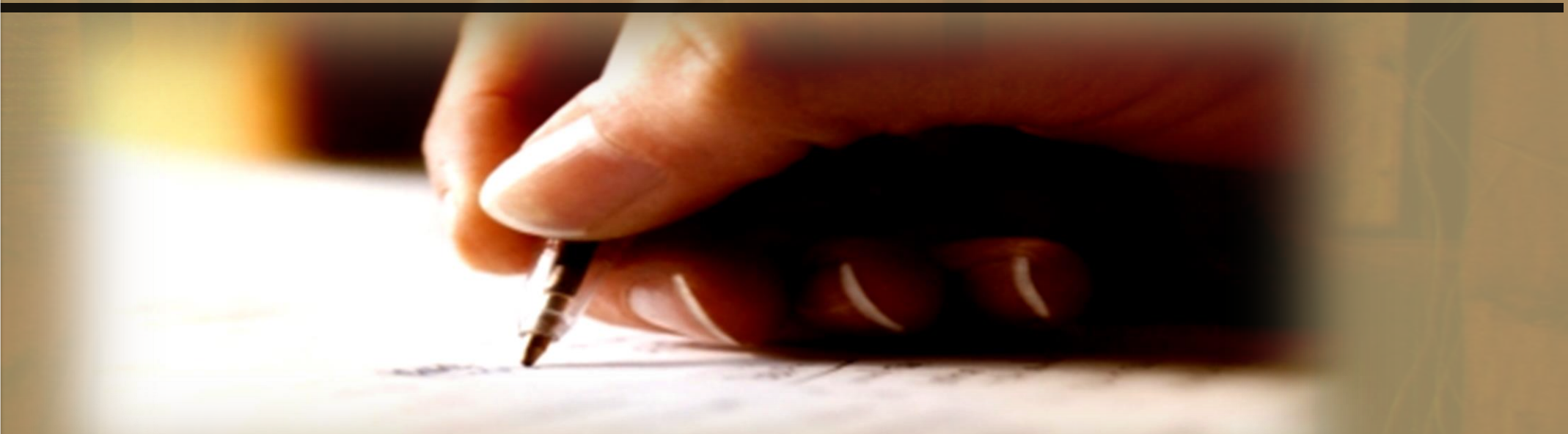




**Institut Teknologi Sepuluh Nopember
Surabaya**

**DEPARTEMEN TEKNIK FISIKA -
FTIRS**



KOEFISIEN KORELASI

Oleh: Syamsul Arifin

Koefisien korelasi

Capaian Pembelajaran:

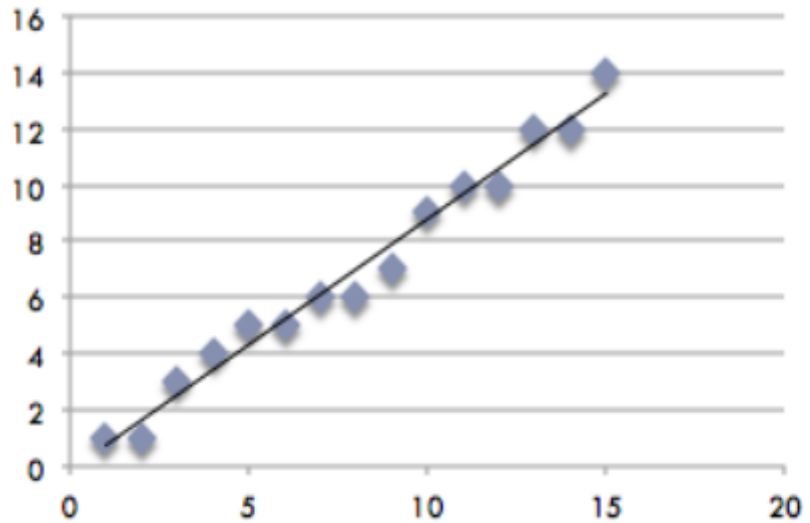
Mampu menentukan besaran nilai koefisien korelasi dan mampu menentukan besar eror prediksi (residual)

Kajian:

