



STATISTICS

Hanung N. Prasetyo

Week 11

REGRESI DAN KORELASI



PENDAHULUAN

Regresi dan korelasi digunakan untuk mengetahui hubungan dua atau lebih kejadian (variabel) yang dapat diukur secara matematis.

Ada dua hal yang diukur atau dianalisis, yaitu :

1. Hubungan fungsional (persamaan matematis)
2. Kekuatan atau keeratan hubungan

ANALISIS REGRESI

PENGERTIAN

- Jenis uji statistika yang dipakai untuk melihat daya prediksi variabel independen (prediktor) terhadap variabel dependen (kriterium)

JENIS ANALISIS REGRESI

- Regresi Linier. Memprediksi peranan prediktor dalam persamaan linier
- Regresi Non Linier. Memprediksi peranan prediktor dalam persamaan non-linier yang dibuat oleh peneliti sendiri

PRASYARAT ANALISIS REGRESI

- Variabel dependen terdistribusi normal
- Korelasi antar prediktor yang rendah (tidak ada multikolinieritas)
- Hubungan antara prediktor dan kriterium adalah linier
- Homokedastisitas
- Jika data prediktor bersifat kualitatif (jender, agama, dsb) maka perlu ditransformasi menjadi variabel dummy

Tujuan Regresi

1. Mengestimasi nilai rata-rata variabel tak bebas dan nilai rata-rata variabel bebas
2. Menguji hipotesis mengenai sifat alamiah ketergantungan
3. Memprediksi atau meramalkan nilai rata-rata variabel tak bebas dan nilai rata-rata variabel bebas tertentu



Analisis Regresi

- Analisis regresi pada dasarnya merupakan kajian yang bertujuan untuk menemukan sampai seberapa besar pengaruh perubahan variabel independen terhadap variabel dependen.
- Analisis regresi linier sederhana berkaitan dengan kajian untuk mengetahui pengaruh satu variabel independen terhadap satu variabel independen
- Besarnya pengaruh variabel independen terhadap variabel dependen dapat dilihat melalui koefisien regresinya



