WRITE(*,500)'X',SUMX
      WRITE(*,500)'Y',SUMY
```

```
      WRITE(*,500)'X*Y',SUMXY
      WRITE(*,500)'SQUARE X',SUMX2
      WRITE(*,500)'SQUARE Y',SUMY2
      WRITE(*,200)
      WRITE(*,600) A0,A1
      WRITE(*,200)
      WRITE(*,700) A0,A1
      WRITE(*,200)
*===============================================================
*     ERROR REPORT
*-------------------------------------------------------
      ST=0.
      YA=SUMY/N
      DO 30 IE=1,N
        ST=ST+(Y(IE)-(YA))**2.
        SR=SR+(Y(IE)-A1*X(IE)-A0)**2.
30     CONTINUE
      WRITE(*,810) ST,SR
      WRITE(*,200)
*     ---------------------------------------------
*     STANDARD DIVIATION
*     ---------------------------------------------
      SY=SQRT(ST/(N-1))
      WRITE(*,900) SY
*     ---------------------------------------------
* STANDARD ERROR OF THE ESTIMATE
*     ---------------------------------------------
      SYX=SQRT(SR/(N-2))
      WRITE(*,1000) SYX
```

```
*        ------------------------------------------
*    COEFFICIENT OF DETERMINATION
*        ------------------------------------------
         R2=(ST-SR)/ST
         WRITE(*,1100) R2
*        ------------------------------------------
*        CORRELATION COEFFICIENT
*        ------------------------------------------
         R=((N*SUMXY)-(SUMX*SUMY))/
    +      (SQRT(N*SUMX2-SUMX**2.)*SQRT(N*SUMY2-SUMY**2.))
         WRITE(*,1200) R
*        ------------------------------------------
*        PERCENT OF THE VARIABILITY OF THE DATA.
*        ------------------------------------------
         P=R2*100
         WRITE(*,1300) P
         WRITE(*,200)
*----------------------------------------------------------
100      FORMAT(5X,'DATA')
200      FORMAT(5X,
    +   '---------------------------------------------------')
300      FORMAT(14X,'ith',9X,'Xi',12X,'Yi')
400      FORMAT(13X,I3,1X,2F14.6)
500      FORMAT(6X,'SUMMATION OF',A9,1X,'=',F12.6)
600      FORMAT(5X,' COEFFICIENT A0 =',F14.10,
    +       /,5X,' COEFFICIENT A1 =',F14.10)
700      FORMAT(6X,'THE LEAST-SQUARE FIT IS:',//,
    +         6X,'Y =',F14.10,' +',F14.10,' * X')
800      FORMAT(/,6X,'PRESS ENTER TO SEE THE RESULT')
```

```
810       FORMAT(6X,'ST = ',F14.10,4X,',     SR = ',F14.10)
900       FORMAT(6X,'STANDARD DIVIATION              = ',F14.10)
1000      FORMAT(6X,'STANDARD ERROR OF THE ESTIMATE = ',F14.10)
1100      FORMAT(6X,'COEFFICIENT OF DETERMINATION   = ',F14.10)
1200      FORMAT(6X,'CORRELATION COEFFICIENT        = ',F14.10)
1300      FORMAT(/,6X,'THE LINE EXPLAINS ',F6.3,
     +              ' % OF THE DATA VARIABILITY.')
*---------------------------------------------------
        STOP
        END
*------------------------------------------------------------------
```

เอาท์พุทจากโปรแกรม คือ

```
DATA
------------------------------------------------------------------
         ith          Xi              Yi
------------------------------------------------------------------
          1        1.000000       0.500000
          2        2.000000       2.500000
          3        3.000000       2.000000
          4        4.000000       4.000000
          5        5.000000       3.500000
          6        6.000000       6.000000
          7        7.000000       5.500000
------------------------------------------------------------------

 PRESS ENTER TO SEE THE RESULT

------------------------------------------------------------------
 SUMMATION OF         X =   28.000000
 SUMMATION OF         Y =   24.000000
 SUMMATION OF       X*Y =  119.500000
 SUMMATION OF SQUARE X =  140.000000
 SUMMATION OF SQUARE Y =  105.000000
------------------------------------------------------------------
 COEFFICIENT A0 =  0.0714285746
 COEFFICIENT A1 =  0.8392857313
------------------------------------------------------------------
 THE LEAST-SQUARE FIT IS:

 Y =  0.0714285746 +  0.8392857313 * X
------------------------------------------------------------------
 ST =  22.7142868042     ,    SR =   2.9910714626
------------------------------------------------------------------
 STANDARD DIVIATION              =   1.9456912279
 STANDARD ERROR OF THE ESTIMATE =   0.7734431624
 COEFFICIENT OF DETERMINATION   =   0.8683176041
 CORRELATION COEFFICIENT        =   0.9318356514

