Metode Numerik
EE221

Bab 6. Pencocokan kurva (Curve Fitting) dengan Regresi

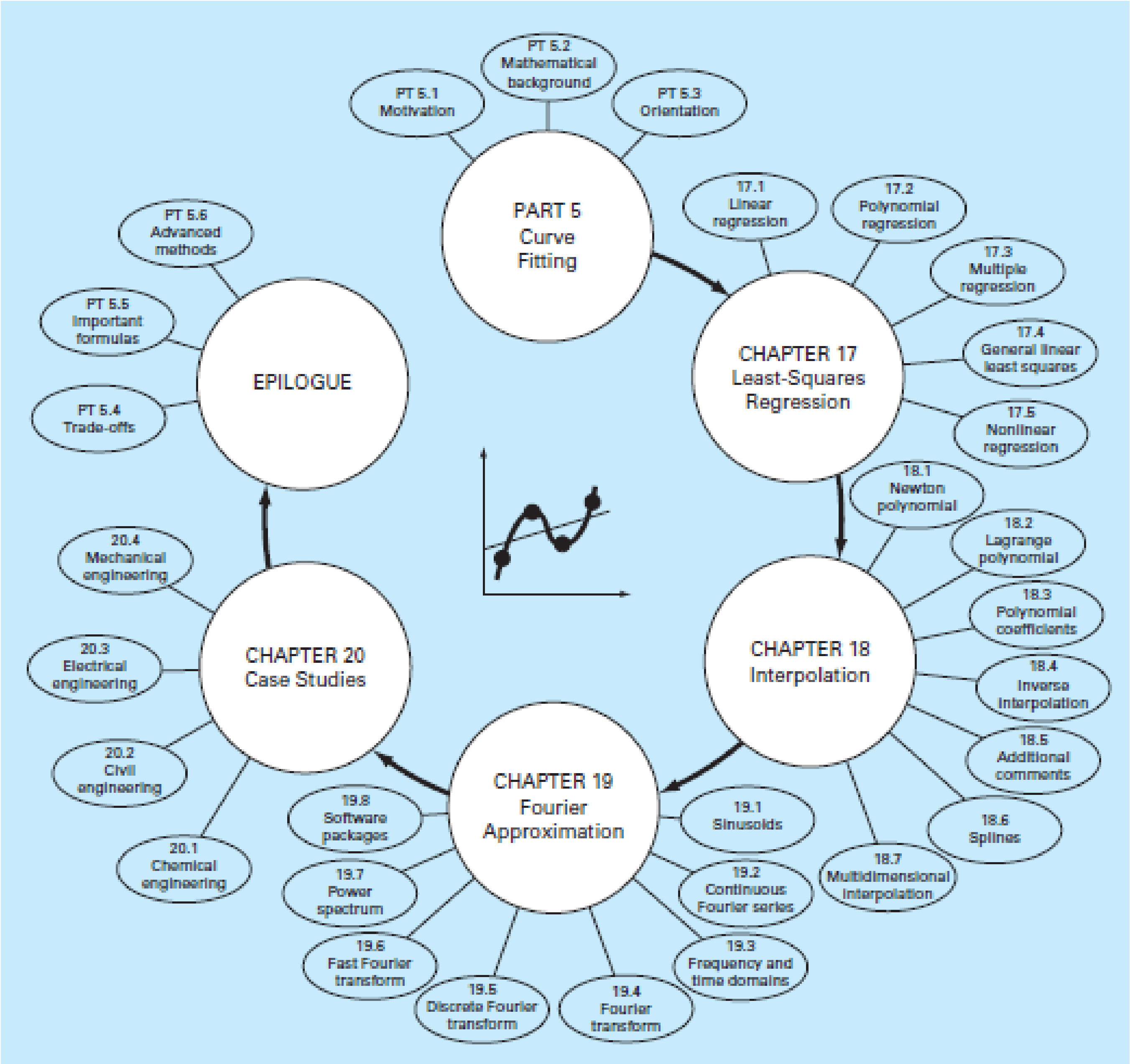
Dirangkum dan diterjemahkan dari : Thomson Brooks Chapra, Steven and Raymond Canale. 2009. Numerical Methods for Engineers 6^{th} Edition, **Chapter 17**

Nabila Husna Shabrina
Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara

nabila.husna@umn.ac.id

Sub Bahasan :

