REGRESI LINEAR

Rita Prasetyowati
Fisika FMIPA UNY
2012

PENCOCOKAN DATA TERHADAP FUNGSI GARIS LURUS DENGAN LEAST SQUARE FITTING (*Linear Regresi*)

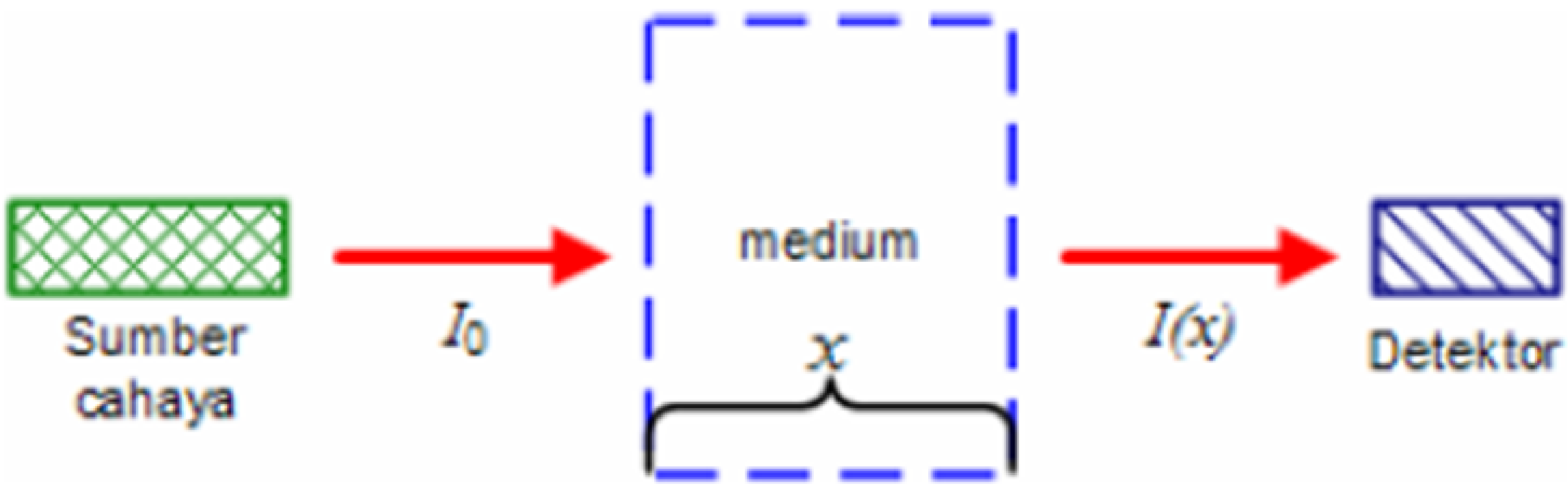
Fungsi matematik ➔ persamaan garis lurus $y = a + bx$.

Persamaan garis lurus :

- sangat sederhana sehingga mudah ditafsirkan secara fisis.
- mudah digunakan sebagai alat dalam analisa hasil eksperimen untuk menentukan nilai tetapan atau variabel fisika yang menjadi tujuan eksperimen

Metode pencocokan data menggunakan *Least Square Fitting* diterapkan dalam persamaan garis lurus ➔ akan dapat ditetapkan nilai-nilai parameter persamaan garis lurus yaitu a dan b tanpa perlu menggunakan grafik.

Contoh sistem fisis yang secara eksperimen dapat diamati dengan persamaan garis lurus : eksperimen untuk menentukan tetapan serapan suatu bahan (misal secara optis).