Perbedaan dengan korelasi

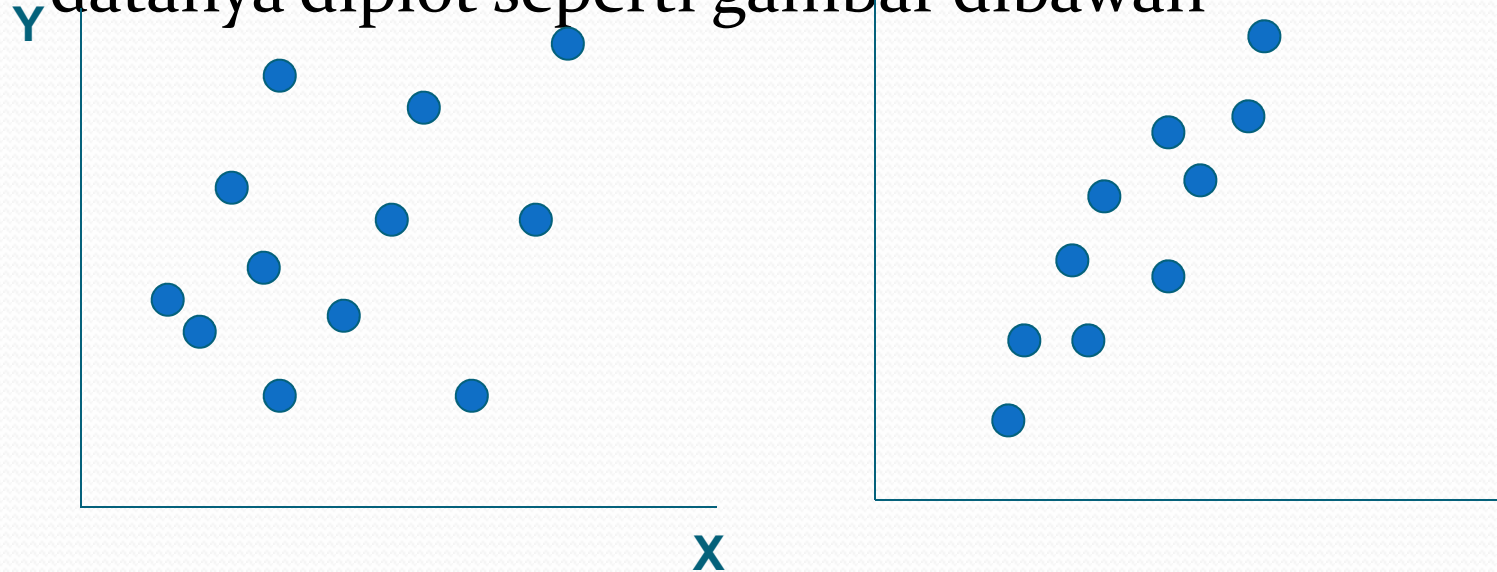
Perbedaan dengan korelasi

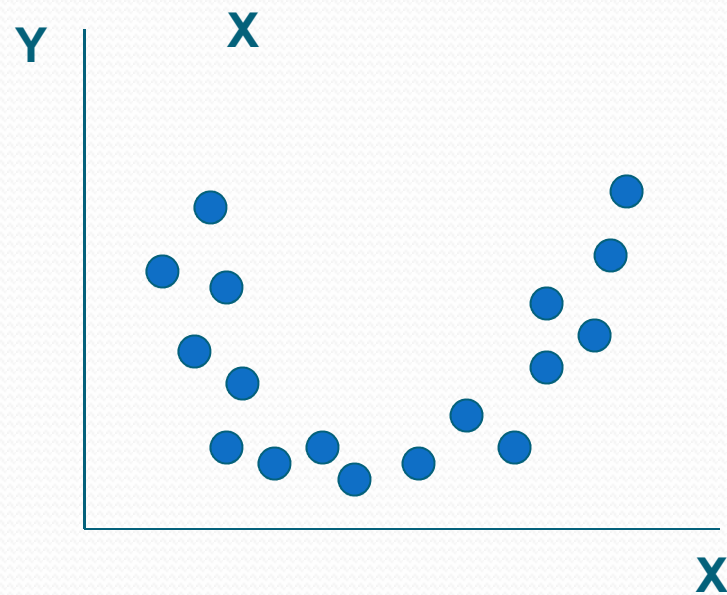
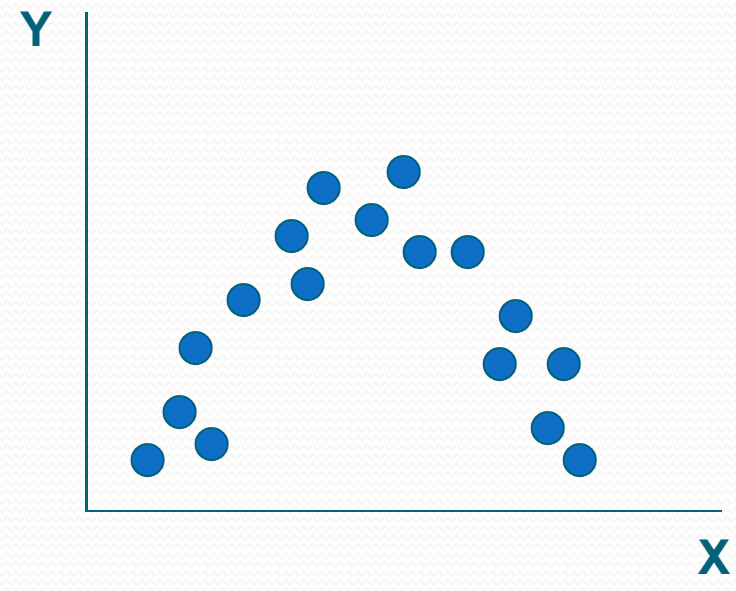
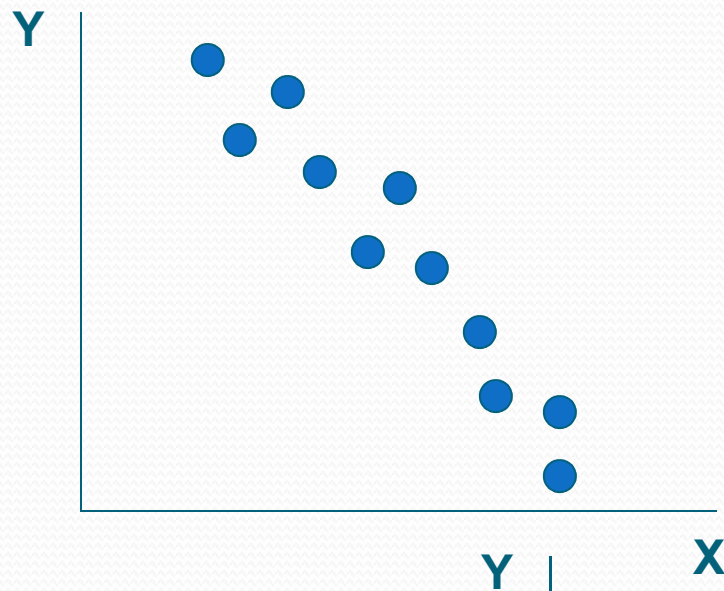
Korelasi : mengukur kekuatan atau tingkat hubungan antara dua variabel (*simple correlation*) dan tiga variabel (*multiple correlation*)

