


STATISTIK LANJUT

Pertemuan 1 sd 2

Oleh ; B. Wisnu Widagdo S.T, M.Sc.IT

POKOK BAHASAN

- 1. Analisis Korelasi**
 - 2. Analisis Regresi Linear Sederhana**
 - 3. Asumsi Asumsi Dalam regresi linear sederhana**
 - 4. Inferensi dalam analisis regresi linear sederhana**
 - 5. Pendekatan Matriks terhadap Analisis Regresi Linear Sederhana**
 - 6. Analisis Regresi Linear Ganda**
 - 7. Asumsi-asumsi dalam Analisis Regresi Linear Ganda**
 - 8. Jumlah Kuadrat Ekstra**
 - 9. Seleksi Model**
 - 10. Analisis Variansi**
 - 11. Asumsi-asumsi dalam Analisis Variansi**
- 

ANALISIS KORELASI

Analisis korelasi adalah analisis statistika yang membahas tentang derajat (kekuatan) hubungan antara peubah-peubah.

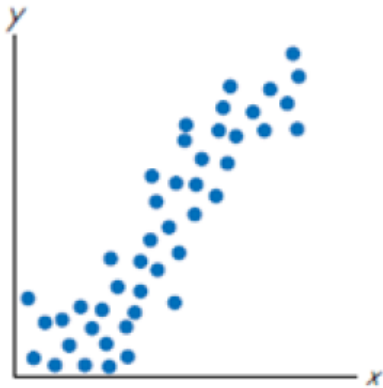
Koefisien korelasi linear mengukur kekuatan hubungan linear antara peubah X dan Y. Koefisien korelasi linear seringkali disebut juga dengan koefisien korelasi Pearson (ditemukan oleh Karl Pearson pada tahun 1857-1936).

Rumus koefisien korelasi linear populasi

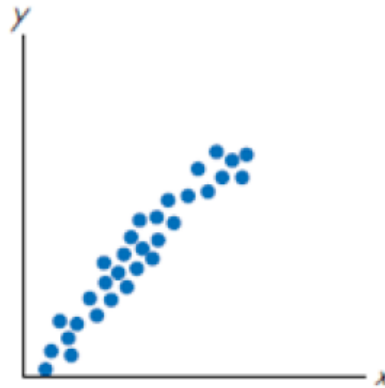
$$\rho = \frac{N \sum_{i=1}^N X_i Y_i - \sum_{i=1}^N X_i \sum_{i=1}^N Y_i}{\sqrt{N \sum_{i=1}^N X_i^2 - (\sum_{i=1}^N X_i)^2} \sqrt{N \sum_{i=1}^N Y_i^2 - (\sum_{i=1}^N Y_i)^2}}$$

Rumus koefisien korelasi linear sampel

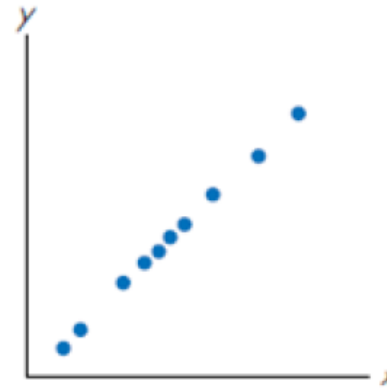
$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}}$$



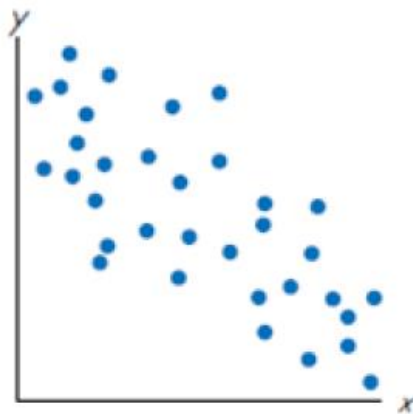
(a) Korelasi positif antara X dan Y



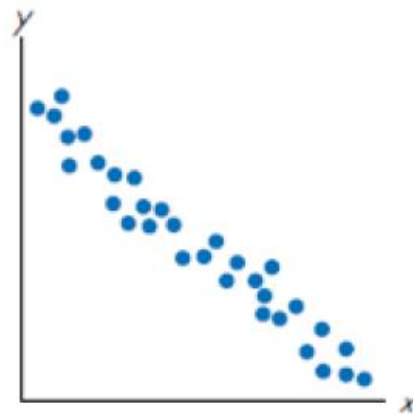
(b) Korelasi positif yang kuat antara X dan Y



