

# **Korelasi dan Regresi Linier**

## **Bahan Kuliah Statistika**

**Sevi Nurafni**

**Fakultas Sains dan Teknologi  
Universitas Koperasi Indonesia 2025**

# Perbedaan Korelasi dan Regresi?

## Korelasi

Korelasi hanya menunjukkan sekedar hubungan.

Dalam korelasi variabel tidak ada istilah tergantung dan variabel bebas

## Regresi

Regresi menunjukkan hubungan pengaruh  
Dalam regresi terdapat istilah tergantung  
dan variabel bebas

# Korelasi Linier Sederhana

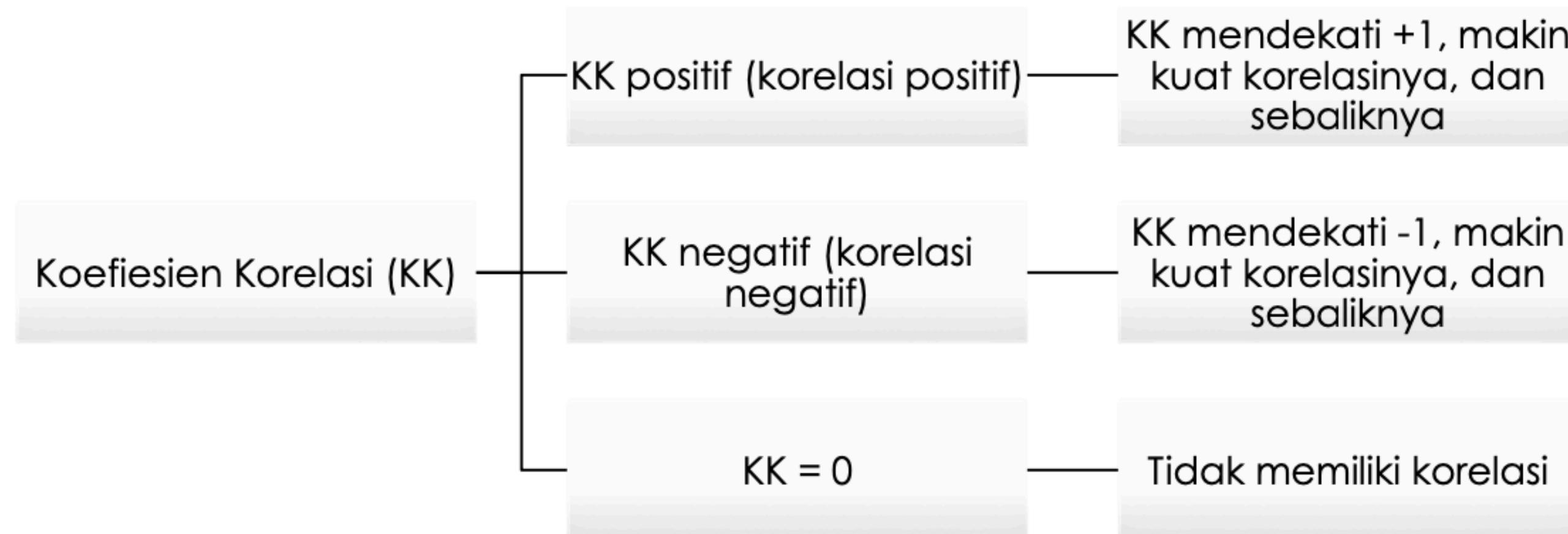
# Contoh Analisis Korelasi

- Mengukur korelasi antar variabel:
  - Fasilitas pendidikan dengan prestasi belajar siswa
  - Kandungan Air dan Umur Simpan Produk Keripik Buah
  - Suhu Penyimpanan dan Kadar Vitamin C pada Jus Jeruk Segar

