

LINEAR REGRESSION (Regresi Linier)

Tim Pengajar Statistika Dasar

- Model regresi linier sederhana adalah model probabilistik yang menyatakan **hubungan linier** antara dua variabel di mana salah satu variabel dianggap memengaruhi variabel yang lain.
- Variabel yang memengaruhi dinamakan **variabel independen** dan variabel yang dipengaruhi dinamakan **variabel dependen**.
- Sebagai contoh, seorang peneliti tertarik untuk menyelidiki pengaruh (hubungan) linier dari *intellegency quotient* (IQ) terhadap hasil belajar statistika mahasiswa. Maka IQ adalah variabel independen, sedangkan hasil belajar statistika adalah variabel dependen.
- Sebuah **diagram *scatter plot*** dapat digunakan untuk menunjukkan hubungan antara dua variabel

- Analisis korelasi digunakan untuk mengukur **kekuatan asosiasi (hubungan linier) antara dua variable**. Korelasi hanya berkaitan dengan **kekuatan hubungan, tidak ada efek kausal** yang tersirat dengan korelasi
- Regresi digunakan untuk mengetahui **hubungan antara satu variabel independen dan satu variabel dependen**, sedangkan korelasi digunakan untuk mengetahui **hubungan antara dua variabel independen**.

- Regresi linier untuk dua variabel didasarkan pada persamaan linier dengan satu variabel independen. Persamaan tersebut memiliki bentuk :

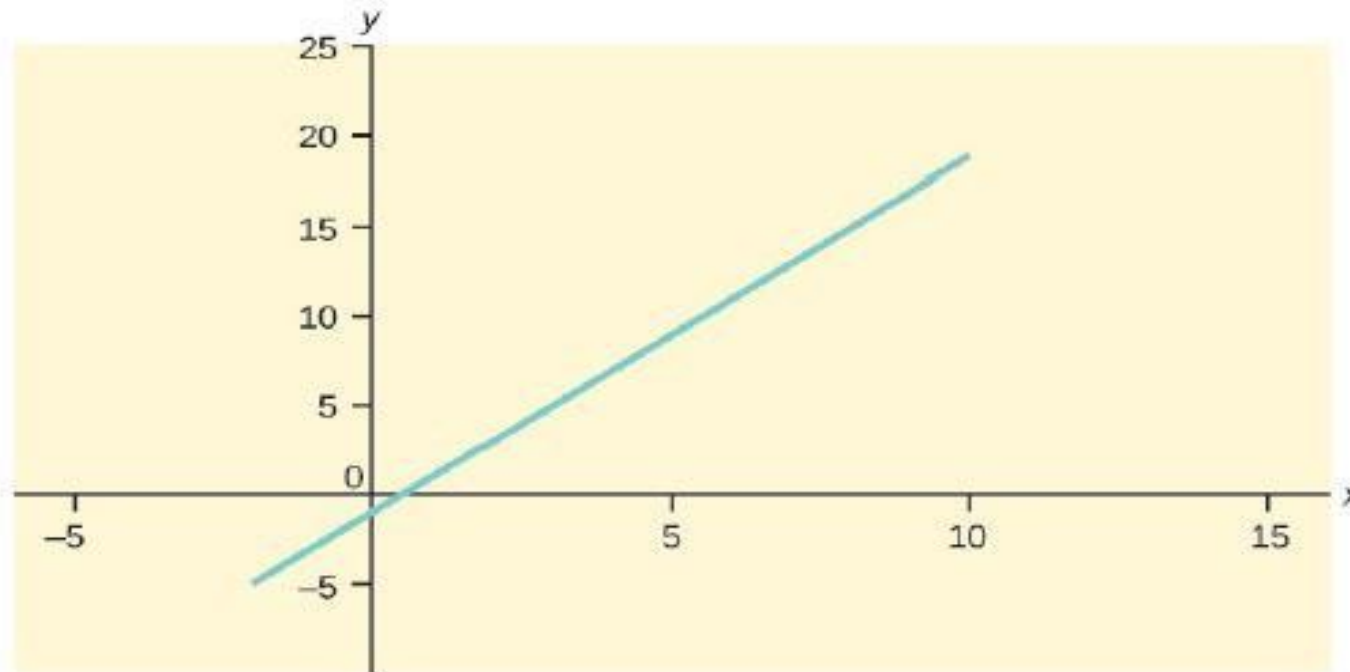
$$y = a + bx$$

dengan :

