



Universitas Gadjah Mada  
Fakultas Teknik  
Departemen Teknik Sipil dan Lingkungan

# REGRESI DAN KORELASI

Statistika dan Probabilitas

# Kurva Regresi

2

- ❑ Mencari garis/kurva yang mewakili serangkaian titik data
- ❑ Ada dua cara untuk melakukannya, yaitu
  - ❑ Regresi
  - ❑ Interpolasi
- ❑ Aplikasi di bidang enjiniring
  - ❑ Pola perilaku data (*trend analysis*)
  - ❑ Uji hipotesis (*hypothesis testing*)

# Kurva Regresi

3

- ❑ Pemakaian regresi
  - ❑ Apabila data menunjukkan tingkat kesalahan yang cukup signifikan atau menunjukkan adanya *noise*
  - ❑ Untuk mencari satu kurva tunggal yang mewakili pola umum perilaku data
  - ❑ Kurva yang dicari tidak perlu melewati setiap titik data

# Kurva Regresi

4

- ❑ Interpolasi
  - ❑ Diketahui bahwa data sangat akurat
  - ❑ Untuk mencari satu atau serangkaian kurva yang melewati setiap titik data
  - ❑ Untuk memperkirakan nilai-nilai di antara titik-titik data
- ❑ Ekstrapolasi
  - ❑ Mirip dengan interpolasi, tetapi untuk memperkirakan nilai-nilai di luar kisaran titik-titik data

# Kurva Regresi terhadap Data Pengukuran

5

- ❑ Analisis pola perilaku data
  - ❑ Pemanfaatan pola data (pengukuran, eksperimen) untuk melakukan perkiraan
  - ❑ Apabila data persis (akurat): interpolasi
  - ❑ Apabila data tak persis (tak akurat): regresi
- ❑ Uji hipotesis
  - ❑ Perbandingan antara hasil teori atau hasil hitungan dengan hasil pengukuran

# Beberapa Besaran Statistik

6

merepresentasikan sebaran data

❑ Rata-rata aritmatik, *mean*



$$\bar{y} = \frac{1}{n} \sum y_i$$

❑ Deviasi standar, simpangan baku, *standard deviation*



$$s_y = \sqrt{\frac{S_t}{n-1}}$$

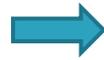
$$S_t = \sum (y_i - \bar{y})^2$$

❑ Varian ('ragam'), *variance*



$$s_y^2 = \frac{S_t}{n-1}$$

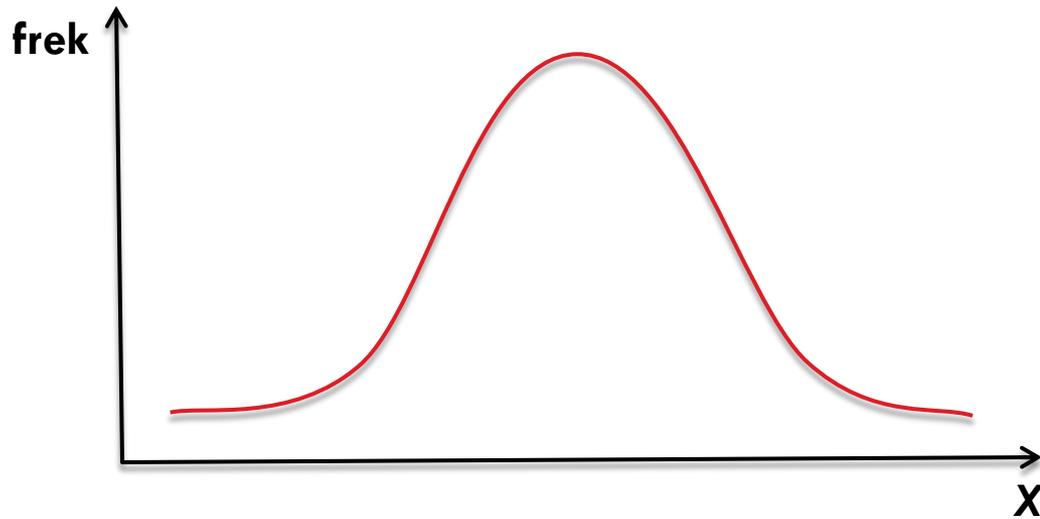
❑ *Coefficient of variation*



$$c_v = \frac{s_y}{\bar{y}} 100\%$$

# Distribusi Probabilitas

7



Distribusi Normal  
salah satu distribusi/sebaran data  
yang sering dijumpai adalah  
distribusi normal