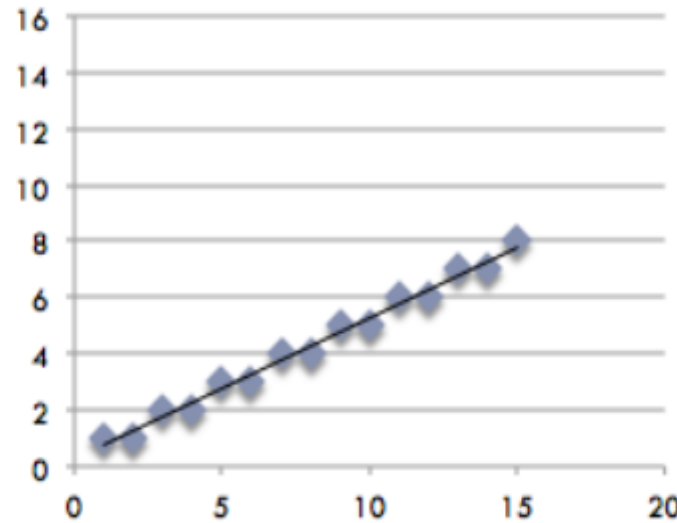
1. Koefisien korelasi
2. Eror prediksi / residual



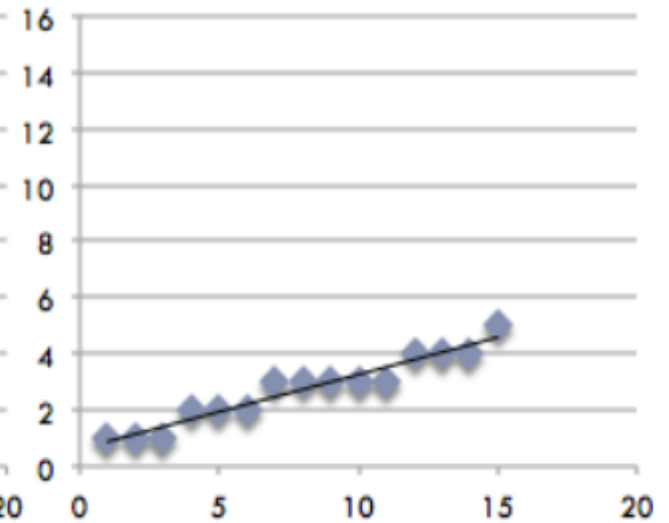
Beberapa pola grafik regresi linier



Hubungan **sangat kuat** antar variabel



Hubungan **“moderate”** antar variabel



Hubungan **“lemah”** antar variabel

Gradien berbeda antara gambar di atas

Koefisien Korelasi Pearson: r

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Standar deviasi variable x (var. bebas)

Standar deviasi variable y (var. terikat)

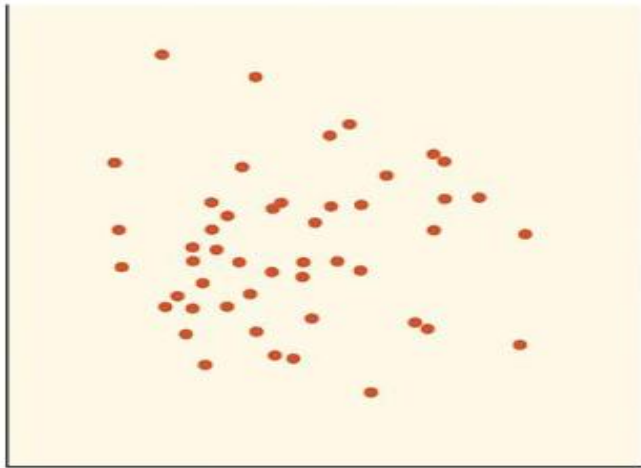
Koefisien korelasi - r

- ✓ r selalu bernilai diantara -1 and 1.
- ✓ $r > 0$ menunjukkan asosiasi positif.
- ✓ $r < 0$ menunjukkan asosiasi negatif.
- ✓ Nilai r yang mendekati 0 menunjukkan hubungan yang sangat lemah
- ✓ Kekuatan hubungan antara variable ditunjukkan dengan semakin naiknya nilai r yang bergerak menjauhi 0 dan menuju ke -1 atau 1
- ✓ Nilai ekstrim $r = -1$ dan $r = 1$ terjadi hanya saat kasus hubungan yang sangat “perfect”