Persamaan serapan

$$I(x) = I_0 e^{-\mu x}$$

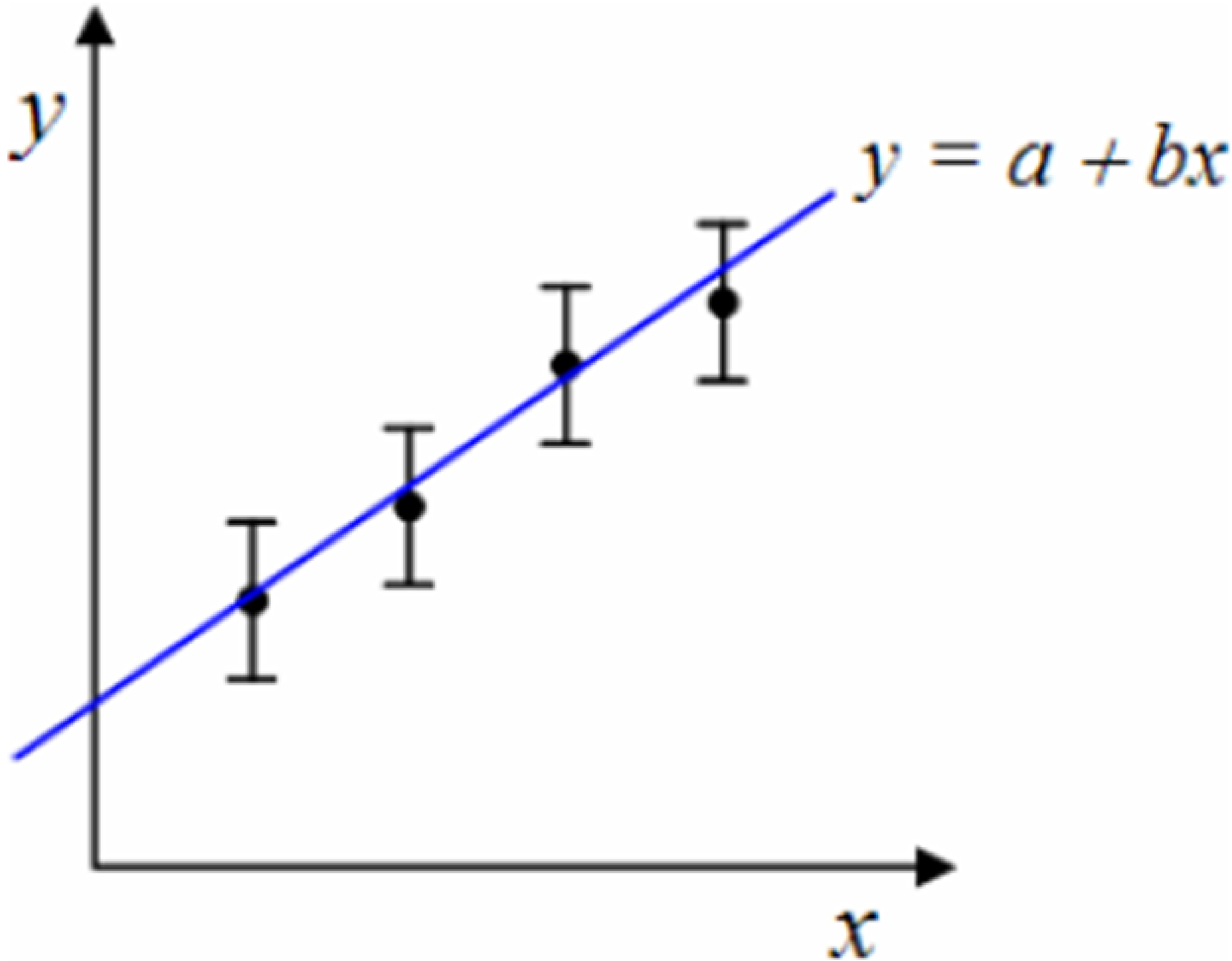
Persamaan tsb harus diubah menjadi bentuk persamaan linear :

$$\ln \frac{I(x)}{I_0} = -\mu x$$

$$\ln I(x) = \ln I_0 - \mu x$$

bentuk terakhir ini sudah sebanding dengan persamaan garis lurus $y = a + bx$, dimana x = variable tak gayut (bebas) dan y = variable gayut (tak bebas). Sehingga dengan memvariasi tebal medium akan dapat diperoleh tetapan μ dari besar gradien garis lurus b.

Suatu persamaan garis lurus berupa $y = a + bx$, a dan b diatur agar P_{Total} = maksimum.



Untuk melakukan regresi atau pencocokan data terhadap fungsi garis lurus maka persamaan garis lurus dimasukkan kedalam fungsi *y(x)* dan harus memenuhi persamaan *least square fitting* :

$$\chi^2 = \sum \left(\frac{y_i - y(x_i)}{S_i} \right)^2 \rightarrow \text{minimum,}$$

dimana y_i , x_i adalah titik data dan $y(x_i)$ adalah fungsi yang dicari. Sehingga :

$$\chi^2 = \sum \left(\frac{y_i - (a + bx_i)}{Si} \right)^2 \rightarrow \text{min}$$

Penyelesaian persamaan di atas dapat dibedakan untuk dua kasus yaitu:

a. Jika $S_1 = S_2 = S_i = S$ (Ketakpastiannya sama), maka penyelesaiannya :

$$\frac{\partial \chi^2}{\partial a} = 0 \quad \text{dan} \quad \frac{\partial \chi^2}{\partial b} = 0, \text{ dimana } \chi^2 = \frac{1}{S^2}\sum_{i=1}^{N}\left(y_i - (a + bx_i)\right)^2$$

Maka diperoleh:

$$\Delta = N\sum x_i^2 - \left(\sum x_i\right)^2$$

$$a = \frac{1}{\Delta}\left(\sum x_1^2 \sum y_i - \sum x_i \sum x_i y_i\right)$$

$$b = \frac{1}{\Delta}\left(N\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)\right)$$

Bagaimana dengan ketidakpastian masing-masing?

$$S_a^{\,2} = \frac{S^2 \sum x_i^{\,2}}{\Delta} \qquad Sb^2 = \frac{NS^2}{\Delta}$$

Nilai S pada rumus Sa dan Sb pada dasarnya adalah sama dengan Sy. Jika nilai Sy tidak diketahui tetapi dapat diperkirakan bahwa besarnya adalah sama (umumnya bersumber dari alat ukur yang sama dan skala yang digunakan juga sama), maka nilai S dapat ditentukan :

$$S^2 = \frac{1}{N-2} \sum (y_i - a - bx_i)^2$$

b. Jika **S1 ≠ S2 ≠ ....≠** Si (ralat tidak sama), maka dengan penurunan yang sama diperoleh :

$$\chi^2 = \sum \frac{y_i - (a + bx_i)^2}{S_i^{\ 2}} \rightarrow \text{minimum}$$

$$\frac{\partial x^2}{\partial a} = 0 \rightarrow a = \frac{1}{\Delta}\left(\sum \frac{x_i^2}{S_i^2} \sum \frac{y_i^2}{S_{i2}} - \sum \frac{x_i}{S_i^2} \sum \frac{x_i y_i}{S_{i2}}\right) \text{ dan } \mathrm{Sa} = \left(\frac{1}{\Delta} \sum \frac{x_i^2}{S_i^2}\right)^2$$