# Regresi

Regresi linear: metode kuadrat terkecil

Regresi hubungan tak-linear yang dilinearkan

# Regresi: Metode Kuadrat Terkecil

9

- ❑ Mencari satu kurva atau satu fungsi (pendekatan) yang sesuai dengan pola umum yang ditunjukkan oleh data
  - ❑ Datanya menunjukkan kesalahan yang cukup signifikan
  - ❑ Kurva tidak perlu memotong setiap titik data
- ❑ Regresi linear
- ❑ Regresi persamaan-persamaan tak-linear yang dilinearkan
- ❑ Regresi tak-linear

# Regresi: Metode Kuadrat Terkecil

10

- ❑ Bagaimana caranya?
  - ❑ Program komputer
  - ❑ *Spreadsheet* (Microsoft Excel)
  - ❑ Program aplikasi: Matlab, Octave, Scilab

# Regresi Linear

12

- ❑ Kesalahan atau residu ( $e$ ) adalah perbedaan antara nilai  $y$  sesungguhnya (data  $y$ ) dan  $y$  nilai pendekatan ( $y_{reg}$ ) menurut persamaan linear  $a_0 + a_1x$ .

$$e = y - y_{reg} = y - (a_0 + a_1x)$$

- ❑ Meminimumkan jumlah kuadrat residu tersebut

$$\min[S_r] = \min \left[ \sum e_i^2 \right] = \min \left[ \sum (y - a_0 - a_1x)^2 \right]$$

# Regresi Linear

13

- Bagaimana cara mencari koefisien  $a_0$  dan  $a_1$ ?
  - Diferensialkan persamaan tersebut dua kali, masing-masing terhadap  $a_0$  dan  $a_1$ .
  - Samakan kedua persamaan hasil diferensiasi tersebut dengan nol.

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

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

# Regresi Linear

14

- Selesaikan persamaan yang didapat untuk mencari  $\alpha_0$  dan  $\alpha_1$

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

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

- dalam hal ini,  $\bar{y}$  dan  $\bar{x}$  masing-masing adalah nilai  $y$  rata-rata x rata-rata