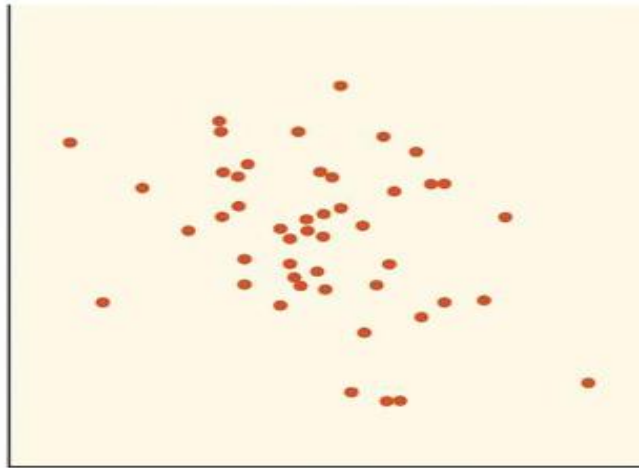
Identifikasi terhadap – hubungan antar variable

Nilai absolut r	Kekuatan hubungan antar variabel
$r < 0.3$	Tidak ada (atau sangat lemah)
$0.3 < r < 0.5$	Lemah
$0.5 < r < 0.7$	Moderate
$r > 0.7$	Kuat