 THE LINE EXPLAINS 86.832 % OF THE DATA VARIABILITY.
------------------------------------------------------------------
```

สมการที่ใช้แทนข้อมูล คือ

$$y = 0.0714285746 + 0.8392857313x$$

โดยมี $S_t = 22.714$

$S_r = 2.991$

ค่าเบียงเบนมาตรฐาน = 1.946

สัมประสิทธิ์สหสัมพันธ์ = 0.932

ในทางปฏิบัติ นักศึกษาสามารถที่จะนำโปรแกรมนี้ไป fit ข้อมูลอื่นที่นักศึกษาได้จากการทดลองได้เป็นอย่างดี

ตัวอย่าง 8.2 จงแสดงวิธีคำนวณ total standard estimate และ correlation coefficient ของข้อมูลในตัวอย่าง 8.1

วิธีทำ

$$S_y = \sqrt{\frac{22.7143}{7-1}} = 1.9457 = \text{ standard deviation}$$

$$S_{y/x} = \sqrt{\frac{2.9911}{7-2}} = 0.7735 = \text{ standard error of estimate}$$

เนื่องจาก $S_{y/x} < S_y$ ดังนั้นจึงสามารถใช้ regression ได้ดี

$$r^2 = \frac{22.7143 - 2.9911}{22.7143} = 0.868$$

หรือ

$$r = \sqrt{0.868} = 0.932$$

ซึ่งมีค่าตรงกับการคำนวณจากโปรแกรม

ตัวอย่างที่ 8.3 จากการคำนวณทางฟิสิกส์

$$v(t) = \frac{gm}{c}\left(1 - e^{(-c/m)t}\right) \tag{8.1}$$

จากสูตรเอ็มไพริคอล

$$v(t) = \frac{gm}{c}\left(\frac{t}{3.75 + t}\right) \tag{8.2}$$

เปรียบเทียบความเร็ว โดยอาศัย Regression

วิธีทำ linear reqression ระหว่างค่าทฤษฎี (สมการ 8.1) กับค่าจากการทดลอง

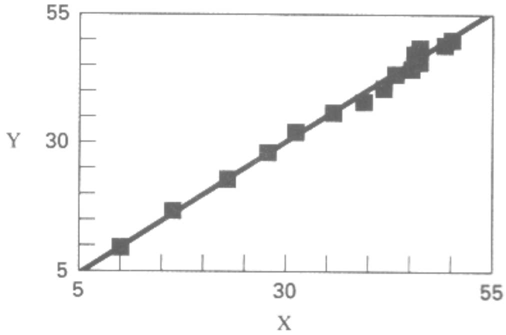
$$v_{\mathrm{mod}el} = -0.859 + 1.032 v_{measure}$$

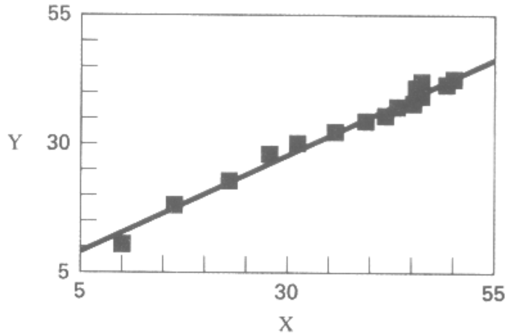
linear reqression ระหว่างค่าเอ็มไพริคอล (สมการ 8.2) กับค่าจากการทดลอง

$$v_{\mathrm{mod}el} = 5.776 + 0.752 v_{measure}$$

ทั้งสองโมเดลได้ correlation coefficient มากกว่า 0.99 ซึ่งแสดงความสัมพันธ์ที่ดีของโมเดลทั้งสอง ดังแสดงในรูปที่ 8.2

Time,s	v_f การทดลอง m/s	v_f คำนวณ m/s Eq(8.1)	v_f คำนวณ m/s Eq(8.2)
1	10.00	8.953	11.240
2	16.30	16.405	18.570
3	23.00	22.607	23.729
4	27.50	27.769	27.556
5	31.00	32.065	30.509
6	35.60	35.641	32.855
7	39.00	38.617	34.766
8	41.50	41.095	36.351
9	42.90	43.156	37.687
10	45.00	44.872	38.829
11	46.00	46.301	39.816
12	45.50	47.490	40.678
13	46.00	48.479	41.437
14	49.00	49.303	42.110
15	50.00	49.988	42.712

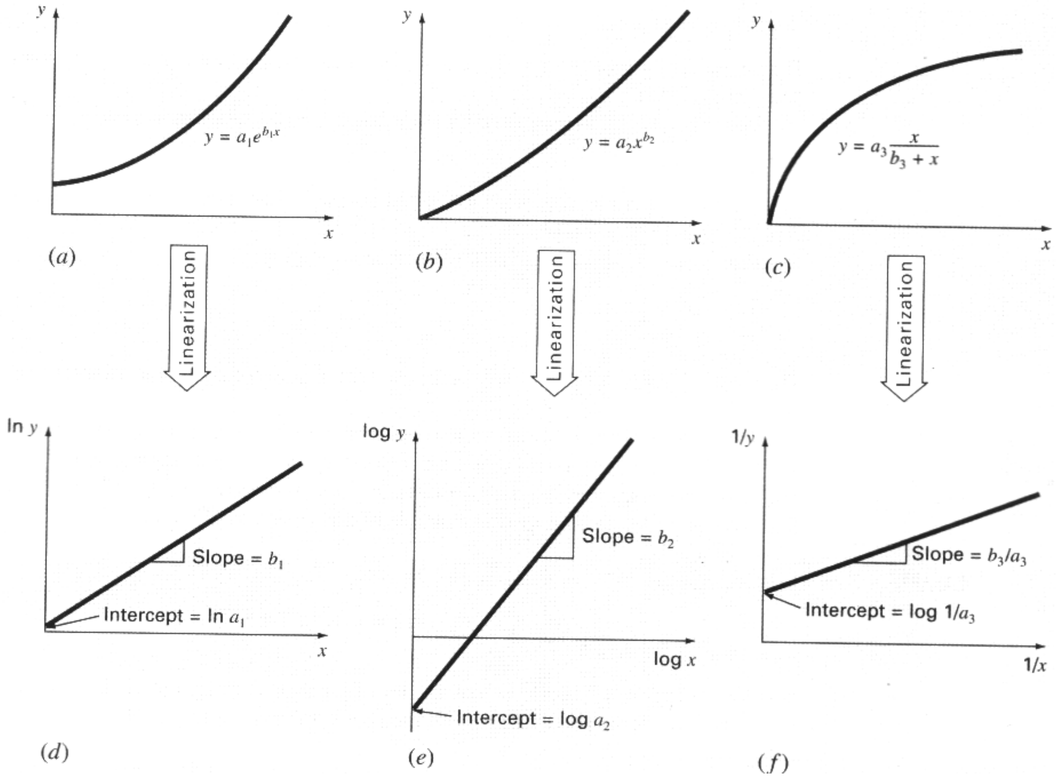




รูปที่ 8.2 a) ใช้ linear regression เปรียบเทียบค่าจากทฤษฎีกับค่าจากการทดลอง b) ใช้ linear regression เปรียบเทียบค่าจากสูตรเอ็มไพริคอลกับค่าจากการทดลอง

8.2 การถดถอยไม่เป็นเชิงเส้น

Linear regression เป็นเครื่องมือที่มีประสิทธิภาพในการ fit ข้อมูล อย่างไรก็ตาม บางครั้งความสัมพันธ์ของข้อมูลไม่ได้เป็นชนิดเชิงเส้น จะต้อง fit ข้อมูลโดยใช้โมเดลอื่น เช่น exponential model, power equation หรือ saturation – growth – rate equation เป็นต้น



รูปที่ 8.3 Model ชนิดอื่นที่ใช้ในการ fit ข้อมูล

สมการที่ใช้สำหรับโมเดลทั้งสาม คือ

1. **Exponential model.**

$$y = a_1 e^{b_1 x} \Rightarrow \ln y = \ln a_1 + \ln b_1$$

2. **Simple power equation.**

$$y = a_2 x^{b_2} \Rightarrow \log y = b_2 \log x + \log a_2$$

3. **Saturation-growth-rate equation.**

$$y = a_3 \frac{x}{b_3 + x} \Rightarrow \frac{1}{y} = \frac{b_3}{a_3}\frac{1}{x} + \frac{1}{a_3}$$

ตัวอย่างที่ 8.4 จงอาศัย least – squares regression เพื่อ fit เส้นตรงกับข้อมูลต่อไปนี้

x	1	3	5	7	10	12	13	16	18	20
y	4	5	6	5	8	7	6	9	12	11

จากนั้นให้สลับ x กับ y แล้ว fit ข้อมูลอีกครั้งหนึ่ง

วิธีทำ ส่วนหนึ่งของโปรแกรมที่ใช้ในการ fit คือ

```
*========================================================
      PROGRAM L_REG
*========================================================
*     SET THIS BLOCK FOR NEW PROBLEM
*--------------------------------------------------
      PARAMETER (N=10)
      DIMENSION X(100),Y(100)
      DATA (X(I),I=1,N) /1.,3.,5.,7.,10.,12.,13.,
     +             16.,18.,20./
      DATA (Y(I),I=1,N) /4.,5.,6.,5.,8.,7.,6.,9.,
     +             12.,11./
*========================================================
```

เอาท์พุทจากโปรแกรม คือ

```
DATA
------------------------------------------------------
         ith          Xi             Yi
------------------------------------------------------
          1        1.000000       4.000000
          2        3.000000       5.000000
          3        5.000000       6.000000
          4        7.000000       5.000000
          5       10.000000       8.000000
          6       12.000000       7.000000
          7       13.000000       6.000000
          8       16.000000       9.000000
          9       18.000000      12.000000
         10       20.000000      11.000000
------------------------------------------------------

 PRESS ENTER TO SEE THE RESULT

------------------------------------------------------
 SUMMATION OF        X =  105.000000
 SUMMATION OF        Y =   73.000000
 SUMMATION OF      X*Y =  906.000000
 SUMMATION OF SQUARE X = 1477.000000
 SUMMATION OF SQUARE Y =  597.000000
------------------------------------------------------
 COEFFICIENT A0 =  3.3887851238
 COEFFICIENT A1 =  0.3724966645
------------------------------------------------------
 THE LEAST-SQUARE FIT IS:

 Y =  3.3887851238 +  0.3724966645 * X
------------------------------------------------------
 ST =  64.1000061035     ,    SR =  12.1367139816
------------------------------------------------------
 STANDARD DIVIATION               =   2.6687493324
 STANDARD ERROR OF THE ESTIMATE =   1.2317017317
 COEFFICIENT OF DETERMINATION    =   0.8106597066
 CORRELATION COEFFICIENT         =   0.9003664255

 THE LINE EXPLAINS 81.066 % OF THE DATA VARIABILITY.
------------------------------------------------------
```

สมการเส้นตรงที่ใช้แทนข้อมูล คือ

$$y = 3.389 + 0.372x$$

โดยมี ความคลาดเคลื่อนมาตรฐาน = 1.232

สัมประสิทธิ์สมสัมพันธ์ = 0.900

เมื่อสลับ x กับ y แล้ว fit เส้นตรง จะได้เอาท์พุท คือ