- **a** dan **b** adalah bilangan konstan
 - variabel **x** adalah variabel bebas/independen, dan
 - variabel **y** adalah variabel terikat/dependen
- Contoh persamaan linear :
 $y = 3 + 2x$
 $y = -0,01 + 1,2 x$

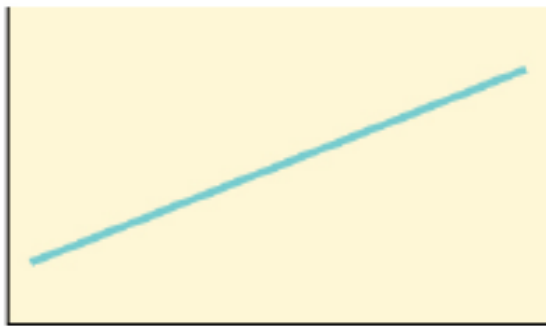
Persamaan Linier: Grafik

- Grafik persamaan linier berbentuk $y = a + bx$ adalah garis lurus
- Contoh : Gambarkan persamaan $y = -1 + 2x$



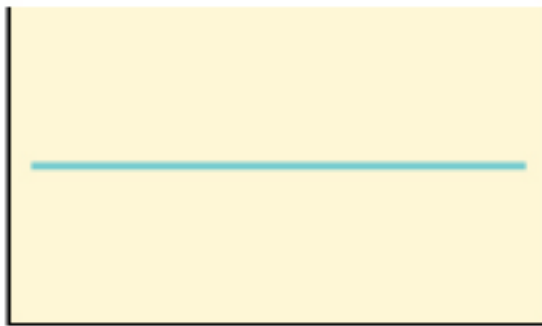
Slope dan Y-intercept pada Persamaan Linier

- Untuk persamaan linier $y = a + bx$, b = kemiringan dan a = titik potong y .
- Tiga kemungkinan grafik dari $y = a + bx$:



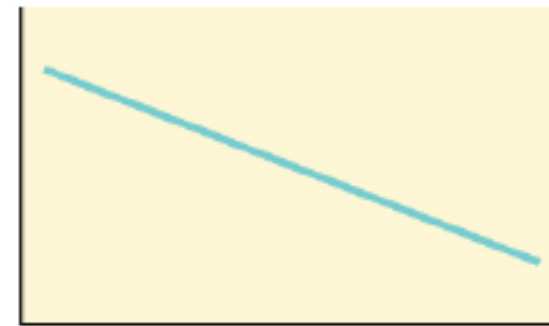
(a)

$$b > 0$$



(b)

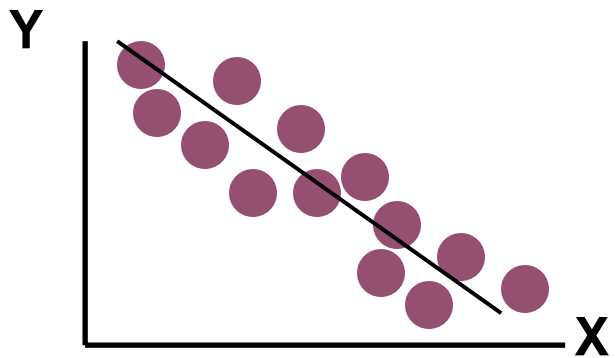
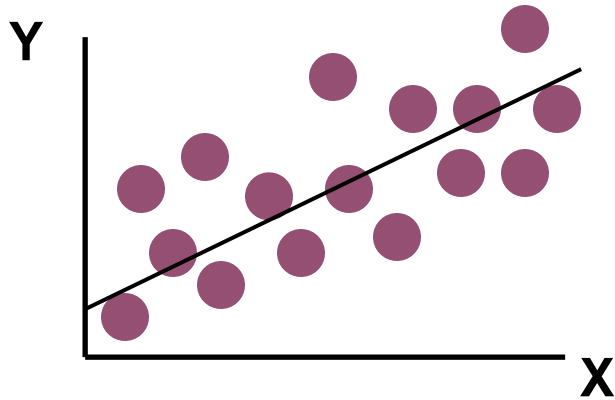
$$b = 0$$