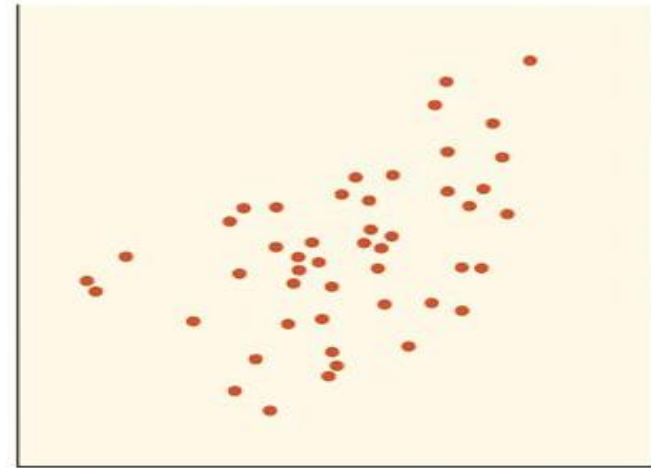
Contoh



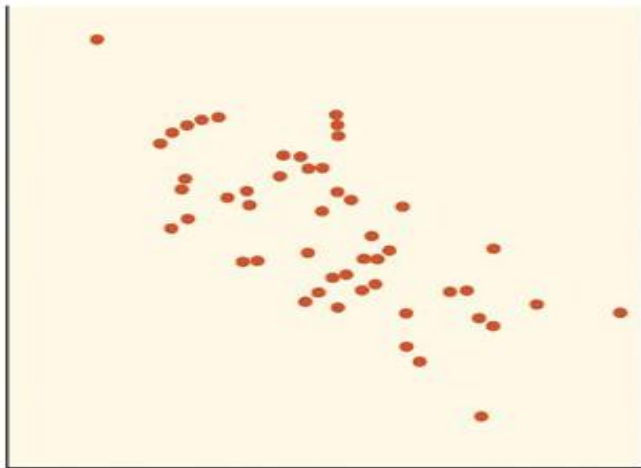
$r = 0$



$r = -0.3$



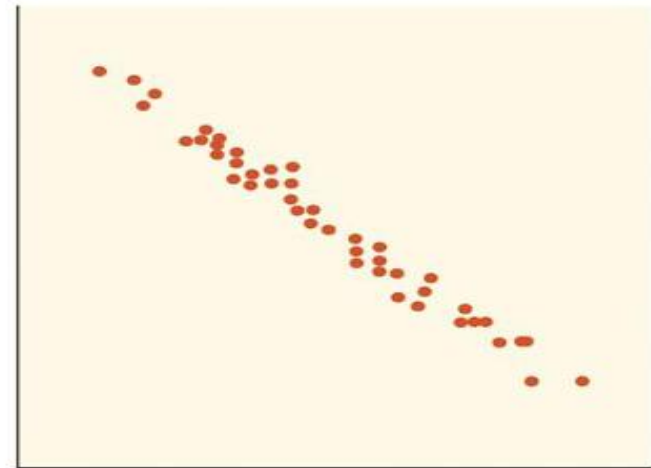
$r = 0.5$



$r = -0.7$

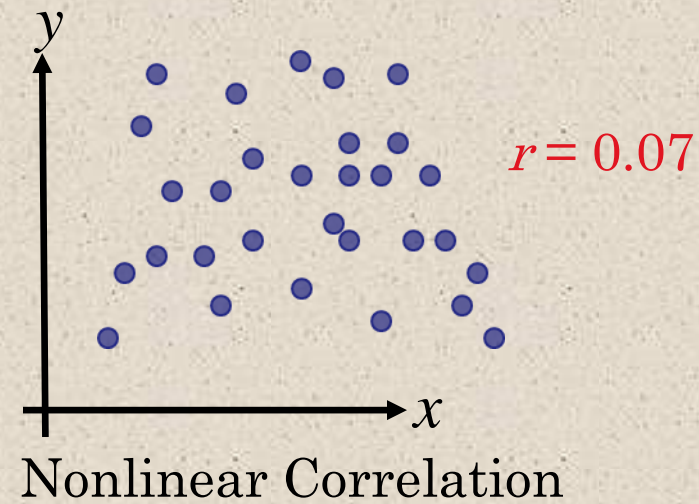
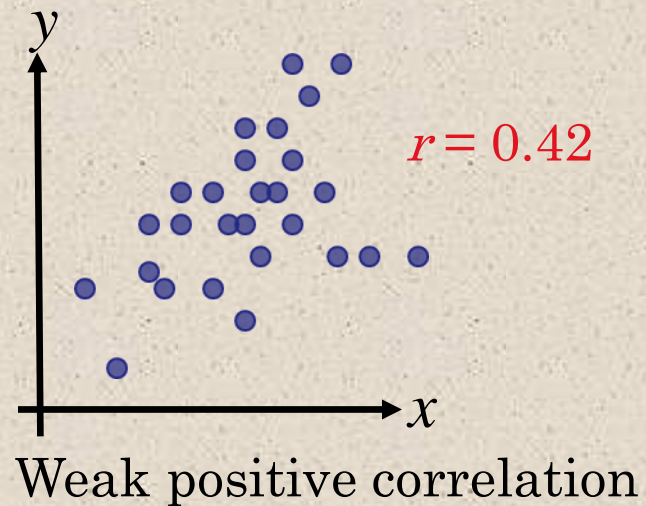
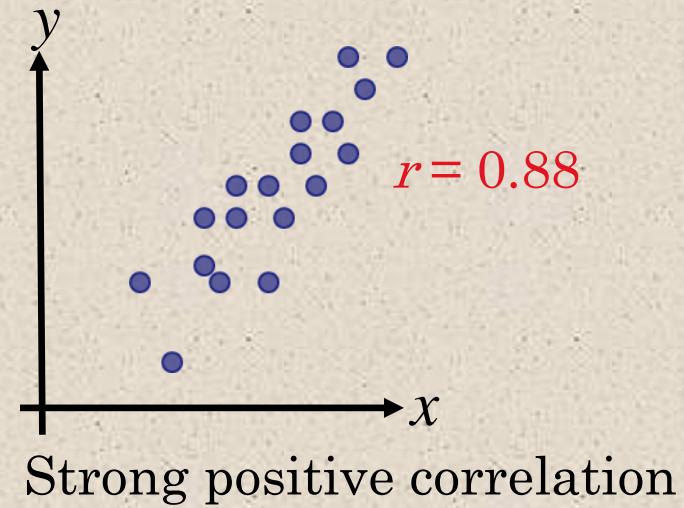
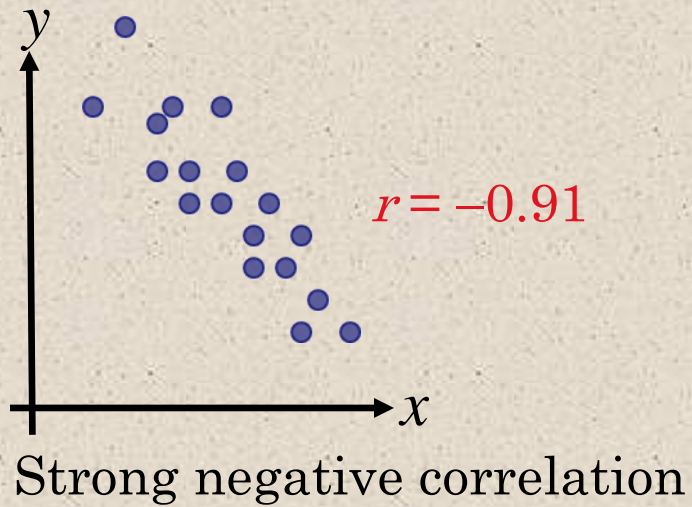