Dalam analisis regresi, ada asimetris antara variabel tak bebas dan variabel bebas. variabel tak bebas bersifat acak atau stokastik dimana variabel bebas diasumsikan mempunyai nilai yang tetap dalam pengambilan sampel berulang

Dalam Korelasi, ada simetris variabel tak bebas dan variabel bebas.

Jika terdapat 2 variabel, misalkan X dan Y yang data-datanya diplot seperti gambar dibawah





Definisi Pengaruh

Maka plot data yang membentuk suatu pola tertentu menunjukkan bahwa variabel X dan Y membentuk suatu hubungan

$X \longleftrightarrow Y$ hubungan

$X \longrightarrow Y$ pengaruh

Regresi Linier Y Terhadap X

Jika pola yang membentuk hubungan X dan Y membentuk suatu garis lurus, maka disebut **Pengaruh Linier**

Dimana :

variabel X \rightarrow variabel bebas (independent)

variabel Y \rightarrow variabel terikat (dependent)

Nilai-nilai Y ditentukan oleh nilai-nilai X

Variabel Y dipengaruhi oleh variabel X

Variabel X mempengaruhi variabel Y



Perbedaan dengan korelasi

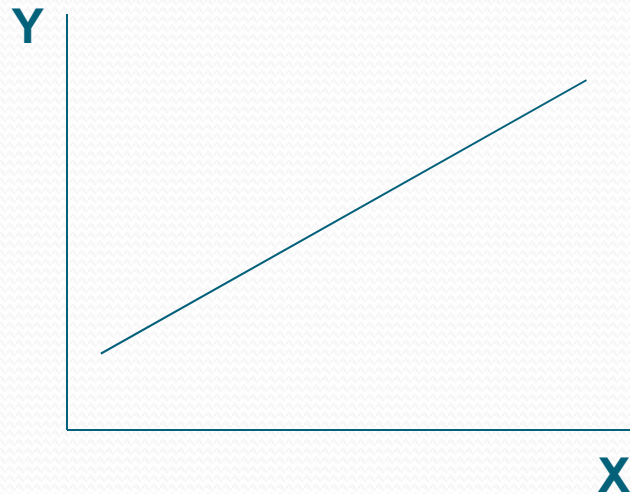
Korelasi : mengukur kekuatan atau tingkat hubungan antara dua variabel (*simple correlation*) dan tiga variabel (*multiple correlation*)

Dalam analisis regresi, ada asimetris antara variabel tak bebas dan variabel bebas. variabel tak bebas bersifat acak atau stokastik dimana variabel bebas diasumsikan mempunyai nilai yang tetap dalam pengambilan sampel berulang

Dalam Korelasi, ada simetris variabel tak bebas dan variabel bebas.