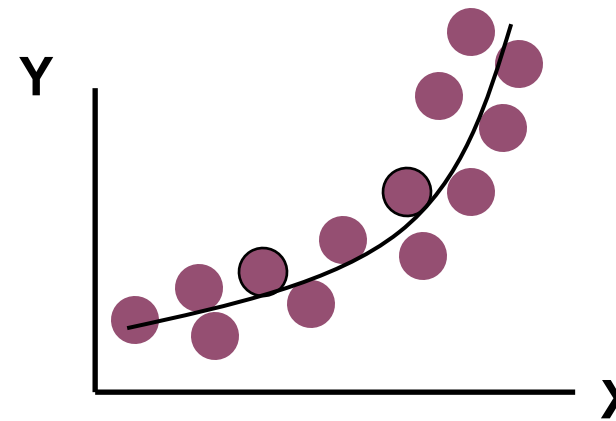
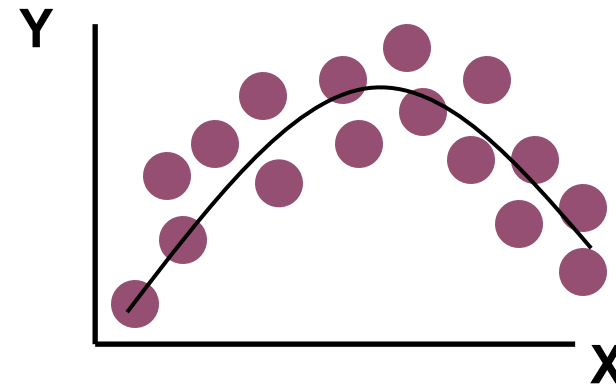
(c)

$$b < 0$$

Linear relationships

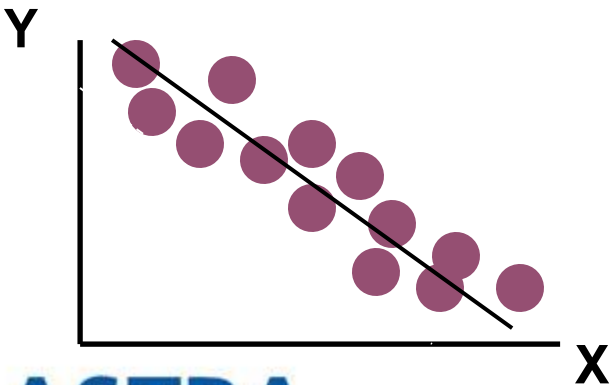
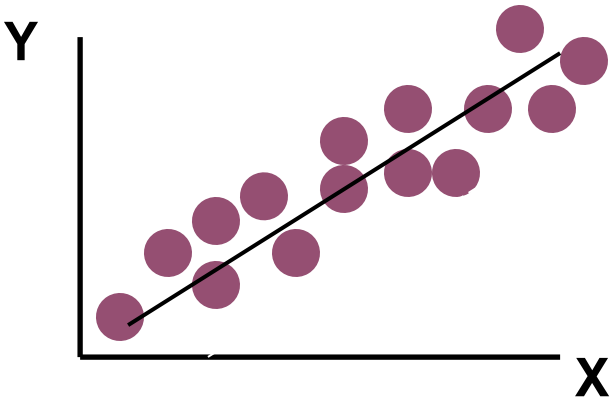


Curvilinear relationships

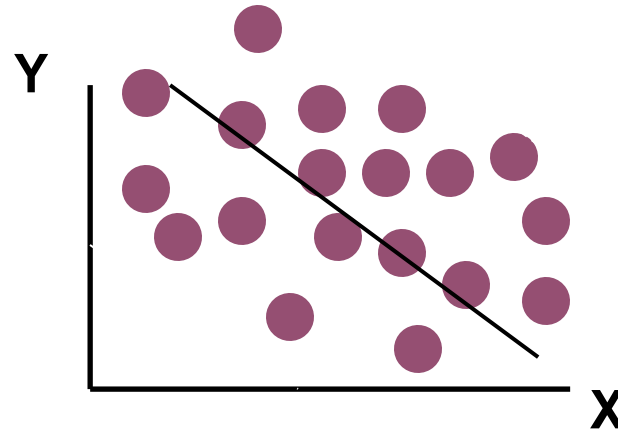
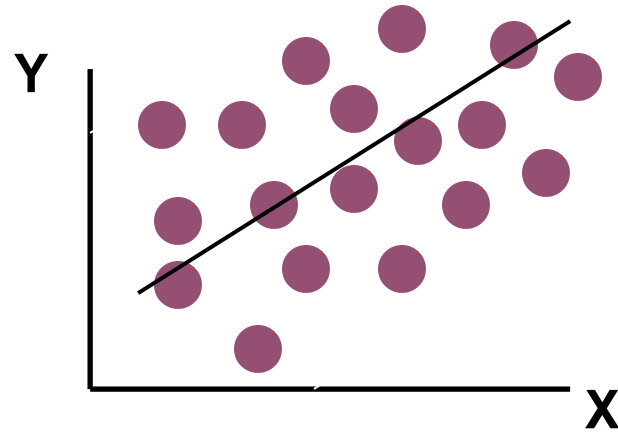