$r = 0.9$



$r = -0.99$

Korelasi linier

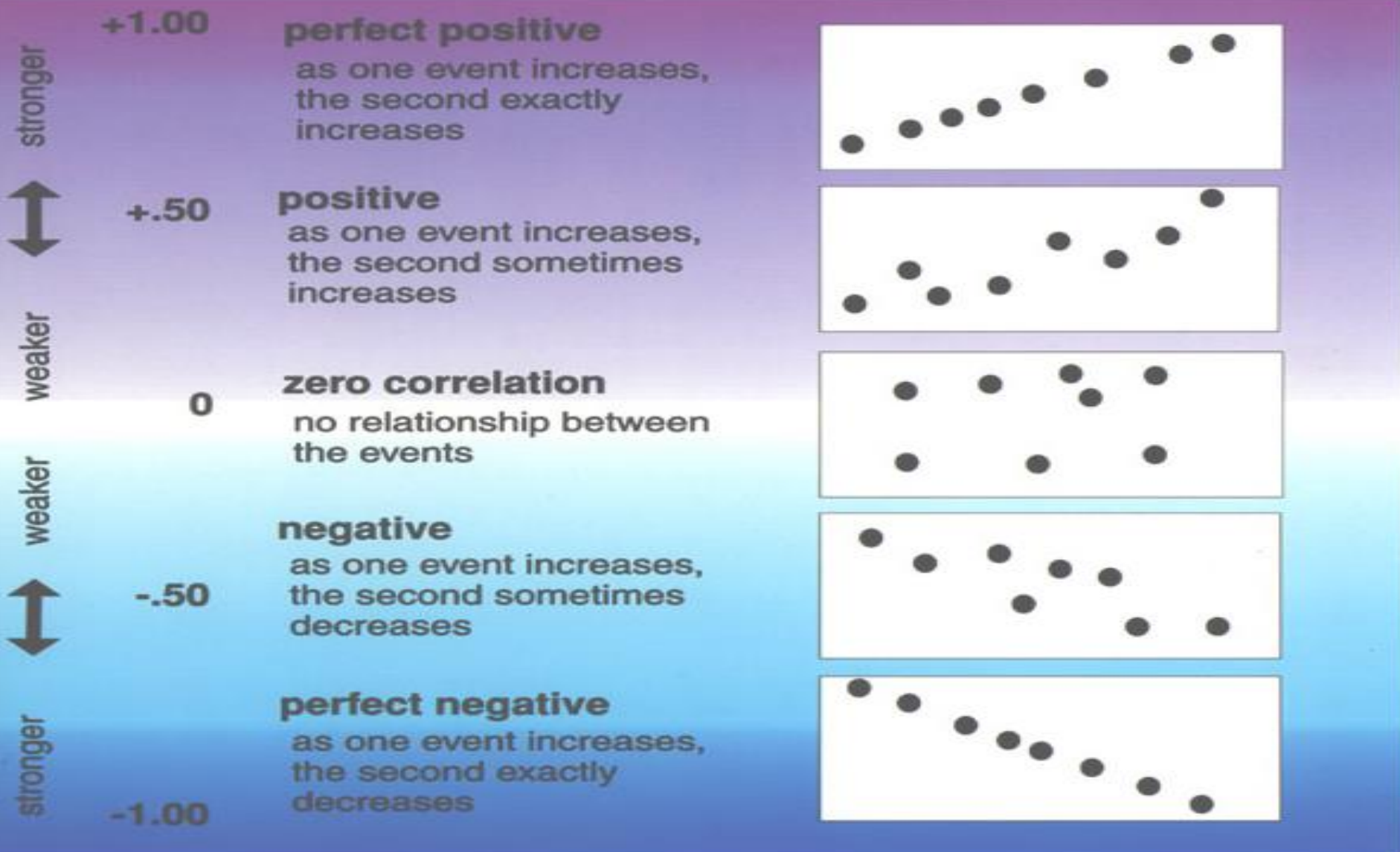


Correlation

High positive correlation

Zero correlation

High negative correlation



Terminologi yang digunakan pada regresi Linier

- Y – variable yang diprediksi
 - Sbg: Variabel terikat
- X – variable yang digunakan untuk memprediksi
 - Sbg: variable bebas
- \hat{Y} - Hasil prediksi (dapat juga dituliskan Y')

Contoh

- Akan dilakukan prediksi.
- Hubungan antara
 - Seberapa cepat kenaikan cacat jantung bawaan (kelahiran) / CHD (Congenital heart defects) terhadap kenaikan jumlah perokok?
 - Catatan: disini adalah melakukan prediksi, tetapi tidak secara actual (sbg hasil prediksi).