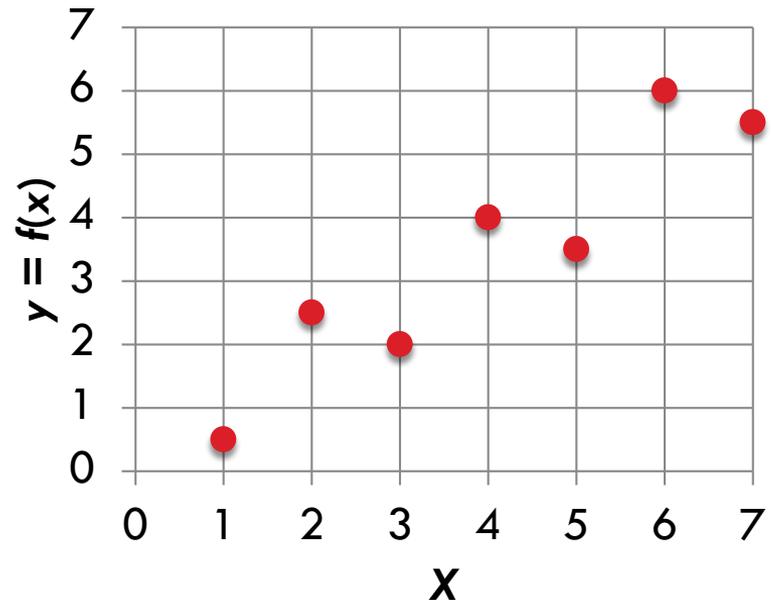
# Contoh Regresi Linear

15

Tabel data

$i$	$x_i$	$y_i = f(x_i)$
0	1	0.5
1	2	2.5
2	3	2
3	4	4
4	5	3.5
5	6	6
6	7	5.5

Grafik/kurva data



# Hitungan Regresi Linear

16

$i$	$x_i$	$y_i$	$x_i y_i$	$x_i^2$	$y_{reg}$	$(y_i - y_{reg})^2$	$(y_i - y_{mean})^2$
0	1	0.5	0.5	1	0.910714	0.168686	8.576531
1	2	2.5	5	4	1.75	0.5625	0.862245
2	3	2.0	6	9	2.589286	0.347258	2.040816
3	4	4.0	16	16	3.428571	0.326531	0.326531
4	5	3.5	17.5	25	4.267857	0.589605	0.005102
5	6	6.0	36	36	5.107143	0.797194	6.612245
6	7	5.5	38.5	49	5.946429	0.199298	4.290816
$\Sigma =$	<b>28</b>	<b>24.0</b>	<b>119.5</b>	<b>140</b>	$\Sigma =$	<b>2.991071</b>	<b>22.71429</b>

# Hitungan Regresi Linear

17

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.839286$$

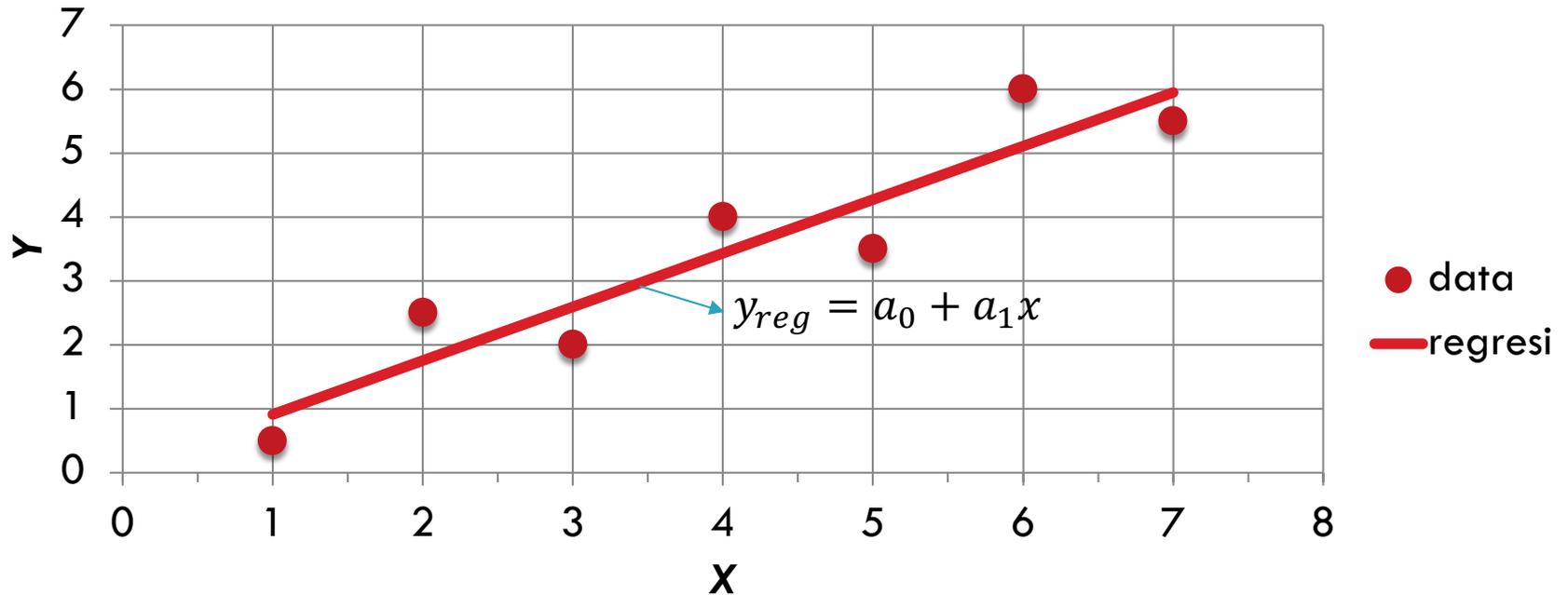
$$\bar{y} = \frac{24}{7} = 3.4$$

$$\bar{x} = \frac{28}{7} = 4$$

$$a_0 = \bar{y} - a_1 \bar{x} = 3.4 - 0.839286(4) = 0.071429$$

# Hitungan Regresi Linear

18



# Regresi Linear

19

- ❑ Kuantifikasi kesalahan
  - ❑ Kesalahan standar

$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} \quad S_r = \sum (y_i - a_0 - a_1 x_i)^2$$

