

PENGANTAR REGRESI SEDERHANA

PERTEMUAN 13



EGI SAFITRI, S.MAT., M.SI

INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA

INTRODUCTION

Regresi linier sering digunakan untuk melihat nilai prediksi atau perkiraan yang akan datang

Apabila X dan Y mempunyai hubungan, maka nilai X yang sudah diketahui dapat digunakan memperkirakan Y

Perkiraan mengenai terjadinya sesuatu kejadian (nilai variabel untuk waktu yang akan datang, seperti prediksi produksi 3 tahun yang akan datang, prediksi harga bulan depan, ramalan jumlah penduduk 10 tahun mendatang, ramalan hasil penjualan tahun depan).

INTRODUCTION

Ramalan mengetahui suatu kejadian baik secara kualitatif (akan turun hujan, akan terjadi perang, akan lulus ujian)

Kuantitatif (produksi padi akan mencapai 16 juta ton, indeks harga 9 bahan pokok naik 10%, penerimaan devisa turun 5%)

Melakukan peramalan adalah dengan menggunakan garis regresi

INTRODUCTION

Ramalan mengetahui suatu kejadian baik secara kualitatif (akan turun hujan, akan terjadi perang, akan lulus ujian)

Kuantitatif (produksi padi akan mencapai 16 juta ton, indeks harga 9 bahan pokok naik 10%, penerimaan devisa turun 5%)

Melakukan peramalan adalah dengan menggunakan garis regresi

KONSEP DASAR

Konsep Dasar Regresi Linier Sederhana

1. Variabel Dependan (Y): Ini adalah variabel yang ingin diprediksi atau dijelaskan. Ini juga dikenal sebagai variabel respons atau target.

2. Variabel Independen (X): Variabel ini digunakan untuk memprediksi atau menjelaskan variabel dependen. Juga dikenal sebagai variabel prediktor atau eksplanatori.

3. Model Regresi Linier Sederhana: Model ini mengasumsikan hubungan linear antara variabel independen (X) dan variabel dependen (Y), yang direpresentasikan oleh persamaan garis lurus

$$Y = a + bX + \varepsilon$$

KONSEP DASAR

$$Y = a + bX + \varepsilon$$

- Y adalah variabel dependen.
- X adalah variabel independen.
- a adalah intercept (titik potong) dari garis regresi pada sumbu Y saat $X = 0$.
- b adalah koefisien regresi yang menggambarkan seberapa banyak Y berubah ketika X berubah satu satuan.
- ε adalah kesalahan acak yang merupakan perbedaan antara nilai yang diamati dan nilai yang diprediksi oleh model.

JENIS ANALISIS REGRESI

I. Regresi linier jika hubungan antara variabel bebas terhadap variabel tak bebas berbentuk linier

- Regresi linier sederhana $\rightarrow \hat{Y} = a + bX$
- Regresi linier berganda $\rightarrow \hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3$

II. Regresi tak linier jika hubungan antara variabel bebas terhadap variabel tak berbentuk linier

- Regresi kuadratik $\rightarrow \hat{Y} = a + bX + cX^2$
 $\hat{Y} = a + bX^2$
- Regresi kubik $\rightarrow \hat{Y} = a + bX + cX^2 + dX^3$
 $\hat{Y} = a + bX^2 + cX^3$
 $\hat{Y} = a + bX^3$

LANGKAH-LANGKAH ANALISIS REGRESI LINIER SEDERHANA

1. **Penentuan Tujuan Penelitian:** Tentukan variabel dependen dan independen, serta tujuan analisis yang ingin dicapai.
2. **Kumpulan Data:** Kumpulkan data untuk variabel dependen dan independen yang relevan.
3. **Visualisasi Data:** Buatlah grafik untuk memahami hubungan kasar antara variabel independen dan dependen.
4. **Estimasi Parameter Model:** Hitung nilai koefisien regresi (a dan b) menggunakan metode seperti metode kuadrat terkecil (OLS - Ordinary Least Squares).
5. **Penilaian Model:** Evaluasi kecocokan model, misalnya, dengan mengukur keakuratan prediksi menggunakan metrik seperti koefisien determinasi (R^2).
6. **Uji Signifikansi:** Lakukan uji hipotesis untuk memastikan signifikansi koefisien regresi.
7. **Prediksi dan Interpretasi:** Gunakan model untuk memprediksi nilai variabel dependen dan interpretasikan hasilnya dalam konteks masalah yang sedang dihadapi.
8. **Validasi Model:** Validasi model menggunakan teknik seperti validasi silang (cross-validation) untuk memastikan generalisasi yang baik pada data baru.

Analisis regresi digunakan untuk memahami hubungan antara dua atau lebih variabel. Hubungan ini bisa bersifat positif atau negatif, tergantung pada arah dan kekuatan hubungan antara variabel-variabel tersebut.

HUBUNGAN SEARAH/POSITIF

Hubungan searah (positif) terjadi ketika perubahan dalam variabel independen berdampak pada perubahan yang searah dalam variabel dependen. Dengan kata lain, jika nilai variabel independen meningkat, maka nilai variabel dependen juga cenderung meningkat, atau sebaliknya, jika nilai variabel independen menurun, nilai variabel dependen cenderung menurun

Contoh :