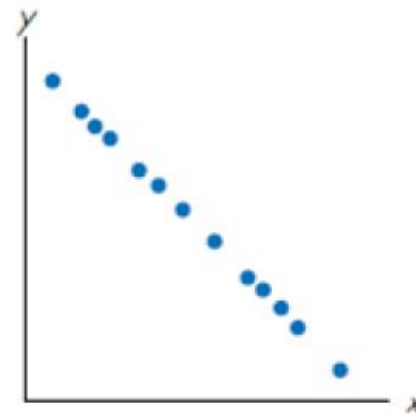
(c) Korelasi positif sempurna antara X dan Y



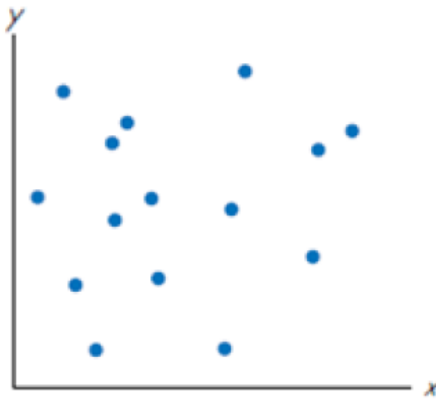
(d) Korelasi negatif antara X dan Y



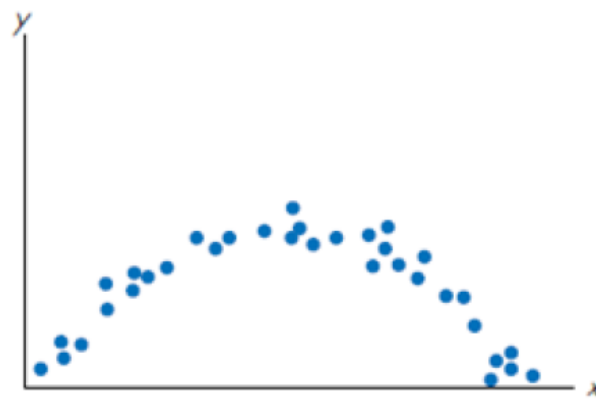
(e) Korelasi negatif yang kuat antara X dan Y



(f) Korelasi negatif sempurna antara X dan Y



(g) Tidak ada korelasi antara X dan Y



(h) Hubungan nonlinear antara X dan Y

Koefisien Determinasi bagi sampel (r^2)

Nilai r^2 menyatakan persentase keragaman Y yang dapat dijelaskan oleh hubungan linear antara X dan Y.

Contoh 1:

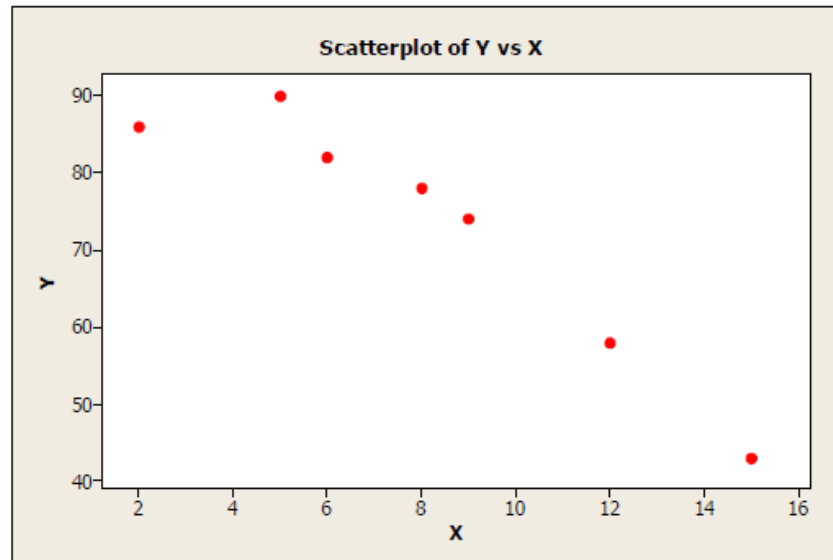
Data berikut adalah tentang banyaknya keketidakhadiran dan nilai akhir dari tujuh mahasiswa yang dipilih secara acak dari suatu kelas Statistika.