$$\frac{\partial x^2}{\partial a} = 0 \rightarrow b = \frac{1}{\Delta}\left(\sum \frac{1}{S_i^2} \sum \frac{x_i y_i}{S_{i2}} - \sum \frac{x_i}{S_i^2} \sum \frac{y_i^{\ 2}}{S_{i2}}\right) \text{ dan } \mathrm{Sb} = \left(\frac{1}{\Delta} \sum \frac{1}{S_i^2}\right)^2$$

$$\text{dengan } \Delta = \sum \frac{1}{S_i^2} \sum \frac{x_i^{\ 2}}{S_i^2} - \left(\sum \frac{x_i}{S_i^2}\right)^2$$

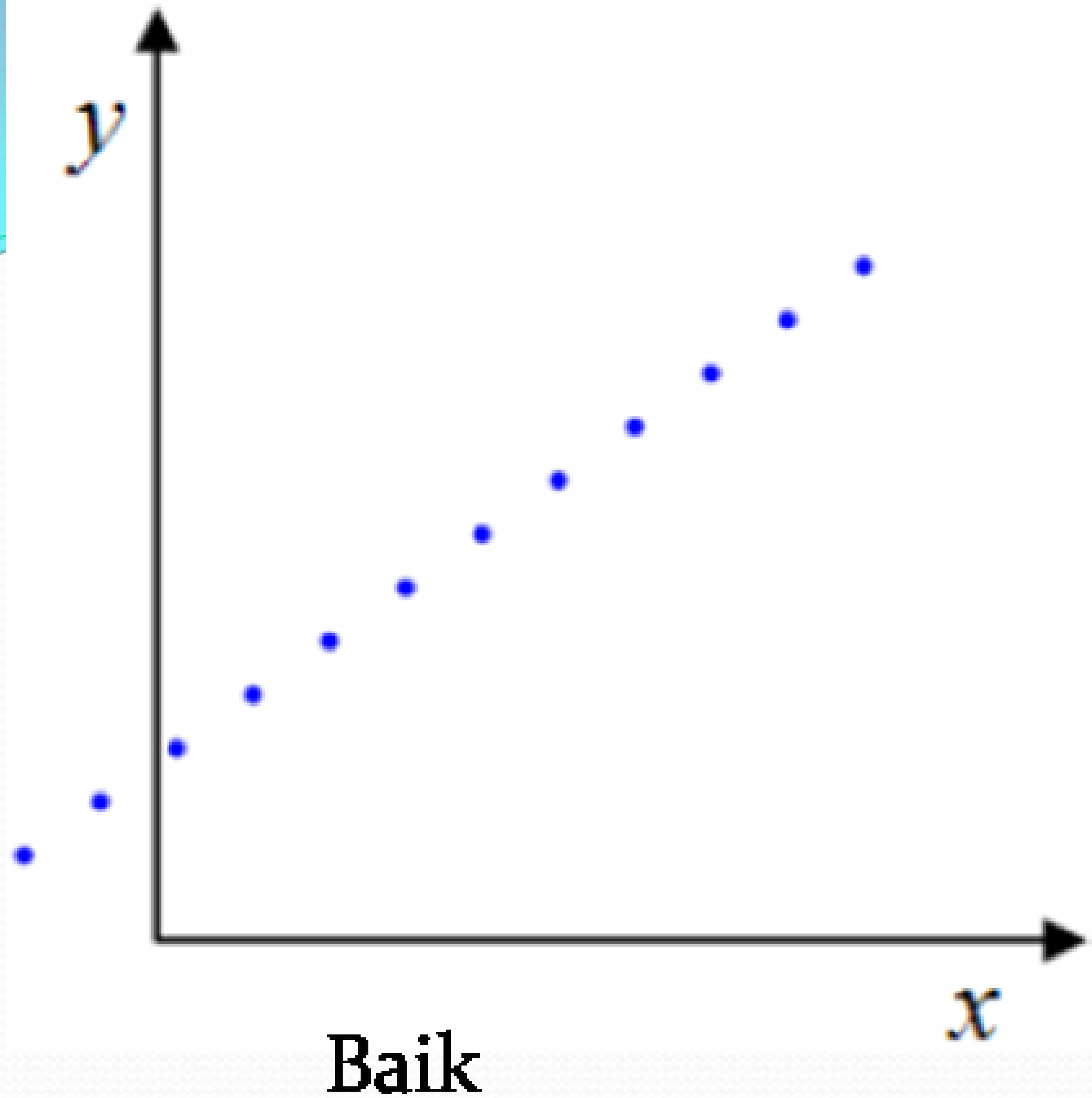
KOEFISIEN KORELASI LINEARITAS

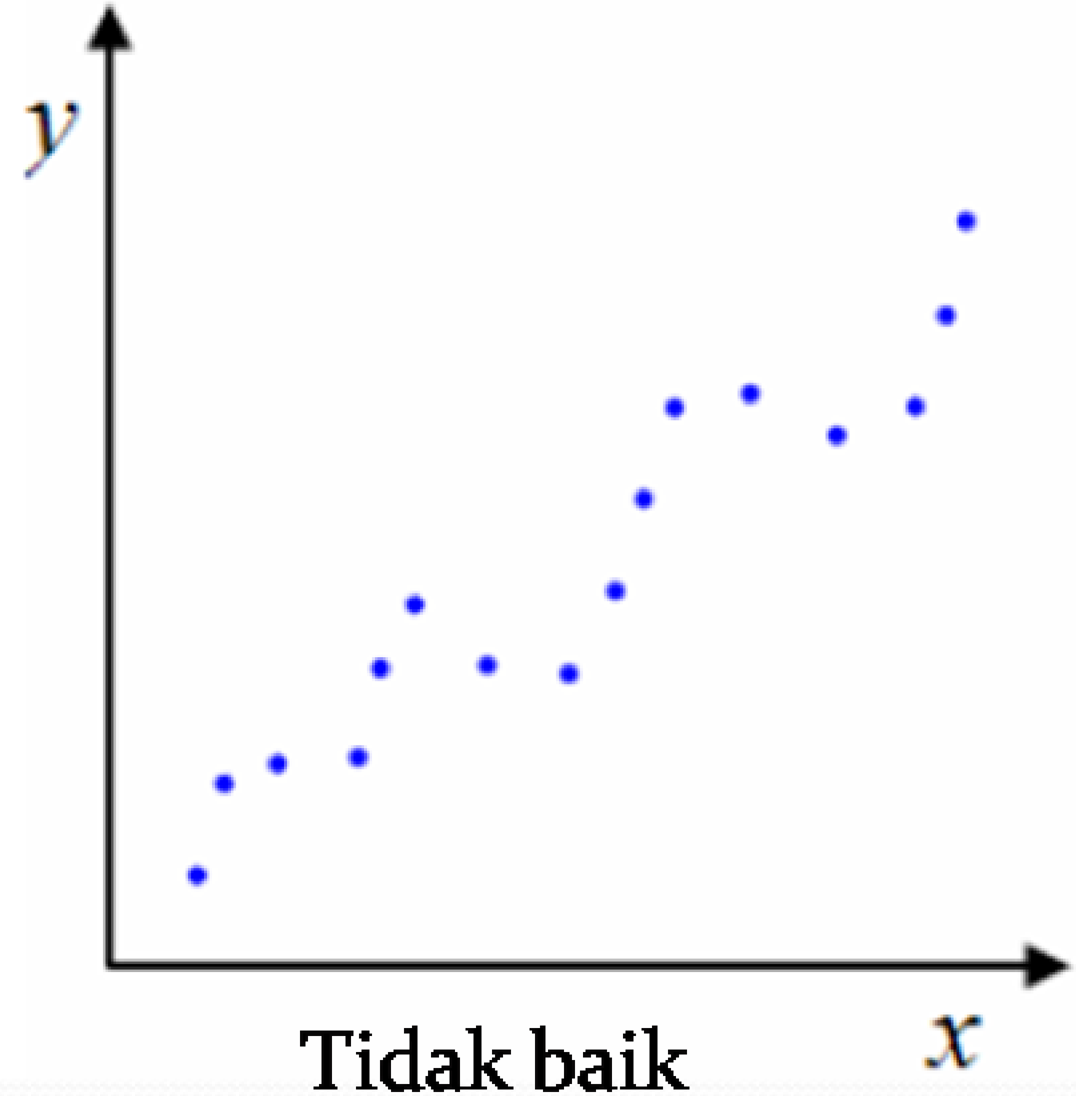
- Parameter persamaan garis lurus $a \pm Sa$ dan $b \pm Sb$ dapat ditentukan dengan metode *least square fitting* tanpa melihat grafik.
- kita tidak dapat memutuskan apakah

  data yang diperoleh cukup “baik” atau tidak
- ”baik” maksudnya bahwa data yang diperoleh jika diplot dalam bentuk grafik akan memperlihatkan bentuk garis lurus

PENCOCOKAN DATA DENGAN METODE KUADRAT TERKECIL (*Least Square Fitting*)

- Salah satu jenis eksperimen yang umum dan menarik adalah yang berkaitan dengan **pengukuran beberapa hasil ukur dari dua variabel fisis yang berbeda**.
- Data yang diperoleh digunakan untuk menyelidiki hubungan (persamaan) matematis dari dua variabel tersebut.