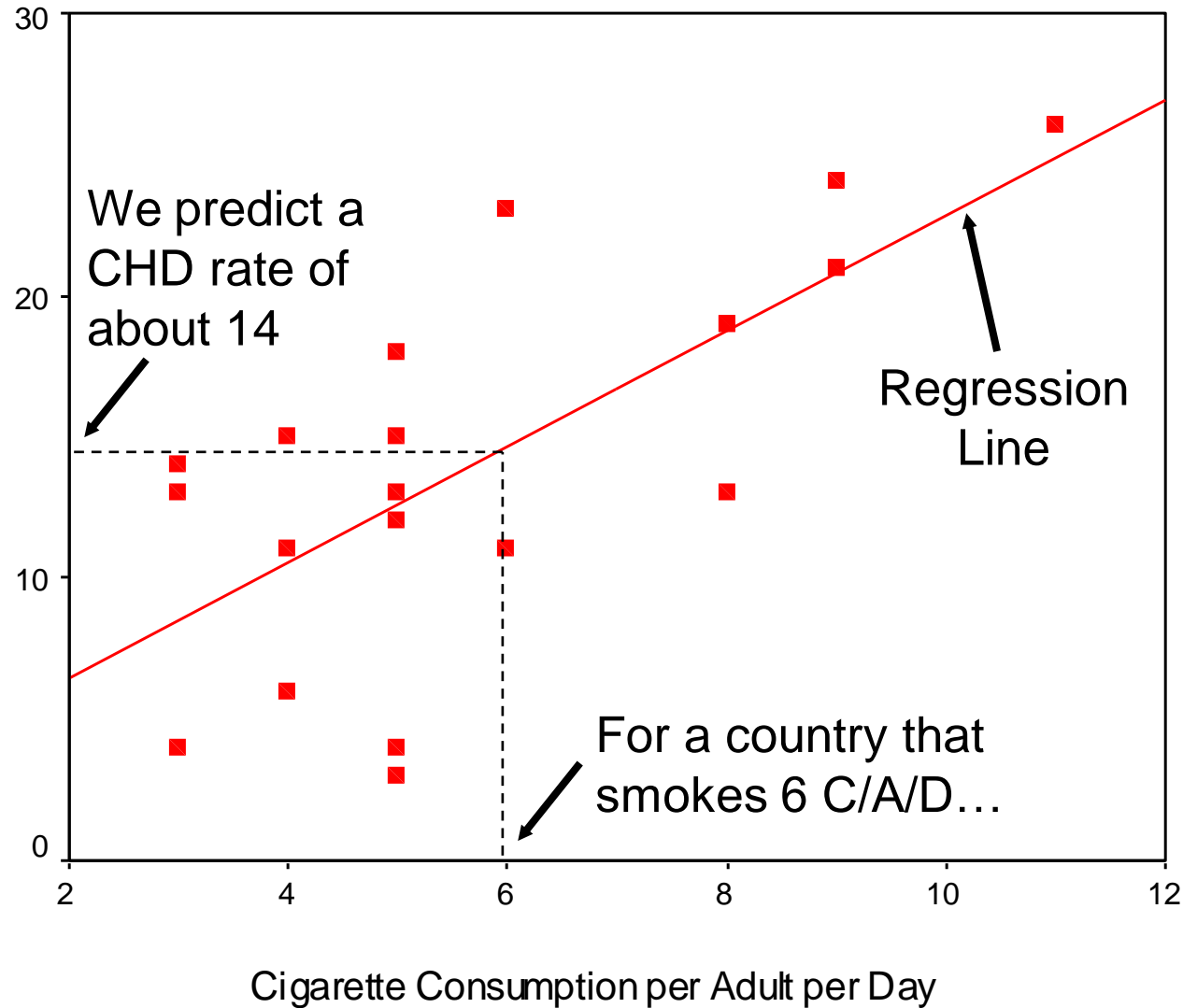
Contoh

- Rokok dan jumlah kelahiran dengan CHD
- Dilakukan prediksi level CHD di sebuah negara ttt dengan jumlah konsumsi rokok per hari (perkulia terhadap 10 pengguna rokok)