Regresi Linier Y Terhadap X

- Plot antara X dan Y



Garis lurus tersebut membentuk persamaan :

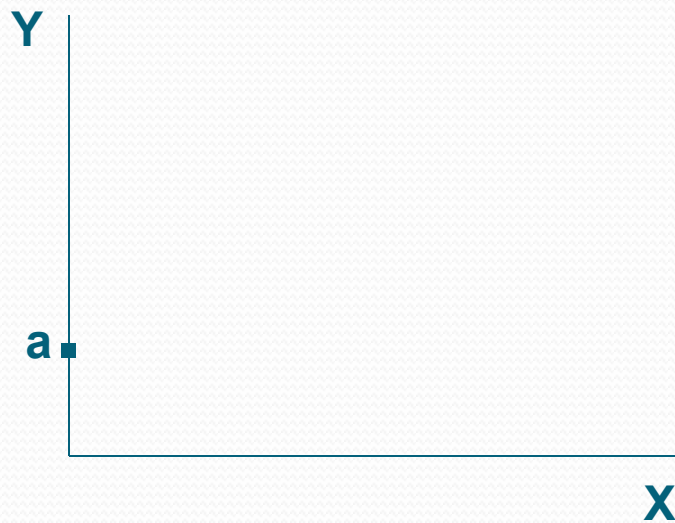
$$Y = a + bX$$

a disebut intersep

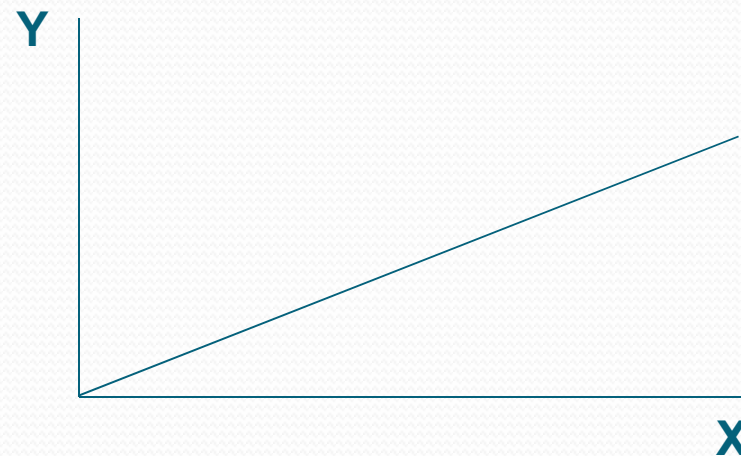
b disebut slope

Intersep

Bila $X = 0$ maka $Y = a$



Bila $a = 0$ maka garis akan melalui titik $(0,0)$

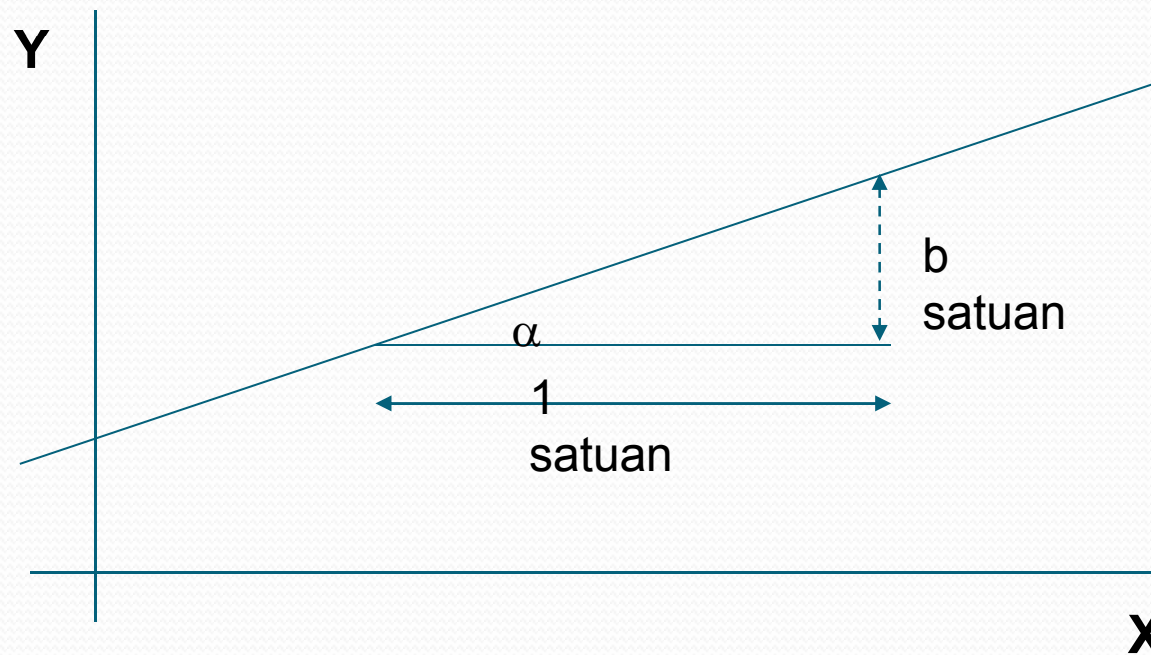


Slope

Slope = kemiringan

$$Y = a + bX$$

Perubahan 1 satuan pada X mengakibatkan perubahan b satuan pada Y, sehingga Y mengukur kemiringan/slope garis tersebut.





Slope

Bila b positif

Bertambahnya nilai X mengakibatkan bertambahnya nilai Y

Bila b negatif

Bertambahnya nilai X mengakibatkan berkurangnya nilai Y

Regresi Linier Sederhana

Model regresi linier yang hanya melibatkan satu variabel bebas (X). Model regresinya sbb:

$$Y = \alpha + \beta X$$

Dimana :

| | |
|-----------------|---------------------|
| Y | = variabel terikat |
| X | = variable bebas |
| α, β | = parameter regresi |

Model Regresi Sederhana