- ❑ Perhatikan kemiripannya dengan simpangan baku

$$s_y = \sqrt{\frac{S_t}{n-1}} \quad S_t = \sum (y_i - \bar{y})^2$$

# Regresi Linear

20

- Beda antara kedua kesalahan tersebut menunjukkan perbaikan atau pengurangan kesalahan

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



koefisien determinasi  
(*coefficient of determination*)

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



koefisien korelasi  
(*correlation coefficient*)



$-1 \leq r \leq +1$

# Hitungan regresi linear

21

$$S_r = \sum (y_i - a_0 - a_1 x_i)^2 = 2.991071$$

$$S_t = \sum (y_i - \bar{y})^2 = 22.71429$$

$$r^2 = \frac{S_t - S_r}{S_t} = \frac{22.71429 - 2.991071}{22.71429} = 0.868318$$

$$r = 0.931836$$

$$-1 \leq r \leq +1$$

## Linearisasi persamaan tak-linear

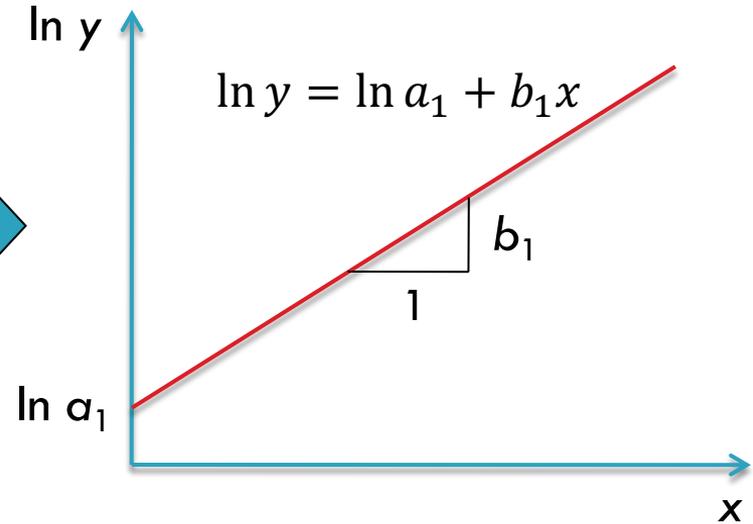
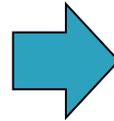
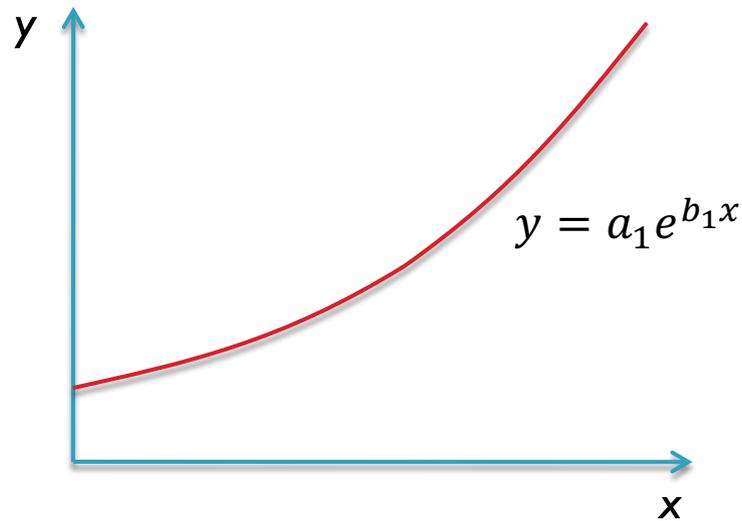
# Regresi Linear

23

- ❑ Linearisasi persamaan-persamaan tak-linear
  - ❑ Logaritmik menjadi linear
  - ❑ Eksponensial menjadi linear
  - ❑ Pangkat (polinomial tingkat  $n > 1$ ) menjadi linear (polinomial tingkat 1)
  - ❑ Dll.