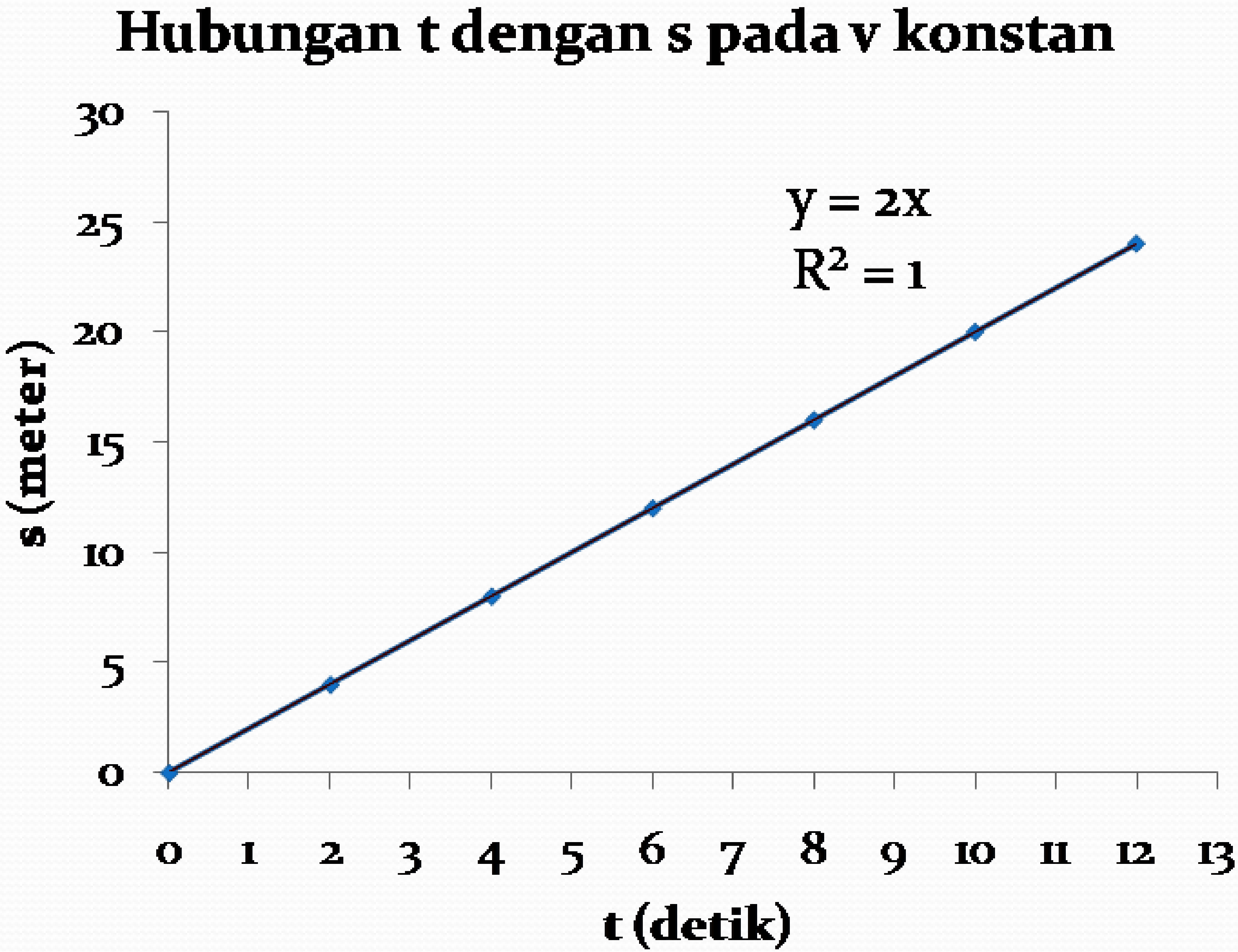
Data yang baik memiliki arti bahwa variabel *x* dan variabel y menunjukkan hubungan persamaan garis lurus, sedangkan data yang tidak baik dapat berarti variabel *x* dan variabel y tidak menunjukkan hubungan garis lurus. "Hubungan" ini didefinisikan sebagai **korelasi linearitas**