Tipe-Tipe Keterkaitan (Relationship)

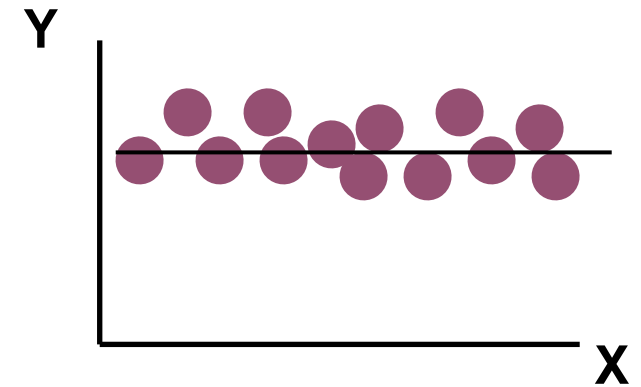
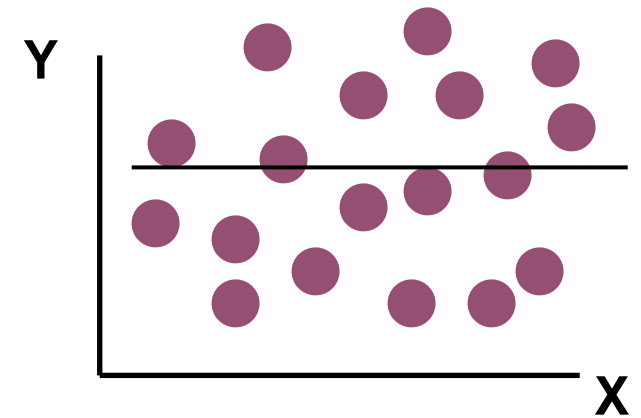
Strong relationships



Weak relationships



No relationship



- Analisis regresi digunakan untuk:
 - Memprediksi nilai variabel dependen berdasarkan nilai dari setidaknya satu variabel independent
 - Menjelaskan dampak perubahan variabel independen terhadap variabel dependen

Variabel dependen: variabel yang ingin kita prediksi atau jelaskan

Variabel independen: variabel yang digunakan untuk memprediksi atau menjelaskan variabel dependen