- Regresi Linier
- Regresi Polinomial
- Regresi Linier Multiple


PT 5.1
Motivation
PT 5.2
Mathematical
background
PT 5.3
Orientation
PART 5
Curve
Fitting
17.1
Linear
regression
17.2
Polynomial
regression
17.3
Multiple
regression
17.4
General linear
least squares
17.5
Nonlinear
regression
CHAPTER 17
Least-Squares
Regression
18.1
Newton
polynomial
18.2
Lagrange
polynomial
18.3
Polynomial
coefficients
18.4
Inverse
interpolation
18.5
Additional
comments
18.6
Splines
18.7
Multidimensional
interpolation
CHAPTER 18
Interpolation
19.1
Sinusoids
19.2
Continuous
Fourier series
19.3
Frequency and
time domains
19.4
Fourier
transform
19.5
Discrete Fourier
transform
19.6
Fast Fourier
transform
19.7
Power
spectrum
19.8
Software
packages
CHAPTER 19
Fourier
Approximation
20.1
Chemical
engineering
20.2
Civil
engineering
20.3
Electrical
engineering
20.4
Mechanical
engineering
CHAPTER 20
Case Studies
PT 5.4
Trade-offs
PT 5.5
Important
formulas
PT 5.6
Advanced
methods
EPILOGUE

Regresi Linier

Regresi adalah metode analisis statistik yang digunakan untuk melihat hubungan antara 2 variabel yang berberda

Regresi Linier

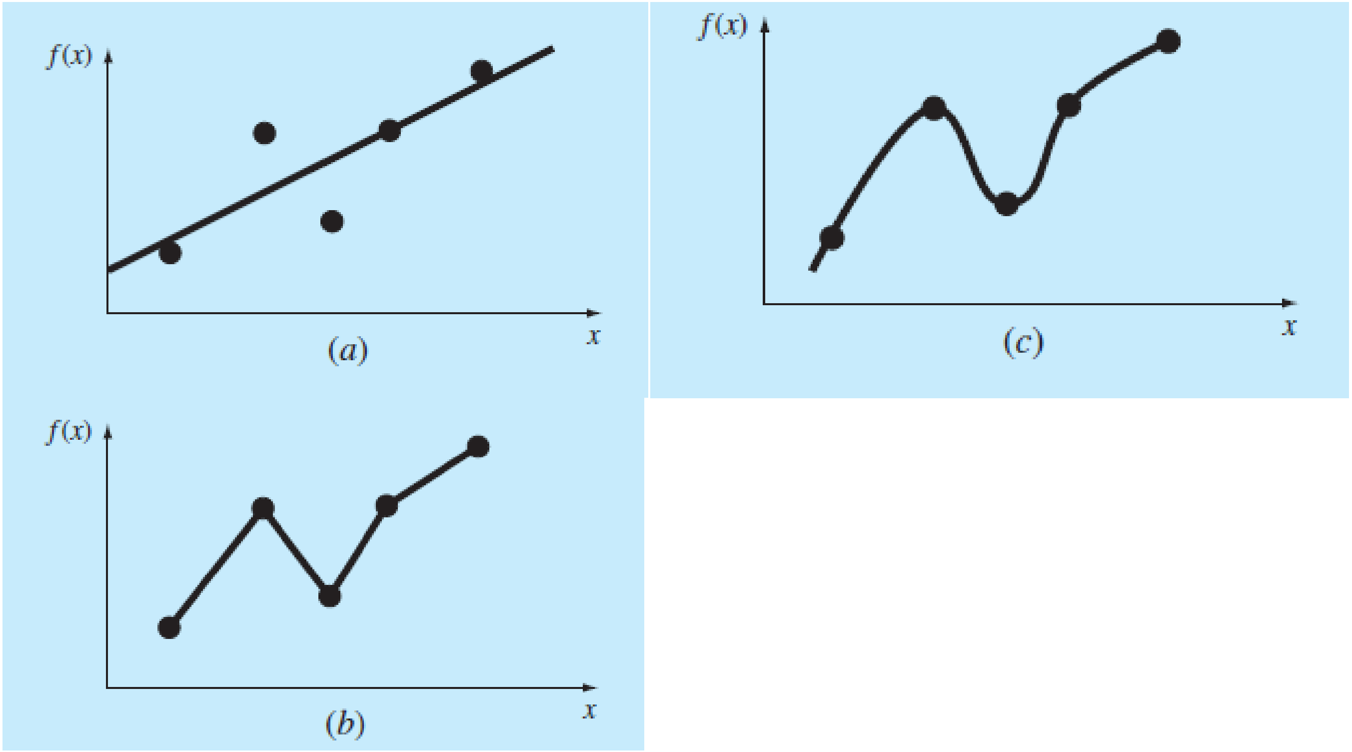


FIGURE PT5.1

Three attempts to fit a "best" curve through five data points. (*a*) Least-squares regression, (*b*) linear interpolation, and (*c*) curvilinear interpolation.