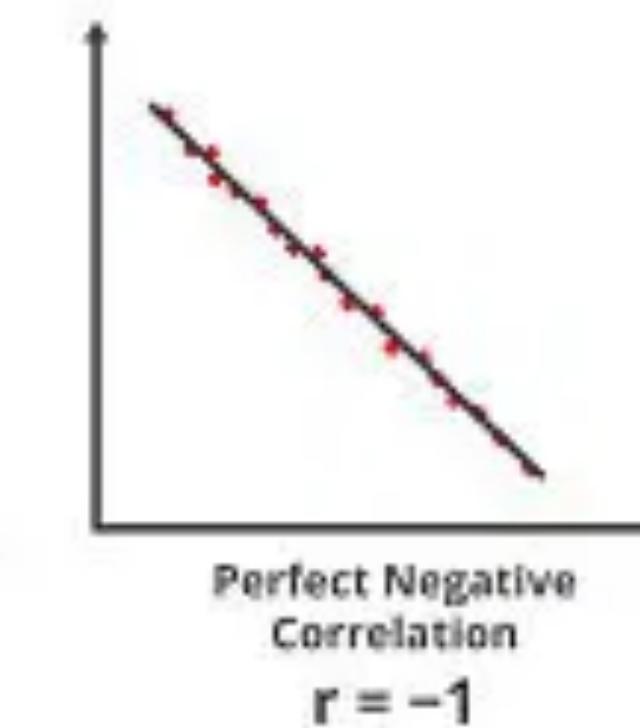
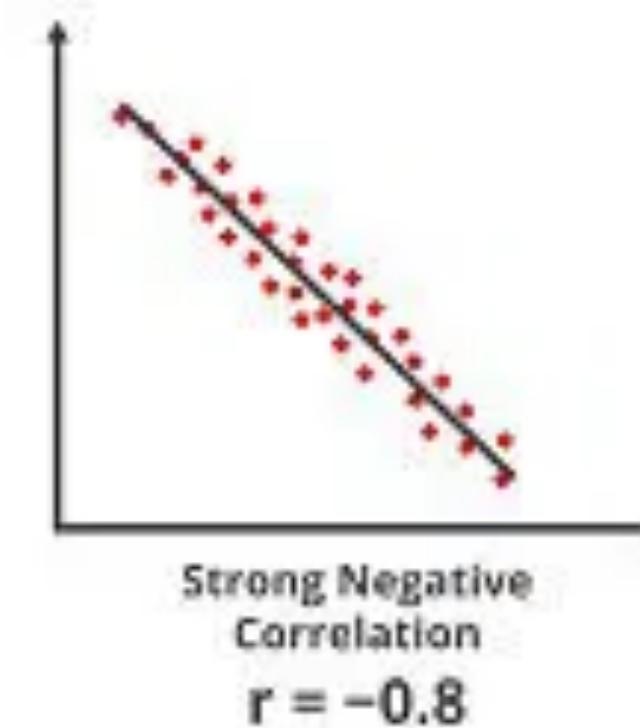
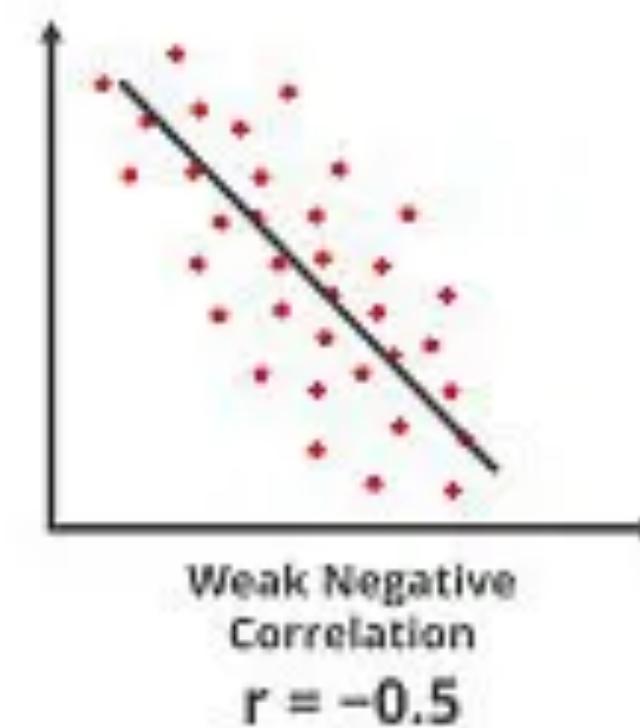
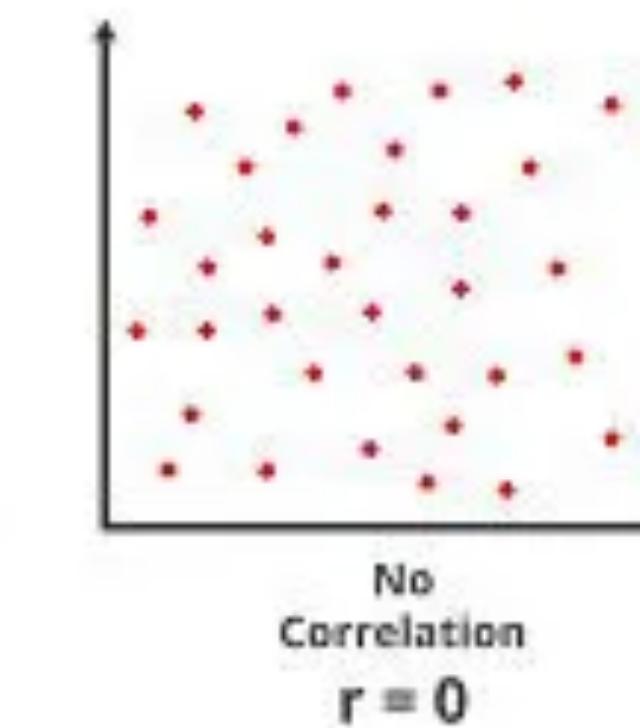
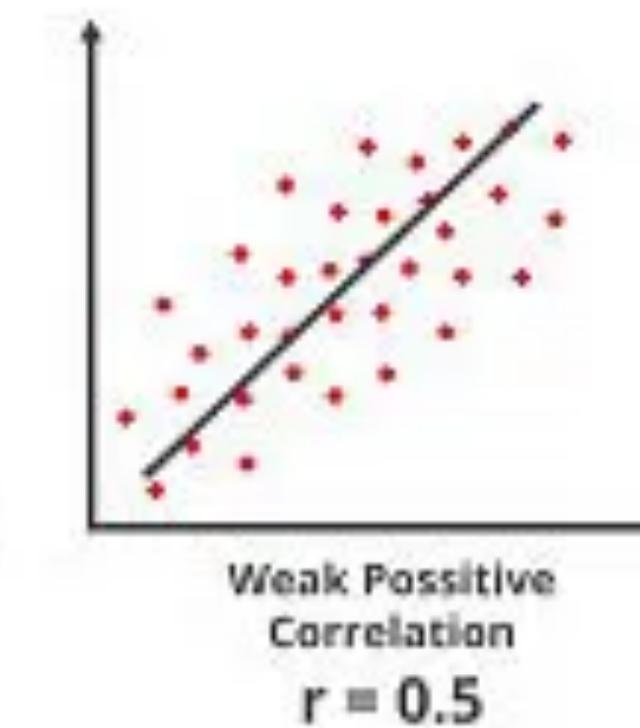
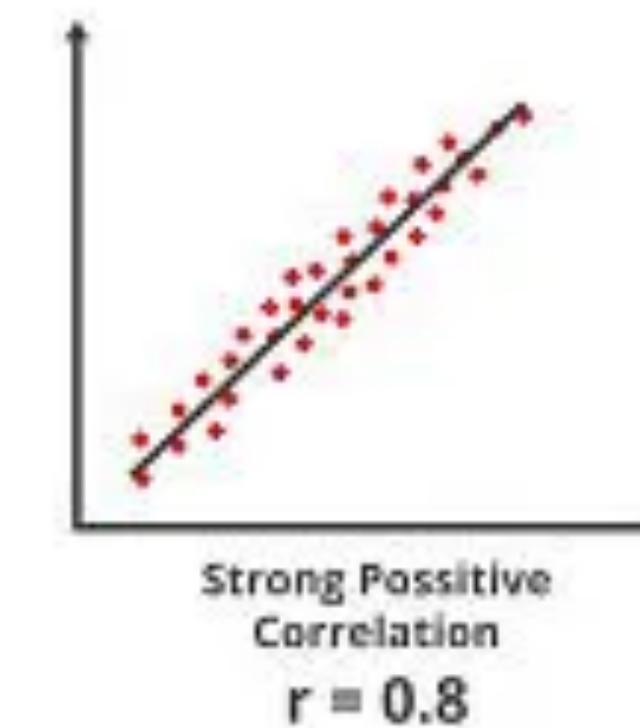
Korelasi tidak secara otomatis menunjukkan adanya hubungan kausalitas.

# Koefisien Korelas Linier Sederhana

- Indeks atau bilangan yang digunakan untuk mengukur keeratan hubungan antar variabel
- Memiliki nilai antara  $-1$  dan  $1 - 1 \leq KK \leq 1$



# Correlation



# Arti dari koefisien relasi r

1. Bila  $0, 90 < r < 1,00$  atau  $-1, 00 < r < -0, 90$  :  
→ hubungan yang **sangat kuat**
2. Bila  $0, 70 < r < 0, 90$  atau  $-0.90 < r < -0. 70$  :  
→ hubungan yang **kuat**
3. Bila  $0, 50 < r < 0, 70$  atau  $-0.70 < r < -0. 50$  :  
→ hubungan yang **moderat**
4. Bila  $0, 30 < r < 0, 50$  atau  $-0.50 < r < -0. 30$  :  
→ hubungan yang **lemah**
5. Bila  $0,0 < r < 0, 30$  atau  $-0.30 < r < 0,0$  :  
→ hubungan yang **sangat lemah**

# Jenis-jenis koefisien korelasi

Pearson

- data interval / rasio

Rank Spearman

- data ordinal

Kontingensi

- data nominal

Koefisien Penentu  
(KP) atau Koefisien  
Determinasi (R)

- Pengaruh antar variabel

# Koefisien Korelasi Pearson

- Digunakan untuk mengukur keeratan hubungan antara 2 variabel yang datanya berbentuk data interbal atau rasio

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

# Contoh

Misal data berikut menggambarkan keuntungan usahatani ( $Y$ ) pada berbagai luas lahan ( $X$ ) padi sawah :

No Petani	Luas Lahan ( $X$ )	Keuntungan ( $Y$ )
1	0,21	0,50
2	0,50	1,10
3	0,14	0,25
4	1,00	1,80
5	0,21	0,40
6	0,07	0,20
7	0,50	0,90
8	1,00	2,00
9	0,70	1,20
10	0,14	0,35
11	0,35	0,70
12	0,28	0,65

No	Luas (X)	Untung (Y)	$X^2$	$Y^2$	$XY$
1	0,21	0,5	0,0441	0,25	0,105
2	0,5	1,1	0,25	1,21	0,55
3	0,14	0,25	0,0196	0,0625	0,035
4	1	1,8	1	3,24	1,8
5	0,21	0,4	0,0441	0,16	0,084
6	0,07	0,2	0,0049	0,04	0,014
7	0,5	0,9	0,25	0,81	0,45
8	1	2	1	4	2
9	0,7	1,2	0,49	1,44	0,84
10	0,14	0,35	0,0196	0,1225	0,049
11	0,35	0,7	0,1225	0,49	0,245
12	0,28	0,65	0,0784	0,4225	0,182
jumlah	5,1	10,05	3,3232	12,2475	6,354