Data

Country	Cigarettes	CHD
1	11	26
2	9	21
3	9	24
4	9	21
5	8	19
6	8	13
7	8	19
8	6	11
9	6	23
10	5	15
11	5	13
12	5	4
13	5	18
14	5	12
15	5	3
16	4	11
17	4	15
18	4	6
19	3	13
20	3	4
21	3	14

Grafik hubungan antara jumlah konsumsi rokok per hari



Regression Line

- Rumus

$$\hat{Y} = b_0 + b_1 X$$

- \hat{Y} = hasil prediksi (yaitu jumlah kelahiran CHD)
- X = variable predictor (yaitu rata-rata rokok yang dikonsumsi org / negara)

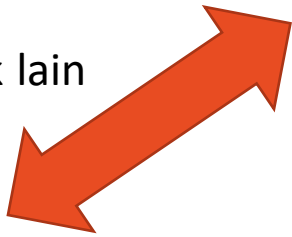
Koefisien regresi

- Koefisien adalah b_0 dan b_1
- b_1 = slope
 - Perubahan nilai prediksi untuk setiap kenaikan 1 satuan X
- b_0 = interpretasi
 - Nilai \hat{Y} saat $X = 0$

Ingat persamaan – parameter regresi

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Persamaan bentuk lain



$$b_1 = \frac{\text{cov}_{XY}}{s_X^2} \text{ atau } b_1 = r \left[\frac{s_y}{s_x} \right]$$

$$b_1 = \frac{N \sum XY - \sum X \sum Y}{[N \sum X^2 - (\sum X)^2]}$$

Cov = Kovarian antara X dg Y

Perhitungan – Gunakan rumus sebelumnya

- Slope

$$b1 = \frac{\text{COV}_{XY}}{s_X^2} \text{ atau } b1 = r \left[\frac{s_y}{s_x} \right]$$

$$b1 = \frac{N \sum XY - \sum X \sum Y}{[N \sum X^2 - (\sum X)^2]}$$

- Interpretasi

$$b0 = \bar{Y} - b1\bar{X}$$

Dalam contoh

- $\text{Cov}_{XY} = 11.12$
- $s^2_X = 2.33^2 = 5.447$
- $b1 = 11.12/5.447 = 2.042$
- $b0 = 14.524 - 2.042*5.952 = 2.32$
- Gunakan SPSS / excel untuk menghitung nilai di atas

SPSS Printout

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.367	2.941		.805	.431
	Cigarette Consumption per Adult per Day	2.042	.461	.713	4.426	.000

a. Dependent Variable: CHD Mortality per 10,000

catatan:

- Nilai yang dihasilkan merupakan bentuk printout SPSS.
- Interpretasi b_0 diberi label “constant”
- Slope b_1 dilabel variable predictor

Berlatihlah dengan SPSS – search free SPSS demo

Menentukan prediksi

- Hasil prediksi

$$\hat{Y} = b_1X + b_0 = 2.042X + 2.367$$

$$\hat{Y} = 2.042 * 10 + 2.367 = 22.787$$

- Dapat kita prediksi bawa untuk 22.77 penduduk / 10000 di sebuah negara ttt akan diprediksi muncul kelahiran 10 yang akan meninggal akibat CHD