Korelasi linearitas dari data yang diperoleh adalah baik atau tidak, dapat diketahui dengan menghitung nilai koefisien korelasi linearis *r* menggunakan Rumus :

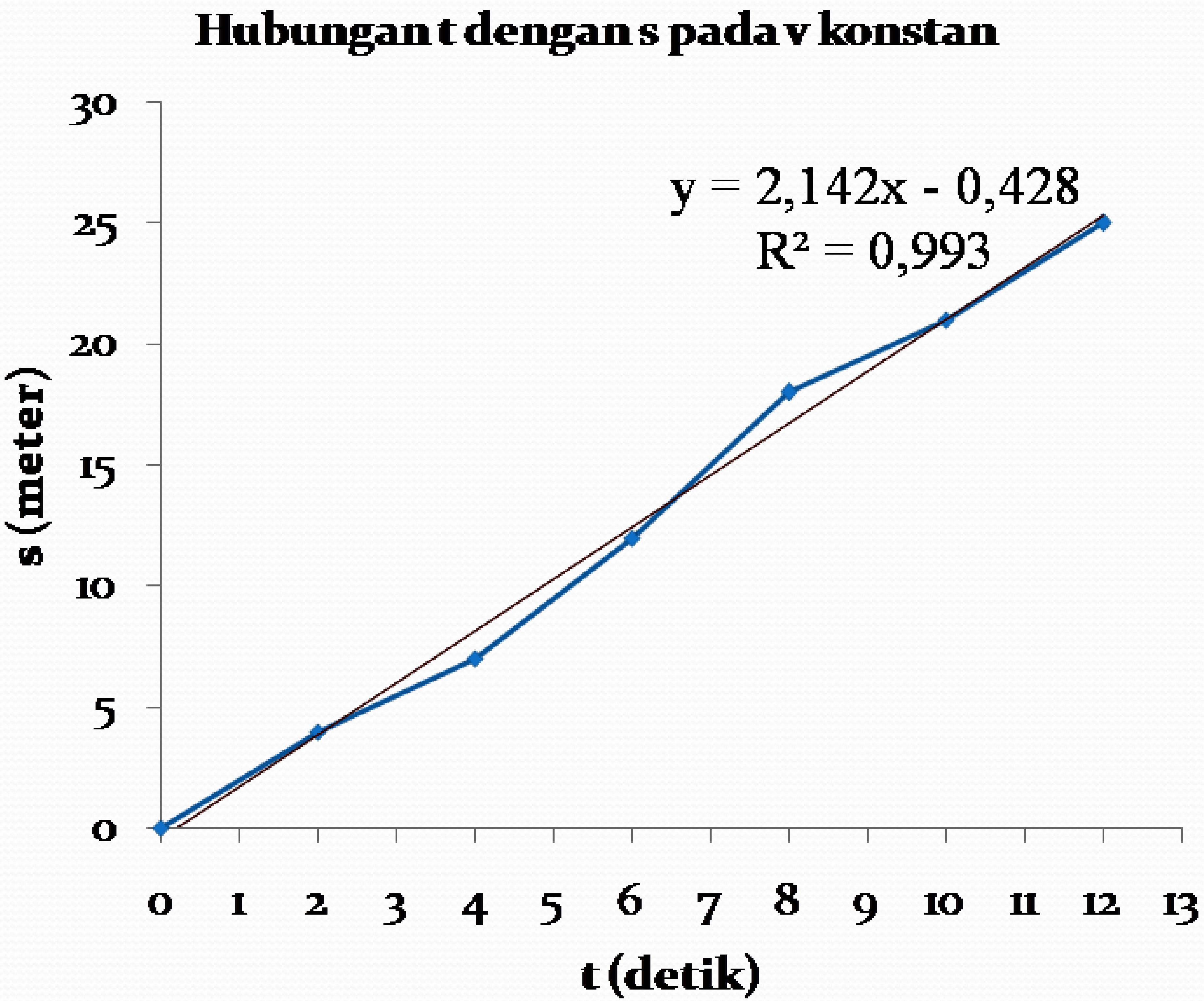
$$r = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\left[N\sum x_i^2 - \left(\sum x_i\right)^2\right]^{1/2}\left[N\sum y_i^2 - \left(\sum y_i\right)^2\right]^{1/2}}$$