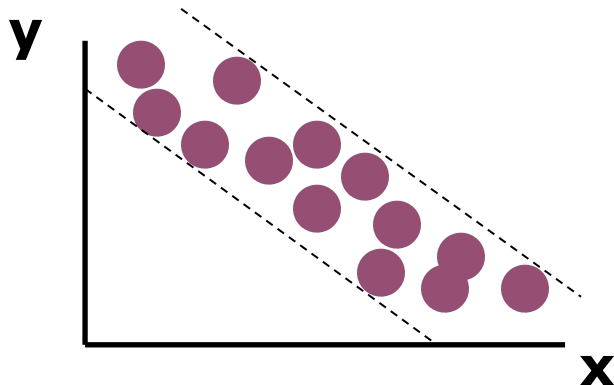
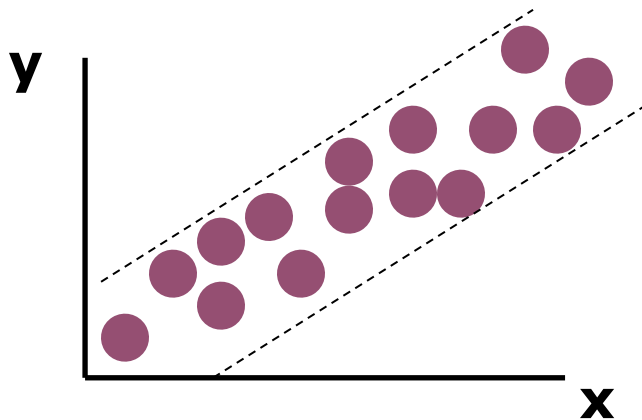
Misalkan Anda melakukan analisis regresi untuk memahami hubungan antara jumlah jam belajar (X) dan nilai ujian (Y). Jika hubungan antara jumlah jam belajar dan nilai ujian bersifat positif, maka semakin banyak waktu yang dihabiskan untuk belajar (peningkatan X), maka nilai ujian (Y) cenderung meningkat. Sebaliknya, jika jumlah jam belajar berkurang, nilai ujian juga cenderung menurun.

STRONG RELATIONSHIP

Ketika dua variabel memiliki hubungan kuat, scatter plot menunjukkan pola di mana titik-titik data cenderung membentuk pola yang jelas atau mengikuti pola garis lurus atau kurva yang ketat.

STRONG RELATIONSHIP

Strong relationships



Misalnya, jika Anda memiliki dua variabel yang saling berkorelasi positif dengan hubungan kuat, scatter plot akan menunjukkan titik-titik yang teratur dan mengikuti pola garis yang jelas ke arah atas. Setiap perubahan dalam variabel independen secara konsisten disertai dengan perubahan yang besar dalam variabel dependen.

HUBUNGAN TIDAK SEARAH/NEGATIF

Hubungan tidak searah (negatif) terjadi ketika perubahan dalam variabel independen berdampak pada perubahan yang berlawanan arah dengan variabel dependen. Dengan kata lain, ketika nilai variabel independen meningkat, nilai variabel dependen cenderung menurun, atau sebaliknya.

Contoh :

Dalam analisis regresi antara jumlah latihan fisik (X) dan berat badan (Y), jika hubungannya bersifat negatif, semakin banyak latihan fisik yang dilakukan (peningkatan X), berat badan (Y) cenderung menurun. Sebaliknya, jika jumlah latihan fisik berkurang, berat badan cenderung meningkat.

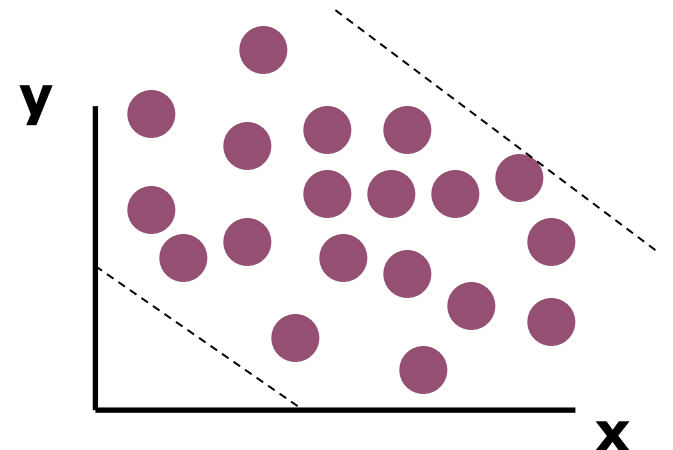
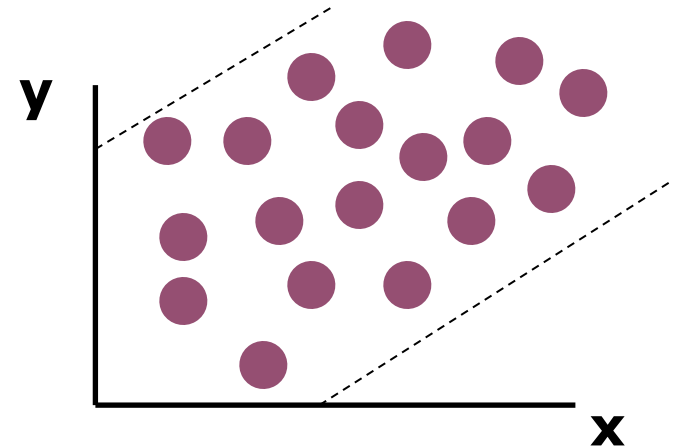
WEAK RELATIONSHIP

Hubungan lemah antara variabel dapat dilihat pada scatter plot di mana titik-titik data tersebar luas di sekitar garis tengah atau tidak mengikuti pola yang jelas. Meskipun terdapat hubungan antara kedua variabel, pola titik-titik data mungkin tidak terlalu teratur atau jelas.

WEAK RELATIONSHIP

Misalnya, pada scatter plot dengan hubungan lemah, titik-titik data tersebar secara acak tanpa pola garis atau pola yang konsisten. Perubahan dalam variabel independen mungkin tidak selalu menghasilkan perubahan yang besar atau konsisten dalam variabel dependen.