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

$$r = \frac{12(6,3540) - (5,10)(10,05)}{\sqrt{(12(3,3232) - (5,10)^2)(12(12,2475) - (10,05)^2)}}$$

$$r = \frac{12(6,3540) - (5,10)(10,05)}{\sqrt{(12(3,3232) - (5,10)^2)(12(12,2475) - (10,05)^2)}}$$

$$r = \frac{76,248 - 51,2550}{\sqrt{(39,8784 - 26,01)(146,97 - 101,0025)}}$$

$$r = \frac{24,993}{\sqrt{(13,8684)(45,9675)}}$$

$$r = 0,9899$$

Nilai  $r = 0,9899$  artinya hubungan antara luas lahan dan keuntungan mempunyai hubungan yang sangat kuat

# Excel



# Koefisien Korelasi Rank Spearman

- Digunakan untuk mengukur keeratan hubungan antara 2 variabel yang datanya berbentuk data ordinal (bertingkat).
- Digunakan untuk mengurutkan objek/tingkatan dari yang paling rendah sampai yang paling tinggi, atau sebaliknya.
- Contoh: Peringkat kelas, skala likert

$$\bullet \quad r_s = 1 - \frac{6\sum d^2}{n^3 - n}$$

- $r_s$  : Koefisien Korelasi Spearman's rank
- $n$  : Jumlah sampel
- $d$  : selisih ranking X dan Y

# Contoh

Contoh data berikut menggambarkan Pengalaman Usahatani (X) dan Penerapan Teknologi (Y) dari 12 petani :

No	X	Y
1	12	85
2	10	74
3	10	78
4	13	90
5	11	85
6	14	87
7	13	94
8	14	98
9	11	81
10	14	91
11	10	76
12	8	74

No	X	Rank
1	8	1
2	10	3
3	10	3
4	10	3
5	11	5,5
6	11	5,5
7	12	7
8	13	8,5
9	13	8,5
10	14	11
11	14	11
12	14	11

No	Y	Rank
1	74	1,5
2	74	1,5
3	76	3
4	78	4
5	81	5
6	85	6,5
7	85	6,5
8	87	8
9	90	9
10	91	10
11	94	11
12	98	12

No	X	Y	Rank-X	Rank-Y	$d_i^2$
1	12	85	7	6,5	0,25
2	10	74	3	1,5	2,25
3	10	78	3	4	1
4	13	90	8,5	9	0,25
5	11	85	5,5	6,5	1
6	14	87	11	8	9
7	13	94	8,5	11	6,25
8	14	98	11	12	1
9	11	81	5,5	5	0,25
10	14	91	11	10	1
11	10	76	3	3	0
12	8	74	1	1,5	0,25
<b>jumlah</b>					<b>22,5</b>

- $r_s = 1 - \frac{6\sum d^2}{n^3 - n}$
- $r_s = 1 - \frac{6(22,5)}{(12)^3 - 12}$
- $r_s = 1 - \frac{135}{1728 - 12}$
- $r_s = 1 - \frac{135}{1716}$

- $r_s = 1 - 0,0787 = 0,9213$

Nilai  $r_s = 0,9213$  artinya hubungan antara pengalaman usaha tani dan penerapan teknologi mempunyai hubungan yang sangat kuat

# Excel

- Rumus yang digunakan untuk menentukan rank =RANK.AVG(B2;\$B\$2:\$B\$13;\$B\$13)

	A	B	C	D	E	F
1	No	X	Y	Rank-X	Rank-Y	$d^2$
2	1	12	85	7	6,5	0,25
3	2	10	74	3	1,5	2,25
4	3	10	78	3	4	1
5	4	13	90	8,5	9	0,25
6	5	11	85	5,5	6,5	1
7	6	14	87	11	8	9
8	7	13	94	8,5	11	6,25
9	8	14	98	11	12	1
0	9	11	81	5,5	5	0,25
1	10	14	91	11	10	1
2	11	10	76	3	3	0
3	12	8	74	1	1,5	0,25
4	jumlah					22,5

B2 = angka yang akan dicari peringkatnya

\$B\$2:\$B\$13 = kolom data yang dicari peringkatnya

\$B\$13 = angka peringkat awal

# Signifikansi

- Secara statisti, signifikan → “benar”
- Memberikan gambaran bagaimana hasil riset memiliki kesempatan untuk benar
  - Signifikansi 0,01, berarti tingkat kepercayaan untuk memperoleh kebenaran dalam riset adalah 99%.
  - Signifikansi 0,05, berarti tingkat kepercayaan untuk memperoleh kebenaran dalam riset adalah 95%.
  - Signifikansi 0,1, berarti tingkat kepercayaan untuk memperoleh kebenaran dalam riset adalah 90%.

# **Regresi Linier Sederhana**

# Analisis Regresi

- Analisis regresi digunakan untuk:
- Meramalkan atau memperkirakan nilai dari satu variabel dalam hubungannya dengan variabel lain
- Mempelajari dan mengukur hubungan statistik yang terjadi antara dua atau lebih variabel
- Menerangkan impak perubahan variabel independen terhadap variabel dependen
- Regresi ada 2 macam yaitu:
  - Regresi sederhana: dikaji dua variabel
  - Regresi majemuk: dikaji lebihd ari dua variabel

# Variabel dalam Analisis Regresi

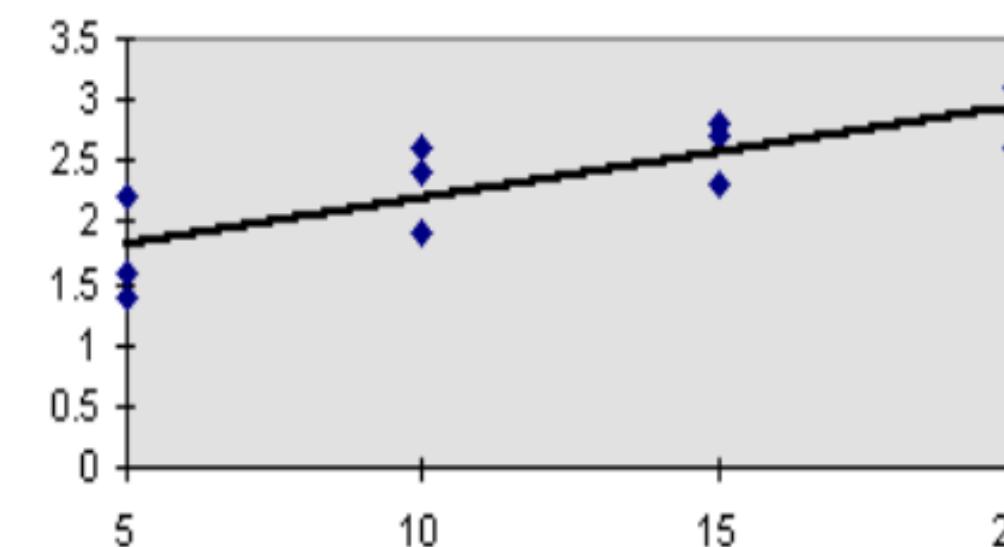
## Variabel Terikat

- Variabel yang akan diestimasi nilainya
- Biasanya diplot pada sumbu- $y$

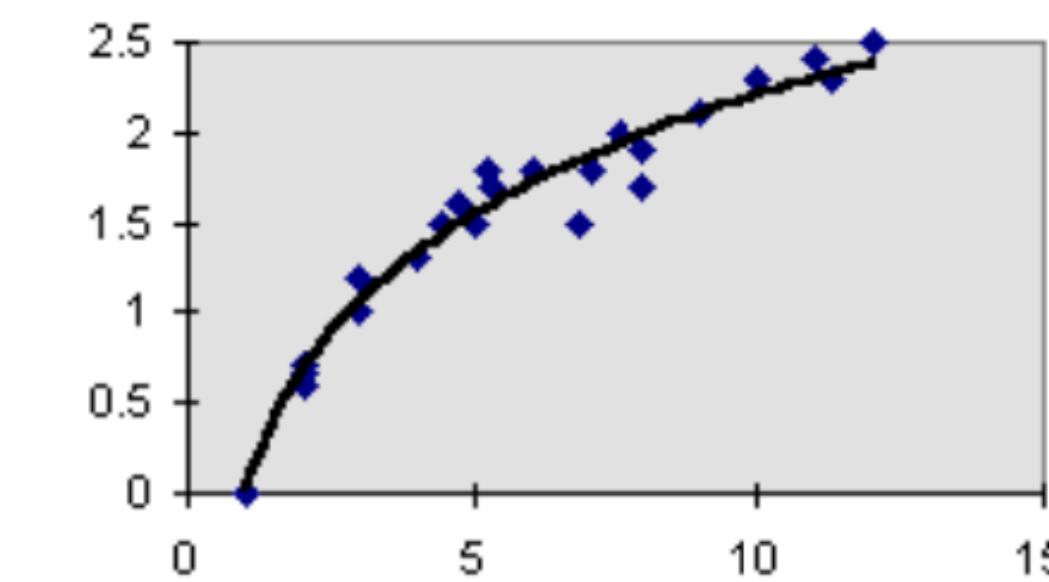
## Variabel Bebas

- Variabel yang diasumsikan memberikan pengaruh terhadap variasi variabel terikat
- Biasanya diplot pada sumbu- $x$