- Variabel Y disebut dengan variabel dependen. Variabel ini disebut pula dengan beberapa nama yang serupa seperti variabel terikat, variabel *regressand*, dan variabel endogen. Perilaku atau variasi dari variabel Y akan dijelaskan oleh model/fungsi regresi sederhana.
- notasi X mewakili variabel independen. Nama yang ekuivalen untuk X adalah variabel bebas, variabel *regressor*, dan variabel eksogen. Secara keseluruhan, variabel X akan menjelaskan variasi dalam variabel Y
- Notasi a dan b keduanya disebut dengan parameter model. Secara lebih spesifik, a adalah konstanta dan b adalah koefisien regresi.

Metode Pendugaan Parameter Regresi

- Nilai dugaan a dan b diperoleh dari proses Metode Kuadrat terkecil sbb :

1. Dilakukan turunan pertama terhadap a dan b

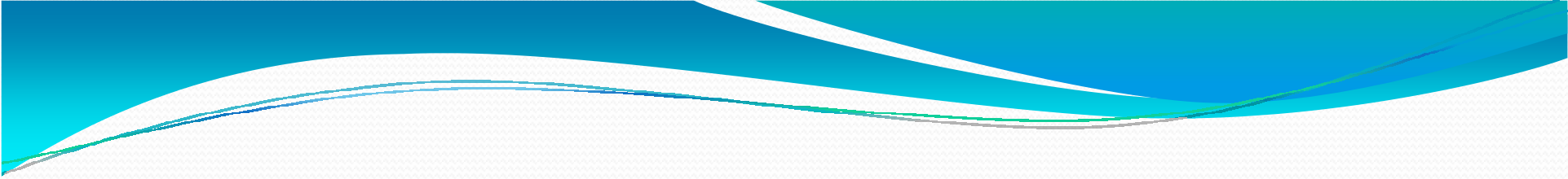
$$\frac{\partial(\sum e_i^2)}{\partial a} = -2 \sum_{i=1}^n (Y_i - a - bX_i)$$

$$\frac{\partial(\sum e_i^2)}{\partial b} = -2 \sum_{i=1}^n (Y_i - a - bX_i)X_i$$

2. Kedua persamaan hasil penurunan disamakan dengan nol

$$na + b \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i$$

$$a \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$


$$b = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}$$
$$a = \bar{Y} - b\bar{X}$$