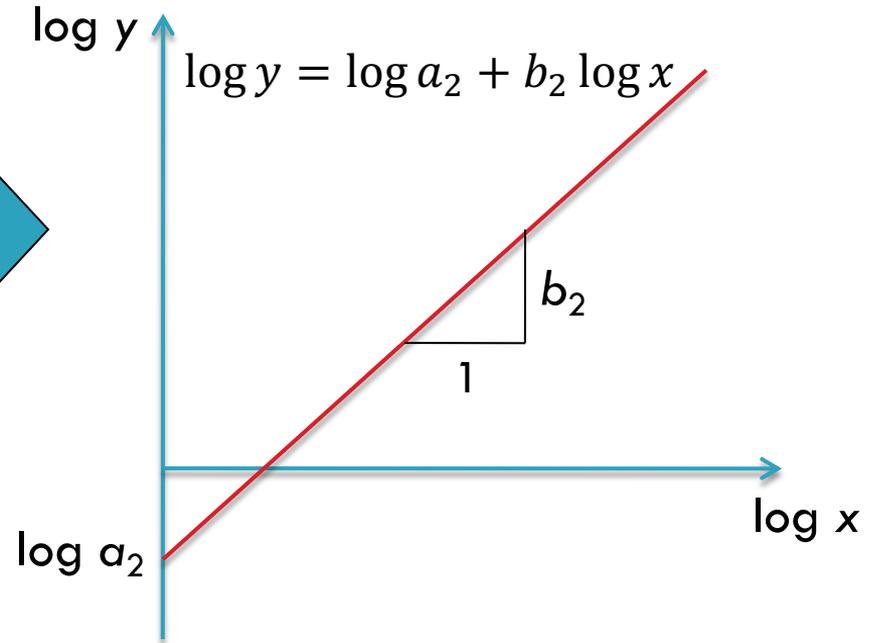
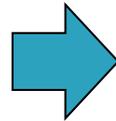
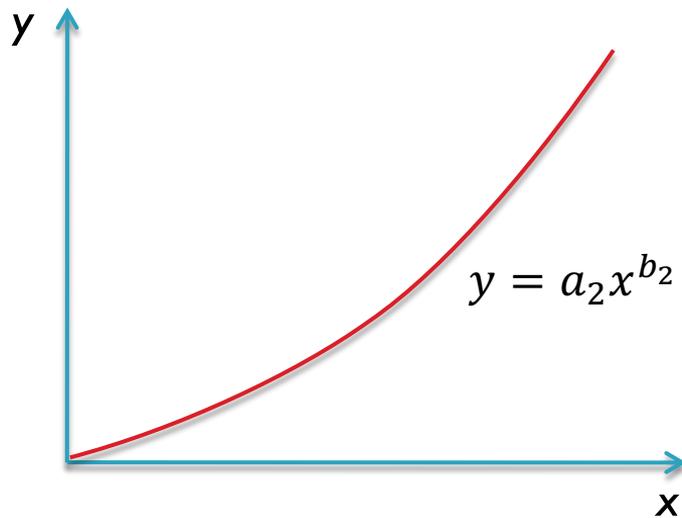
# Linearisasi Persamaan Tak-Linear