Weak relationships



NO RELATIONSHIP

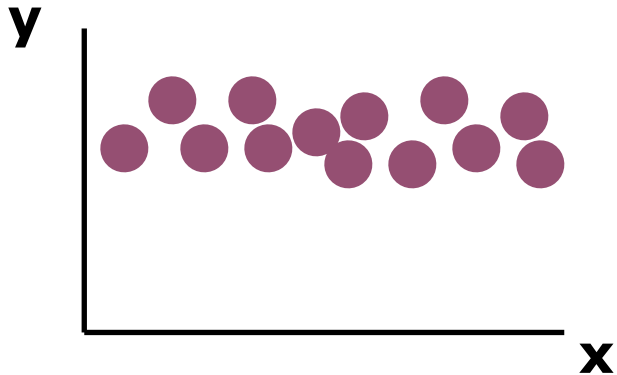
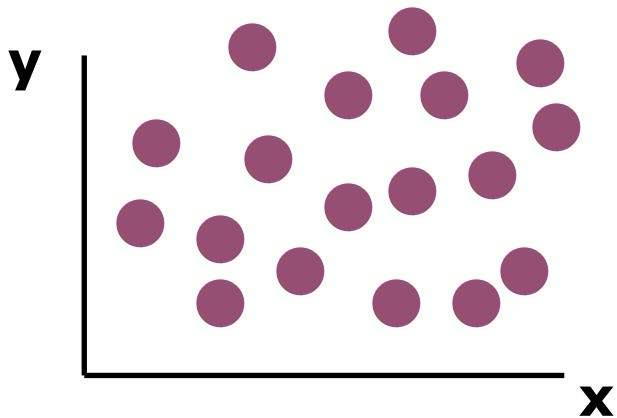
Tidak ada hubungan (no relationship) terjadi ketika tidak terdapat pola atau keterkaitan yang jelas antara variabel independen dan variabel dependen. Dalam konteks regresi linier, ini menunjukkan bahwa perubahan dalam variabel independen tidak memiliki pengaruh yang konsisten terhadap variabel dependen.

NO RELATIONSHIP

Misalnya, jika Anda menganalisis hubungan antara tinggi badan seseorang (X) dan kecerdasan (Y), dan tidak ada hubungan yang dapat diidentifikasi, ini berarti tidak ada pola yang dapat diprediksi di mana tinggi seseorang memengaruhi kecerdasan mereka secara signifikan. Dalam kasus ini, tidak ada hubungan yang dapat dijelaskan oleh regresi linier antara kedua variabel tersebut.

NO RELATIONSHIP

No relationship

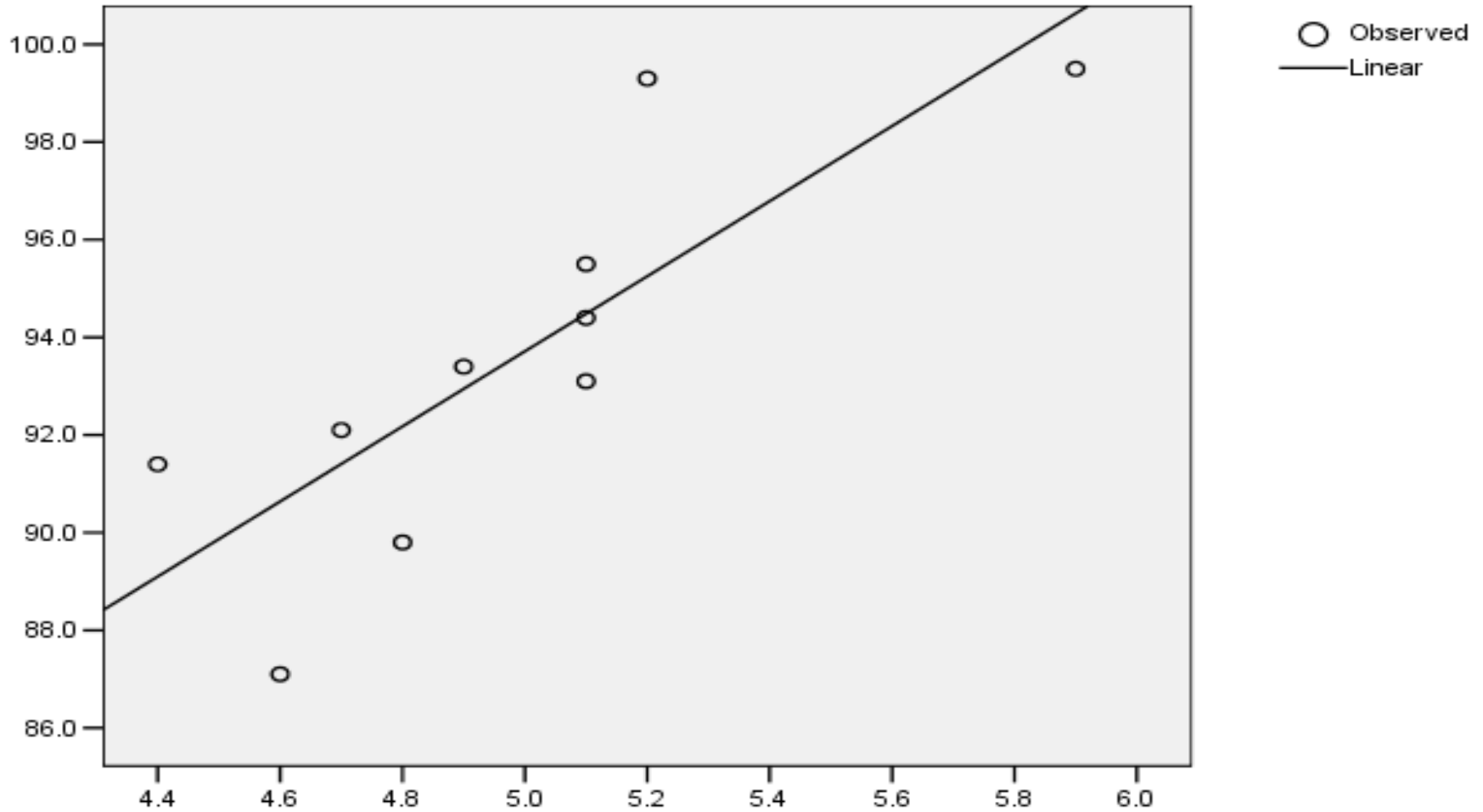


Misalnya, pada scatter plot tanpa hubungan, titik-titik data tersebar secara acak di sepanjang sumbu tanpa arah atau pola yang dapat diidentifikasi. Perubahan dalam variabel independen tidak memberikan petunjuk yang pasti tentang perubahan dalam variabel dependen.