Atau bila di rangkum

Persamaan garis regresi : $\hat{Y} = a + b X$

Dimana :

$$a = \frac{\Sigma Y \Sigma X^2 - \Sigma X \Sigma XY}{n \Sigma X^2 - (\Sigma X)^2}$$

$$b = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2}$$

Atau :

$$b = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2}$$

$$a = \frac{\Sigma Y}{n} - b \left(\frac{\Sigma X}{n} \right)$$

Persamaan Regresi Sederhana

Contoh 1

- Manajemen musik ingin mengetahui hubungan fungsional antara nilai penjualan album (Y) dengan biaya promosi (X) yang diyakini bahwa biaya biaya promosi dapat mempengaruhi nilai penjualan. Buatlah persamaan regresi linear yang menunjukkan hubungan antara nilai penjualan album (Y) dengan biaya promosi (X) menggunakan data berikut dengan metode kuadrat terkecil

Data nilai penjualan dan biaya promosi

| Nilai penjualan (Y) | Biaya promosi (X) |
|---------------------|-------------------|
| 64 | 20 |
| 61 | 16 |
| 84 | 34 |
| 70 | 23 |
| 88 | 27 |
| 92 | 32 |
| 72 | 18 |
| 77 | 22 |

Penyelesaiannya

| Y | X | XY | X ² |
|-----------|-----------|-------------|----------------|
| 64 | 20 | 1280 | 400 |
| 61 | 16 | 976 | 256 |
| 84 | 34 | 2856 | 1156 |
| 70 | 23 | 1610 | 529 |
| 88 | 27 | 2376 | 729 |
| 92 | 32 | 2944 | 1024 |
| 72 | 18 | 1296 | 324 |
| <u>77</u> | <u>22</u> | <u>1694</u> | <u>484</u> |
| 608 | 192 | 15032 | 4902 |

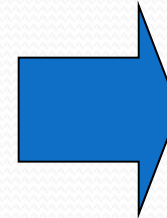
Mencari nilai a dan b

$$a = \bar{Y} - b \bar{X}$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{608}{8} = 76$$

$$\bar{X} = \frac{\sum X}{n} = \frac{192}{8} = 24$$

$$a = 76 - 1,5 (24) = 40$$

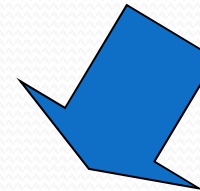


$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$= \frac{8(15.032) - (192)(608)}{8(4.902) - (192)^2}$$

$$= \frac{3520}{2352}$$

$$= 1,4965 = 1,5$$



$$\hat{Y} = 40 + 1,5X$$

KESALAHAN BAKU

Persamaan regresi mempunyai total kuadrat error sebesar : $\Sigma e^2 = \Sigma \left(Y - \hat{Y} \right)^2$

Maka kesalahan bakunya :

$$S_{\hat{Y}.X} = \sqrt{\frac{\Sigma \left(Y - \hat{Y} \right)^2}{n}} \quad \text{atau} \quad S_{\hat{Y}.X} = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n}}$$

Contoh 2

Tabel berikut menunjukkan tinggi badan (in) dan berat badan (lb) dari 12 mahasiswa.

| | | | | | | | | | | | | |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Tinggi Badan (X) | 70 | 63 | 72 | 60 | 66 | 70 | 74 | 65 | 62 | 67 | 65 | 68 |
| Berat Badan (Y) | 155 | 150 | 180 | 135 | 156 | 168 | 178 | 160 | 132 | 145 | 139 | 152 |

a. Tentukan persamaan regresi dari data tersebut!

b. Hitung kesalahan baku penaksiran \hat{Y}

JAWAB

| | | | | | | | | | | | | | |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| Tinggi badan (X) | 70 | 63 | 72 | 60 | 66 | 70 | 74 | 65 | 62 | 67 | 65 | 68 | 802 |
| Berat badan (Y) | 155 | 150 | 180 | 135 | 156 | 168 | 178 | 160 | 132 | 145 | 139 | 152 | 1850 |
| X ² | 4900 | 3969 | 5184 | 3600 | 4356 | 4900 | 5476 | 4225 | 3844 | 4489 | 4225 | 4624 | 53792 |
| Y ² | 24025 | 22500 | 32400 | 18225 | 24336 | 28224 | 31684 | 25600 | 17424 | 21025 | 19321 | 23104 | 287868 |
| XY | 10850 | 9450 | 12960 | 8100 | 10296 | 11760 | 13172 | 10400 | 8184 | 9715 | 9035 | 10336 | 124258 |