Model Regresi Linier Sederhana

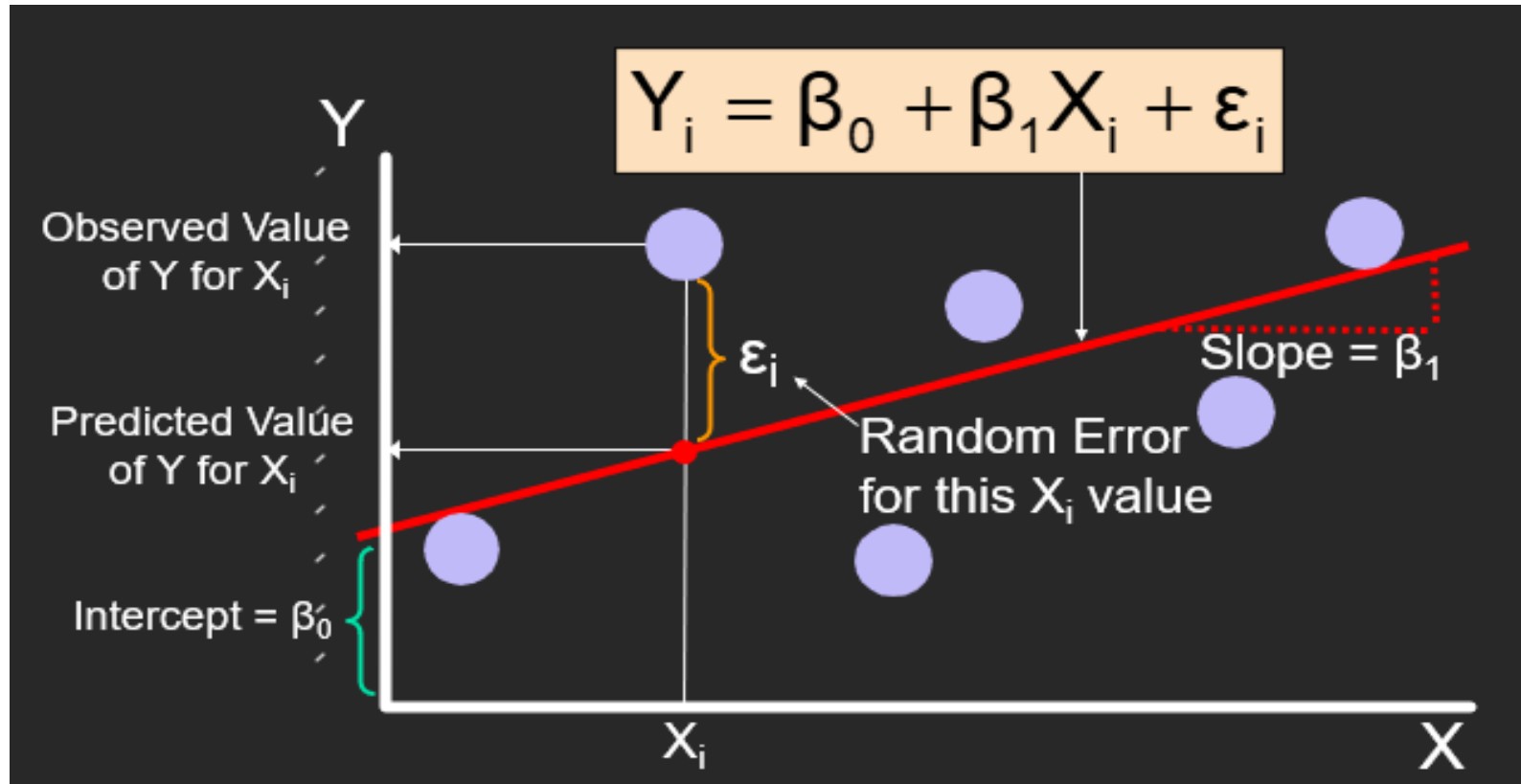
- Model regresi linier sederhana (*simple linear regression model*) melibatkan hanya satu variabel bebas yaitu, X
- Hubungan antara X dan Y digambarkan dengan fungsi linier
- Perubahan Y diasumsikan berhubungan dengan perubahan X

The diagram shows the equation $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ inside a light orange box. Red arrows point from labels to specific parts of the equation: 'Dependent Variable' points to Y_i , 'Population Y intercept' points to β_0 , 'Population Slope Coefficient' points to β_1 , 'Independent Variable' points to X_i , and 'Random Error term' points to ε_i . Below the box, two purple curly braces group the terms: the first brace under $\beta_0 + \beta_1 X_i$ is labeled 'Linear component', and the second brace under ε_i is labeled 'Random Error component'.

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

Linear component Random Error component

Model Regresi Linier Sederhana



Model Regresi Linier Sederhana

Persamaan regresi linier sederhana memberikan perkiraan garis regresi populasi

Estimasi (atau
prediksi) nilai Y
untuk
observasi i

Estimasi
intercept regresi

Estimasi kemiringan
(slope) regresi

Nilai X untuk
observasi i

$$\hat{Y}_i = b_0 + b_1 X_i$$

- b_0 : estimasi nilai rata-rata Y ketika nilai X adalah nol
- b_1 : perkiraan perubahan nilai rata-rata Y sebagai akibat dari peningkatan satu unit di X