DALAM KONTEKS REGRESI LINIER

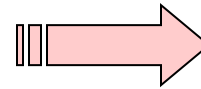
- Strong relationship atau hubungan kuat dapat terjadi baik dalam hubungan positif maupun negatif, di mana variabel independen memiliki pengaruh yang besar dan konsisten terhadap variabel dependen.
- Weak relationship atau hubungan lemah dapat terjadi baik dalam hubungan positif maupun negatif, di mana pengaruh variabel independen terhadap variabel dependen tidak begitu signifikan atau konsisten

CONTOH REGRESI LINIER SEDERHANA



PENDUGAAN TERHADAP KOEFISIEN REGRESI

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$



Metode
Kuadrat Terkecil

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

Bagaimana Pengujian terhadap model regresi ??

- parsial (per koefisien) → uji-t
- bersama → uji-F (Anova)

Bagaimana menilai kesesuaian model ??

R^2 (Koef. Determinasi: % keragaman Y yang mampu dijelaskan oleh X)

y	x	xy	x^2	y^2
.
.
Σy	Σx	Σxy	Σx^2	Σy^2

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \quad a = \bar{y} - b\bar{x} \quad \bar{y} = \frac{\sum y}{n} \quad \bar{x} = \frac{\sum x}{n}$$

PENDUGAAN TERHADAP KOEFISIEN REGRESI

Carilah persamaan regresi Y pada X dari data Tabel :

Mat (X)	Fis (Y)	XY	X ²	Y ²
60	80	4800	3600	6400
45	69	3105	2025	4761
50	71	3550	2500	5041
60	85	5100	3600	7225
50	80	4000	2500	6400
65	82	5330	4225	6724
60	89	5340	3600	7921
65	93	6045	4225	8649
50	76	3800	2500	5776
65	86	5590	4225	7396
45	71	3195	2025	5041
50	69	3450	2500	4761
665	951	53305	37525	76095

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$
$$= \frac{53305 - \frac{(665)(951)}{12}}{37525 - \frac{(665)^2}{12}} = 0.8972$$

$$a = \bar{y} - b_1 \bar{x} = 29.53$$

jadi diperoleh persamaan regresi :

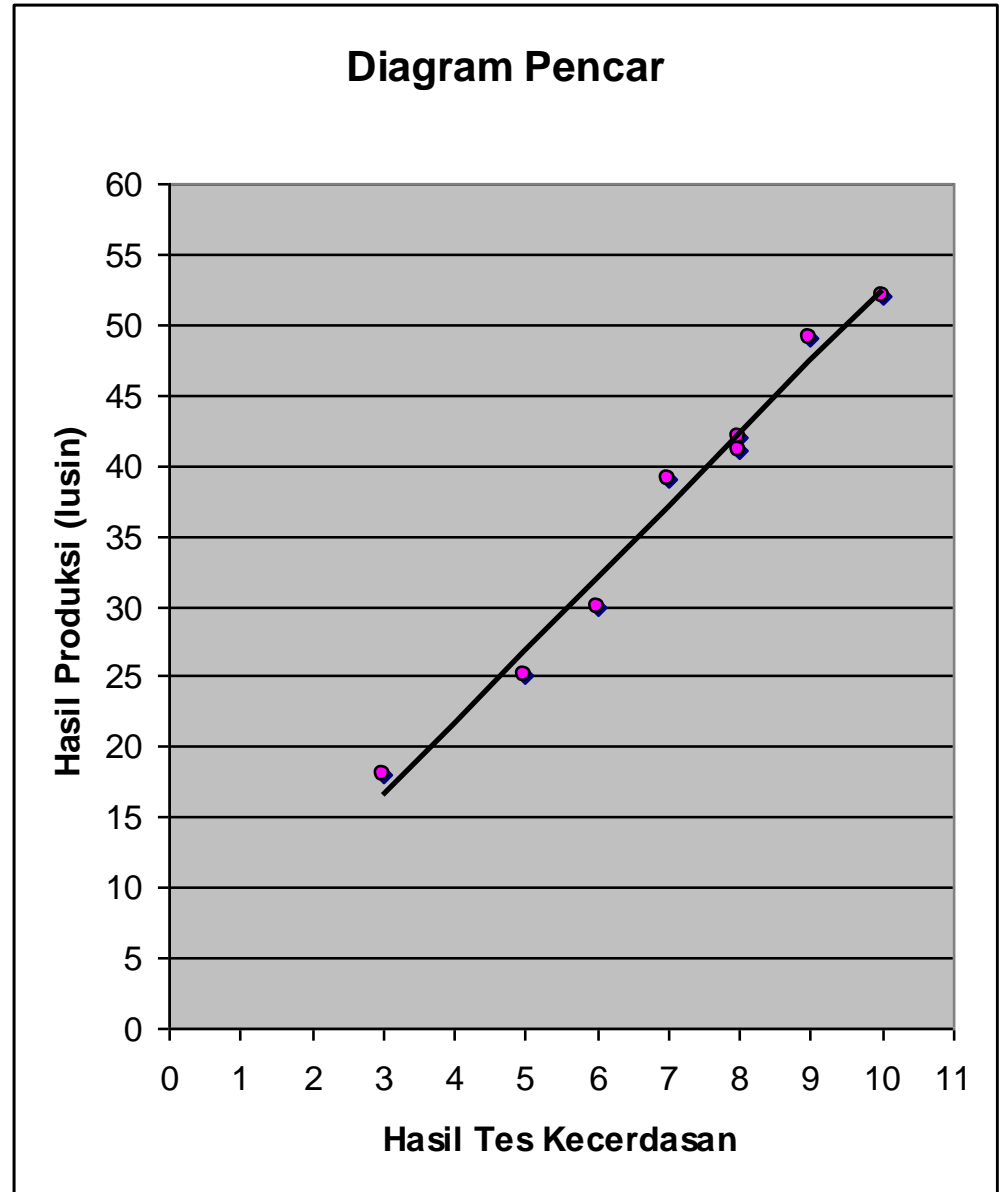
$$\hat{Y}_i = 29.5294 + 0.8972 X_i$$

CONTOH

Diketahui

Karyawan	Hasil Produksi (lusin) (Y)	Skor Tes Kecerdasan (X)
A	30	6
B	49	9
C	18	3
D	42	8
E	39	7
F	25	5
G	41	8
H	52	10

Karyawan	Hasil Produksi (lusin) (Y)	Skor Tes Kecerdasan (X)
A	30	6
B	49	9
C	18	3
D	42	8
E	39	7
F	25	5
G	41	8
H	52	10



METODE JUMLAH KUADRAT TERKECIL

- Regresi merupakan suatu alat ukur yang juga digunakan untuk mengukur ada atau tidaknya korelasi antar variabelnya.
- Istilah **regresi** itu sendiri berarti **ramalan** atau **taksiran**.
- Persamaan yang digunakan untuk mendapatkan garis regresi pada data diagram pencar disebut persamaan regresi.