```
DATA
--------------------------------------------------------------------
        ith          Xi              Yi
--------------------------------------------------------------------
         1        4.000000        1.000000
         2        5.000000        3.000000
         3        6.000000        5.000000
         4        5.000000        7.000000
         5        8.000000       10.000000
         6        7.000000       12.000000
         7        6.000000       13.000000
         8        9.000000       16.000000
         9       12.000000       18.000000
        10       11.000000       20.000000
--------------------------------------------------------------------

 PRESS ENTER TO SEE THE RESULT

--------------------------------------------------------------------
 SUMMATION OF          X =   73.000000
 SUMMATION OF          Y =  105.000000
 SUMMATION OF        X*Y =  906.000000
 SUMMATION OF SQUARE X =  597.000000
 SUMMATION OF SQUARE Y = 1477.000000
--------------------------------------------------------------------
 COEFFICIENT A0 = -5.3868956566
 COEFFICIENT A1 =  2.1762869358
--------------------------------------------------------------------
 THE LEAST-SQUARE FIT IS:

 Y = -5.3868956566 +  2.1762869358 * X
--------------------------------------------------------------------
 ST  = 374.5000000000    ,    SR =  70.9079513550
--------------------------------------------------------------------
 STANDARD DIVIATION                =   6.4506673813
 STANDARD ERROR OF THE ESTIMATE =   2.9771621227
 COEFFICIENT OF DETERMINATION    =   0.8106596470
 CORRELATION COEFFICIENT          =   0.9003664255

 THE LINE EXPLAINS 81.066 % OF THE DATA VARIABILITY.
--------------------------------------------------------------------
```

ซึ่งจะได้สมการเส้นตรงใหม่ที่ใช้แทนข้อมูล คือ

$$y = -5.387 + 2.176x$$

โดยมี ความคลาดเคลื่อนมาตรฐาน = 2.977

สัมประสิทธิ์สมสัมพันธ์ = 0.900

ตัวอย่างที่ 8.5 จง Fit Simple power equation กับข้อมูลในตาราง

วิธีทำ ได้ $\log y = 1.75 \log x - 0.3$

ข้อมูลที่ใช้ในการ fit คือ

x	y	Log x	Log y
1	0.5	0	-0.301
2	1.7	0.301	0.226
3	3.4	0.477	0.534
4	5.7	0.602	0.753
5	8.4	0.699	0.922

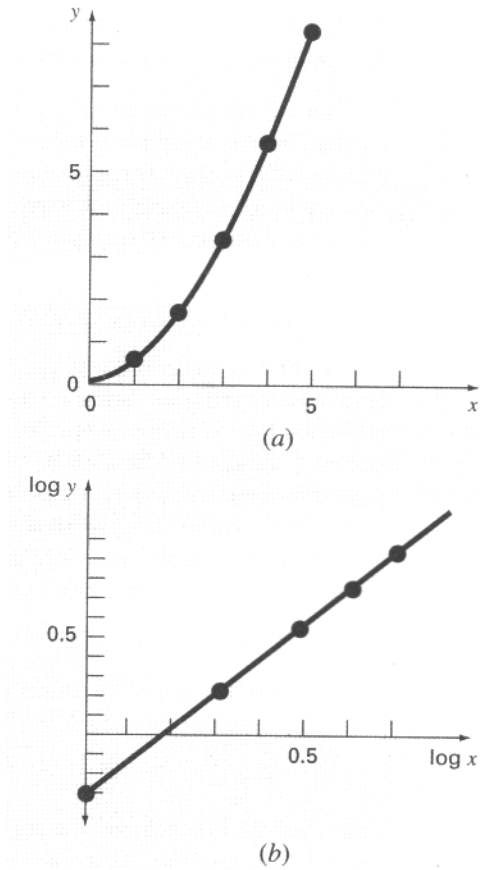
จุดตัดบนแกน $\log a_2 = -0.3$

ดังนั้น $a_2 = 10^{-0.3}$

$= 0.5$

ความชันของกราฟ คือ $b_2 = 1.75$

ดังนั้น $y = 0.5x^{1.75}$



รูปที่ 8.4 การ fit โดยใช้ power equation

8.3 การถดถอยโพลีโนเมียล

สมมุติว่าต้องการ fit ด้วย second-order polynomial

$$y = a_0 + a_1 x + a_2 x^2 + e$$

Sum ของ error ยกกำลังสอง คือ

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^{\,2} \right)^2$$

หา derivative เทียบกับแต่ละ unknown

$$\frac{\partial S_r}{\partial a_0} = -2\sum\left(y_i - a_0 - a_1 x_i - a_2 x_i^{\,2}\right)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_i\left(y_i - a_0 - a_1 x_i - a_2 x_i^{\,2}\right)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_i^{\,2}\left(y_i - a_0 - a_1 x_i - a_2 x_i^{\,2}\right)$$

Set ทั้งสามสมการ =0 จะได้ normal eqs

$$(n)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^{\,2}\right)a_2 = \sum y_i$$

$$\left(\sum x_i\right)a_0 + \left(\sum x_i^{\,2}\right)a_1 + \left(\sum x_i^{\,3}\right)a_2 = \sum x_i y_i$$

$$\left(\sum x_i^{\,2}\right)a_0 + \left(\sum x_i^{\,3}\right)a_1 + \left(\sum x_i^{\,4}\right)a_2 = \sum x_i^{\,2} y_i$$

โดย $\sum$ จาก i=1 ถึง n สมการทั้งสามมี unknown ที่ต้องหาค่า คือ a_0, a_1 และ a_2 โดยสัมประสิทธิ์หาได้จากข้อมูล จะเห็นได้ว่า โจทย์ที่หา least-squares ของ second-orderpolynomial จะเหมือนกับการแก้ระบบ 3 สมการ

ในกรณีของ $m^{th} -$ order polynomial

$$y = a_0 + a_1 x + a_2 x^2 + ... + a_m x^m + e$$

จะเหมือนกับการแก้ระบบ m+1 สมการ

$$S_{y/x} = \sqrt{\frac{S_r}{n - (\mathrm{m+1}\ )}}$$

ตัวอย่างที่ 8.6 จง Fit second-order polynomial กับข้อมูลในตาราง

x_i	y_i	$\left(y_i - \overline{y_i}\right)^2$	$\left(y_i - a_0 - a_1 x_i + a_2 x_i^{\ 2}\right)$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.2	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1272.11	0.09439
ε	152.6	2513.39	3.74657

และจงเขียนโปรแกรมเพื่อ fit ข้อมูลนี้

วิธีทำ จากข้อมูลที่ให้

$$m = 2, \sum x_i = 15, \sum x_i^{\ 4} = 979$$
$$n = 6, \sum y_i = 152.6, \sum x_i y_i = 585.6$$
$$\overline{x} = 2.5, \sum x_i^{\ 2} = 55, \sum x_i^{\ 2} y_i = 2488$$
$$\overline{y} = 25.433, \sum x_i^{\ 3} = 225$$

ดังนั้นจะเขียน simultaneous linear eqs. ได้เป็น

$$\begin{pmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{pmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2488.8 \end{Bmatrix}$$

แก้สมการโดยวิธี Gauss elimination จะได้

$$a_0 = 2.47857, a_1 = 2.35929, a_2 = 1.86071$$

ดังนั้น

$$y = 2.47857 + 2.35929x + 1.86071$$

$$S_{y/x} = \sqrt{\frac{3.74657}{6-3}} = 1.12$$

$$r^2 = \frac{2513.39 - 3.74657}{2513.39} = 0.99851$$

$r = 0.9925$ correlation coefficient

โปรแกรมที่ใช้ในการ fit ข้อมูล คือ