Mahasiswa	A	B	C	D	E	F	G
Banyaknya ketidakhadiran (X)	6	2	15	9	12	5	8
Nilai Akhir (Y)	82	86	43	74	58	90	78

- Buatlah diagram pencar dari data tersebut.
- Tentukan koefisien korelasi dan maknanya.
- Tentukan koefisien determinasi dan maknanya.

Penyelesaian:

- a) Diagram pencar bagi X dan Y, terlihat bahwa titik-titik data mengikuti arah garis lurus.



- b) Koefisien korelasi $r = -0,944$ artinya ada korelasi negatif yang kuat antara banyaknya ketidakhadiran dan nilai akhir, semakin banyak ketidakhadiran maka semakin menurun nilai akhirnya
- c) Koefisien determinasi $r^2 = 0,891$, artinya sebesar 89,1% keragaman nilai akhir yang dapat dijelaskan oleh hubungan linear antara banyaknya ketidakhadiran dan nilai akhir.

Pengujian Korelasi Populasi

Nilai koefisien korelasi antara -1 dan $+1$. Bila nilai r dekat $+1$ atau -1 maka ada hubungan linear yang kuat. Bila nilai r dekat 0 maka hubungan linear itu lemah. Bila r samadengan 0 maka tidak ada hubungan linear antara dua peubah tersebut.



Analisis Regresi Linear Sederhana

Analisis regresi adalah analisis statistika yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah dapat diramalkan dari peubah lainnya.

Model Regresi Linear Sederhana

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

dengan


Y_i adalah nilai peubah tak bebas dalam pengamatan ke- i

β_0 dan β_1 adalah parameter

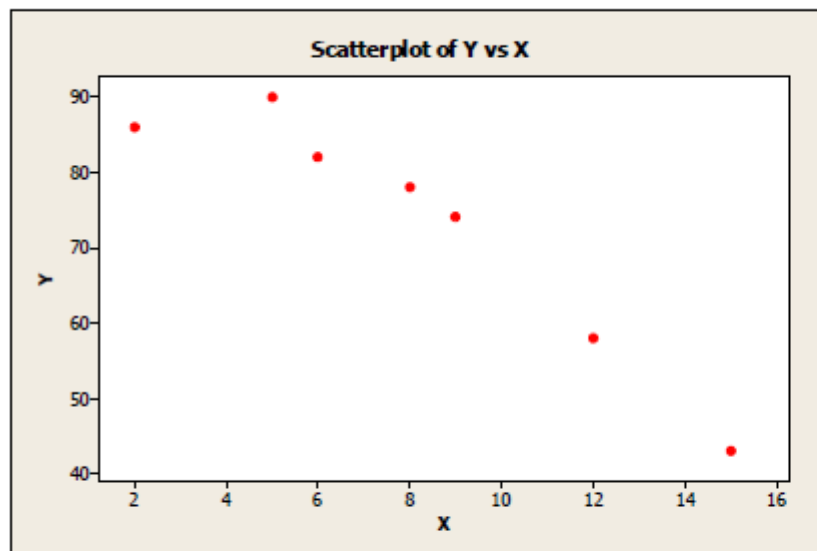
X_i adalah konstanta yang diketahui, yaitu nilai peubah bebas dari pengamatan ke- i