Regresi Linier

Regresi Linier

- Regresi linier dapat dilakukan dengan memasangkan dua variable yang berbeda (x_1, y_1) , (x_2, y_2) , …. , (x_n, y_n) dan menyatakannya dalam persamaan linier

$$y = a_0 + a_1 x + e$$

kemiringan

error

Titik potong

Regresi Linier

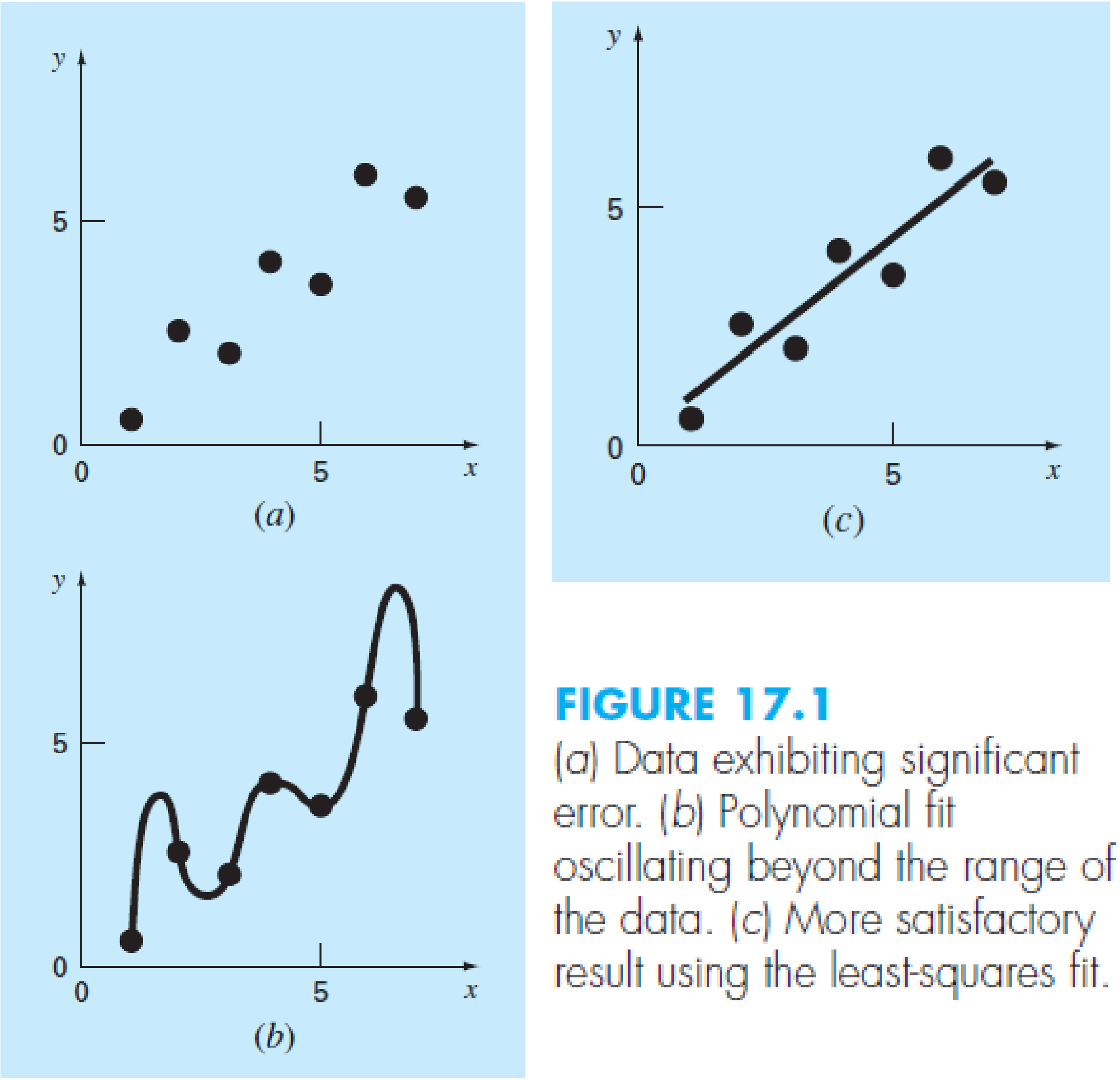


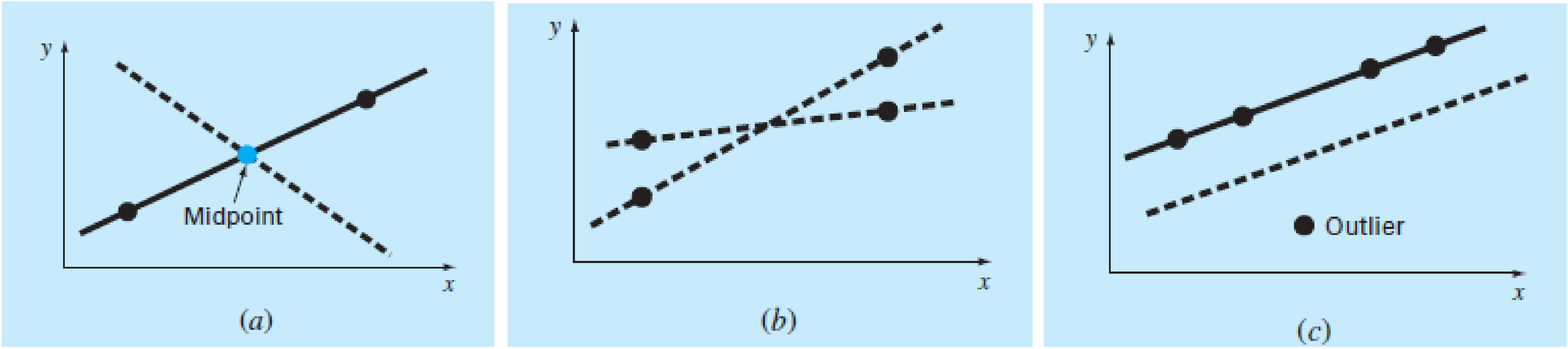
FIGURE 17.1
(*a*) Data exhibiting significant error. (*b*) Polynomial fit oscillating beyond the range of the data. (*c*) More satisfactory result using the least-squares fit.