- Untuk menempatkan garis regresi pada data yang diperoleh maka digunakan metode kuadrat terkecil, sehingga bentuk persamaan regresi adalah sebagai berikut:

$$Y' = a + b X$$

Dimana:

- Y' : nilai estimate variabel terikat
- a : titik potong garis regresi pd sumbu y (nilai estimate Y' bila $x=0$)
- b : gradien garis regresi (perub nilai estimate Y' per satuan perubahan nilai x)
- X : nilai variabel bebas

- Nilai dari a dan b pada persamaan regresi dapat dihitung dengan rumus berikut :

$$b = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

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

Tabel perhitungan.

Karyawan	Hasil Produksi (lusin) (Y)	Skor Tes (X)	y (Y - \bar{Y})	x (X - \bar{X})	xy	x ²	y ²
A	30	6	-7	-1	7	1	49
B	49	9	12	2	24	4	144
C	18	3	-19	-4	76	16	361
D	42	8	5	1	5	1	25
E	39	7	2	0	0	0	4
F	25	5	-12	-2	24	4	144
G	41	8	4	1	4	1	16
H	52	10	15	3	45	9	225
	296	56	0	0	185	36	968

$$\bar{Y} = \frac{\sum Y}{N} = \frac{296}{8} = 37$$

$$\bar{X} = \frac{\sum X}{N} = \frac{56}{8} = 7$$

$$b = \frac{\sum xy}{\sum x^2} = \frac{185}{36} = 5,138 \sim 5,14$$

$$a = \bar{Y} - b\bar{X} = 37 - 5,14(7) = 1,02$$

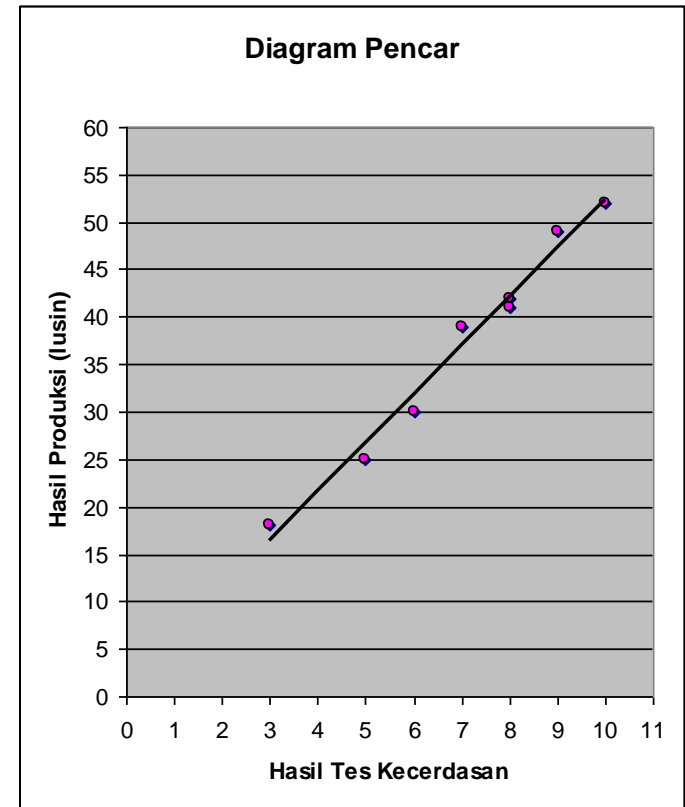
$$Y' = 1,02 + 5,14X$$

$$X = 3 \rightarrow Y' = 1,02 + 5,14(3) = 16,44$$

$$X = 5 \rightarrow Y' = 1,02 + 5,14(5) = 26,72$$

$$X = 6 \rightarrow Y' = 1,02 + 5,14(6) = 31,86$$

$$X = 10 \rightarrow Y' = 1,02 + 5,14(10) = 52,42$$



PEMBACAAN GARIS REGRESI

Contoh:

Jika garis regresi ditunjukkan dengan persamaan :

$$y^1 = 2,94 + 0,95x$$

- maka dapat diartikan bahwa : hasil penjualan akan bertambah sebesar 0,95 pada setiap kenaikan pengeluaran iklan sebanyak 1 satuan

KOEFISIEN REGRESI

- Gradien garis regresi adalah nilai b
- Nilai b positif, menunjukkan hubungan antara variabel x dan y searah atau hubungannya positif.
- Nilai b negatif, menunjukkan hubungan antara variabel x dan y berlawanan arah atau hubungannya negatif
- Besar kecilnya perubahan variabel x terhadap variabel y ditentukan besar kecilnya koefisien regresi.

KOEFISIEN DETERMINASI

- Adalah alat utama untuk mengetahui sejauh mana tingkat hubungan antara variabel x dan y.

$$r^2 = \frac{a(\sum Y) + b(\sum XY) - n(\bar{Y})^2}{\sum (Y)^2 - n(\bar{Y})^2}$$

- Nilai koefisien determinasi antara $0 \leq r^2 \leq 1$
- Nilai koefisien determinasi = 1 menunjukkan hubungan sempurna.
- Nilai koefisien determinasi = 0 menunjukkan tidak ada hubungan.
- $r^2 = 81$ artinya 81% perubahan dari variabel y ditentukan oleh variabel x.