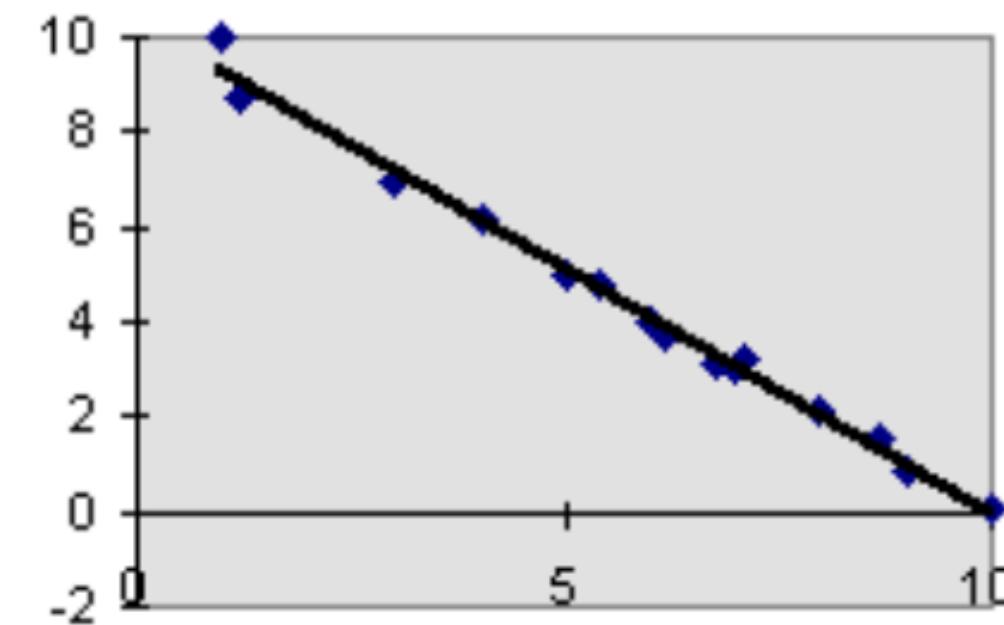
# Tipe Model Regresi



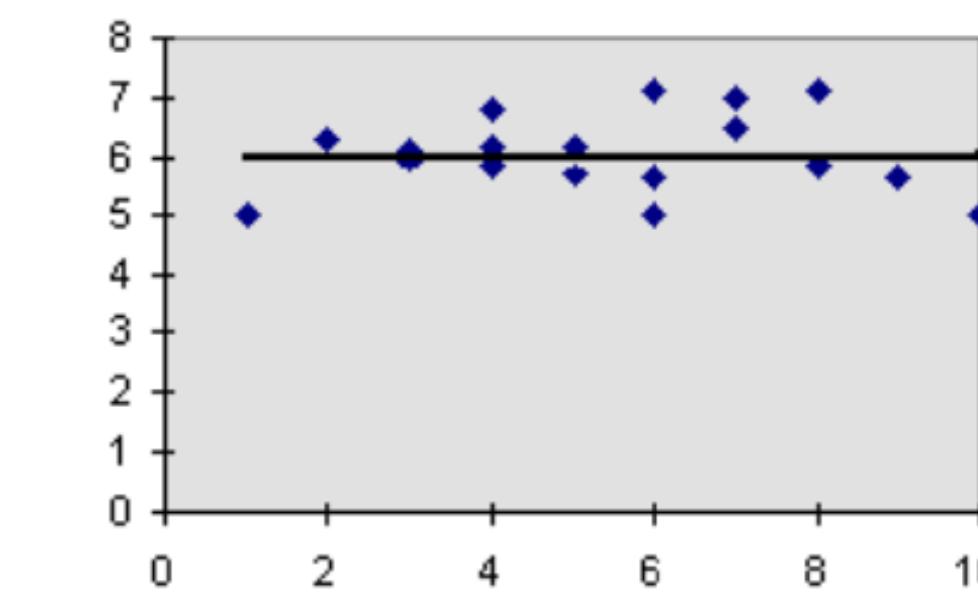
**Positive Linear Relationship**



**Relationship NOT Linear**



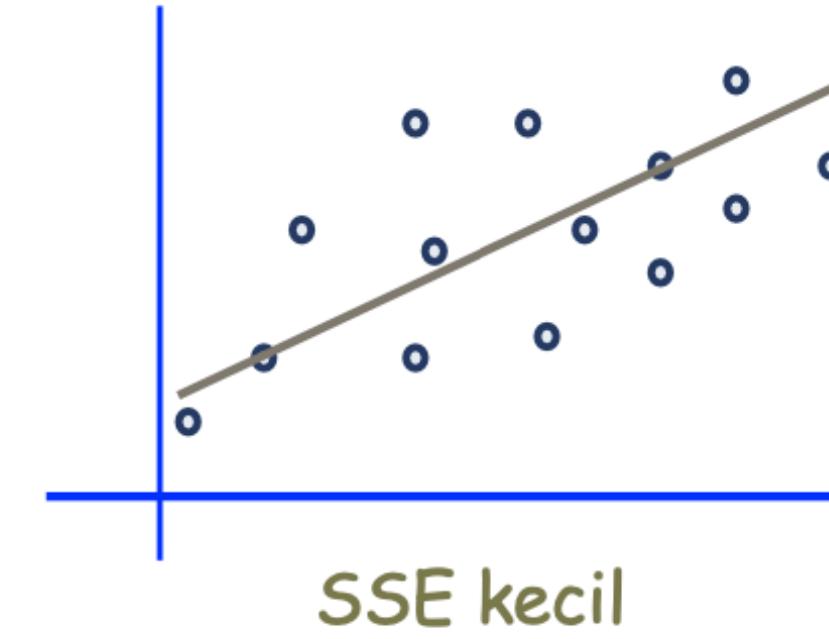
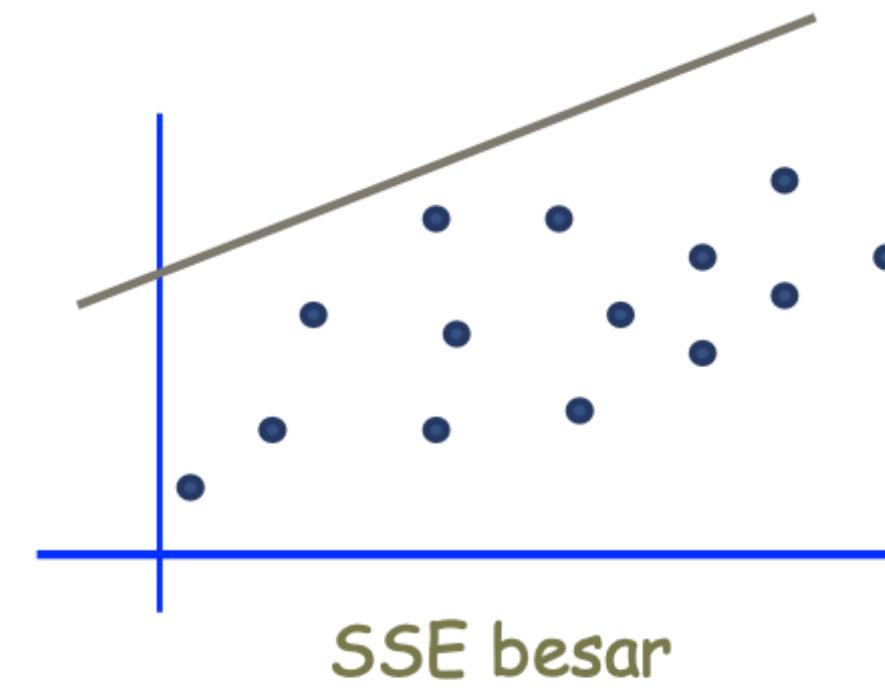
**Negative Linear Relationship**