Regresi Linier

Kriteria 'best fit'

- 'best fit' → perbedaan antara nilai y sebenarnya dan y prediksi bernilai minimum
- Dinyatakan dengan persamaan $$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

FIGURE 17.2
Examples of some criteria for "best fit" that are inadequate for regression: (*a*) minimizes the sum of the residuals, (*b*) minimizes the sum of the absolute values of the residuals, and (*c*) minimizes the maximum error of any individual point.



Regresi Linier

Least-Squares Fit of a Straight Line

$$y = a_0 + a_1 x + e$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

$$a_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

Regresi Linier

Contoh 1.

Lakukan pencocokkan untuk nilai x dan y pada kolom pertama dan kedua dari tabel di bawah.

TABLE 17.1 Computations for an error analysis of the linear fit.

x_i	y_i	$(y_i - \bar{y})$	$(y_i - a_0 - a_1 x_i)^2$
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
Σ	24.0	22.7143	2.9911

$$n = 7 \qquad \sum x_i y_i = 119.5 \qquad \sum x_i^2 = 140$$

$$\sum x_i = 28 \qquad \bar{x} = \frac{28}{7} = 4$$

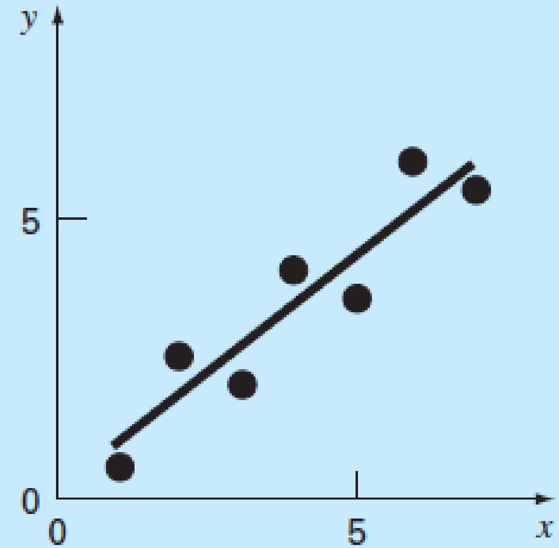
$$\sum y_i = 24 \qquad \bar{y} = \frac{24}{7} = 3.428571$$

$$a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$

$$a_0 = 3.428571 - 0.8392857(4) = 0.07142857$$

Regresi Linier

$$y = 0.07142857 + 0.8392857x$$



(c)