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{12(124258) - (802)(1850)}{12(53792) - (802)^2} = 3,22$$

$$a = \frac{\sum Y}{n} - b \left(\frac{\sum X}{n} \right) = \frac{1850}{12} - (3,22) \left(\frac{802}{12} \right) = -61,04$$

Jadi persamaan regresinya $\hat{Y} = -61,04 + 3,22 X$

JAWAB (lanjutan)

| | | | | | | | | | | | | | |
|-------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| Tinggi badan (X) | 70 | 63 | 72 | 60 | 66 | 70 | 74 | 65 | 62 | 67 | 65 | 68 | 802 |
| Berat badan (Y) | 155 | 150 | 180 | 135 | 156 | 168 | 178 | 160 | 132 | 145 | 139 | 152 | 1850 |
| X ² | 4900 | 3969 | 5184 | 3600 | 4356 | 4900 | 5476 | 4225 | 3844 | 4489 | 4225 | 4624 | 53792 |
| Y ² | 24025 | 22500 | 32400 | 18225 | 24336 | 28224 | 31684 | 25600 | 17424 | 21025 | 19321 | 23104 | 287868 |
| XY | 10850 | 9450 | 12960 | 8100 | 10296 | 11760 | 13172 | 10400 | 8184 | 9715 | 9035 | 10336 | 124258 |
| Y | 164,4 | 141,8 | 170,8 | 132,2 | 151,5 | 164,4 | 177,2 | 148,3 | 138,6 | 154,7 | 148,3 | 157,9 | |
| (Y - \hat{Y}) ² | 88,4 | 67,2 | 84,6 | 7,8 | 20,3 | 13 | 0,6 | 136,9 | 43,6 | 94,1 | 86,5 | 34,8 | 677,8 |

$$S_{\hat{Y}.X} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n}} = \sqrt{\frac{677,8}{12}} = 7,52$$

KOEFISIEN KORELASI

Variasi total adalah $\Sigma(Y - \bar{Y})^2$

Dimana $\Sigma(Y - \bar{Y})^2 = \Sigma(Y - \hat{Y})^2 + \Sigma(\hat{Y} - \bar{Y})^2$

Perbandingan antara variasi yang dijelaskan dengan variasi total adalah koefisien determinasi, yaitu :

$$r^2 = \frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2}$$

Dimana $0 < r^2 < 1$

KOEFISIEN KORELASI (lanjutan)

Jadi koefisien korelasinya adalah :

$$r = \pm \sqrt{\frac{\sum \left(\hat{Y} - \bar{Y} \right)^2}{\sum \left(Y - \bar{Y} \right)^2}}$$

Dimana $-1 < r < 1$ dan jika :

1. $r = -1$ maka disebut korelasi linear negatif
2. $r = 1$ maka disebut korelasi linear positif
3. $r = 0$ maka disebut tidak berkorelasi secara linear

KOEFISIEN KORELASI (lanjutan)

Koefisien korelasi juga dapat dinyatakan dengan rumus :

$$r = \sqrt{1 - \frac{S^2_{\hat{Y}.X}}{S^2_Y}}$$

Dimana :

$$S^2_{\hat{Y}.X} = \frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n} \quad (\text{kuadrat dari kesalahan baku})$$

$$S^2_Y = \frac{\Sigma (Y - \bar{Y})^2}{n} \quad (\text{variansi } Y)$$

KOEFISIEN KORELASI (lanjutan)

Bila hubungan antara variable X dan Y linear, maka koefisien korelasinya disebut koefisien korelasi produk momen:

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}}$$

Dimana :

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

KOEFISIEN KORELASI (lanjutan)

Koefisien korelasi produk momen juga dapat dihitung dengan rumus :

$$r = \frac{S_{XY}}{S_X S_Y}$$

Dimana :

$$S_{XY} = \frac{\sum XY}{n} \quad ; \quad S_X = \sqrt{\frac{\sum X^2}{n}} \quad ; \quad S_Y = \sqrt{\frac{\sum Y^2}{n}}$$

KOEFISIEN KORELASI (lanjutan)

Atau bentuk yang lebih sederhana :

