



Linier Regression

Kuswanto-2012

REGRESSION



- Regresi → menganalisis bentuk hubungan antara 2 variabel, dimana salah satu variabel mempengaruhi yang lain
- Contoh hubungan antara dosis pupuk N dengan hasil panen padi
- We can say - how much change in one for each unit change in the other.
- We call the dependant variable Y and the independent variable X.

Dependent VS independent variable

Indep. Var. (X)	Dep. Var (Y)	Regression Equation
Fertilizer doses	Yield	$y = b_0 + b_1x$
Evaporation	Rain fall	$y = b_0+b_1x+b_2x^2$
Sum of Leave	pod yield	
Etc.....		

Regression vs Correlation

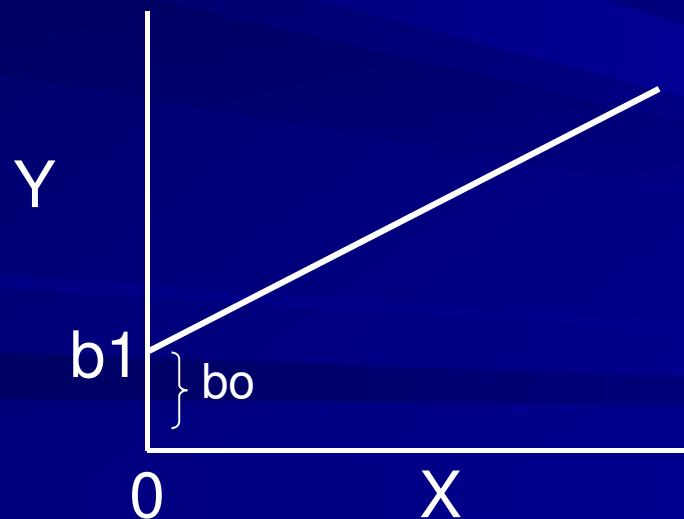
- The relationship between correlation and regression analysis is intertwined (saling terkait).
- Strictly speaking regression analysis is done where one variable is dependent on the other variable.
- In cases where both variables are independent (leave number, plant height, example), correlation analysis is appropriate.

Bentuk umum persamaan regresi

- In general, the equation for a straight line is :