Regresi Linier

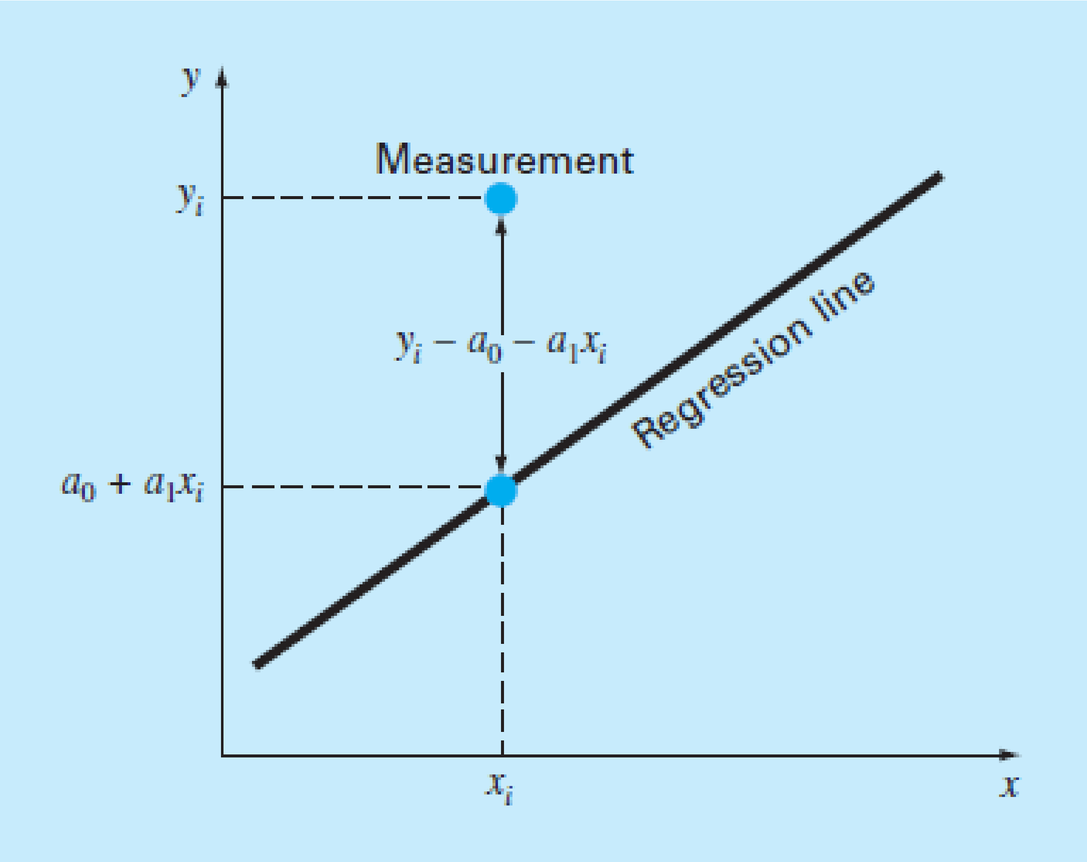


FIGURE 17.3
The residual in linear regression represents the vertical distance between a data point and the straight line.

Regresi Linier

Error pada regresi linier

Standar error untuk nilai estimasi

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

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

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

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

Koefisien determinan

Koefisien korelasi

Regresi Linier

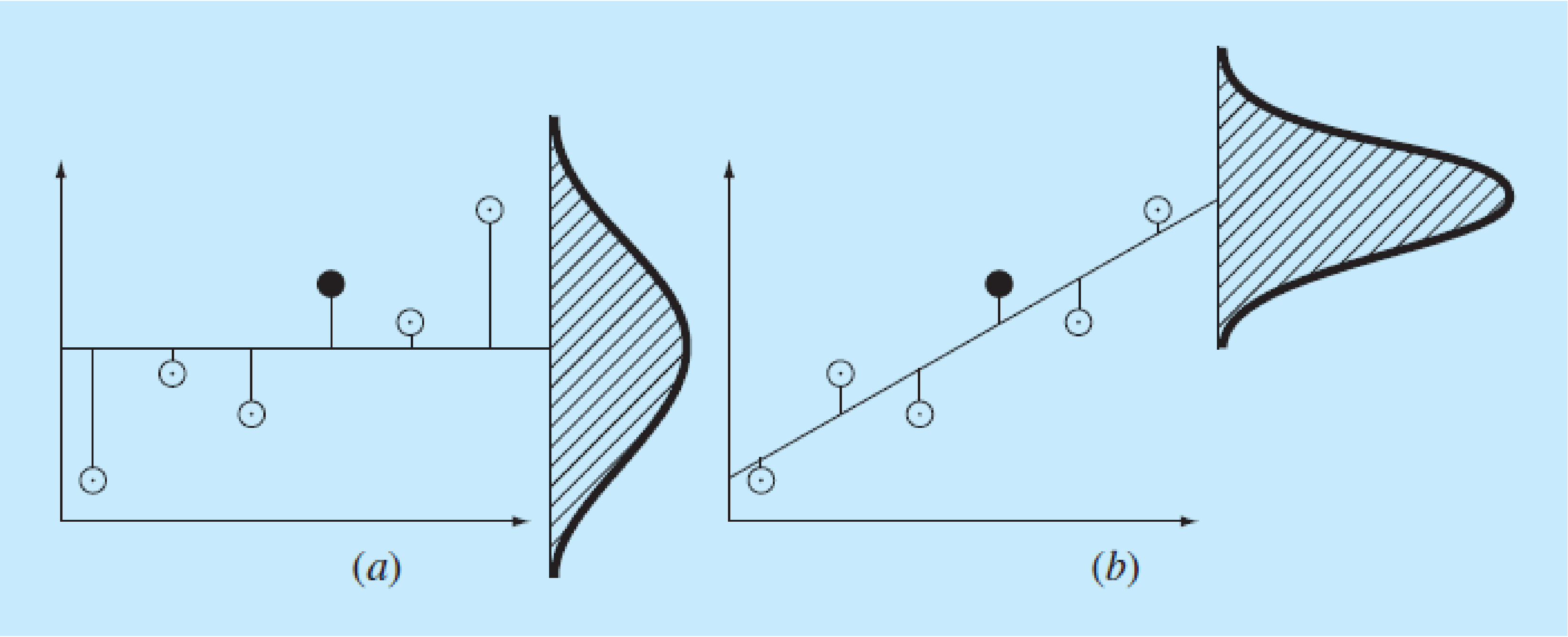


FIGURE 17.4
Regression data showing (a) the spread of the data around the mean of the dependent variable and (b) the spread of the data around the best-fit line. The reduction in the spread in going from (a) to (b), as indicated by the bell-shaped curves at the right, represents the improvement due to linear regression.