```
*=========================================================
        PROGRAM POLY_REG
*=========================================================
*       SET THIS BLOCK FOR NEW PROBLEM
*---------------------------------------------------
        PARAMETER (N=6)
        DIMENSION X(100),Y(100)
        DIMENSION A(10,10),B(10),XX(10),SR(10)
        DATA (X(I),I=1,N) /0.,1.,2.,3.,4.,5./
        DATA (Y(I),I=1,N) /2.1,7.7,13.6,27.2,40.9,61.1/
*=========================================================
        WRITE(*,*)' PLEASE ENTER ORDER OF POLYNOMIAL',
     +            ' (MAXIMUM ORDER IS 10)'
        READ(*,*) M
        DO 10 IR=1,10
          B(IR)=0.
          DO 10 IC=1,10
            A(IR,IC)=0.
10      CONTINUE
```

```
      DO 20 IR=1,M+1
        DO 30 IC=1,M+1
          K=IR+IC-2
          DO 40 I=1,N
            A(IR,IC)=A(IR,IC)+X(I)**K
40          CONTINUE
30        CONTINUE
        DO 50 I=1,N
          B(IR)=B(IR)+Y(I)*(X(I)**(IR-1))
50        CONTINUE
20      CONTINUE
      MP1=M+1
      CALL GAUSS(MP1,A,B,XX)
*     ---------------------------------
      WRITE(*,100) M
100     FORMAT(/,
   +  ' COEFFICIENTS OF FITTED POLYNOMIAL ORDER ',I2,' ARE:', /)
      DO 60 I=1,M+1
        IM1=I-1
        WRITE(*,300) IM1,XX(I)
300       FORMAT(5X,'A(',I2,') =',F16.10)
60      CONTINUE
*     ----------------------------------
*=============================================================
*       ERROR REPORT
*-----------------------------------------------------
      SUMSR=0.
      DO 70 I=1,N
        SR(I)=0.
```

```
          P=0.
          DO 80 J=1,M
            P=P+(XX(J+1)*(X(I)**J))
80         CONTINUE
          SR(I)=(Y(I)-XX(1)-P)**2.
          SUMSR=SUMSR+SR(I)
70       CONTINUE
        YK=0.
        DO 90 K=1,N
          YK=YK+Y(K)
90       CONTINUE
        YAV=YK/N
        SUMYY=0.
        DO 91 L=1,N
          SUMYY=SUMYY+(Y(L)-YAV)**2.
91       CONTINUE
        SYX=SQRT(SUMSR/(N-(M+1)))
        R2=(SUMYY-SUMSR)/SUMYY
        R=SQRT(ABS(R2))
        PC=R2*100.
        WRITE(*,400) SUMYY,SUMSR,SYX,R2,R,PC
400      FORMAT(/,5X,'ST = ',F16.7,
    +        /,5X,'SR = ',F16.7,
    +        /,5X,'THE STANDARD ERROR IS :',F12.7
    +        /,5X,'THE COEFFICIENT OF DETERMINATION IS :',F12.7,
    +        /,5X,'THE CORRELATION COEFFICIENT IS :',F12.7,
    +        /,5X,'THE RESULT EXPLAINS ',F9.3,
    +             ' % OF THE DATA VARIABILITY.')
        STOP
```

```
      END
*-----------------------------------------------------
      SUBROUTINE GAUSS(NN,A,B,XX)
      DIMENSION A(10,10),B(10),XX(10)
      CALL SCL(NN,A,B)
      DO 100 IP=1,NN-1
      CALL PIVOT(NN,A,B,IP)
      DO 200 IE=IP+1,NN
      RATIO = A(IE,IP)/A(IP,IP)
      DO 300 IC=IP+1,NN
      A(IE,IC) = A(IE,IC) - RATIO*A(IP,IC)
300     CONTINUE
      B(IE)=B(IE)-RATIO*B(IP)
200     CONTINUE
      DO 400 IE=IP+1,NN
      A(IE,IP)=0.
400     CONTINUE
100     CONTINUE
      XX(NN)=B(NN)/A(NN,NN)
      DO 500 IE=NN-1,1,-1
      SUM=0.
      DO 600 IC=IE+1,NN
      SUM=SUM+A(IE,IC)*XX(IC)
600     CONTINUE
      XX(IE)=(B(IE)-SUM)/A(IE,IE)
500     CONTINUE
      RETURN
      END
*-----------------------------------------------------
```

```
      SUBROUTINE SCL(NN,A,B)
      DIMENSION A(10,10),B(10)
      DO 10 IE=1,NN
        BIG=ABS(A(IE,1))
        DO 20 IC=2,NN
          AMAX=ABS(A(IE,IC))
          IF(AMAX.GT.BIG) BIG=AMAX
20        CONTINUE
        DO 30 IC=1,NN
          A(IE,IC)=A(IE,IC)/BIG
30        CONTINUE
        B(IE)=B(IE)/BIG
10      CONTINUE
      RETURN
      END
*-------------------------------------------------------
      SUBROUTINE PIVOT(NN,A,B,IP)
      DIMENSION A(10,10),B(10)
      JP=IP
      BIG=ABS(A(IP,IP))
      DO 10 I=IP+1,NN
        AMAX=ABS(A(I,IP))
        IF(AMAX.GT.BIG) THEN
          BIG=AMAX
          JP=I
        ENDIF
10      CONTINUE
      IF(JP.NE.IP) THEN
        DO 20 J=IP,NN
```

```
          DUMY   = A(JP,J)
          A(JP,J)= A(IP,J)
          A(IP,J)= DUMY
20        CONTINUE
        DUMY  = B(JP)
        B(JP) = B(IP)
        B(IP) = DUMY
        ENDIF
        RETURN
        END
*-----------------------------------------------------------
```

เอาท์พุทที่ได้จากโปรแกรม คือ

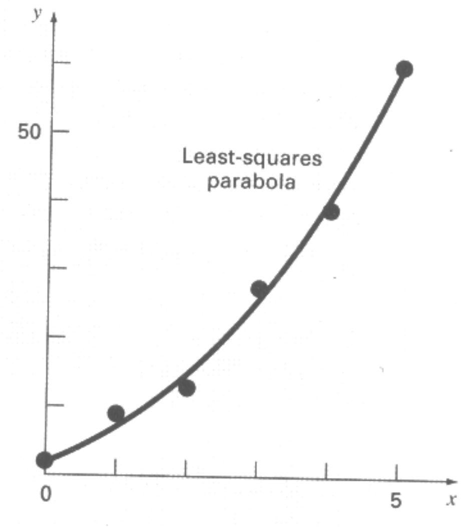
```
 PLEASE ENTER ORDER OF POLYNOMIAL (MAXIMUM ORDER IS 10)
2

 COEFFICIENTS OF FITTED POLYNOMIAL ORDER  2 ARE:

    A( 0) =    2.4785928726
    A( 1) =    2.3592605591
    A( 2) =    1.8607189655

    ST =     2513.3933105
    SR =        3.7465670
    THE STANDARD ERROR IS :   1.1175221
    THE COEFFICIENT OF DETERMINATION IS :   0.9985093
    THE CORRELATION COEFFICIENT IS :   0.9992544
    THE RESULT EXPLAINS    99.851 % OF THE DATA VARIABILITY.
```

สมการที่ fit ข้อมูล คือ

$$y = 2.47859 + 2.35926x + 1.860719x^2$$

ซึ่งตรงกับการคำนวณด้วยคือ มีค่าความคลาดเคลื่อนมาตรฐาน $S_{y/x}$ = 1.12 และสัมประสิทธิ์สหสัมพันธ์ = 0.9992544 แสดงว่า 99.851% ของข้อมูลอยู่ภายในเส้นโค้งนี้ การ fit ข้อมูลด้วยสมการ quadration จึงมีประสิทธิภาพมาก ดังแสดงในรูปที่ 8.5

> โปรแกรมที่ใช้ในการ fit ข้อมูลด้วยโพลีโนเมียลนี้ สามารถใช้ได้ถึงโพลีโนเมียลลำดับ 10 จึงมีประสิทธิภาพในการ fit ข้อมูลมาก



รูปที่ 8.5 การ fit ด้วย second – order polynomial

ตัวอย่างที่ 8.7 ก) จง fit ข้อมูลต่อไปนี้ด้วยเส้นตรงและจงเขียนกราฟของข้อมูล และกราฟเส้นตรงเพื่อการเปรียบเทียบ

ข) เช่นเดียวกับข้อ ก) แต่ fit ด้วยเส้นพาราโบลา

x	2	3	4	7	8	9	5	5
y	9	6	5	10	9	11	2	3

วิธีทำ ก) ส่วนหนึ่งโปรแกรมที่ใช้ในการ fit ด้วยเส้นตรง คือ

