

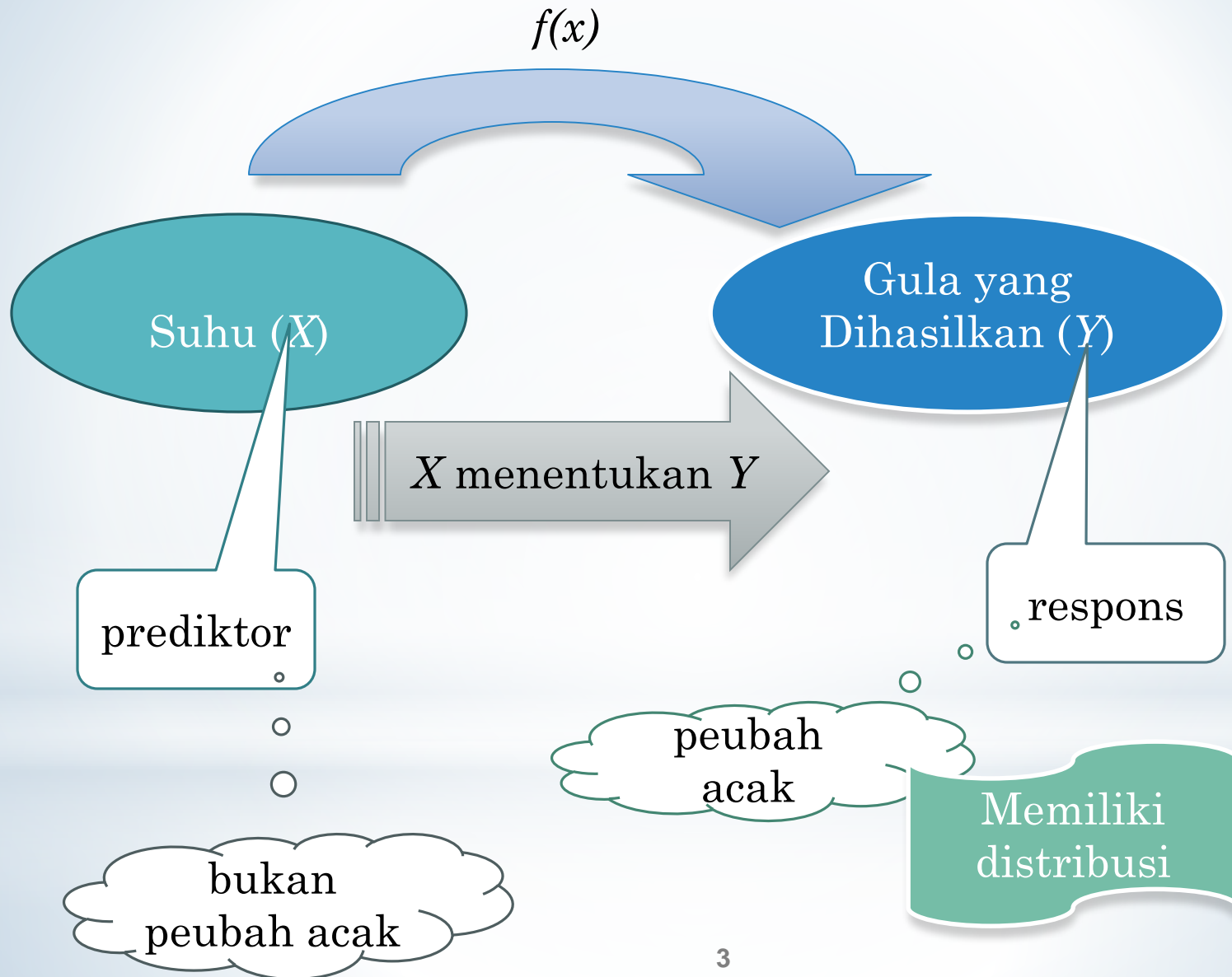
* Regresi Linear Sederhana dan Korelasi

1. Model Regresi Linear dan Penaksir Kuadrat Terkecil
2. Prediksi Nilai Respons
3. Inferensi Untuk Parameter-parameter Regresi
4. Kecocokan Model Regresi
5. Korelasi

Utriweni Mukhaiyar

1. Menentukan/menaksir parameter-parameter yang terlibat dalam suatu model matematik yang linear terhadap
2. Melakukan prediksi terhadap nilai suatu variabel, misalkan Y , berdasarkan nilai variabel yang lain, misalkan X , dengan menggunakan model regresi linier (**interpolasi**).