ε_i adalah galat yang bersifat acak dengan rata-rata $E[\varepsilon_i]=0$ dan ragam $\text{Var} [\varepsilon_i]=\sigma^2$; ε_i dan ε_j tidak berkorelasi sehingga peragam/kovariansi $\sigma \{ \varepsilon_i, \varepsilon_j \} = 0$ untuk semua i, j ; $i \neq j$

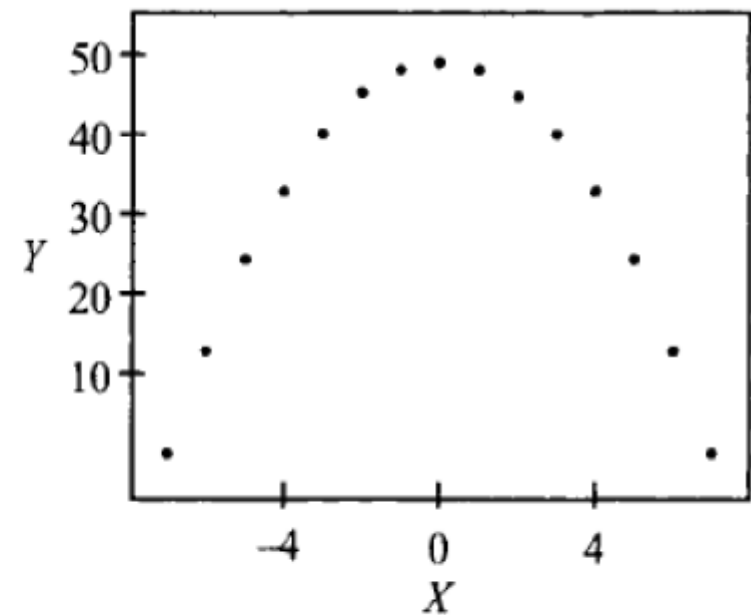
Model regresi linear sederhana:

- Dikatakan “ **sederhana**” karena hanya ada satu peubah bebas.
 - Dikatakan “**linear dalam parameter**” karena tidak ada parameter yang muncul sebagai suatu eksponen atau dikalikan atau dibagi oleh parameter lain.
 - Dikatakan “**linear dalam peubah bebas**” karena peubah dalam model tersebut berpangkat satu.
 - Model yang linear dalam parameter dan linear dalam peubah bebas juga dinamakan **model ordo-pertama**.
- 

Bila sudah diperoleh data sampel (X_i, Y_i) , selanjutnya hal yang penting adalah membuat **diagram pencar** antara X dan Y untuk mengetahui pola dari data. Bila pola data menunjukkan linear maka model regresi linear sederhana dapat digunakan. Perhatikan gambar berikut.



(a)



(b)

Makna dugaan koefisien regresi

Misalkan ingin mengetahui hubungan jarak tempuh kendaraan mobil dalam km (X) dengan tingkat emisinya dalam ppm (Y).

- Plot data ternyata menunjukkan ada hubungan linear antara X dan Y
- Dicobakan model linear $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, diperoleh persamaan regresi $\hat{Y}_i = 364 + 5,47 X_i$.
- Apa makna b_0 dan b_1 pada konteks ini ?

Makna dari b_1 yaitu rata-rata emisi meningkat 5,47 ppm untuk setiap kenaikan jarak tempuh kendaraan mobil 1 km (atau kenaikan jarak tempuh kendaraan mobil 1 km akan meningkatkan rata-rata emisi yang dihasilkan mobil sebesar 5,47 ppm).

Makna dari b_0 yaitu untuk mobil dengan jarak tempuh kendaraan mobil 0 km (mobil baru) maka rata-rata tingkat emisi yang dihasilkan sebesar 364 ppm.

b_0 tidak selalu bermakna

Tugas dan Latihan pertemuan 1 dan 2

1 . Diketahui data sebagai berikut :

Sistolik	138	130	135	140	120	125	120	130	130	144	143	140	130	150
Diastolik	82	91	100	100	80	90	80	80	80	98	105	85	70	100

- Buatlah diagram pencar dari data tersebut.
- Tentukan koefisien korelasi dan maknanya.
- Tentukan koefisien determinasi dan maknanya.