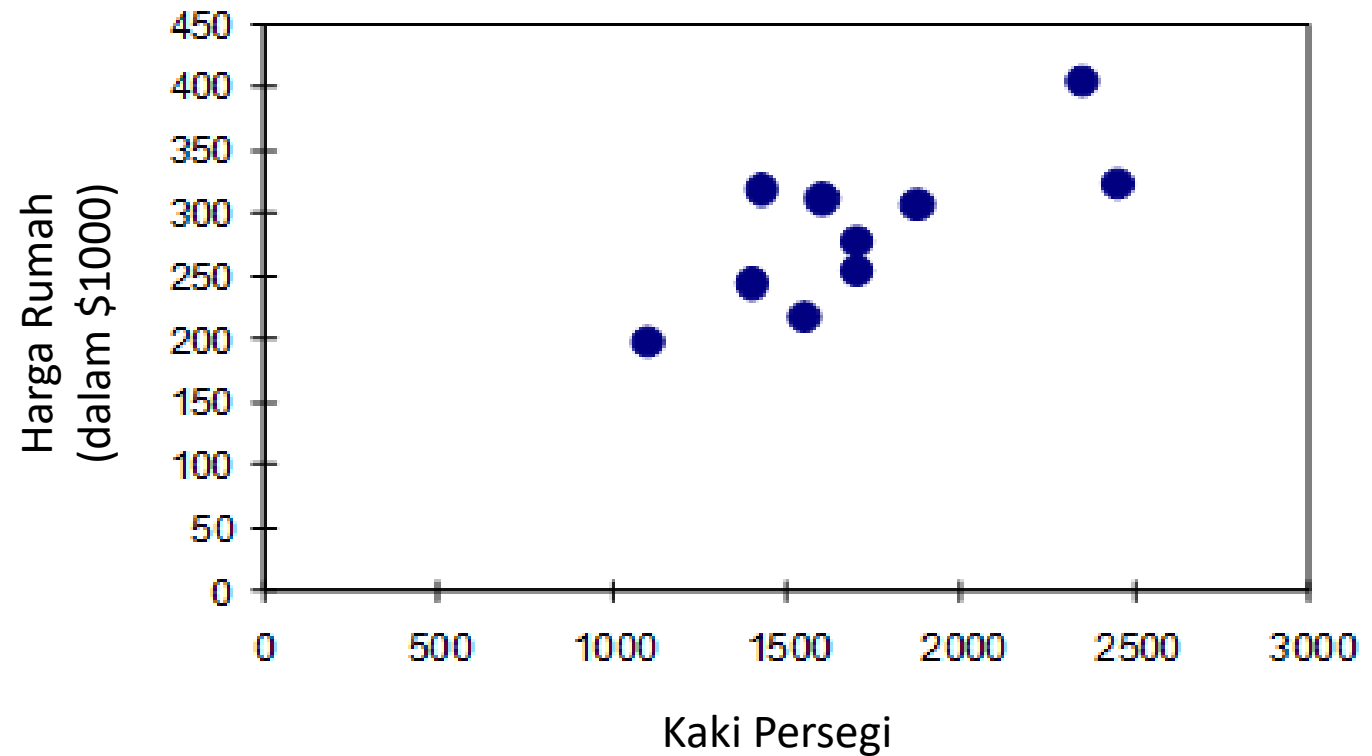
Contoh Soal :

- Seorang agen *real estate* ingin memeriksa hubungan antara harga jual rumah dan luasnya (diukur dalam kaki persegi (*square feet*))
- Sebuah sampel acak dari 10 rumah dipilih
- Variabel dependen (Y) = harga rumah (dalam *range* \$1000)
- Variabel independen (X) = luas rumah (dalam kaki persegi (*square feet*))

Contoh Soal :

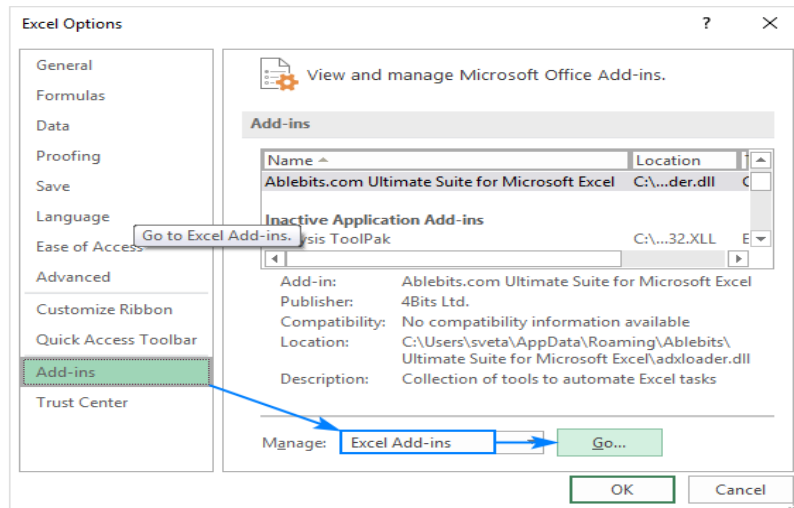
Harga Rumah (dalam \$1000) (Y)	Luas Rumah (dalam kaki persegi) (X)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