ILUSTRASI



Observasi	1	2	3	...	n
X	X_1	X_2	X_3	...	X_n
Y	Y_1	Y_2	Y_3	...	Y_n

1 Variabel yang nilainya mempengaruhi variabel yang lainnya.

2 Variabel yang terjadiannya lebih dahulu terjadi.

3 Variabel yang variansinya terkecil

Mana yang merupakan prediktor ??

MODEL REGRESI LINEAR SEDERHANA

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

- β_1 dan β_0 merupakan parameter-parameter model yang akan ditaksir
- e_i adalah galat pada observasi ke- i (acak)

1. Ketidakmampuan model regresi dalam memodelkan hubungan prediktor dan respons dengan tepat
2. Ketidakmampuan peneliti dalam melakukan pengukuran dengan tepat
3. Ketidakmampuan model untuk melibatkan semua variabel prediktor

*Sumber Galat

*Penaksir Kuadrat Terkecil

- β_1 dan β_0 ditaksir dengan metode kuadrat terkecil (*least square*)
- Asumsi-asumsi :
 1. Ada pengaruh X terhadap Y
 2. $Y_i = \beta_0 + \beta_1 X_i + e_i$ untuk $i = 1, 2, \dots, n$
 3. Nilai harapan dari e_i adalah 0, atau $E[e_i] = 0$
 4. Variansi dari e_i , sama untuk semua $i = 1, 2, \dots, n$
 5. e_i berdistribusi normal untuk semua $i = 1, 2, \dots, n$
 6. e_1, e_2, \dots, e_n saling bebas (independen)

Misalkan b_1 adalah taksiran bagi β_1 dan b_0 adalah taksiran bagi β_0 . Maka taksiran bagi model regresi adalah

$$\hat{Y}_i = b_0 + b_1 X_i$$

Kriteria penaksiran kuadrat terkecil adalah meminimumkan

$$\sum_{i=1}^n e_i^2$$

terhadap b_0 dan b_1 , dengan $e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i$