KOEFISIEN KORELASI

- Mengukur seberapa kuat atau derajat kedekatan suatu relasi antar variabel
- Koefisien korelasi memiliki nilai $-1 \leq KK \leq +1$
- Untuk menentukan keeratan korelasi antar variabel diberikan patokan KK
- $0 < KK \leq 0,2$, korelasi sgt lemah
- $0,2 < KK \leq 0,4$, korelasi lemah tp pasti
- $0,4 < KK \leq 0,7$, korelasi yg cukup berarti
- $0,7 < KK \leq 0,9$, korelasi sgt kuat
- $0,9 < KK < 1$, korelasi kuat sekali
- $KK = 1$, korelasi sgt sempurna

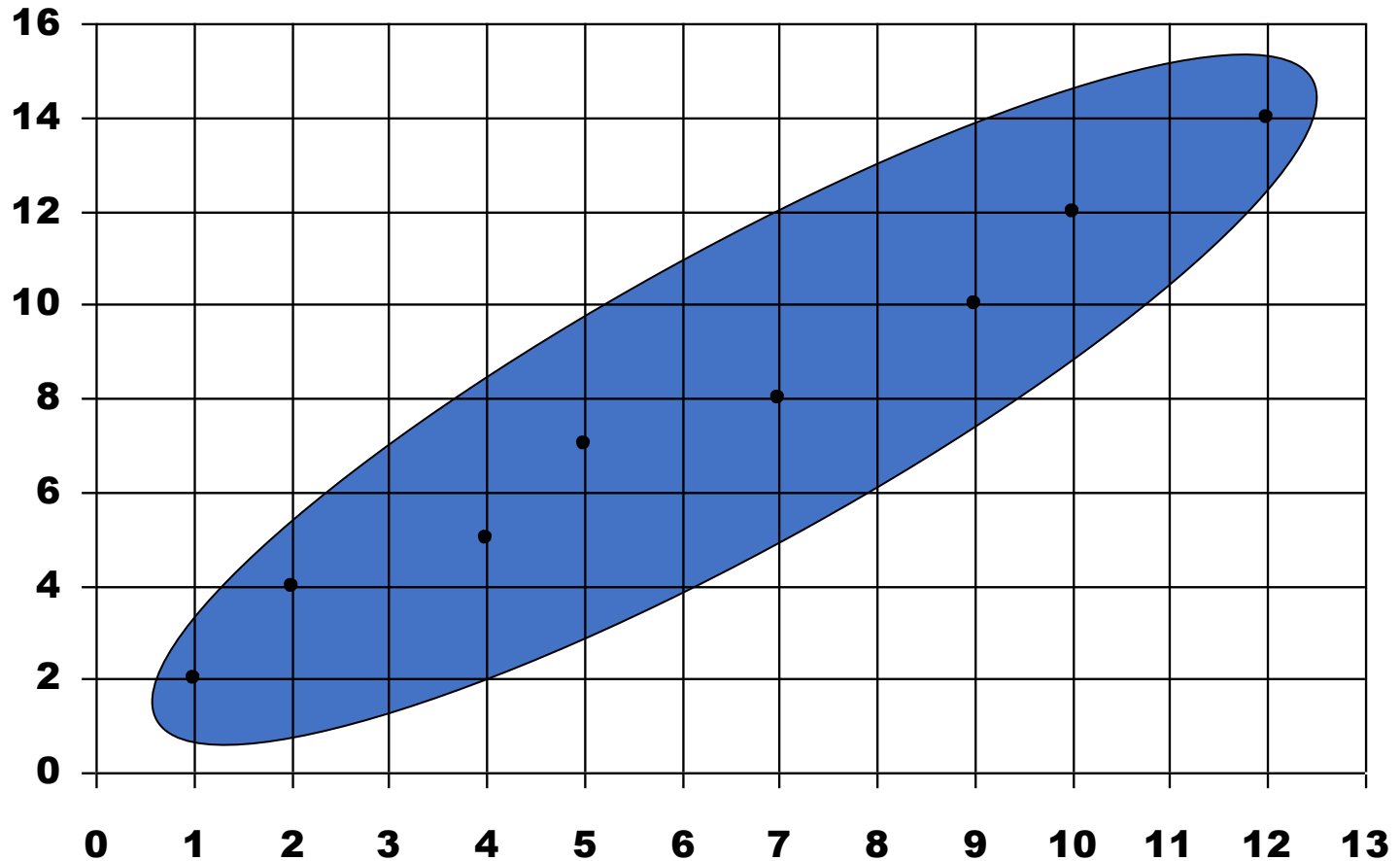
Koefisien Determinasi:
$$r^2 = \frac{a(\sum Y) + b(\sum XY) - n(\bar{Y})^2}{\sum (Y)^2 - n(\bar{Y})^2}$$