- Model harga rumah: Scatter Plot

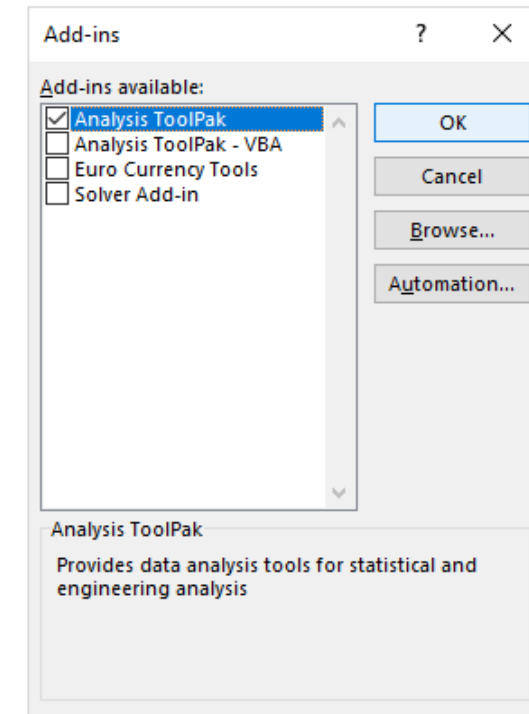


Menambahkan Add-Ins Toolpack di Ms. Excel

- Buka Ms. Excel, kemudian :
 - Klik File → Options
 - Pilih Add-ins di sidebar kiri, pastikan Excel Add-ins dipilih di kotak Manage, dan klik Go.



- Pada kotak dialog Add-ins, centang Analysis Toolpak dan klik OK



- *Skip Langkah ini apabila Analysis Toolpak sudah terinstall*

Penyelesaian :

1. Pilih (klik) Data

2. Pilih Data Analysis

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. A red arrow points to the 'Data Analysis' button in the 'Analysis' group. Below the ribbon, the 'Data Analysis' task pane is open, showing a list of analysis tools. A red arrow points to 'Regression' at the bottom of the list.

no	harga rumah	Kaki Persegi
1	245	1400
2	312	1600
3	279	1700
4	308	1875
5	199	1100
6	219	1550
7	405	2350
8	324	2450
9	319	1425
10	255	1700

3. Pilih (klik) Regression

Masukkan rentang Y dan rentang X dan opsi yang diinginkan

The screenshot shows the 'Regression' dialog box. The 'Input Y Range' is set to '\$B\$1:\$B\$11' and the 'Input X Range' is set to '\$C\$1:\$C\$11'. The 'Labels' checkbox is checked. The 'Confidence Level' is set to 95%. Under 'Output options', 'New Worksheet Ply:' is selected. Under 'Residuals', the 'Residuals' checkbox is checked. The 'OK' button is highlighted with a red box.

no	harga rumah	Kaki Persegi
1	245	1400
2	312	1600
3	279	1700
4	308	1875
5	199	1100
6	219	1550
7	405	2350
8	324	2450
9	319	1425
10	255	1700



Penyelesaian : Hasil Analisis Regresi

	A	B	C	D	E	F	G	H	I	J
1	SUMMARY OUTPUT									
2										
3	Regression Statistics									
4	Multiple R	0,762114								
5	R Square	0,580817								
6	Adjusted R S	0,528419								
7	Standard Err	41,33032								
8	Observation	10								
9										
10	ANOVA									
11		df	SS	MS	F	Significance F				
12	Regression	1	18934,93	18934,93	11,08476	0,010394016				
13	Residual	8	13665,57	1708,196						
14	Total	9	32600,5							
15										
16		Coefficients	Standard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%	
17	Intercept	98,24833	58,03348	1,69296	0,128919	-35,57711198	232,0738	-35,5771	232,0738	
18	Kaki Persegi	0,109768	0,032969	3,329378	0,010394	0,033740065	0,185795	0,03374	0,185795	
19										