Diperoleh

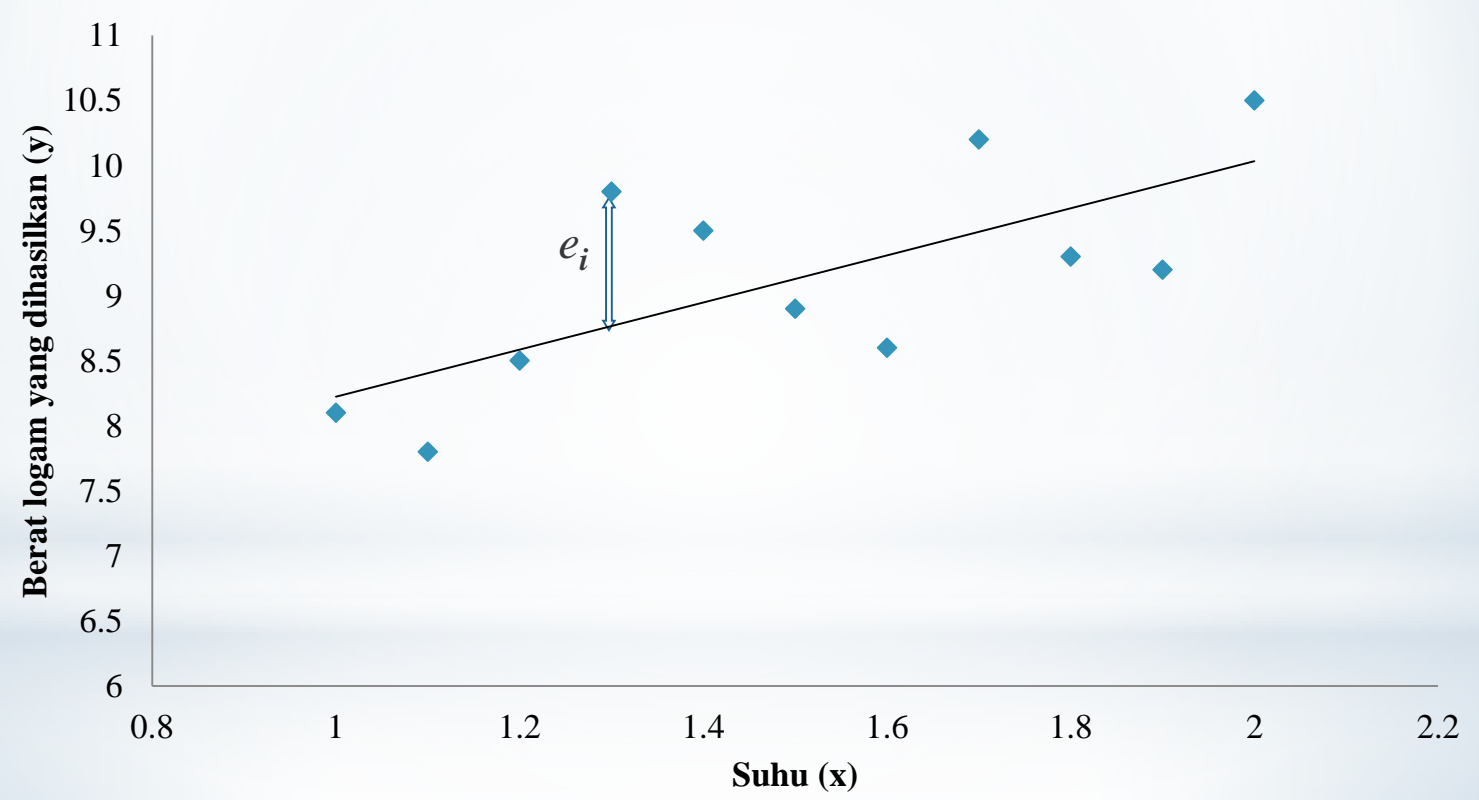
$$b_1 = \frac{JK_{XY}}{JK_{XX}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Sedangkan taksiran untuk variansi galat acak adalah

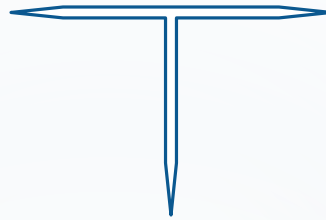
$$\hat{\sigma}^2 = s^2 = \frac{JK_G}{n-2} = \frac{(y_i - \hat{y}_i)^2}{n-2} = \frac{JK_{YY} - b_1 JK_{XY}}{n-2}$$

Suhu (X)	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Logam yg dihasilkan (Y)	8.1	7.8	8.5	9.8	9.5	8.9	8.6	10.2	9.3	9.2	10.5



$$n = 11$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{11} X_i = 1,5$$



$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{11} Y_i = 9,13$$

$$b_1 = \frac{\sum_{i=1}^{11} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{11} (X_i - \bar{X})^2} = 1,8091$$



$$b_0 = \bar{Y} - b_1 \bar{X} = 6,4136$$



$$\hat{Y}_i = 6,4136 + 1,8091X_i$$

Model persamaan regresi

*Prediksi Nilai Respons

Suhu (x_i)	Logam yg dihasilkan (y_i)	Prediksi model \hat{y}_i	$e_i = y_i - \hat{y}_i$
1	8.1	8.22	-0.12
1.1	7.8	8.40	-0.60
1.2	8.5	8.58	-0.08
1.3	9.8	8.77	1.03
1.4	9.5	8.95	0.55
1.5	8.9	9.13	-0.23
1.6	8.6	9.31	-0.71
1.7	10.2	9.49	0.71
1.8	9.3	9.67	-0.37
1.9	9.2	9.85	-0.65
2	10.5	10.03	0.47

Taksiran variansi galat acak $s^2 = \frac{JK_G}{n-2} = \frac{(y_i - \hat{y}_i)^2}{9} = 0,4$

Prediksi Nilai Respons

Misalkan suhu proses (X) adalah 1.55 satuan suhu.