**No Relationship**

# Analisis Regresi

- Gagasannya adalah meminimalkan penyebaran total nilai  $y$  dari garis.



- Regresi atau garis kuadrat terkecil → adalah garis dengan SSE yang terkecil

# Analisis Regresi Linear

- $\hat{Y} = A + BX$
- $\hat{Y}$ = penduga (bagi rata-rata Y untuk X tertentu)
- $X$ =variabel bebas
- $A, B$ =koefisien regresi sampel

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

# Contoh

- Dari suatu pengujian diperoleh data yang menghubungkan variabel bebas X dan variabel terikat Y seperti pada tabel berikut:

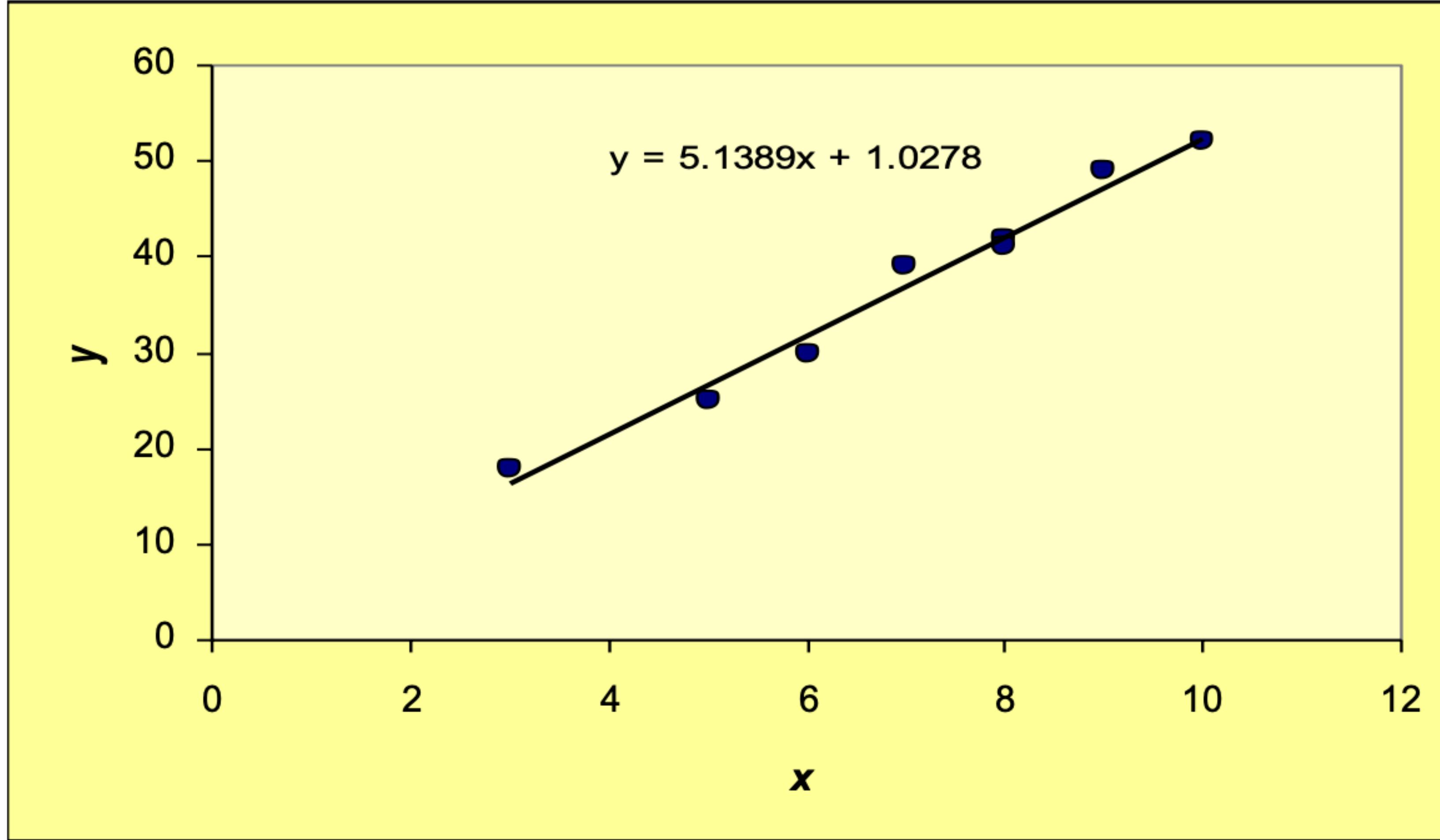
Uji ke-	x	y
1	6	30
2	9	49
3	3	18
4	8	42
5	7	39
6	5	25
7	8	41
8	10	52
$\Sigma$	56	296

### Tabel perhitungan:

Uji ke-	x	y	xy	$x^2$	$y^2$
1	6	30	180	36	900
2	9	49	441	81	2401
3	3	18	54	9	324
4	8	42	336	64	1764
5	7	39	273	49	1521
6	5	25	125	25	625
7	8	41	328	64	1681
8	10	52	520	100	2704
$\Sigma$	<b>56</b>	<b>296</b>	<b>2257</b>	<b>428</b>	<b>11920</b>
$\bar{x} = \frac{\sum x}{n} = \frac{56}{8} = 7$ $\bar{y} = \frac{\sum y}{n} = \frac{296}{8} = 37$					

# Rumus

- $b = \frac{(n\Sigma XY - \Sigma X \Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2}$
- $b = \frac{8(2257) - (56)(296)}{8(428) - (56)^2}$
- $b = \frac{1480}{288}$
- $b = 5,1389$
- $a = \bar{Y} - b\bar{X} = \frac{(\Sigma Y \Sigma X^2) - \Sigma X \Sigma XY}{n\Sigma X^2 - (\Sigma X)^2}$
- $a = 37 - (5,1389)(7)$
- $a = 1,0277$
- $\hat{y} = a + bx$
- $\hat{y} = 1,0277 + 5,1389x$



# Standard Error of Estimate

- Merupakan indeks yang digunakan untuk mengukur tingkat ketepatan regresi dan koefisien regresi atau mengukur variasi titik-titik observasi di sekitar garis regresi.
- Dengan kesalahan baku, batasan seberapa jauh melesatnya perkiraan dalam meramal data dapat diketahui
- Bila semua titik oberservasi tepat di garis regresi maka kesalahan baku = nol

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

# Kesalahan baku koefisien regresi $a$ dan $b$

- Koefisien regresi  $a$

$$S_a = \sqrt{\frac{\Sigma X^2 - S_{y,x}}{n\Sigma X^2 - (\Sigma X)^2}}$$

- Koefisien regresi  $b$

$$S_b = \frac{S_{y,x}}{\sqrt{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}}$$

# Contoh SSE

- Dengan menggunakan data dan tabel perhitungan dapa contoh sebelumnya maka SSE dari garis regresi yang diperoleh adalah