Jika nilai *r* semakin mendekati nilai *0* ➔ variabel *x* dan variabel *y* semakin **tak terkolerasi linear.**
Jika nilai *r* semakin mendekati ± 1 ➔ variabel *x* dan variabel *y* semakin **terkolerasi secara linear.**

t (detik)	S (meter)
0	0
2	4
4	8
6	12
8	16
10	20
12	24

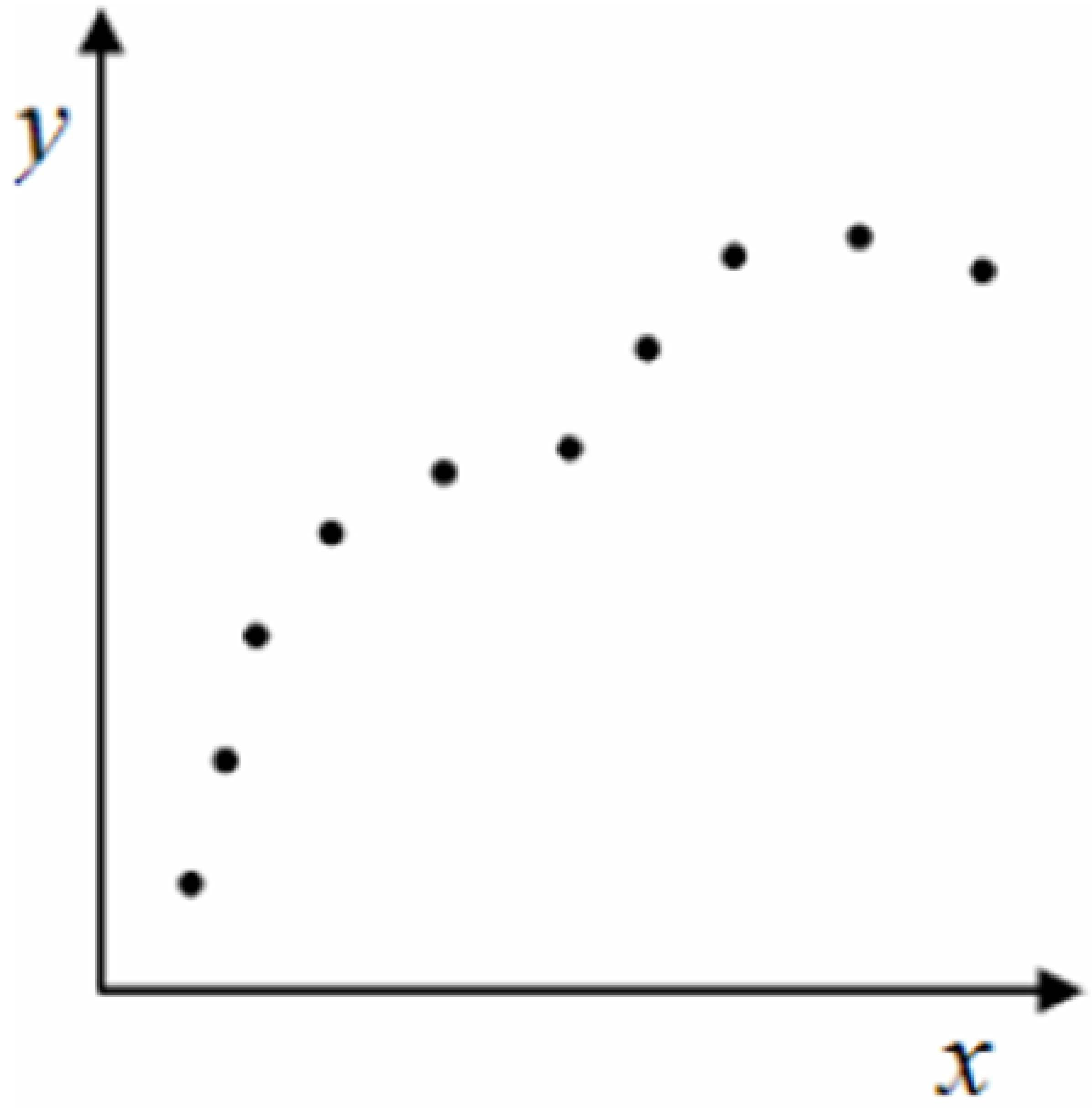


t (detik)	s (meter)
0	0
2	4
4	7
6	12
8	18
10	21
12	25

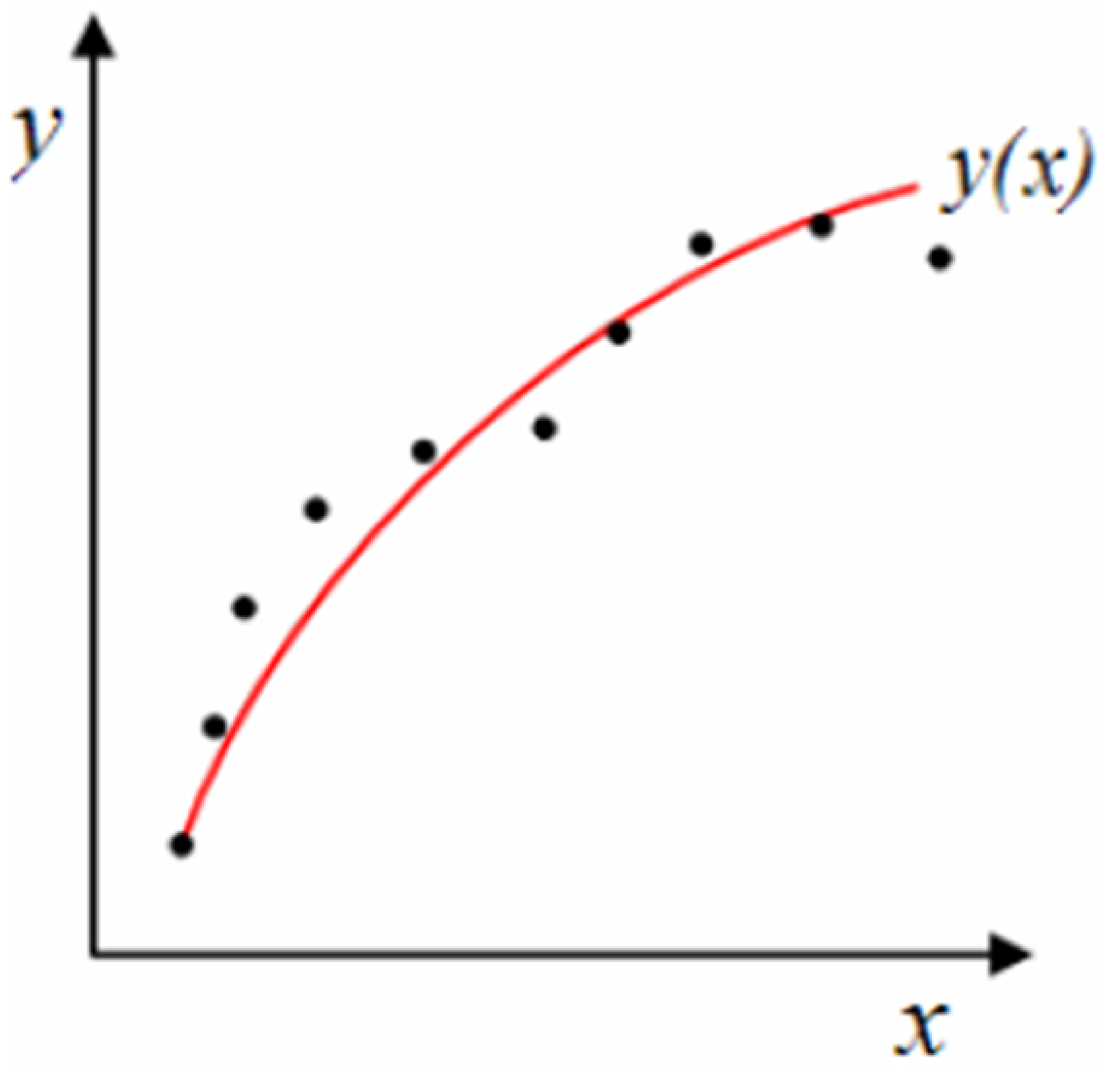


Ilustrasi berikut memperlihatkan titik-titik data yang diperoleh diplot dalam bentuk grafik.