Maka prediksi berat logam yang dihasilkan pada suhu tersebut adalah

$$\begin{aligned}\hat{Y} &= 6,4136 + 1,8091X \\ &= 6,4136 + 1,8091(1,55) \\ &= 9,2177\end{aligned}$$

ASUMSI KENORMALAN

1

- Asumsi e_i berdistribusi normal untuk semua $i = 1, 2, \dots, n$

2

- Y_i berdistribusi normal untuk semua $i = 1, 2, \dots, n$

3

- b_0 dan b_1 berdistribusi normal

INFERENSI UNTUK PARAMETER β_0

$$T_0 = \frac{b_0 - \beta_0}{s \sqrt{\sum_{i=1}^n x_i^2 / nJK_{XX}}}$$

berdistribusi t dengan derajat kebebasan $n-2$.

Selang kepercayaan $(1-\alpha)$ untuk β_0 :

$$b_0 - t_{\frac{\alpha}{2}, n-2} s \sqrt{\sum_{i=1}^n x_i^2 / nJK_{XX}} < \beta_0 < b_0 + t_{\frac{\alpha}{2}, n-2} s \sqrt{\sum_{i=1}^n x_i^2 / nJK_{XX}}$$

$t_{\alpha/2; n-2}$ adalah nilai distribusi t dengan derajat kebebasan $n-2$

INFERENSI UNTUK PARAMETER β_1

$$T_1 = \frac{b_1 - \beta_1}{s / \sqrt{JK_{XX}}}$$

berdistribusi t dengan derajat kebebasan $n-2$.

Selang kepercayaan $(1-\alpha)$ untuk β_1 :

$$b_1 - \frac{t_{\alpha/2; n-2} s}{\sqrt{JK_{XX}}} < \beta_1 < b_1 + \frac{t_{\alpha/2; n-2} s}{\sqrt{JK_{XX}}}$$

$t_{\alpha/2; n-2}$ adalah nilai distribusi t dengan derajat kebebasan $n-2$

PENGUJIAN PARAMETER REGRESI

Tujuan : menentukan apakah parameter-parameter tersebut dapat diabaikan atau tidak.

Rumusan Hipotesis

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Titik & Daerah Kritis

$$t_0 < t_{\alpha/2} \\ \text{atau } t_0 > t_{\alpha/2}$$

$$t_1 < t_{\alpha/2} \\ \text{atau } t_1 > t_{\alpha/2}$$

Statistik uji

$$t_0 = \frac{b_0}{\hat{\sigma} \sqrt{\frac{\sum_{i=1}^n x_i^2}{nJK_{xx}}}}$$

$$t_1 = \frac{b_1}{\hat{\sigma} / \sqrt{JK_{xx}}}$$

SELANG PREDIKSI

Misalkan nilai respons Y untuk $X = X_0$ adalah Y_0 , dan misalkan \hat{Y}_0 adalah prediksi model regresi bagi Y_0 . Maka

$$\tilde{T} = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma} \sqrt{1 + (1/n) + [(x_0 - \bar{x})^2 / JK_{xx}]}}$$

berdistribusi t dengan derajat kebebasan $n-2$.

Selang prediksi $(1 - \alpha)$ bagi y_0 adalah

$$\hat{y}_0 - t_{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK_{xx}}} < y_0 < \hat{y}_0 + t_{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK_{xx}}}$$

CONTOH 1 SELANG KEPERCAYAAN $1-\alpha$

TINJAU CONTOH SEBELUMNYA

$$1.8091 - \frac{(2.26)(0.4)}{\sqrt{1.1}} < \beta_1 < 1.8091 + \frac{(2.26)(0.4)}{\sqrt{1.1}}$$

$$JK_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 1.1$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2 = 0.4$$

$$b_1 = 1,8091$$

$$t_{\alpha/2} = t_{0.025} = 2.26$$

$$b_0 = 6,4136$$

$$0.947 < \beta_1 < 2.671$$

Selang kepercayaan
95% untuk β_1 :

$$5.092 < \beta_0 < 7.735$$

Selang kepercayaan
95% untuk β_0 :

$$6.4136 - (2.26)(0.4) \sqrt{\frac{25.85}{(11)(1.1)}} < \beta_0 < 6.4136 + (2.26)(0.4) \sqrt{\frac{25.85}{(11)(1.1)}}$$