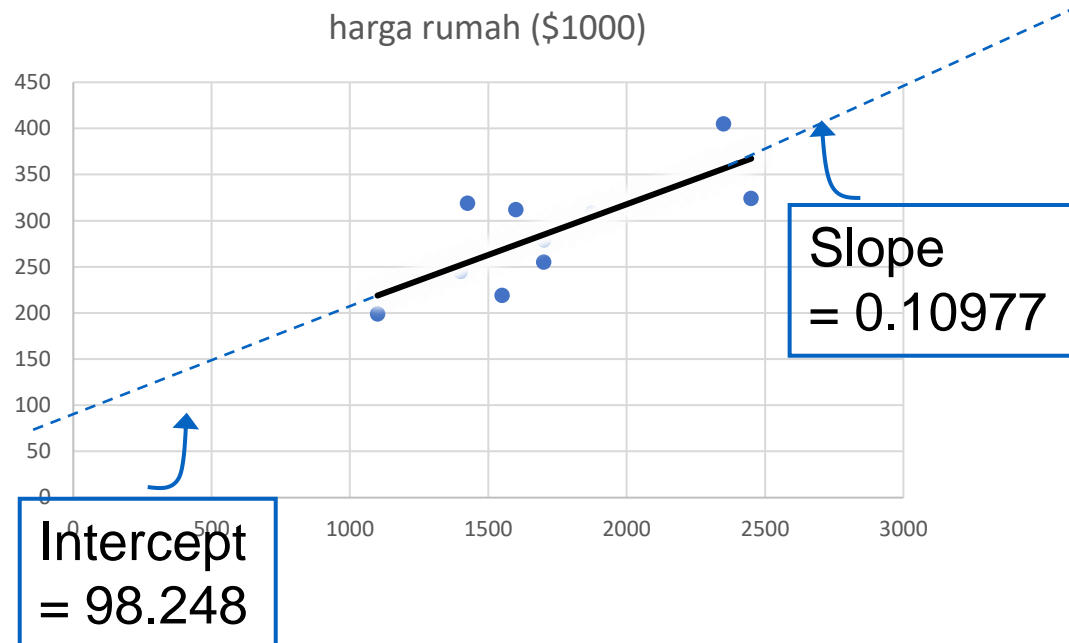
Persamaan regresi adalah:

$$Y = 98.24833 + 0.10977X$$

$$\text{Harga Rumah} = 98.24833 + 0.10977 (\text{Luas Rumah})$$



Scatter Plot and Prediction Line



$$\text{Harga Rumah} = 98.24833 + 0.10977 (\text{Luas Rumah})$$

Interpretasi dari b_0

- b_0 estimasi nilai rata-rata Y ketika nilai X adalah nol (jika $X = 0$ berada dalam rentang nilai X yang diamati)
- $\text{Harga Rumah} = 98.24833 + 0.10977 \times 0 = 98.24833$
- Nilai b_0 menggambarkan harga awal rumah sebelum mempertimbangkan luas rumah. Namun karena sebuah rumah tidak dapat memiliki luas persegi 0, b_0 tidak memiliki aplikasi praktis

Interpretasi dari b_1

- b_1 memperkirakan perubahan nilai rata-rata Y sebagai akibat dari peningkatan satu unit X
- Di sini, $b_1 = 0.10977$ memberi tahu kita bahwa nilai rata-rata sebuah rumah meningkat sebesar $0.10977(\$1000) = \$109,77$, secara rata-rata, untuk setiap tambahan ukuran luas 1 kaki persegi.

Prediksi harga rumah dengan luas 2000 kaki persegi:

$$\begin{aligned}\text{Harga rumah} &= 98.25 + 0.1098 (\text{kaki persegi}) \\ &= 98.25 + 0.1098 (2000) \\ &= 317.85\end{aligned}$$

Prediksi harga rumah dengan luas 2000 kaki persegi adalah $317,85(\$1.000) = \317.850



Menafsirkan Keluaran

Akurasi Model

- **Multiple R** adalah Koefisien Korelasi antara Y dan X.
- **Rsquare** adalah Koefisien Determinasi, yang digunakan sebagai indikator kecocokan. Ini menunjukkan berapa banyak poin yang jatuh pada garis regresi.
 - Pada output, R Square adalah 0,58 (dibulatkan), artinya model hanya 58% dari nilai kita yang sesuai dengan model analisis regresi..
 - Dengan kata lain, 58% variabel dependen (nilai y) dijelaskan oleh variabel independen (nilai x), sehingga cukup baik.
- **Adjusted R Square** adalah kuadrat R yang disesuaikan dengan jumlah variabel independen dalam model. Nilai ini Anda gunakan sebagai ganti R square untuk analisis regresi berganda.
- **Standar Error** menunjukkan ketepatan analisis regresi. Semakin kecil angkanya, semakin baik.

Regression Statistics	
Multiple R	0,76211371
R Square	0,58081731
Adjusted R Square	0,52841948
Standard Error	41,3303237
Observations	10

Latihan :

Data Jumlah Biaya Iklan dan Pendapatan adalah seperti berikut:

Hitunglah persamaan regresi linier dari data tabel di samping dengan menggunakan fungsi data analysis di Ms. Excel dan gambarkan diagram scatter plot-nya!

No	Biaya Iklan (dalam juta) (x)	Pendapatan (dalam milyar) (y)
1	15	65
2	13	61
3	14	60
4	15	63
5	13	59
6	14	62
7	12	56
8	14	61
9	13	60
10	12	58