Koefisien Korelasi :
$$r = \pm \sqrt{r^2}$$

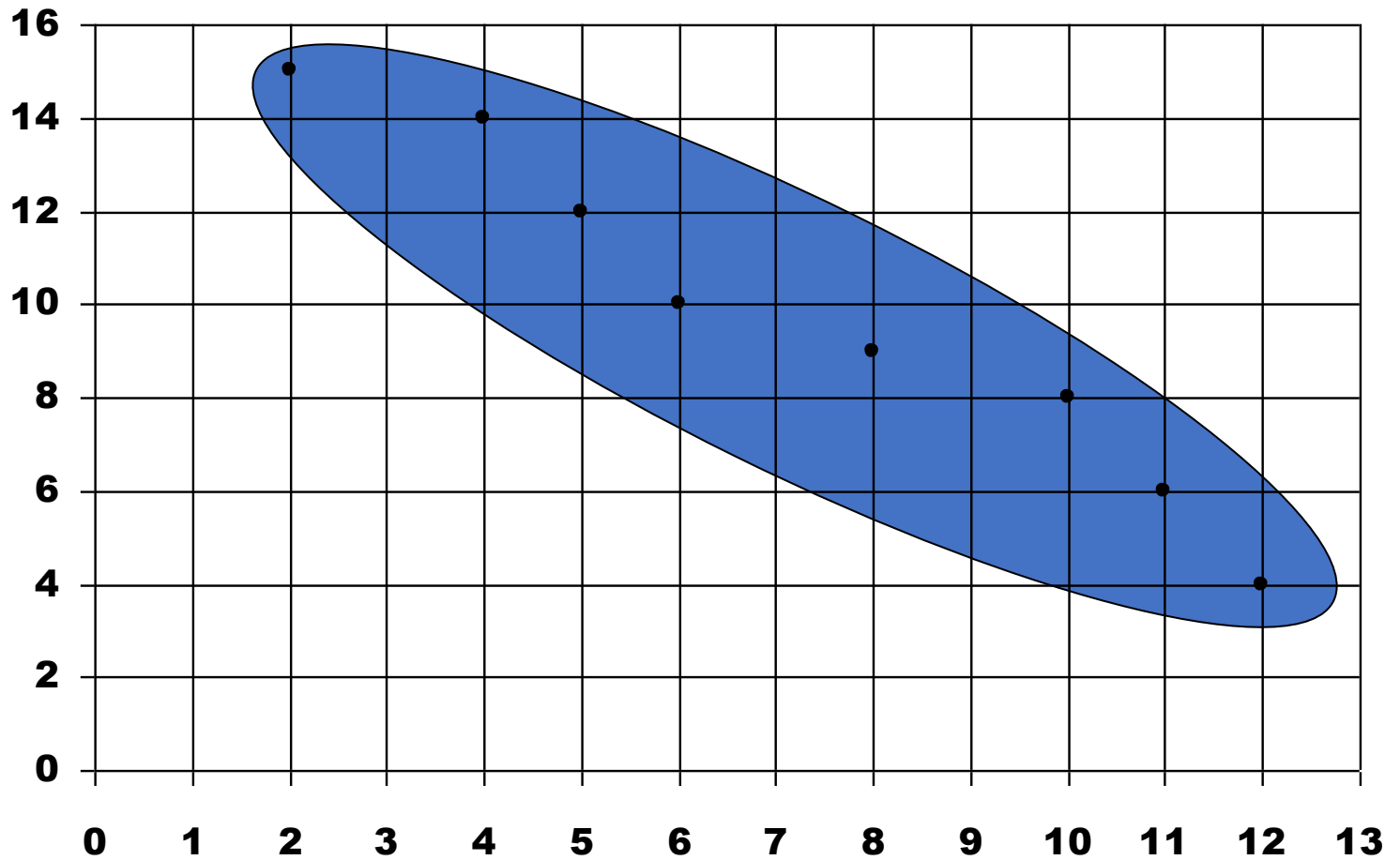
Jenis-jenis koefisien korelasi

1. Koefisien korelasi pearson
2. Koefisien korelasi rank spearman
3. Koefisien korelasi kontingensi
4. Koefisien penentu

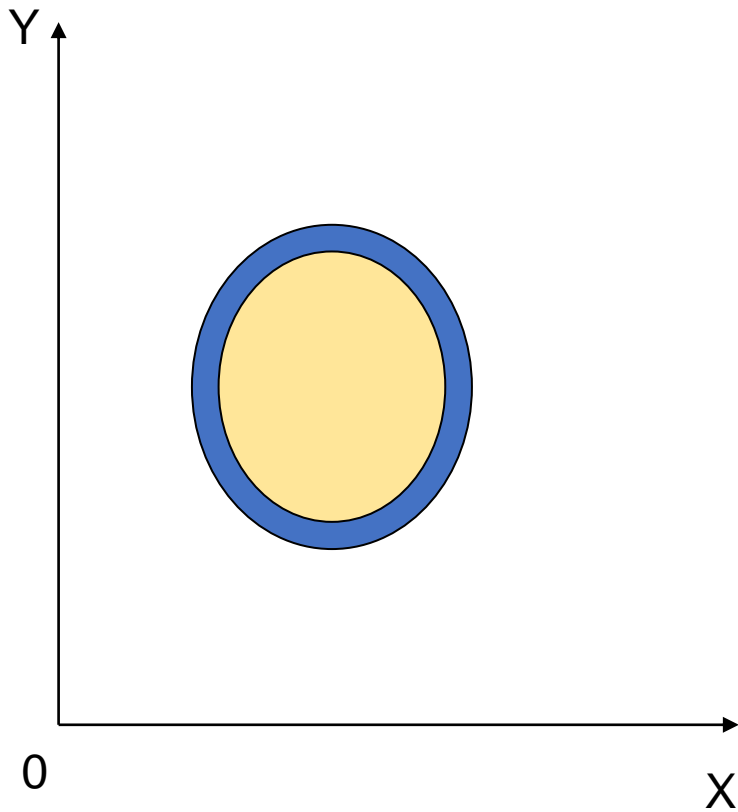
Koefisien korelasi (x dan y) mempunyai hubungan positif



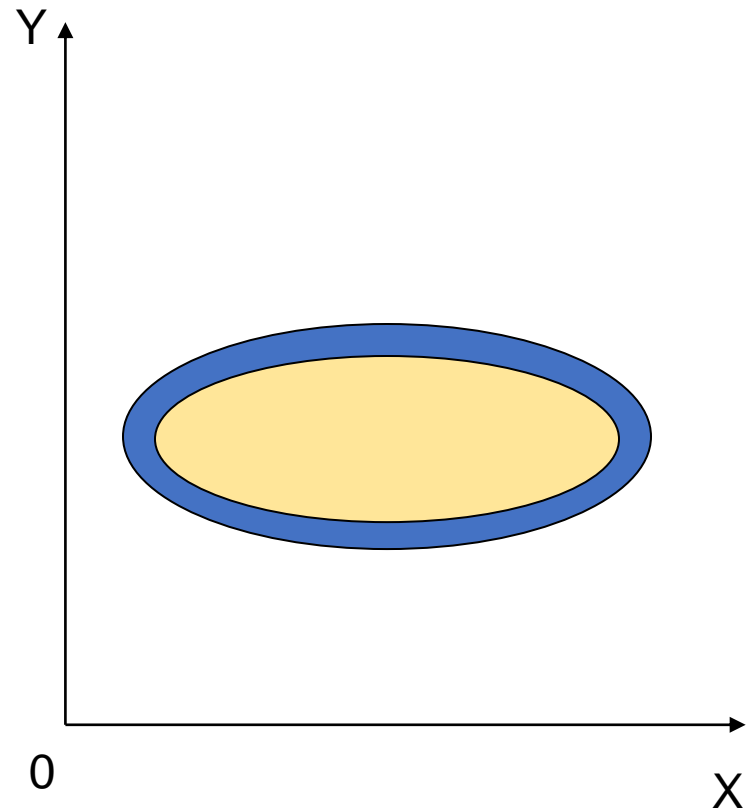
Koefisien korelasi (x dan y) mempunyai hubungan negatif