CONTOH 2 UJI HIPOTESIS

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$

$$t_0 = \frac{b_0}{\hat{\sigma} \sqrt{\frac{\sum_{i=1}^n x_i^2}{nJK_{xx}}}} = \frac{6.4136}{0.4 \sqrt{\frac{25.85}{(11)(1.1)}}} = 10.97$$

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

$$t_1 = \frac{b_1}{\hat{\sigma} / \sqrt{JK_{xx}}} = \frac{1.8091}{0.4 / \sqrt{1.1}} = 4.31$$

derajat kebebasan
 $n - 2 = 9$,
nilai kritis
 $t_{0.025} = 2.26$

$t_0 > t_{0.025}$ &
 $t_1 > t_{0.025}$
maka masing-masing H_0 ditolak

Kesimpulan
 β_0 dan β_1 tidak
dapat diabaikan

*Kecocokan Model Regresi

Salah satu alat ukur untuk melihat apakah model regresi yang diperoleh sudah memadai adalah koefisien determinasi yaitu

$$R^2 = \frac{JK_R}{JK_T} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}, \quad \text{dengan} \quad 0 \leq R^2 \leq 1$$

Besaran R^2 menunjukkan proporsi variasi total dalam respons Y yang diterangkan oleh model regresi yang diperoleh

UJI KEBAIKAN MODEL

H_0 : Model regresi yang diperoleh tidak memadai

H_1 : Model memadai

Statistik uji

$$f = \frac{JK_R}{s} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{s}$$

Tolak H_0 pada tingkat keberartian α jika $f > f_{\alpha, (1, n-2)}$,
dimana $f_{\alpha, (1, n-2)}$ adalah nilai distribusi F dengan
derajat kebebasan 1 dan $n - 2$.

CONTOH 3

Untuk contoh sebelumnya diperoleh $R^2 = 0,499$.

Artinya proporsi variasi total dalam respons Y yang diterangkan oleh model regresi yang diperoleh adalah 49.9%

Uji kebaikan model

$$f = \frac{JK_R}{s} = \frac{\sum_{i=1}^{11} (\hat{y}_i - \bar{y}_i)^2}{s} = 8,99$$

Untuk $\alpha = 5\%$, titik kritis $f_{0.05,(1,9)} = 5,12$

$f > f_{0.05,(1,9)}$, model memadai.

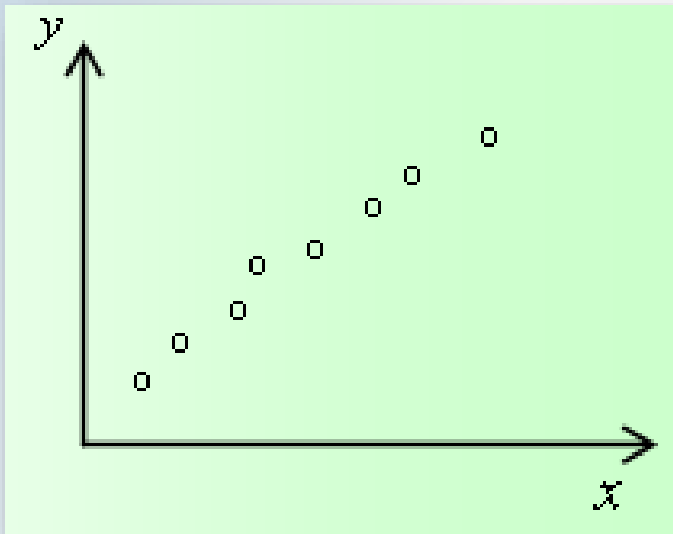
* Korelasi

- * Mengukur hubungan linear dua peubah acak
- * Misalkan X dan Y adalah dua peubah acak, maka korelasi antara X dan Y dinyatakan dengan