Regresi Linier

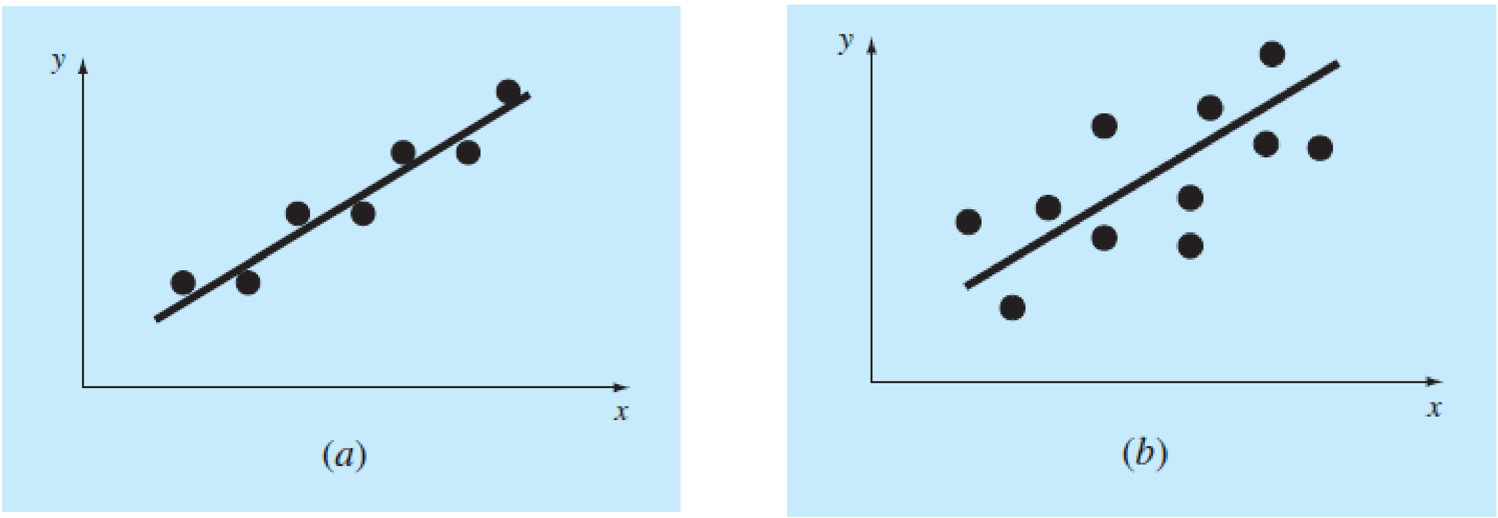


FIGURE 17.5
Examples of linear regression with (*a*) small and (*b*) large residual errors.

Regresi Linier

Contoh 2.

Tentukan standar deviasi, estimasi error dan koefisien korelasi dari soal contoh 1.

$$s_y = \sqrt{\frac{22.7143}{7-1}} = 1.9457$$

$$s_{y/x} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

$$r^2 = \frac{22.7143 - 2.9911}{22.7143} = 0.868$$ atau $$r = \sqrt{0.868} = 0.932$$

Regresi Linier

Pseudocode

```
SUB Regress(x, y, n, a1, a0, syx, r2)

  sumx = 0: sumxy = 0: st = 0
  sumy = 0: sumx2 = 0: sr = 0
  DOFOR i = 1, n
    sumx = sumx + x_i
    sumy = sumy + y_i
    sumxy = sumxy + x_i*y_i
    sumx2 = sumx2 + x_i*x_i
  END DO
  xm = sumx/n
  ym = sumy/n
  a1 = (n*sumxy − sumx*sumy)/(n*sumx2 − sumx*sumx)
  a0 = ym − a1*xm
  DOFOR i = 1, n
    st = st + (y_i − ym)^2
    sr = sr + (y_i − a1*x_i − a0)^2
  END DO
  syx = (sr/(n − 2))^0.5
  r2 = (st − sr)/st

END Regress
```

Regresi Polinomial

Regresi Polinomial

- Regresi polinomial dinyatakan dalam $y = a_0 + a_1 x + a_2 x^2 + e$
- Persamaan regresi polinomial untuk derajat m

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

$$\begin{aligned}
(n)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 &= \sum y_i \\
\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 &= \sum x_i y_i \\
\left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 &= \sum x_i^2 y_i
\end{aligned}$$

- Persamaan standar error

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

dengan

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

Regresi Polinomial

Contoh 3.

Tentukan regresi polinomial untuk nilai x dan y pada tabel di bawah.