Koefisien korelasi (x dan y) tidak mempunyai hubungan atau hubungan lemah sekali



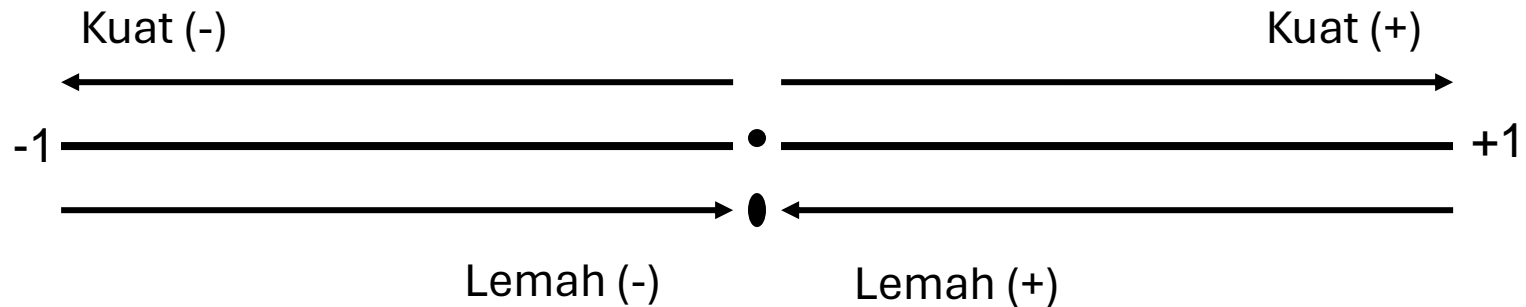
atau



Kuat dan tidaknya hubungan antara X dan Y apabila dapat dinyatakan dengan fungsi linear (paling tidak mendekati), diukur dengan suatu nilai yang disebut koefisien korelasi. Nilai koefisien korelasi ini paling sedikit -1 dan paling besar $+1$.

Jadi jika $r =$ koefisien korelasi, maka r dapat dinyatakan sebagai berikut :

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



Jika $r = +1$, hubungan X dan Y sempurna dan positif,
 $r = -1$, hubungan X dan Y sempurna dan negatif,
 r mendekati $+1$, hubungan sangat kuat dan positif,
 r mendekati -1 , hubungan sangat lemah dan negatif.

PERBEDAAN REGRESI DAN KORELASI

- Regresi menunjukkan hubungan antara variabel satu dengan variabel lainnya.
- Sifat hubungan dapat dijelaskan: variabel yang satu sebagai penyebab, variabel yang lain sebagai akibat.
- Korelasi tidak menunjukkan hubungan sebab akibat, akan tetapi menunjukkan hubungan antara variabel satu dengan yang lain.

CONTOH SOAL

- Berikut ini data mengenai pengalaman kerja dan penjualan
- X =pengalaman kerja (tahun)
- Y =omzet penjualan (ribuan)

X	2	3	2	5	6	1	4	1
Y	5	8	8	7	11	3	10	4

- Tentukan nilai a dan b
- Buatkan persamaan regresinya!
- Berapa omzet penjualan dari seorang karyawan yg pengalaman kerjanya 3,5 tahun

Penyelesaian :

X	Y	X ²	Y ²	XY
2	5	4	25	10
3	8	9	64	24
2	8	4	64	16
5	7	25	49	35
6	11	36	121	66
1	3	1	9	3
4	10	16	100	40
1	4	1	16	4
24	56	96	448	198

$$\bar{X} = \frac{24}{8} = 3 \quad \bar{Y} = \frac{56}{8} = 7$$

$$a = \frac{(56)(96) - (24)(198)}{(8)(96) - (24)^2}$$

$$a = \frac{5.376 - 4.752}{768 - 576} = 3,25$$

$$b = \frac{(8)(198) - (24)(56)}{(8)(96) - (24)^2}$$

$$b = \frac{1.584 - 1.344}{768 - 576} = 1,25$$

- a. Diperoleh nilai $a = 3,25$ dan nilai $b = 1,25$
- b. Persamaan regresi linearnya adalah $Y = 3,25 + 1,25X$
- c. Nilai duga Y , jika $X = 3,5$ adalah $Y = 3,25 + 1,25X$

$$Y = 3,25 + 1,25(3,5)$$

$$= 7,625$$

KOEFISIEN DETERMINASI (r^2)

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

$$r^2 = \frac{((8)(198) - (24)(56))^2}{(8(96) - (24)^2)(8(448) - (56)^2)}$$

$$r^2 = \frac{(1.584 - 1.344)^2}{(768 - 576)(3.584 - 3.136)}$$

$$r^2 = \frac{(240)^2}{(192)(448)} = \frac{57.600}{86.016} = 0,6696$$

Nilai determinasi (r^2) sebesar **0,6696**, artinya sumbangan atau pengaruh pengalaman kerja terhadap naik turunnya omzet penjualan adalah sebesar 66,96%. Sisanya 33,04% disebabkan oleh faktor lain yang tidak dimasukkan dalam model.

Nilai koef korelasi = **0,818** artinya ada hubungan yang kuat dan positif antara pengalaman kerja dan naik turunnya penjualan