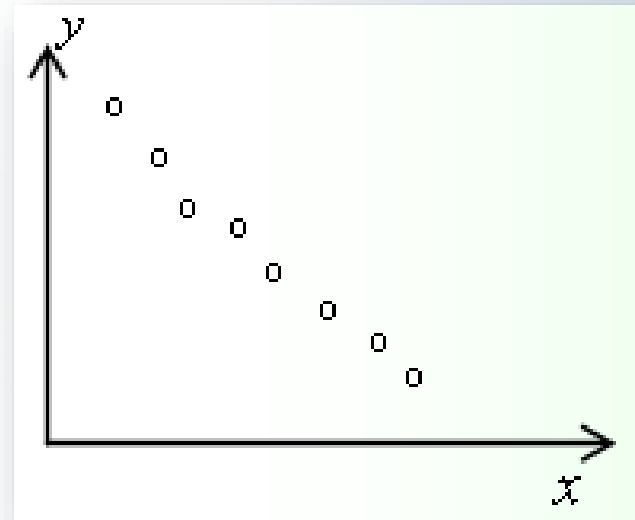
$$\rho_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_X)^2]E[(Y - \mu_Y)^2]}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Jika nilai korelasi mendekati 1 maka hubungan kedua peubah “sangat erat” dan searah sedangkan jika nilai korelasi mendekati -1 maka hubungan kedua peubah “sangat erat” dan berlawanan arah.

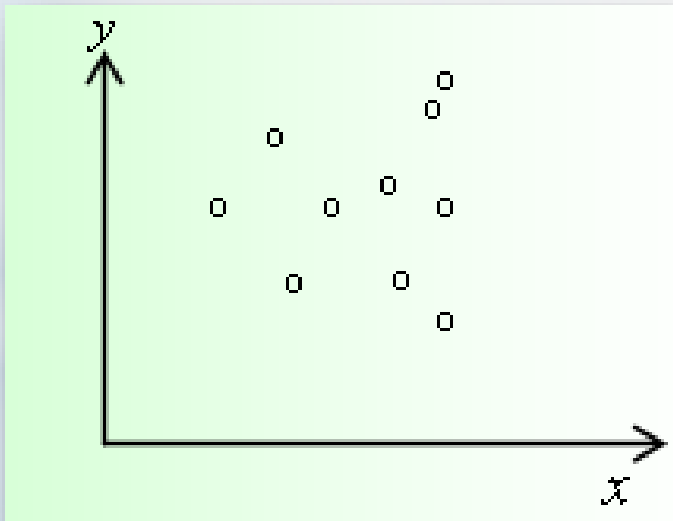
Jika nilai korelasi sama dengan nol berarti tidak terdapat hubungan linear antara kedua peubah acak.



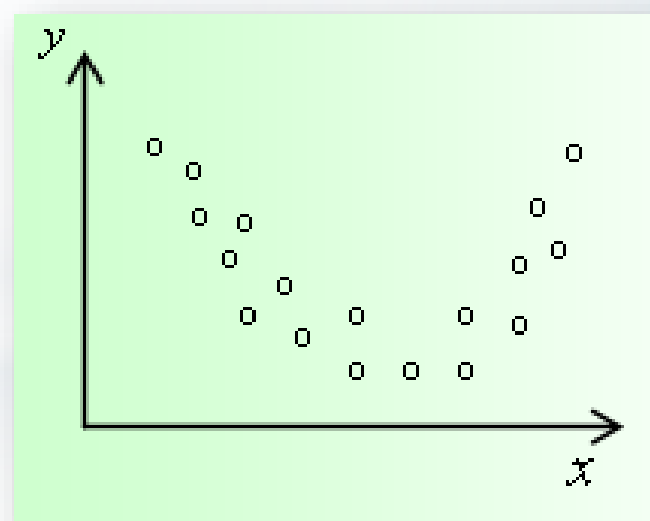
Gambar 1 Korelasi positif



Gambar 2 Korelasi negatif



Gambar 3 Korelasi nol



Gambar 4 Korelasi nol

KORELASI SAMPEL

Korelasi dapat ditaksir dengan koefisien korelasi sampel, yaitu

$$r = \frac{JK_{XY}}{\sqrt{JK_{XX} JK_{YY}}}$$
$$= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) \left(\sum_{i=1}^n (Y_i - \bar{Y})^2\right)}}$$