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\{n\sum X^2 - (\sum X)^2\} \{n\sum Y^2 - (\sum Y)^2\}}}$$

Bila nilai r :

1. $0,9 < r < 1,0$ atau $-1,0 < r < -0,9$ (hubungan yang sangat kuat)
2. $0,7 < r < 0,9$ atau $-0,9 < r < -0,7$ (hubungan kuat)
3. $0,5 < r < 0,7$ atau $-0,7 < r < -0,5$ (hubungan moderat)
4. $0,3 < r < 0,5$ atau $-0,5 < r < -0,3$ (hubungan lemah)
5. $0,0 < r < 0,3$ atau $-0,3 < r < 0,0$ (hubungan yang sangat lemah)

KOEFISIEN KORELASI (lanjutan)

Bila data variabel X dan variabel Y merupakan data yang dikelompokkan dalam bentuk distribusi frekuensi, maka koefisien korelasinya dihitung dengan rumus :

$$r = \frac{n(\Sigma fXY) - (\Sigma fX)(\Sigma fY)}{\sqrt{\{n\Sigma fX^2 - (\Sigma fX)^2\} \{n\Sigma fY^2 - (\Sigma fY)^2\}}}$$

CONTOH 3

Dari contoh 2, tentukan :

- a. Koefisien korelasi r dan artinya!
- b. Koefisien determinasi r^2 dan artinya!

JAWAB

| | | | | | | | | | | | | | |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| Tinggi badan (X) | 70 | 63 | 72 | 60 | 66 | 70 | 74 | 65 | 62 | 67 | 65 | 68 | 802 |
| Berat badan (Y) | 155 | 150 | 180 | 135 | 156 | 168 | 178 | 160 | 132 | 145 | 139 | 152 | 1850 |
| X ² | 4900 | 3969 | 5184 | 3600 | 4356 | 4900 | 5476 | 4225 | 3844 | 4489 | 4225 | 4624 | 53792 |
| Y ² | 24025 | 22500 | 32400 | 18225 | 24336 | 28224 | 31684 | 25600 | 17424 | 21025 | 19321 | 23104 | 287868 |
| XY | 10850 | 9450 | 12960 | 8100 | 10296 | 11760 | 13172 | 10400 | 8184 | 9715 | 9035 | 10336 | 124258 |

$$r = \frac{12(124258) - (802)(1850)}{\sqrt{\{12(53792) - (802)^2\} \{12(287868) - (1850)^2\}}} = 0,86$$

Karena r terletak antara 0,7 dan 0,9, maka terdapat hubungan positif yang kuat antara tinggi badan dan berat badan mahasiswa.

Koefisien Determinasi(KD)

Koefisien Determinasi adalah analisis untuk memperlihatkan besar pengaruh variabel X terhadap variabel Y.

$$KD = r^2 \times 100 \%$$

Berdasarkan contoh sebelumnya dapat dihitung

Koefisien Determinasinya

$$r^2 = (0,86)^2 = 0,7396 = 73,96 \%$$

Artinya variasi berat badan yang dapat dijelaskan oleh variasi tinggi badan mahasiswa (\hat{X}) oleh persamaan regresi

$$Y = -61,04 + 3,22 X$$

adalah sebesar 73,96 %. Sisanya sebesar 26,04 % dijelaskan oleh faktor lain di luar variabel pada persamaan regresi tersebut.

LATIHAN 1

Berikut disajikan data tinggi badan ayah dan tinggi badan putra yang diperoleh dari suatu survei dengan sampel 12 orang ayah dan putra (dalam in).

| | | | | | | | | | | | | |
|--------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| Tinggi badan ayah | 65 | 63 | 67 | 64 | 68 | 62 | 70 | 66 | 68 | 67 | 69 | 71 |
| Tinggi badan putra | 68 | 66 | 68 | 65 | 69 | 66 | 68 | 65 | 71 | 67 | 68 | 70 |

- Tentukan persamaan regresinya!
- Bila tinggi seorang ayah 7,4 in, berapa kira-kira tinggi badan putranya?
- Tentukan kesalahan baku penaksir \hat{Y}
- Tentukan koefisien korelasi dan artinya!
- Tentukan koefisien determinasi dan artinya!