24



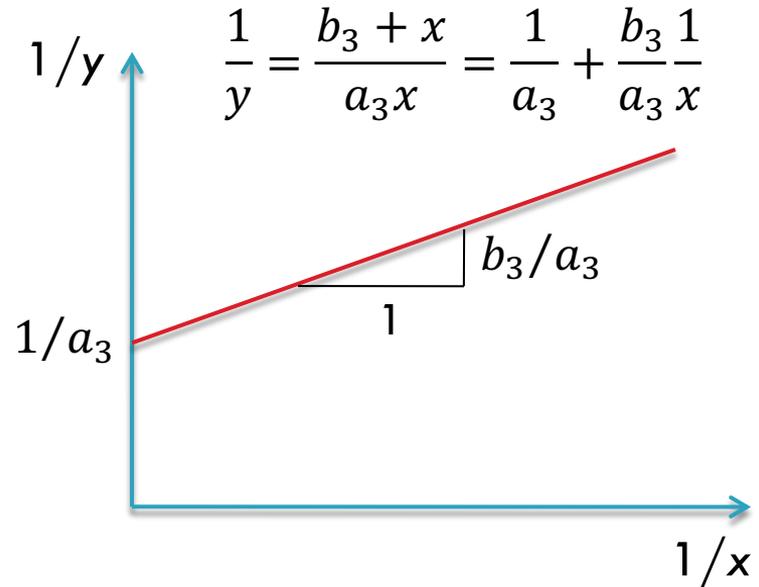
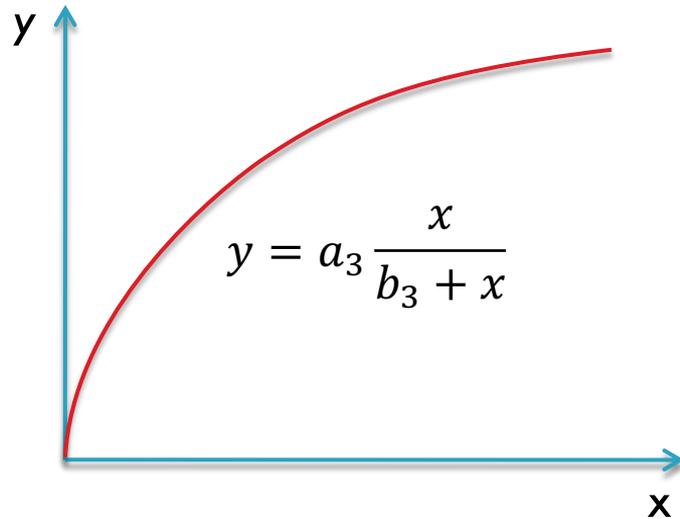
# Linearisasi Persamaan Tak-Linear

25



# Linearisasi Persamaan Tak-Linear

26



# Korelasi

Korelasi

Korelasi serial (auto korelasi)

# Koefisien Korelasi

28

$$r^2 = \frac{S_t - S_r}{S_t} = 1 - \frac{S_r}{S_t} \quad \rightarrow \quad r = \sqrt{\frac{S_t - S_r}{S_t}} = \sqrt{1 - \frac{S_r}{S_t}} \quad \text{koefisien korelasi}$$

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

$$S_r = \sum (y_i - a_0 - a_1 x_i)^2$$

$$r = \sqrt{1 - \frac{\sum (y_i - a_0 - a_1 x_i)^2}{\sum (y_i - \bar{y})^2}}$$

# Koefisien Korelasi

29

$$r = \sqrt{1 - \frac{\sum(y_i - a_0 - a_1x_i)^2}{\sum(y_i - \bar{y})^2}} \quad \rightarrow \quad r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

# Koefisien Korelasi

30

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

$$s_{X,Y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$



kovarian X dan Y

$$s_X = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$



simpangan baku X

$$s_Y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n - 1}}$$



simpangan baku Y

# Koefisien Korelasi

31

$$r = r_{X,Y} = \frac{S_{X,Y}}{S_X S_Y} \quad \rightarrow \quad r_{X,Y} \leftarrow \frac{= \text{COVARIANCE.S}(X,Y)}{= \text{STDEV.S}(X) \times \text{STDEV.S}(y)}$$
$$r_{X,Y} \leftarrow = \text{CORREL}(X,Y)$$

} MS Excel

- koefisien korelasi antara variabel random X dan Y

# Koefisien Korelasi

32

- ❑ Pengertian koefisien korelasi
  - ❑ Koefisien korelasi menunjukkan tingkat **keeratan hubungan linear** antara suatu variabel random  $Y$  dan suatu variabel kedua yang merupakan fungsi linear dari satu atau lebih variabel(-variabel)  $X$ 
    - Setiap variabel  $X$  dapat berupa variabel random atau bukan variabel random