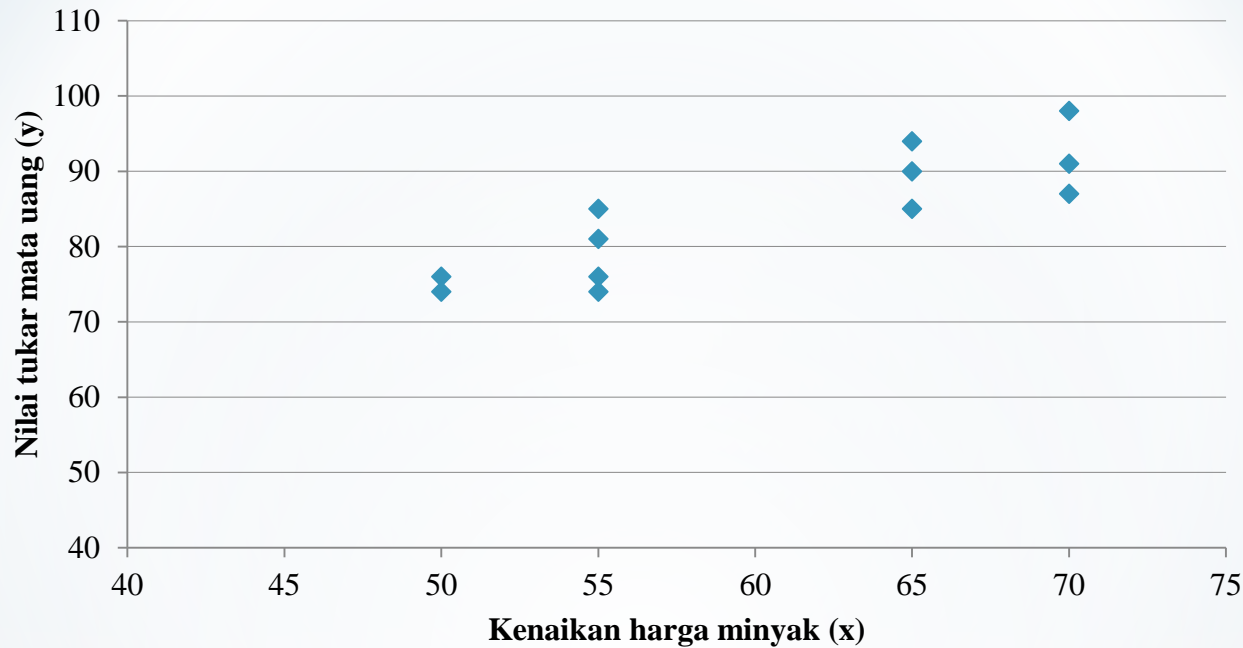
CONTOH 4

Data berikut menggambarkan kenaikan harga minyak dan nilai tukar mata uang di 12 negara terhadap US dollar pada suatu tahun.

Negara	Kenaikan harga minyak (x)	Nilai tukar mata uang terhadap US dollar (y)
1	65	85
2	50	74
3	55	76
4	65	90
5	55	85
6	70	87
7	65	94
8	70	98
9	55	81
10	70	91
11	50	76
12	55	74

Rata-rata kenaikan harga minyak = 60,42 ,

Rata-rata nilai tukar mata uang terhadap US dollar = 84,25



$$r = \frac{JK_{XY}}{\sqrt{JK_{XX} JK_{YY}}} = \frac{\sum_{i=1}^{12} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{12} (X_i - \bar{X})^2 \sum_{i=1}^{12} (Y_i - \bar{Y})^2}} = 0,863$$

TUGAS B

Lanjutan Tugas A (Kelompok)

- Terapkan minimal satu topik berikut, yang sudah dipelajari dalam perkuliahan Statdas, ke dalam data kelompok Anda seperti pada Tugas A.
- Topik Bahasan :
 - Uji Hipotesis
 - ANOVA
 - Regresi Linier dan Korelasi
 - Analisis Deret Waktu
 - Analisis Spasial (Geostatistik)
- Tugas diketik rapi dan lengkap (data dan analisisnya) dalam bentuk laporan (*style* masing-masing) dalam format Mic. Word. Dengan penamaan file :

“Tugas B - Statdas02 - II.2014 – Kelompok <nomor kelompok>”
- Tugas dikumpulkan via email ke utriweni@math.itb.ac.id paling lambat Senin, 5 Mei 2014.

Referensi

- Pasaribu, U.S., 2007, *Catatan Kuliah Biostatistika*.
- Walpole, Ronald E. Dan Myers, Raymond H., *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*, Edisi 4, Bandung: Penerbit ITB, 1995.
- Walpole, Ronald E., et.al, *Statistic for Scientist and Engineering*, 8th Ed., 2007.