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

$$S_{y,x} = \sqrt{\frac{(11920) - 1,0277(296) - 5,1389(2257)}{8 - 2}}$$

$$S_{y,x} = 1,698$$

# Contoh Kesalahan baku

- Koefisien regresi  $a$

$$S_a = \sqrt{\frac{\Sigma X^2 - S_{y,x}}{n\Sigma X^2 - (\Sigma X)^2}}$$

$$S_a = \sqrt{\frac{428 - 1,698}{8(428) - (56)^2}}$$

$$S_a = 1,217$$

- Koefisien regresi  $b$

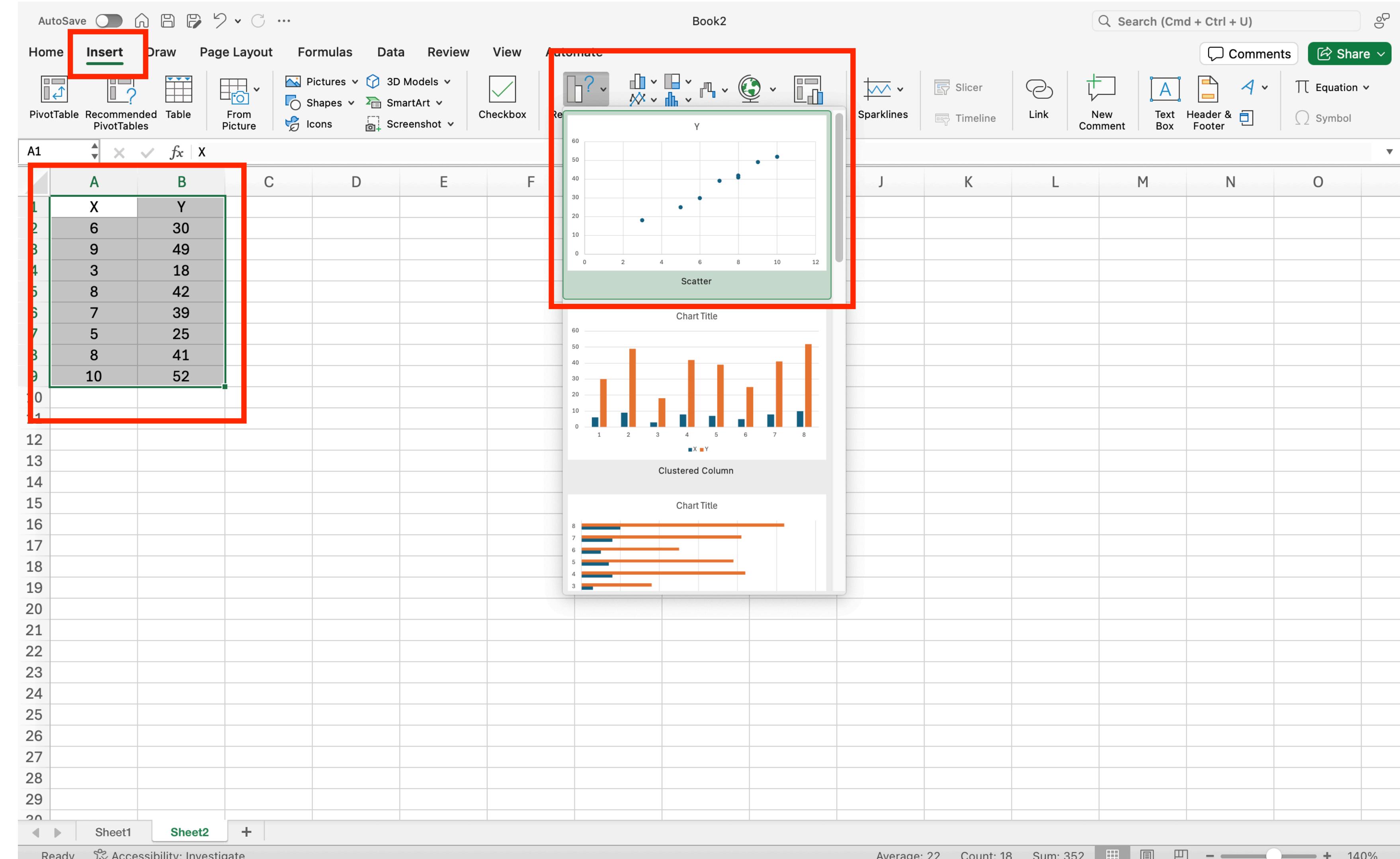
$$S_b = \frac{S_{y,x}}{\sqrt{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}}$$