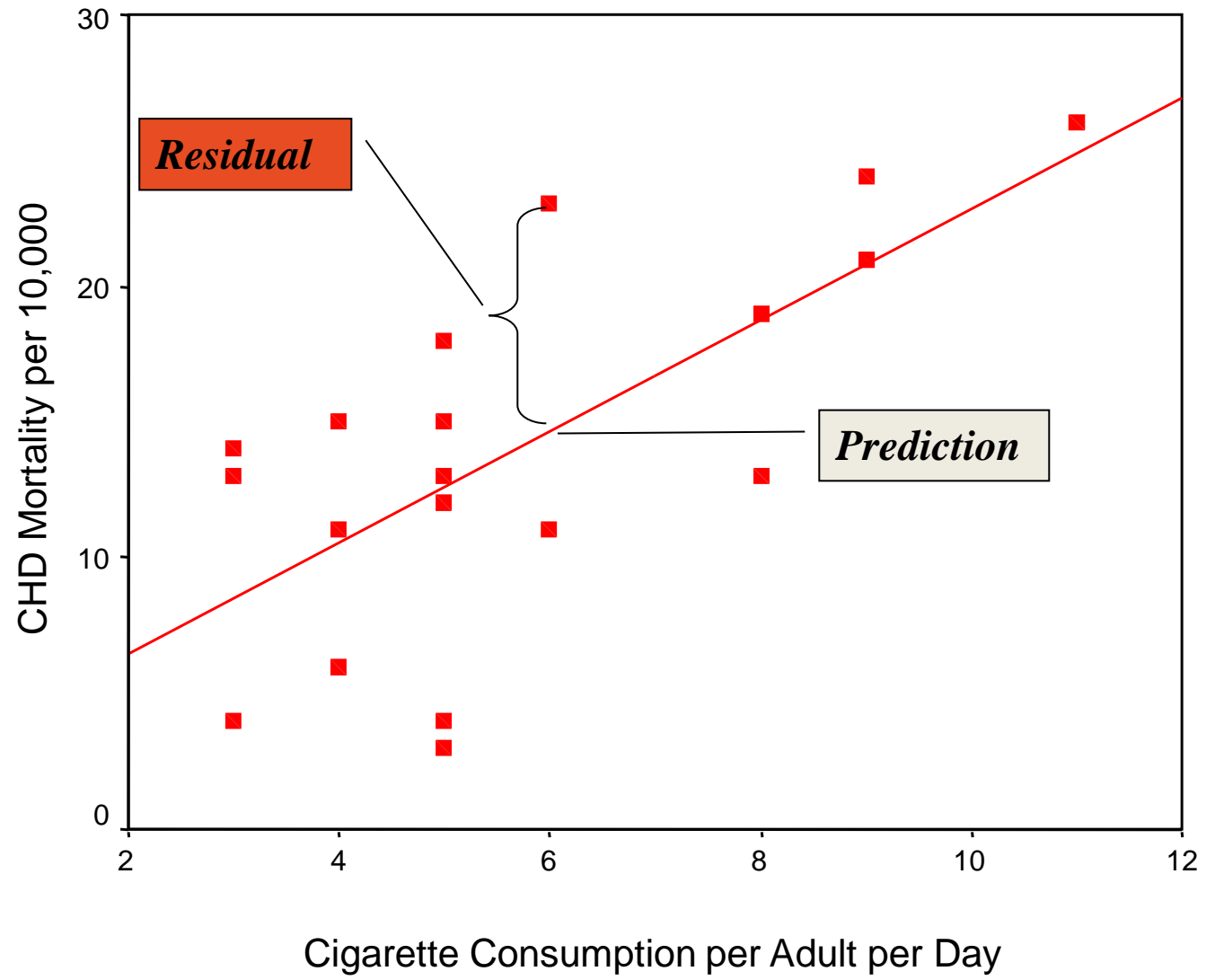
Akurasi hasil prediksi

- Saat perokok dengan 6 C/A/D
- diprediksi

$$\hat{Y} = b_1X + b_0 = 2.042X + 2.367$$

$$\hat{Y} = 2.042 * 6 + 2.367 = 14.619$$

- Dan data actual menunjukkan 23 kematian/10.000
- Error (“residual”) = $23 - 14.619 = 8.38$
 - Ini kategori error yang besar



Koefisien Korelasi

Tugas – gunakan excel.

Tugas no 1
Tentukan koefisien
korelasi - r

x	y	xy	x^2	y^2
1	-3			
2	-1			
3	0			
4	1			
5	2			
7	2.3			
8	4.9			
9	5.2			
10	5.8			
11	6.8			
12	7.1			
13	7.9			
14	8.4			
15	9.3			

Tugas No 2

Cari Data riil (dengan syarat jumlah min data 30): dapat melalui

1. Searching di internet
2. Hasil penelitian – peneliti lain
3. Hasil rekaman data saat melakukan pratikum
4. dll

Berdasarkan data:

- a. Tulsiskan dalam bentuk Tabel yang standard
- b. Tentukan persamaan regresi linier
- c. Tentukan koefisien korelasi

Tugas di upload di My Classroom masing-masing
Paling lambat, Rabu 18 Nop. 2020 – jam 24.00

Terimakasih