# Koefisien Korelasi

33

- ❑ Nilai koefisien korelasi adalah  $-1 < r_{X,Y} < 1$ 
  - ❑  $r_{X,Y} = \pm 1$  menunjukkan hubungan linear sempurna antara  $X$  dan  $Y$
  - ❑  $r_{X,Y} = 0$  menunjukkan independensi (ketidak-bergantungan) linear, namun dapat saja keduanya memiliki hubungan (kebergantungan) yang lain, yang tidak linear
  - ❑ Jika  $X$  dan  $Y$  tidak saling bergantung (*independent*), maka  $r_{X,Y} = 0$
- ❑ Koefisien korelasi sampel dan populasi
  - ❑  $r_{X,Y}$  koefisien korelasi sampel
  - ❑  $\rho_{X,Y}$  koefisien korelasi populasi

# Inferensi terhadap Koefisien Korelasi Populasi

34

- ❑ Dua variabel random
  - ❑ tak berkorelasi,  $\rho_{X,Y} = 0$
  - ❑ berkorelasi,  $\rho_{X,Y} \neq 0$
- ❑ Situasi
  - ❑ Sampel yang diperoleh dari variabel random yang tidak berkorelasi
    - jarang menunjukkan nilai  $r_{X,Y} = 0$
    - koefisien korelasi sering  $r_{X,Y} \neq 0$ , karena kebetulan
  - ❑ Oleh karena itu perlu pengujian
    - untuk mengetahui penyimpangan koefisien korelasi dari nol tersebut disebabkan oleh kebetulan, atau
    - penyimpangan tersebut terlalu besar untuk dikatakan sebagai akibat kebetulan

# Inferensi terhadap $\rho$

35

- Uji hipotesis
  - $H_0: \rho_{X,Y} = 0$
  - $H_1: \rho_{X,Y} \neq 0$

statistik uji  $T = r \left[ \frac{n-2}{1-r^2} \right]^{1/2} \rightarrow |T| > t_{1-\alpha/2, n-2} \quad H_0 \text{ ditolak}$

# Inferensi terhadap $\rho$

36

## Uji hipotesis

□  $H_0: \rho_{X,Y} = \rho^*$  ( $\rho^*$  konstanta)

ukuran sampel  $n > 25$

□  $H_1: \rho_{X,Y} \neq \rho^*$

statistik uji  $Z = (W - \omega)(n - 3)^{1/3}$    $|Z| > z_{1-\alpha/2}$   $H_0$  ditolak

$$W = \frac{1}{2} \ln \left[ \frac{1+r}{1-r} \right] = \tanh^{-1} r$$

$$\omega = \frac{1}{2} \ln \left[ \frac{1+\rho^*}{1-\rho^*} \right] = \tanh^{-1} \rho^*$$

□ Rentang keyakinan  $\rho$ :  $\ell = \tanh \left[ W - \frac{z_{1-\alpha/2}}{(n-3)^{1/2}} \right]$   $u = \tanh \left[ W + \frac{z_{1-\alpha/2}}{(n-3)^{1/2}} \right]$

# Korelasi Serial

37

- ❑ Korelasi serial (*serial correlation*)
  - ❑ dikenal pula sebagai autokorelasi (*autocorrelation*)
  - ❑ yaitu korelasi antara data hasil pengukuran pada suatu waktu dengan data hasil pengukuran pada waktu sebelumnya
  - ❑ elemen dalam sampel yang memiliki korelasi serial **bukan** elemen random (ingat definisi variabel random)

# Korelasi Serial

38

- ❑ Pada korelasi serial, dengan demikian
  - ❑ sampel berukuran  $n$  yang memiliki korelasi serial akan memberikan informasi yang lebih sedikit dibandingkan dengan informasi yang dimiliki oleh sampel random berukuran  $n$
  - ❑ sebagian informasi pada sampel yang memiliki korelasi serial dapat diperoleh dari atau telah diketahui dalam data hasil pengukuran pada waktu sebelumnya