$$S_b = \frac{1,698}{\sqrt{428 - \frac{(56)^2}{8}}}$$

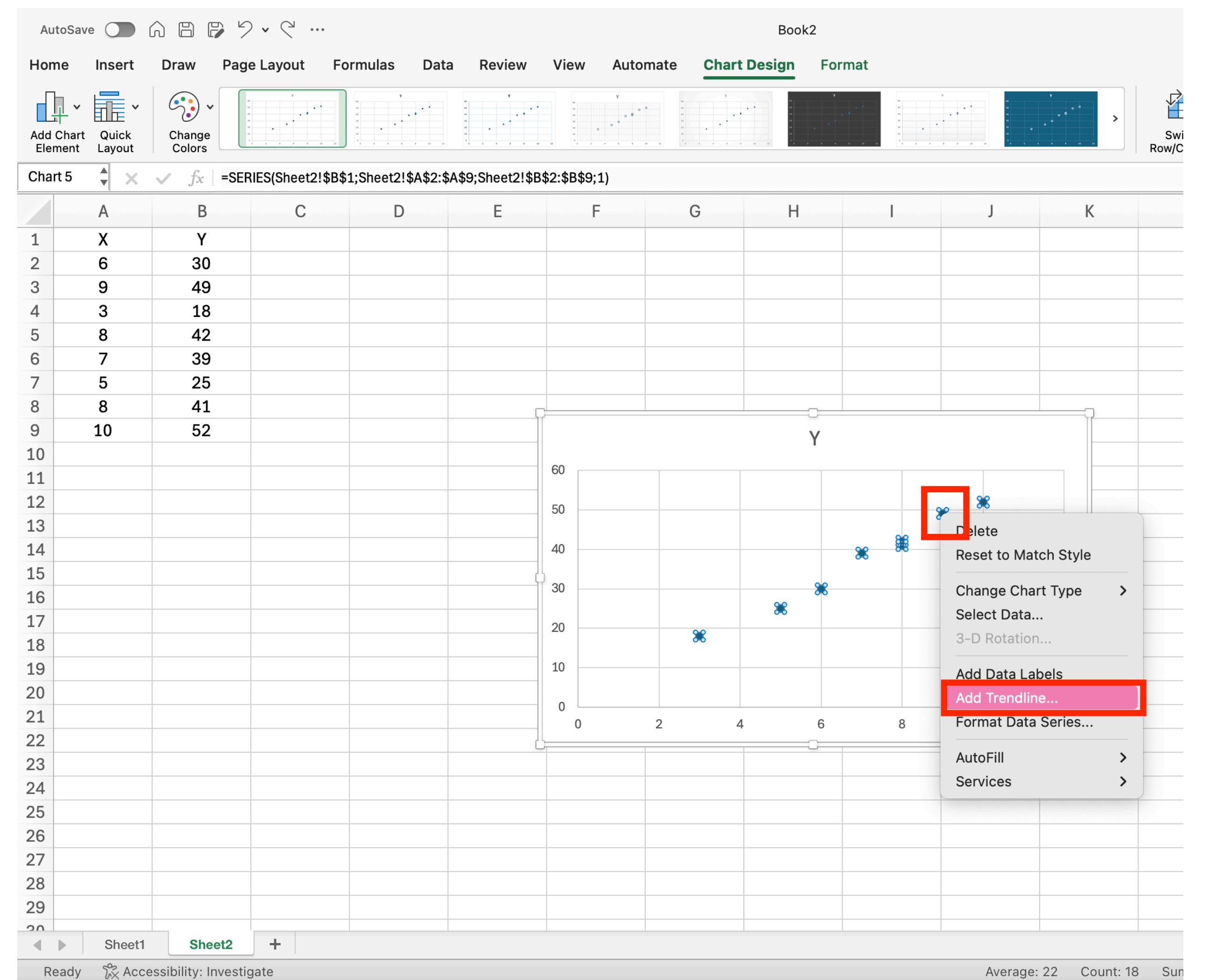
$$S_b = \frac{1,698}{6}$$

$$S_b = 0,283$$

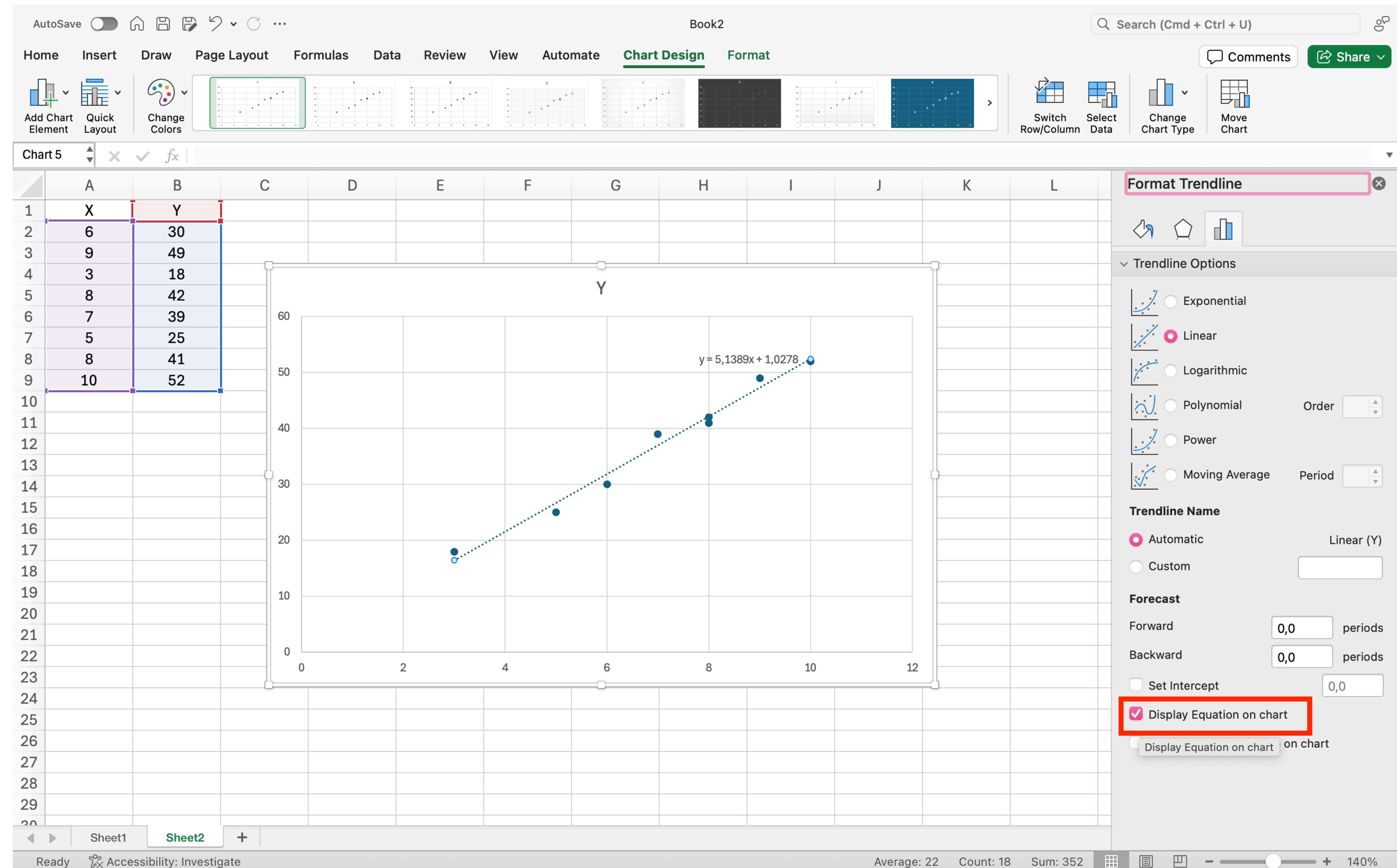
# **Membuat Grafik Linier di Excel**



- Block Kolom A dan B
- Pilih insert → Recommended chart → Scatter



- Klik kanan pada dot data di grafik
- Pilih Add Trendline



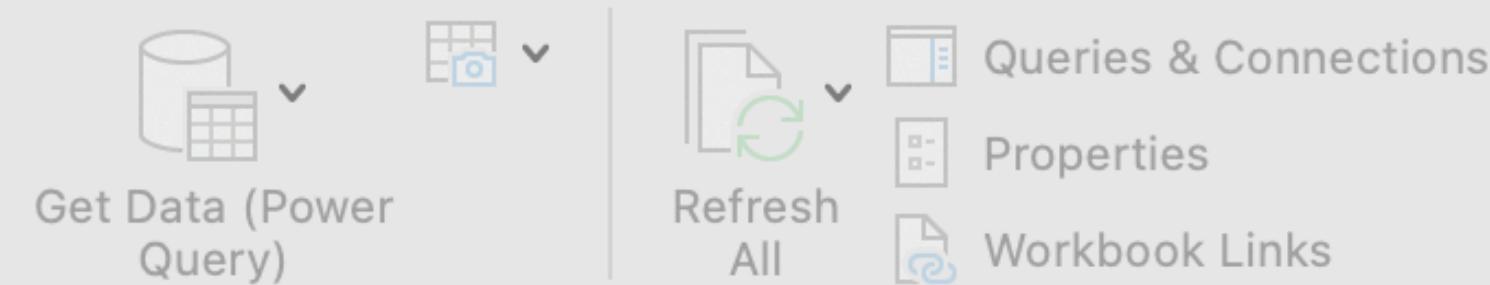
- Pilih Display Equation on Chart

# Regressi di Excel

The screenshot shows a Microsoft Excel interface with the following details:

- Top Menu Bar:** Home, Insert, Draw, Page Layout, Formulas, **Data**, Review, View, Automate.
- Data Tab Options:** Get Data (Power Query), Refresh All, Properties, Stocks, Currencies, Sort, Filter, Advanced, Text to Columns, Flash Fill, Data Validation, Remove Duplicates, Consolidate.
- Analysis Tools:** Group, Ungroup, Subtotal, What-If Analysis, Solver, **Data Analysis**.
- Excel Window:** Book2, Sheet1, Sheet2, Ready, Accessibility: Investigate.
- Data Analysis Dialog Box:** Title: Data Analysis. Category: Analysis Tools. Option: Regression (highlighted with a red box). Other options include Sampling, t-Test: Paired Two Sample for Means, t-Test: Two-Sample Assuming Equal Variances, t-Test: Two-Sample Assuming Unequal Variances, and z-Test: Two Sample for Means. Buttons: OK (highlighted with a red box) and Cancel.
- Table Data:** A1 contains X and Y. Rows 2 through 9 show data pairs (X, Y): (6, 30), (9, 49), (3, 18), (8, 42), (7, 39), (5, 25), (8, 41), (10, 52).