```
*=========================================================
      PROGRAM LIN_REG
*=========================================================
*     SET THIS BLOCK FOR NEW PROBLEM
*---------------------------------------------------
      PARAMETER (N=8)
      DIMENSION X(100),Y(100)
```

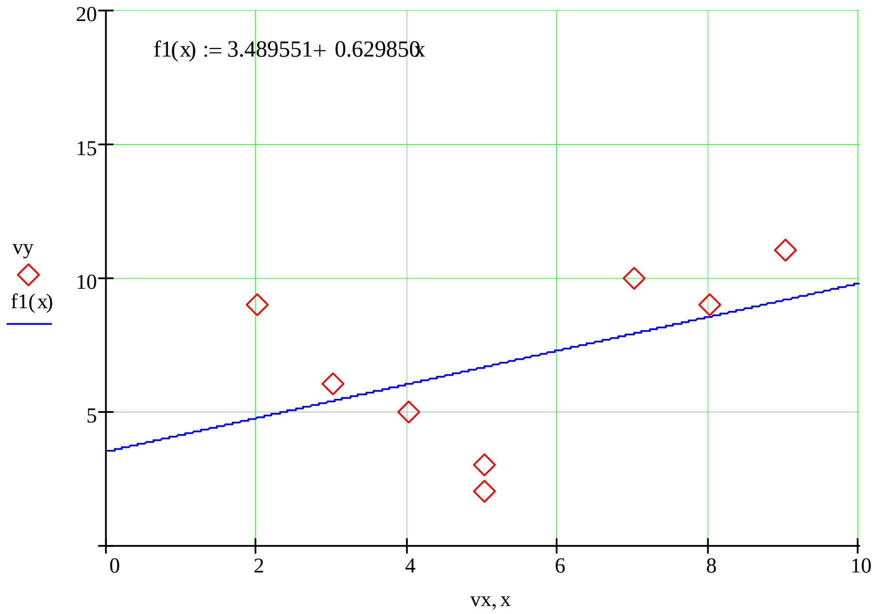
```
      DATA (X(I),I=1,N) /2.,3.,4.,7.,8.,9.,5.,5./
      DATA (Y(I),I=1,N) /9.,6.,5.,10.,9.,11.,2.,3./
*==========================================================
*     LINEAR REGRESSION PROCESS
*--------------------------------------------------
```

เอาท์พุทจากโปรแกรม คือ

```
DATA
------------------------------------------------------------
          ith          Xi             Yi
------------------------------------------------------------
           1        2.000000       9.000000
           2        3.000000       6.000000
           3        4.000000       5.000000
           4        7.000000      10.000000
           5        8.000000       9.000000
           6        9.000000      11.000000
           7        5.000000       2.000000
           8        5.000000       3.000000
------------------------------------------------------------

 PRESS ENTER TO SEE THE RESULT

------------------------------------------------------------
 SUMMATION OF        X =   43.000000
 SUMMATION OF        Y =   55.000000
 SUMMATION OF      X*Y =  322.000000
 SUMMATION OF SQUARE X =  273.000000
 SUMMATION OF SQUARE Y =  457.000000
------------------------------------------------------------
 COEFFICIENT A0 =  3.4895522594
 COEFFICIENT A1 =  0.6298507452
------------------------------------------------------------
 THE LEAST-SQUARE FIT IS:

 Y =  3.4895522594 +  0.6298507452 * X
------------------------------------------------------------
 ST =  78.8750000000     ,    SR =  62.2626876831
------------------------------------------------------------
 STANDARD DIVIATION              =   3.3567628860
 STANDARD ERROR OF THE ESTIMATE =   3.2213528156
 COEFFICIENT OF DETERMINATION    =   0.2106156796
 CORRELATION COEFFICIENT         =   0.4589288831

 THE LINE EXPLAINS 21.062 % OF THE DATA VARIABILITY.
------------------------------------------------------------
```

สมการที่ fit คือ $y = 3.48955 + 0.62985x$

โดยมีสัมประสิทธิ์สหสัมพันธ์ = 0.45893 แสดงถึงความสัมพันธ์ระหว่างข้อมูลและกราฟเส้นตรงที่เลวมาก เพราะว่าข้อมูลเพียง 21.062 % เท่านั้น ที่แทนด้วยกราฟเส้นตรง ดังแสดงในรูปต่อไปนี้



ข) ส่วนหนึ่งโปรแกรมที่ใช้ในการ fit ด้วยเส้นพาราโบลา คือ

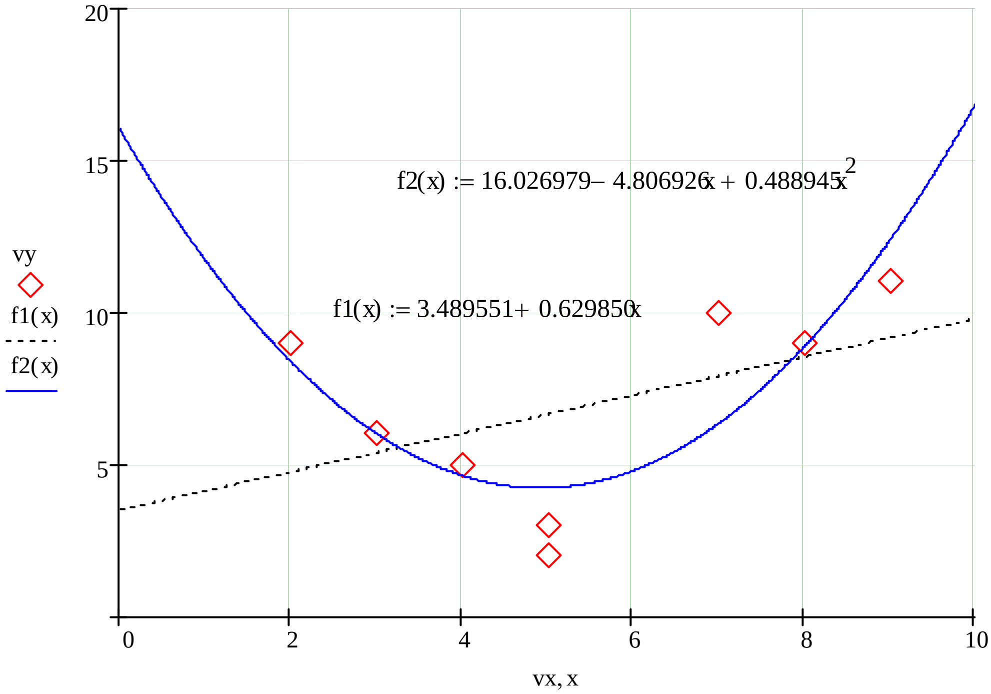
```
*=====================================================
      PROGRAM POLY_REG
*=====================================================
*     SET THIS BLOCK FOR NEW PROBLEM
*-----------------------------------------------------
      PARAMETER (N=8)
      DIMENSION X(100),Y(100)
      DIMENSION A(10,10),B(10),XX(10),SR(10)
      DATA (X(I),I=1,N) /2.,3.,4.,7.,8.,9.,5.,5./
      DATA (Y(I),I=1,N) /9.,6.,5.,10.,9.,11.,2.,3./
```

```
*=======================================================
        WRITE(*,*)' PLEASE ENTER ORDER OF POLYNOMIAL',
     +            ' (MAXIMUM ORDER IS 10)'
        READ(*,*) M
```

เอาท์พุทจากโปรแกรม คือ

```
  PLEASE ENTER ORDER OF POLYNOMIAL (MAXIMUM ORDER IS 10)
2

 COEFFICIENTS OF FITTED POLYNOMIAL ORDER  2 ARE:

     A( 0) =   16.0269794464
     A( 1) =   -4.8069262505
     A( 2) =    0.4889450371

     ST =       78.8750000
     SR =       22.2422161
     THE STANDARD ERROR IS :   2.1091332
     THE COEFFICIENT OF DETERMINATION IS :   0.7180068
     THE CORRELATION COEFFICIENT IS :   0.8473528
     THE RESULT EXPLAINS    71.801 % OF THE DATA VARIABILITY.
```

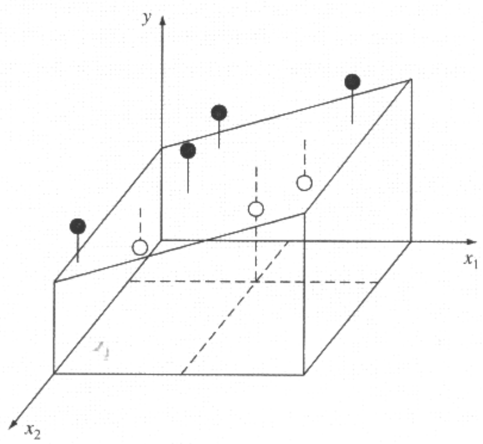
สมการที่ fit คือ $y = 166.02698 - 4.80693x + 0.48894x^2$ โดยมีสัมประสิทธิ์สหสัมพันธ์ = 0.8473528 ซึ่งแสดงว่าข้อมูลสามารถอธิบายได้ด้วยสมการถึง 71.801 % ดีกว่าการ fit ด้วยกราฟเส้นตรงมาก ดังแสดงในรูปต่อไปนี้



8.4 การถดถอยเชิงเส้นหลายตัวแปร

ถ้าหากว่า y เป็นฟังก์ชันเชิงเส้นกับตัวแปรสองตัว เช่น y เป็นฟังก์ชันของ x_1 และ x_2

$$y = a_0 + a_1 x_1 + a_2 x_2 + e$$



รูปที่ 8.6 การถดถอยเชิงเส้นหลายตัวแปร

เช่นเดียวกับวิธีการที่ผ่านมา ค่าสัมประสิทธิ์หาจากผลรวมของกำลังสองของความคลาดเคลื่อน

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} \right)^2$$

$$\frac{\partial S_r}{\partial a_0} = -2\sum \left(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} \right)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i} \left(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} \right)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i} \left(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} \right)$$

เซต partial derivatives = 0

$$\begin{pmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^{2} & \sum x_{1i} x_{2i} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^{2} \end{pmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{Bmatrix}$$

ในกรณี m dimension

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots\ldots + a_m x_m + e$$

เมื่อ standard error

$$S_{y/x} = \sqrt{\frac{S_r}{n - (\mathrm{m}+1\ \ )}}$$

สำหรับกรณี power equation

$$y = a_0 x_1^{a_1} x_2^{a_2} \ldots\ldots x_m^{a_m}$$

ตัวอย่างที่ 8.8 ข้อมูลต่อไปนี้คำนวณจากสมการ

$$y = 5 + 4x_1 - 3x_2$$

x_1	x_2	y
0	0	5
2	1	10
2.5	2	9
1	3	0
4	6	3
7	2	27

จงอาศัย multiple linear regression เพื่อ fit ข้อมูลชุดนี้ และจงเขียนโปรแกรมเพื่อ fit ข้อมูล

วิธีทำ หาค่าสัมประสิทธิ์ต่าง ๆ ได้

$$\begin{pmatrix} 6 & 16.5 & 14 \\ 16.5 & 76.25 & 48\ 54 \\ 14 & 48 & 979 \end{pmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 54 \\ 243.5 \\ 100 \end{Bmatrix}$$

แก้สมการโดยวิธี Gauss elimination จะได้

$$a_0 = 5, a_1 = 4, a_2 = -3$$

ซึ่งตรงกับสมการข้างบน

โปรแกรมที่ใช้ในการ fit ข้อมูล คือ

```
*=========================================================
      PROGRAM MULTI_REG
*=========================================================
*     SET THIS BLOCK FOR NEW PROBLEM
*---------------------------------------------------------
      PARAMETER (N=6,K=2)
      DIMENSION X(100,10),Y(100)
      DIMENSION A(10,10),B(10),XX(10)
      DATA ((X(I,J),J=1,K),I=1,N)/0.,0.,
   +                        2.,1.,
   +                        2.5,2.,
   +                        1.,3.,
   +                        4.,6.,
   +                        7.,2./
      DATA (Y(I),I=1,N) /5.,10.,9.,0.,3.,27./
*=========================================================
*     PROCESS
*---------------------------------------------------------
      DO 10 IR=1,10
        B(IR)=0.
        DO 10 IC=1,10
          A(IR,IC)=0.
10     CONTINUE
      DO 20 I=1,N
        DO 30 IR=1,K+1
          IF(IR.EQ.1) FR=1.
          IF(IR.GT.1) FR=X(I,IR-1)
          DO 40 IC=1,K+1
            IF(IC.EQ.1) FC=1.
```

```
              IF(IC.GT.1) FC=X(I,IC-1)
              A(IR,IC)=A(IR,IC)+FR*FC
40          CONTINUE
            B(IR)=B(IR)+FR*Y(I)
30        CONTINUE
20      CONTINUE
        KP1=K+1
        CALL GAUSS(KP1,A,B,XX)
*=====================================================
*       OUTPUT REPORT
*-----------------------------------------------------
100     FORMAT(5X,
     +  '-----------------------------------------')
200     FORMAT(6X,
     +  ' COEFFICIENTS OF FITTED POLYNOMIAL ARE:')
300     FORMAT(7X,'A(',I2,') =',F16.10)
*       -------------------------------
        WRITE(*,100)
        WRITE(*,200)
        WRITE(*,100)
        DO 50 I=1,K+1
          IM1=I-1
          WRITE(*,300) IM1,XX(I)
50      CONTINUE
        WRITE(*,100)
*-----------------------------------------------------
        STOP
        END
*=====================================================
```

```
      SUBROUTINE GAUSS(NN,A,B,XX)
      DIMENSION A(10,10),B(10),XX(10)
      CALL SCL(NN,A,B)
      DO 100 IP=1,NN-1
      CALL PIVOT(NN,A,B,IP)
      DO 200 IE=IP+1,NN
      RATIO = A(IE,IP)/A(IP,IP)
      DO 300 IC=IP+1,NN
      A(IE,IC) = A(IE,IC) - RATIO*A(IP,IC)
300     CONTINUE
      B(IE)=B(IE)-RATIO*B(IP)
200     CONTINUE
      DO 400 IE=IP+1,NN
      A(IE,IP)=0.
400     CONTINUE
100     CONTINUE
      XX(NN)=B(NN)/A(NN,NN)
      DO 500 IE=NN-1,1,-1
      SUM=0.
      DO 600 IC=IE+1,NN
      SUM=SUM+A(IE,IC)*XX(IC)
600     CONTINUE
      XX(IE)=(B(IE)-SUM)/A(IE,IE)
500     CONTINUE
      RETURN
      END
*--------------------------------------------------
      SUBROUTINE SCL(NN,A,B)
      DIMENSION A(10,10),B(10)
```

```
      DO 10 IE=1,NN
        BIG=ABS(A(IE,1))
        DO 20 IC=2,NN
         AMAX=ABS(A(IE,IC))
         IF(AMAX.GT.BIG) BIG=AMAX
20       CONTINUE
        DO 30 IC=1,NN
          A(IE,IC)=A(IE,IC)/BIG
30       CONTINUE
        B(IE)=B(IE)/BIG
10     CONTINUE
      RETURN
      END
*----------------------------------------------------
      SUBROUTINE PIVOT(NN,A,B,IP)
      DIMENSION A(10,10),B(10)
      JP=IP
      BIG=ABS(A(IP,IP))
      DO 10 I=IP+1,NN
        AMAX=ABS(A(I,IP))
        IF(AMAX.GT.BIG) THEN
          BIG=AMAX
          JP=I
        ENDIF
10     CONTINUE
      IF(JP.NE.IP) THEN
        DO 20 J=IP,NN
          DUMY   = A(JP,J)
          A(JP,J)= A(IP,J)
```