# Korelasi Serial

39

- ❑ Korelasi serial (*serial correlation*)
  - ❑ Dapat pula dijumpai antara suatu pengukuran pada waktu tertentu dengan pengukuran pada waktu  $k$  periode waktu sebelumnya (terdahulu),  $k = 1, 2, \dots$
  - ❑ Asumsi
    - Selang waktu antar pengukuran adalah sama (seragam)
    - Sifat-sifat statistis proses atau peristiwa yang diukur tidak berubah terhadap waktu (bersifat permanen)
    - $\rho(k)$  koefisien korelasi serial populasi
    - $r(k)$  koefisien korelasi serial sampel

# Korelasi Serial

40

$$r(k) = \frac{\sum_{i=1}^{n-k} x_i x_{i+k} - \sum_{i=1}^{n-k} x_i \sum_{i=1}^{n-k} x_{i+k} / (n - k)}{\left[ \sum_{i=1}^{n-k} x_i^2 - \left( \sum_{i=1}^{n-k} x_i \right)^2 / (n - k) \right]^{1/2} \left[ \sum_{i=1}^{n-k} x_{i+k}^2 - \left( \sum_{i=1}^{n-k} x_{i+k} \right)^2 / (n - k) \right]^{1/2}}$$

$$r(k) = r_{X_i, X_{i+k}} = \frac{S_{X_i, X_{i+k}}}{S_{X_i} S_{X_{i+k}}} \quad r_{X,Y} \leftarrow \frac{= \text{COVARIANCE.S}(X,Y)}{= \text{STDEV.S}(X) \times \text{STDEV.S}(Y)}$$

$$r_{X,Y} \leftarrow = \text{CORREL}(X,Y)$$

Regresi linear ganda (*multiple linear regression*)

# Korelasi Serial

42

- ❑  $r(0) = 1$  korelasi suatu elemen data dengan dirinya sendiri adalah sama dengan satu
- ❑ semakin besar  $k$ , jumlah pasangan data untuk menghitung  $r(k)$  semakin sedikit;  $r(k)$  adalah nilai estimasi  $\rho(k)$
- ❑ oleh karena itu,  $k \ll n$
- ❑ jika  $\rho(k) = 0$  untuk semua  $k$ , maka proses atau peristiwa atau populasi tersebut bersifat random murni

# Regresi Linear Ganda

43

- Misal variabel  $y$  adalah fungsi linear dua variabel bebas  $x_1$  dan  $x_2$

$$y = a_0 + a_1x_1 + a_2x_2$$

- Koefisien  $a_0, a_1, a_2$  pada persamaan di atas dapat ditemukan dengan metode kuadrat terkecil kesalahan (*error*)

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

# Regresi Linear Ganda

44

- Diferensial parsial persamaan tersebut terhadap masing-masing koefisien adalah sbb.

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

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

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

- Samakan persamaan diferensial tsb dengan nol dan atur suku-suku dalam persamaan

$$n a_0 + \sum_{i=1}^n x_{1i} a_1 + \sum_{i=1}^n x_{2i} a_2 = \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n x_{1i} a_0 + \sum_{i=1}^n x_{1i}^2 a_1 + \sum_{i=1}^n x_{1i} x_{2i} a_2 = \sum_{i=1}^n x_{1i} y_i$$

$$\sum_{i=1}^n x_{2i} a_0 + \sum_{i=1}^n x_{1i} x_{2i} a_1 + \sum_{i=1}^n x_{2i}^2 a_2 = \sum_{i=1}^n x_{2i} y_i$$

# Regresi Linear Ganda

- Persamaan-persamaan linear tersebut dapat dituliskan dalam bentuk persamaan matriks sbb.

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

# Regresi Linear Ganda

46

## □ Inversi matriks

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}x_{1i} & \sum_{i=1}^n x_{2i}x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{2i}x_{1i} & \sum_{i=1}^n x_{2i}x_{2i} \end{bmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i}y_i \\ \sum_{i=1}^n x_{2i}y_i \end{pmatrix}$$

# Contoh

47

- Temukanlah persamaan linear yang mewakili pola sebaran data dalam tabel di samping ini.
- Jawab

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

$$r^2 = 1$$

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

# Terima kasih