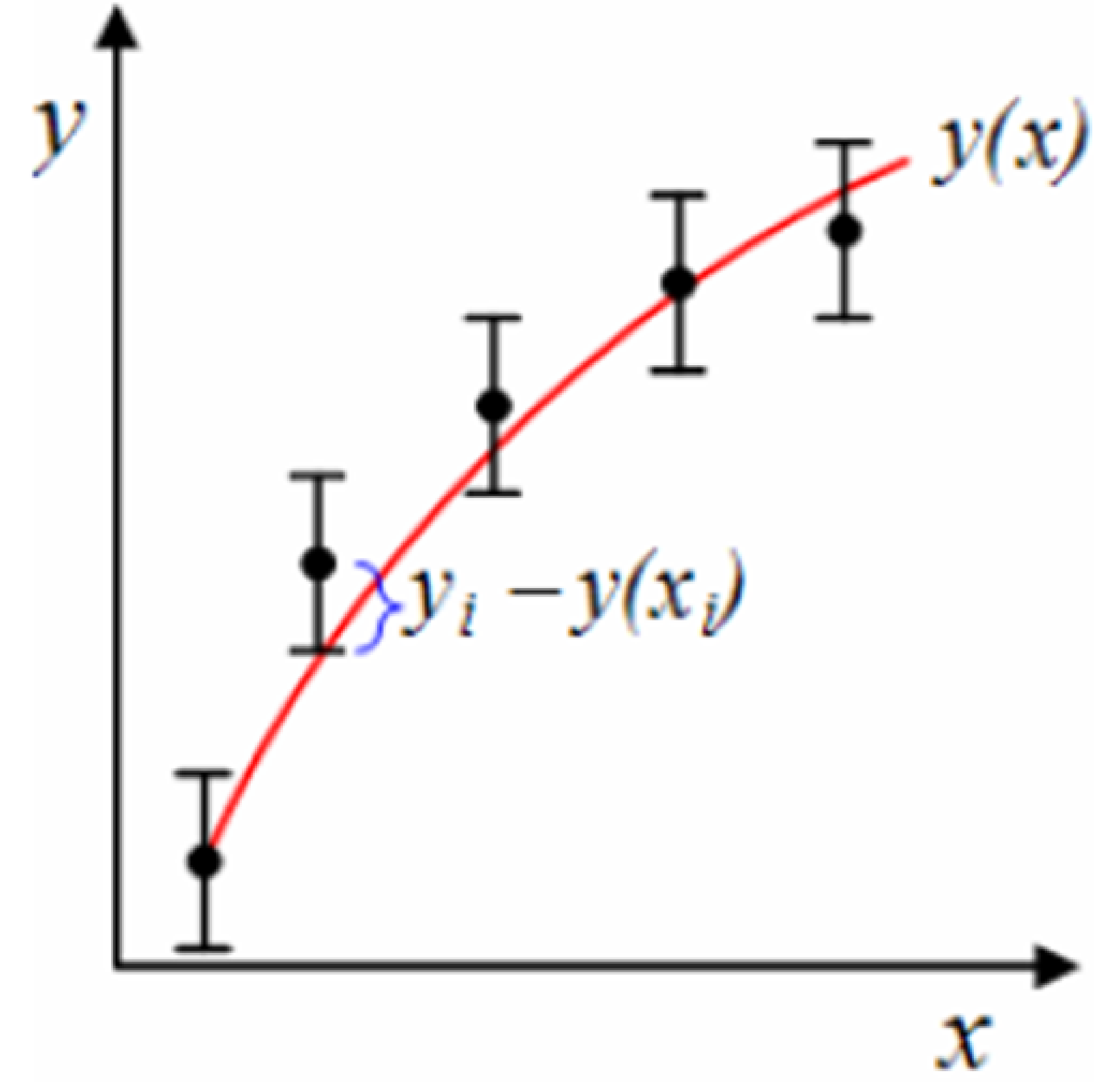
$f(x)$.



Titik-titik data tersebut jika dihubungkan dengan garis akan membentuk suatu fungsi $y = f(x)$



$y = f(x)$ secara teori dapat diketahui, tetapi terdapat perbedaan antara nilai eksperimen y_i dengan nilai teori $y(x_i)$ sebesar $y_i - y(x_i)$



Jika data yang dihasilkan cocok dengan teori, maka :

$$y_i - y(x_i) \rightarrow 0$$

Pada N buah titik data terdapat simpangan masing-masing sebesar:

$$d_1 = y_1 - y(x_1)$$

$$d_2 = y_2 - y(x_2)$$

$$d_3 = y_i - y(x_i)$$

Tetapi :

✓Untuk mengumpulkan seluruh titik data tadi tidak dapat dilakukan dengan menguji deviasi titik demi titik.

✓ $\sum_{i=1}^{n} d_i$ tidak dapat dipakai sebagai kriteria kecocokan

Solusi

untuk mengetahui perbedaan antara seluruh hasil eksperimen dengan teori, diusahakan dengan suatu metode tertentu

Kebolehjadian menemukan titik y_i, diketahui sebesar :

$$P = \frac{1}{S_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_i - y(x_i)}{S_i}\right)^2}$$

kebolehjadian menemukan set data (N buah) yang bernilai y_1 , y_2 , ... Adalah :

$$P_{\text{titik}} = P_1 , P_2 , P_3 \ldots . P_N$$

$$= \prod P_i = \frac{1}{\sqrt{2\pi}} \left(\prod_{i=1}^{N} S_i \right) e^{-\frac{1}{2}\sum\left(\frac{y_i - y(x_i)}{S}\right)^2}$$

Jika y_i cocok dengan $y(x_i)$ untuk seluruh N data maka PTOT akan maksimal, yang akan dipenuhi jika:

$$\chi^2 = \sum\left(\frac{y_i - y(x_i)}{S_i}\right)^2 \text{ minimal, yaitu } \frac{\partial x^2}{\partial y} = 0$$

Metode pencocokan ini dinamakan metode ***Least Square Fitting***
(metode kuadrat terkecil)