- Pilih Data → Data Analysis → Regression



	A	B	C	D	E
1	X	Y			
2	6	30			
3	9	49			
4	3	18			
5	8	42			
6	7	39			
7	5	25			
8	8	41			
9	10	52			
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					

**Regression**

**Input**

Input Y Range:

Input X Range:

Labels  Constant is Zero

Confidence Level: 95 %

**Output options**

Output Range:

New Worksheet Ply:

New Workbook

**Residuals**

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

**Normal Probability**

Normal Probability Plots

**OK** **Cancel**

Home Insert Draw Page Layout Formulas Data Review View Automate

Get Data (Power Query) Refresh All Workbook Links

Queries & Connections Properties

Stocks Currencies

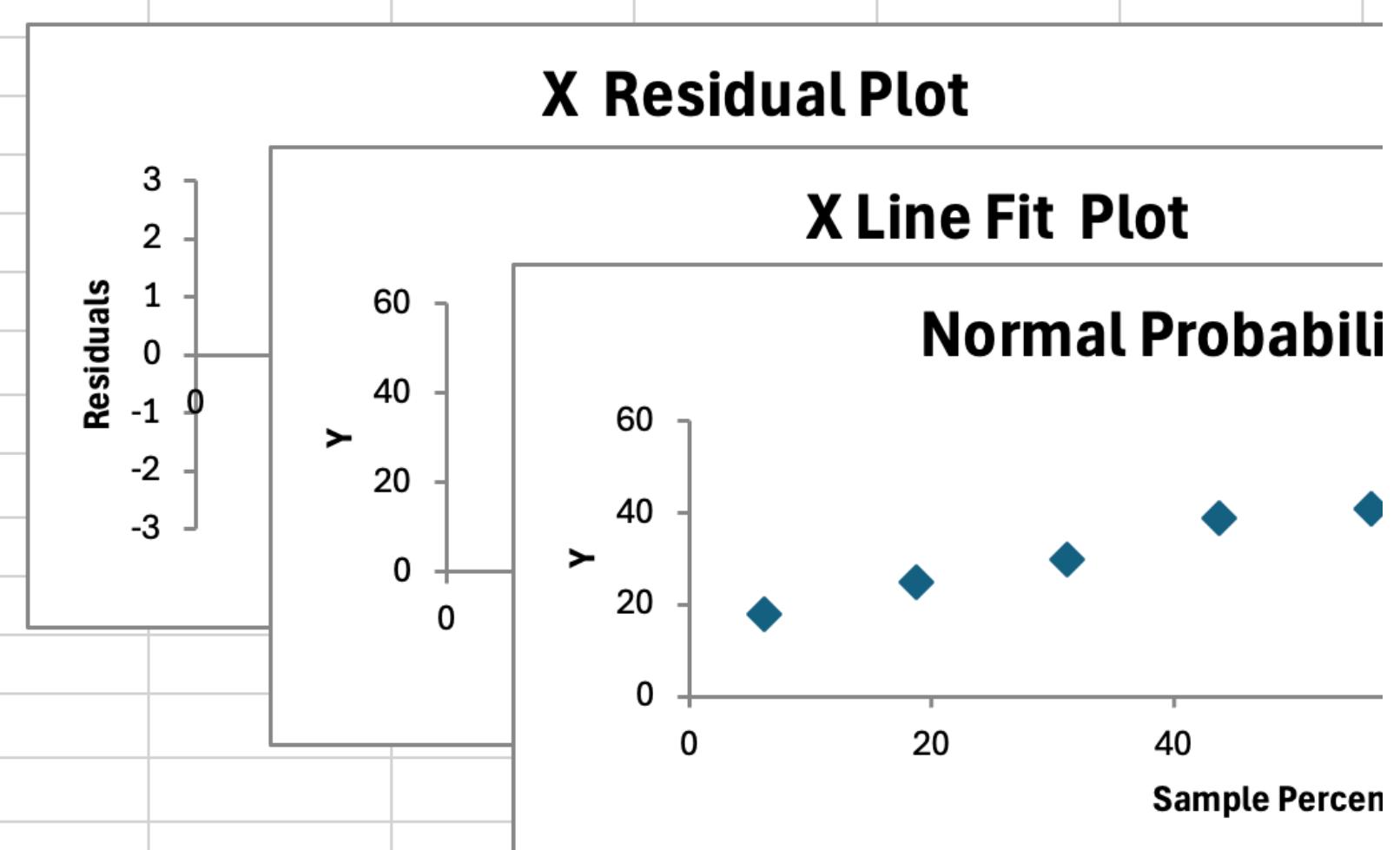
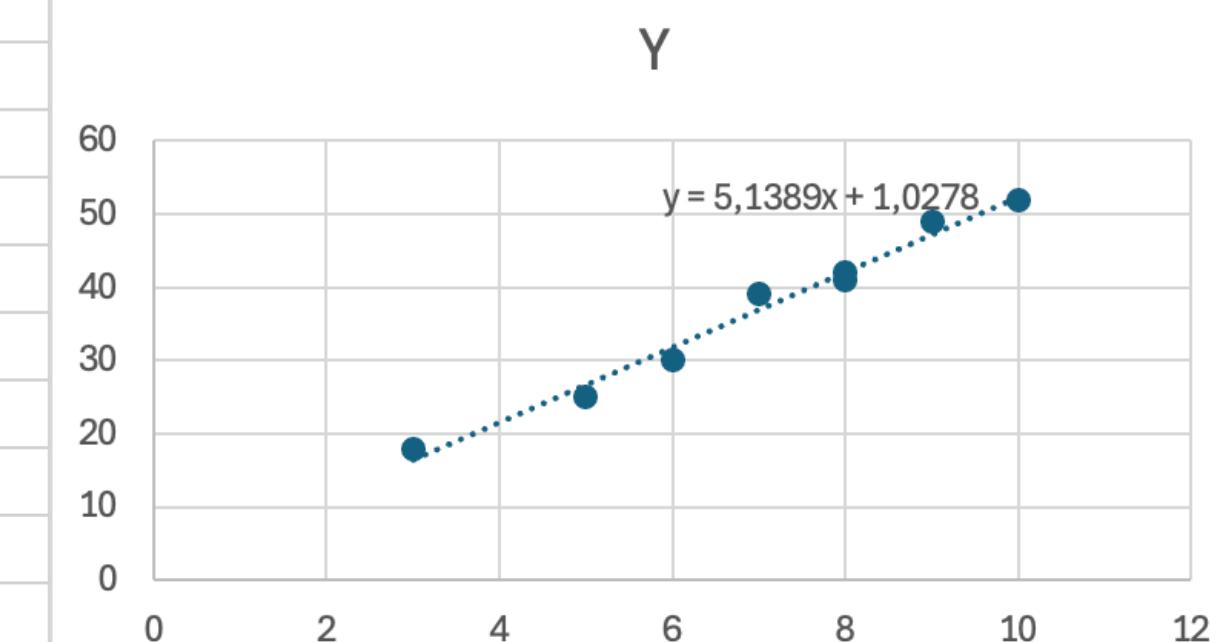
A Z Sort Filter Advanced

Text to Columns Flash Fill Data Validation

Remove Duplicates Consolidate What-If Analysis

Group Ungroup Subtotal Analysis Tools Data Analysis Solver

A11	X	Y	SUMMARY OUTPUT									
1	X	Y										
2	6	30										
3	9	49										
4	3	18										
5	8	42										
6	7	39										
7	5	25										
8	8	41										
9	10	52										
10												
11	SUMMARY OUTPUT											
12												
13	Regression Statistics											
14	Multiple R	0,9910209										
15	R Square	0,9821224										
16	Adjusted R Square	0,9791428										
17	Standard Error	1,6983107										
18	Observations	8										
19												
20	ANOVA											
21		df	SS	MS	F	Significance F						
22	Regression	1	950,69444	950,69444	329,61477	1,798E-06						
23	Residual	6	17,305556	2,8842593								
24	Total	7	968									
25												
26		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%			
27	Intercept	1,0277778	2,0703453	0,4964282	0,6372524	-4,038175	6,0937303	-4,038175	6,0937303			
28	X	5,1388889	0,2830518	18,155296	1,798E-06	4,4462861	5,8314917	4,4462861	5,8314917			
29												



**SELAMAT  
BELAJAR**