```
          A(IP,J)= DUMY
20        CONTINUE
        DUMY  = B(JP)
        B(JP) = B(IP)
        B(IP) = DUMY
        ENDIF
        RETURN
        END
*--------------------------------------------------
```

เอาท์พุทจากโปรแกรม คือ

```
-----------------------------------------
  COEFFICIENTS OF FITTED POLYNOMIAL ARE:
-----------------------------------------
  A( 0) =     5.0000000000
  A( 1) =     4.0000000000
  A( 2) =    -3.0000000000
-----------------------------------------
```

ซึ่งจะได้ $y = 5 + 4x_1 - 3x_2$ ตรงกับสมการเริ่มต้น

ตัวอย่างที่ 8.9 จง Fit ฟังก์ชัน $f(x, a_0, a_1) = a_0\left(1 - e^{-a_1 x}\right)$ กับข้อมูล

x	0.25	0.75	1.25	1.75	2.25
y	0.28	0.57	0.68	0.74	0.79

ใช้ค่าเริ่มต้น $a_0 = 1.0$ และ $a_1 = 1.0$ สำหรับค่าเริ่มจะมีค่า ผลรวมของกำลังสองของ error = 0.0248

วิธีทำ

$$\frac{\partial f}{\partial a_0} = 1 - e^{-a_1 x} \qquad \text{และ} \qquad \frac{\partial f}{\partial a_1} = a_0 x e^{-a_1 x}$$

อาศัยสมการที่ผ่านมาหาค่าเมทริกซ์

$$[Z_0] = \begin{pmatrix} 0.2212 & 0.1947 \\ 0.5276 & 0.3543 \\ 0.7135 & 0.3581 \\ 0.8262 & 0.3041 \\ 0.8946 & 0.2371 \end{pmatrix}$$

$$[Z_0]^T [Z_0] = \begin{pmatrix} 2.3193 & 0.9489 \\ 0.9489 & 0.4404 \end{pmatrix}$$

$$\left[[Z_0]^T [Z_0]\right]^{-1} = \begin{pmatrix} 3.6397 & -7.8421 \\ -7.8421 & 19.1678 \end{pmatrix}$$

คูณด้วย $[Z_0]^T$ จะได้

$$[Z_0]^T \{D\} = \begin{pmatrix} -0.1533 \\ -0.0365 \end{pmatrix}$$

หาค่า vector $\{\Delta A\}$ โดยอาศัยสมการ (17.35)

$$\Delta A = \begin{Bmatrix} -0.2714 \\ 0.5019 \end{Bmatrix}$$

นำไปรวมกับพารามิเตอร์ซึ่งคาดไว้ จะได้

$$\begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 1.0 \\ 1.0 \end{Bmatrix} + \begin{Bmatrix} -0.2714 \\ 0.5019 \end{Bmatrix} = \begin{Bmatrix} 0.7286 \\ 1.5019 \end{Bmatrix}$$

ซึ่งค่านี้ จะได้ sum ของ error ยกกำลังสอง = 0.0243

ทำเช่นนี้อีก จะได้

$$a_0 = 0.79186 \qquad \text{และ} \qquad a_1 = 1.6751$$

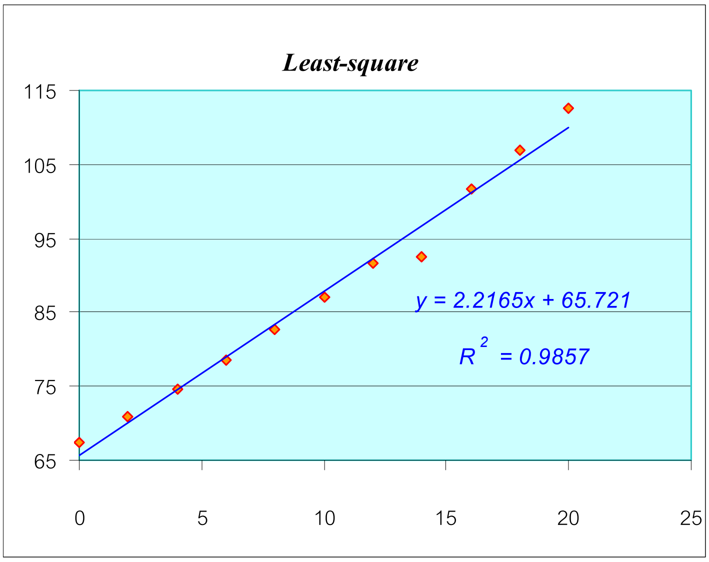
ซึ่งมีค่า sum ของ error ยกกำลังสอง = 0.000662

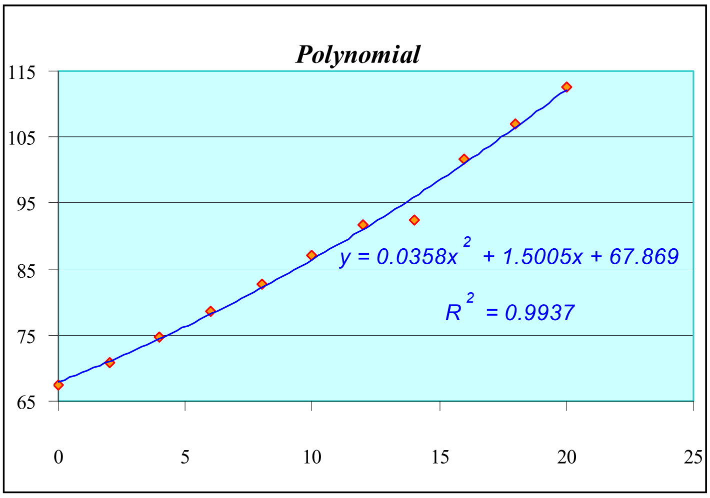
ตัวอย่างที่ 8.10 ข้อมูลต่อไปนี้ได้จากการเก็บตัวอย่างในห้องปฏิบัติการชีววิทยา โดยทำการนับจำนวนแบคทีเรียที่เพาะได้ในแต่ละวัน

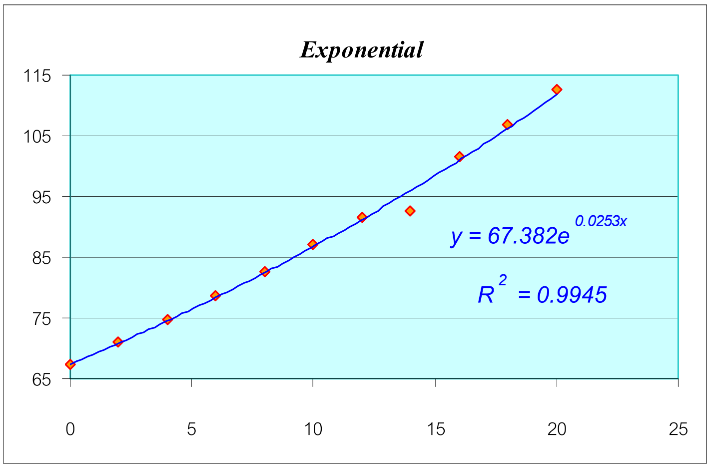
วัน	0	2	4	6	8	10	12	14	16	18	20
จำนวน x 10^6	67.38	70.93	74.67	78.60	82.74	87.10	91.69	92.51	101.60	106.95	112.58

จง fit สมการกับข้อมูลนี้โดยใช้วิธี fit ต่าง ๆ กัน เพื่อหาว่าวิธีใดดีที่สุด

วิธีทำ







จะเห็นได้ว่าการ fit ด้วยกราฟเอ็กโพเน็นเชียลมีค่าสัมประสิทธิ์สหสัมพันธ์มีค่ามากที่สุด จึงเป็นวิธีการที่ดีที่สุดในการ fit ข้อมูลชุดนี้

8.5 การประมาณค่า

เมื่อทำการ fit ข้อมูลแล้ว บางครั้งเราอาจจะต้องการหาค่าที่อยู่ระหว่างชุดข้อมูล เช่น เมื่อทำการทดลอง x อาจจะเท่ากับ 1, 5, 10,… ต้องการทราบค่า เมื่อ x = 3 จะสามารถใช้วิธีการประมาณได้ วิธีการง่ายที่สุดก็คือวิธี Newton’s divided – difference interpolating polynomials โดยที่

$$f_n(x) = f(x_0) + (x - x_0)f[x_1,x_0] + \ldots + (x - x_0)\ldots f[x_n,\ldots, x_0]$$

ตัวอย่างที่ 8.11 จง fit โพลิโนเมียลลำดับสองกับข้อมูล

$x_0 = 1$ $f(x_0) = 0$

$x_1 = 4$ $f(x_1) = 1.386294$

$x_2 = 5$ $f(x_2) = 1.609438$

$x_3 = 6$ $f(x_3) = 1.791759$

อาศัยโพลีโนเมียลประมาณค่า ln(2)

วิธีทำ โปรแกรมที่ใช้ในการคำนวณ คือ