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i - a_2x_i^2)^2$
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
Σ	152.6	2513.39	3.74657

$m = 2 \qquad \sum x_i = 15 \qquad \sum x_i^4 = 979$

$n = 6 \qquad \sum y_i = 152.6 \qquad \sum x_i y_i = 585.6$

$\bar{x} = 2.5 \qquad \sum x_i^2 = 55 \qquad \sum x_i^2 y_i = 2488.8$

$\bar{y} = 25.433 \qquad \sum x_i^3 = 225$

Regresi Polinomial

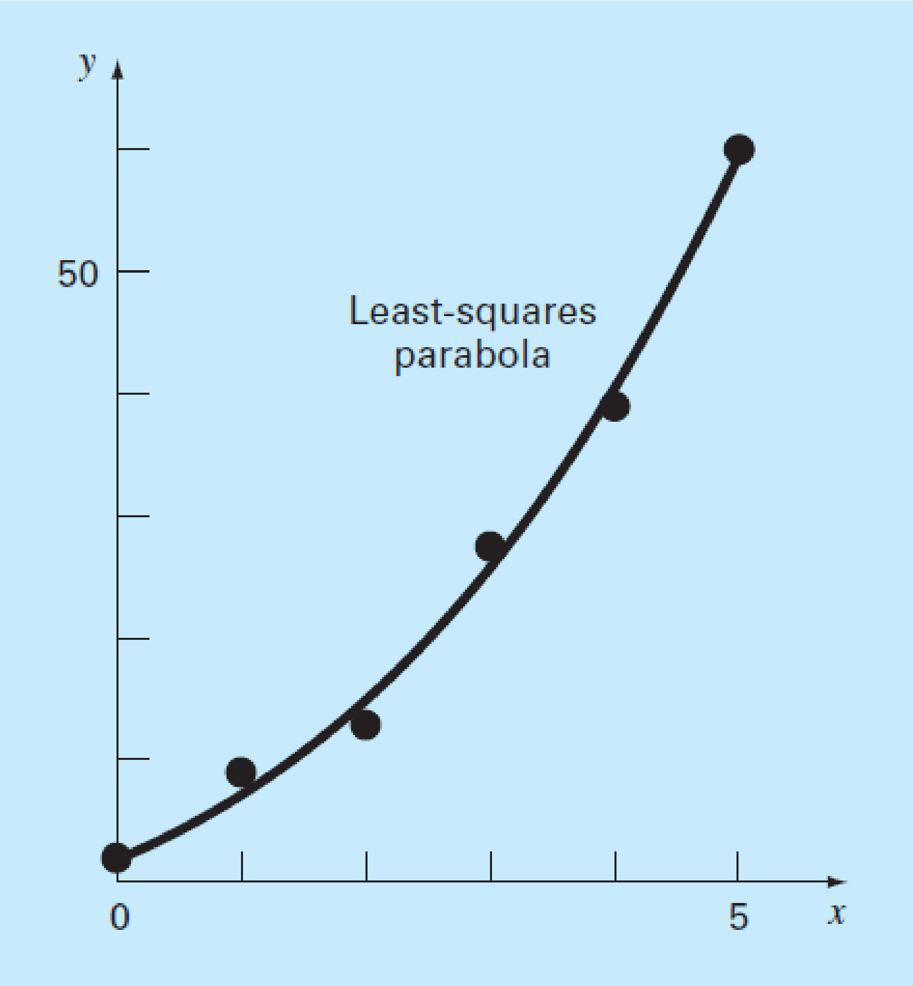


FIGURE 17.11
Fit of a second-order polynomial.