```
*=====================================================
      PROGRAM NT_ITP
      DIMENSION X(10),FX(10,10)
*-----------------------------------------------------
      PARAMETER (N=4)
      DATA (X(I),I=1,N) /1.,4.,5.,6./
      DATA (FX(I,1),I=1,N) /0.,1.386294,1.609438,1.791759/
      XX=2.
*-----------------------------------------------------
      M=N
      DO 10 IC=2,N
        M=M-1
        DO 20 IR=1,M
          FX(IR,IC)=FX(IR+1,IC-1)-FX(IR,IC-1)
          FX(IR,IC)=FX(IR,IC)/(X(IR+IC-1)-X(IR))
20        CONTINUE
10      CONTINUE
      FF=FX(1,1)
      FAC=1.
```

```
      DO 30 I=2,N
        FAC=FAC*(XX-X(I-1))
        FF=FF+FX(1,I)*FAC
30     CONTINUE
*======================================================
*      OUTPUT REPORT
*------------------------------------------------
100    FORMAT(5X,'DATA')
200    FORMAT(5X,
     +  '------------------------------------------')
300    FORMAT(10X,'ith',9X,'Xi',11X,'F(Xi)')
400    FORMAT(9X,I3,1X,2F14.6)
500    FORMAT(5X,'VALUE OF F(X) AT X=',F10.6,
     +        ' IS', F12.8)
      WRITE(*,100)
      WRITE(*,200)
      WRITE(*,300)
      WRITE(*,200)
      DO 40 ITH=1,N
        WRITE(*,400) ITH,X(ITH),FX(ITH,1)
40     CONTINUE
      WRITE(*,200)
      WRITE(*,500) XX,FF
      WRITE(*,200)
      STOP
      END
*------------------------------------------------
```

เอาท์พุทจากโปรแกรม คือ

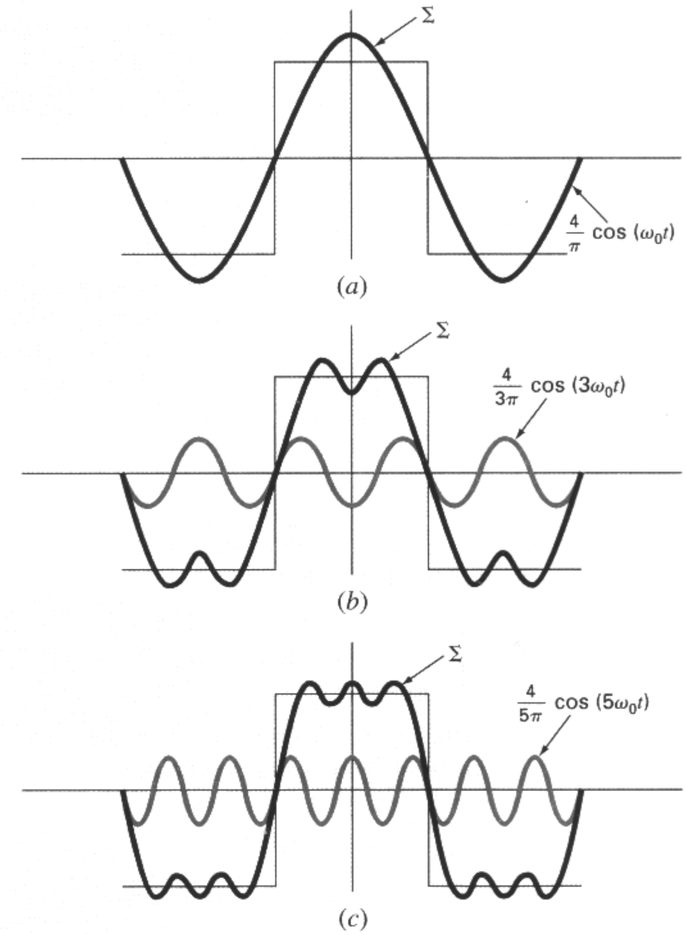
```
DATA
---------------------------------------------
     ith         Xi            F(Xi)
---------------------------------------------
      1       1.000000       0.000000
      2       4.000000       1.386294
      3       5.000000       1.609438
      4       6.000000       1.791759
---------------------------------------------
VALUE OF F(X) AT X=  2.000000 IS  0.62876755
---------------------------------------------
```

จะได้ ln(2) = 0.62876755

8.6 อนุกรมฟูเรียร์

เราได้ศึกษาการปรับเส้นโค้งโดยอาศัยโพลีโนเมียลในหัวข้อที่ผ่านมา ปัญหาทางด้านฟิสิกส์บางครั้งเกี่ยวข้องกับการออสซิลเลต ซึ่งอาจจะต้องใช้ฟังก์ชันเรขาคณิต เช่น cos x, cos 2x, …, sin x, sin 2x, … มาใช้แทนการประมาณอนุกรมฟูเรียร์จะอาศัยฟังก์ชันเหล่านี้มาใช้แทนโมเดล สำหรับฟังก์ชันที่มีคาบ T อนุกรมฟูเรียร์สามารถเขียนได้เป็น

$$f(x) = a_0 + a_1 \cos(\omega_0 t) + b_1 \sin(\omega_0 t) + a_2 \cos(2\omega_0 t) + b_2 \sin(2\omega_0 t) + \ldots$$



รูปที่ 8.7 การประมาณอนุกรมฟูเรียร์กับคลื่นรูปสี่เหลี่ยม

แบบฝึกหัดบทที่ 8

1. จง fit ข้อมูลชุดนี้ด้วยกราฟเส้นตรง

x	5	6	10	14	16	20	22	28	28	36	38
y	30	22	28	14	22	16	8	8	14	0	4

และจงประมาณค่าที่ x = 12, 24 และ 31

2. จง fit ข้อมูลชุดนี้ด้วยโมเดล saturation – growth – rate

x	0.75	2	2.5	4	6	8	8.5
y	0.8	1.3	1.2	1.6	1.7	1.8	1.7

และจงเขียนกราฟข้อมูลและสมการบนกราฟเดียวกัน

3. จงใช้โมเดลเอ็กโพเนนเชียล fit ข้อมูลนี้

x	0.4	0.8	1.2	1.6	2.0	2.3
y	750	1000	1400	2000	2700	3750

4. จงใช้ multiple linear regression เพื่อ fit ข้อมูล

x_1	0	1	1	2	2	3	3	4	4
x_2	0	1	2	1	2	1	2	1	2
y	1.5	1.8	12.8	25.7	20.6	35.0	29.8	45.5	40.3

5. จงอาศัยการถดถอยไม่เป็นเชิงเส้นเพื่อ fit กราฟพาราโบลากับข้อมูลนี้

x	0.075	0.5	1	1.2	1.7	2.0	2.3
y	600	800	1200	1400	2050	2650	3750