Regresi Polinomial

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

$$y = 2.47857 + 2.35929x + 1.86071x^2$$

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

$$r^2 = \frac{2513.39 - 3.74657}{2513.39} = 0.99851 \qquad r = 0.99925.$$

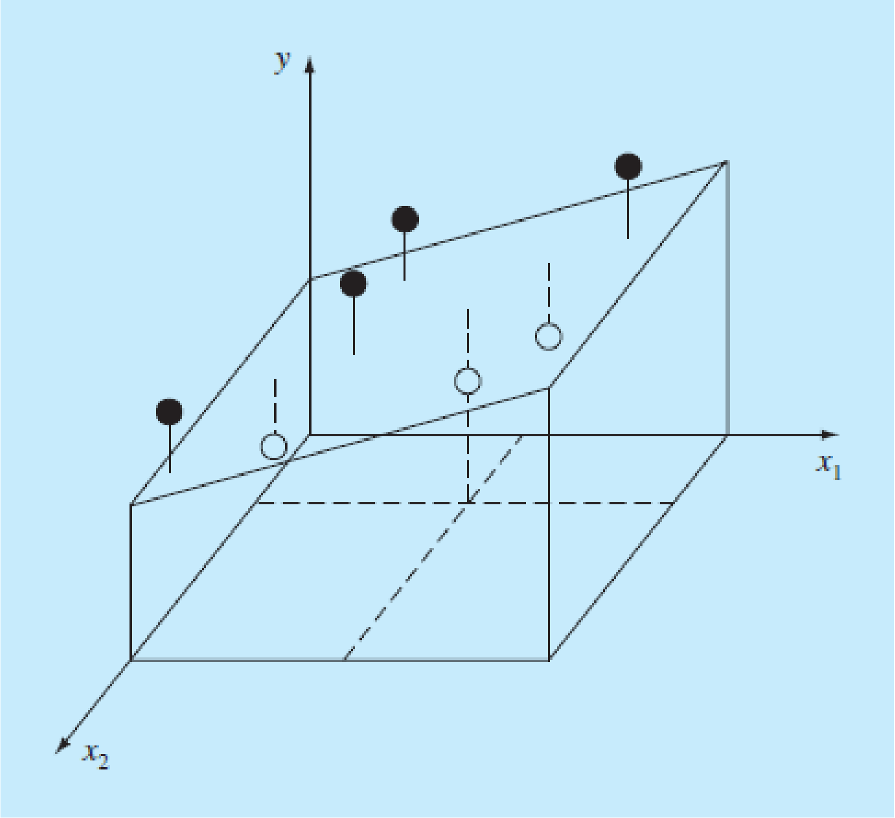
Regresi Multiple Linier

Regresi Multiple Linier

Bentuk umum

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

FIGURE 17.14
Graphical depiction of multiple linear regression where y is a linear function of x_1 and x_2.



Regresi Multiple Linier

- Regresi multiple linier memiliki bentuk umum

$$y = a_0 + a_1x_1 + a_2x_2 + e$$

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_mx_m + e$$

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

- Persamaan standar error

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

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1x_{1i} - a_2x_{2i})^2$$

Regresi Multiple Linier

Contoh 4.

Data pada table di bawah diperoleh dari persamaan $y = 5 + 4x_1 - 3x_2$:

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

Gunakan regresi multiple linier untuk mencocokkan data berikut.

Regresi Multiple Linier

TABLE 17.5 Computations required to develop the normal equations for Example 17.6.

	y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
	5	0	0	0	0	0	0	0
	10	2	1	4	1	2	20	10
	9	2.5	2	6.25	4	5	22.5	18
	0	1	3	1	9	3	0	0
	3	4	6	16	36	24	12	18
	27	7	2	49	4	14	189	54
Σ	54	16.5	14	76.25	54	48	243.5	100

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

$$a_0 = 5 \quad a_1 = 4 \quad a_2 = -3$$

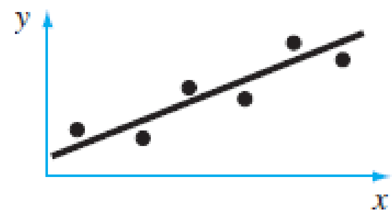
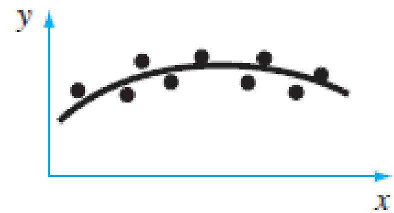
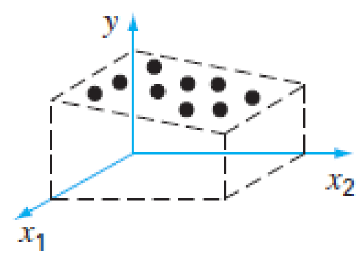
Regresi Multiple Linier

Pseudocode

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    sum = 0
    DOFOR ℓ = 1, n
      sum = sum + x_{i-1,ℓ} · x_{j-1,ℓ}
    END DO
    a_{i,j} = sum
    a_{j,i} = sum
  END DO
  sum = 0
  DOFOR ℓ = 1, n
    sum = sum + y_ℓ · x_{i-1,ℓ}
  END DO
  a_{i,order+2} = sum
END DO
```

Rangkuman

TABLE PT5.5 Summary of important information presented in Part Five.

Method	Formulation	Graphical Interpretation	Errors
Linear regression	$y = a_0 + a_1 x$ where $a_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$ $a_0 = \bar{y} - a_1 \bar{x}$		$s_{y/x} = \sqrt{\dfrac{S_r}{n-2}}$ $r^2 = \dfrac{S_t - S_r}{S_t}$
Polynomial regression	$y = a_0 + a_1 x + \cdots + a_m x^m$ (Evaluation of a's equivalent to solution of $m + 1$ linear algebraic equations)		$s_{y/x} = \sqrt{\dfrac{S_r}{n-(m+1)}}$ $r^2 = \dfrac{S_t - S_r}{S_t}$
Multiple linear regression	$y = a_0 + a_1 x_1 + \cdots + a_m x_m$ (Evaluation of a's equivalent to solution of $m + 1$ linear algebraic equations)		$s_{y/x} = \sqrt{\dfrac{S_r}{n-(m+1)}}$ $r^2 = \dfrac{S_t - S_r}{S_t}$

Latihan Soal

Latihan

Problem Chapter 17

Nomor 17.4 , 17.5 , 17.6