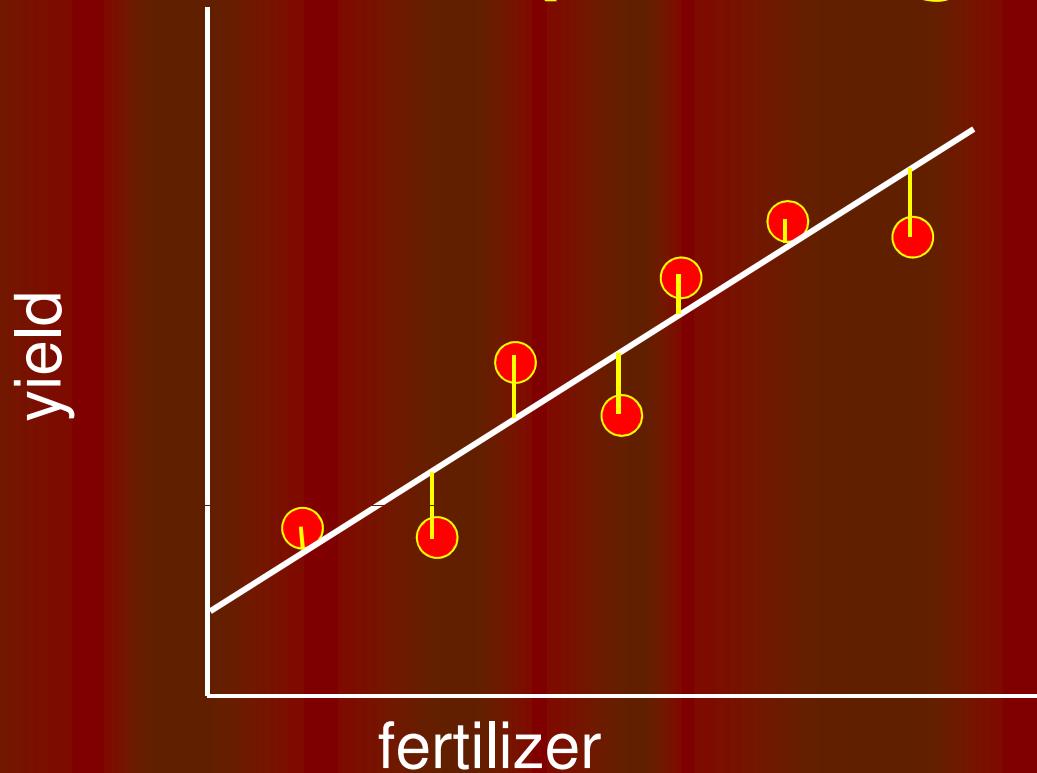
$$Y = b_0 + b_1 X + e$$

- where b_0 = intercept (set $X = 0$) (some books written a)
- b_1 = slope of line Or amount of change in Y with each unit change in X , Also called regression coefficient



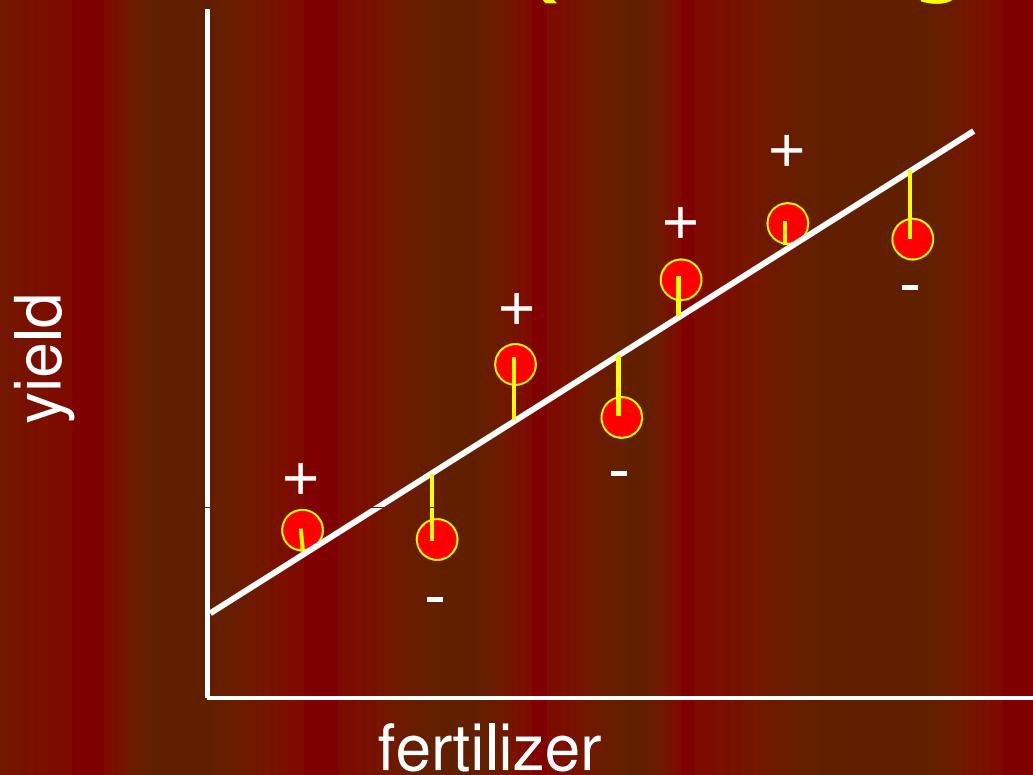
$$Y = b_0 + b_1 X$$

Error (Galat regresi)



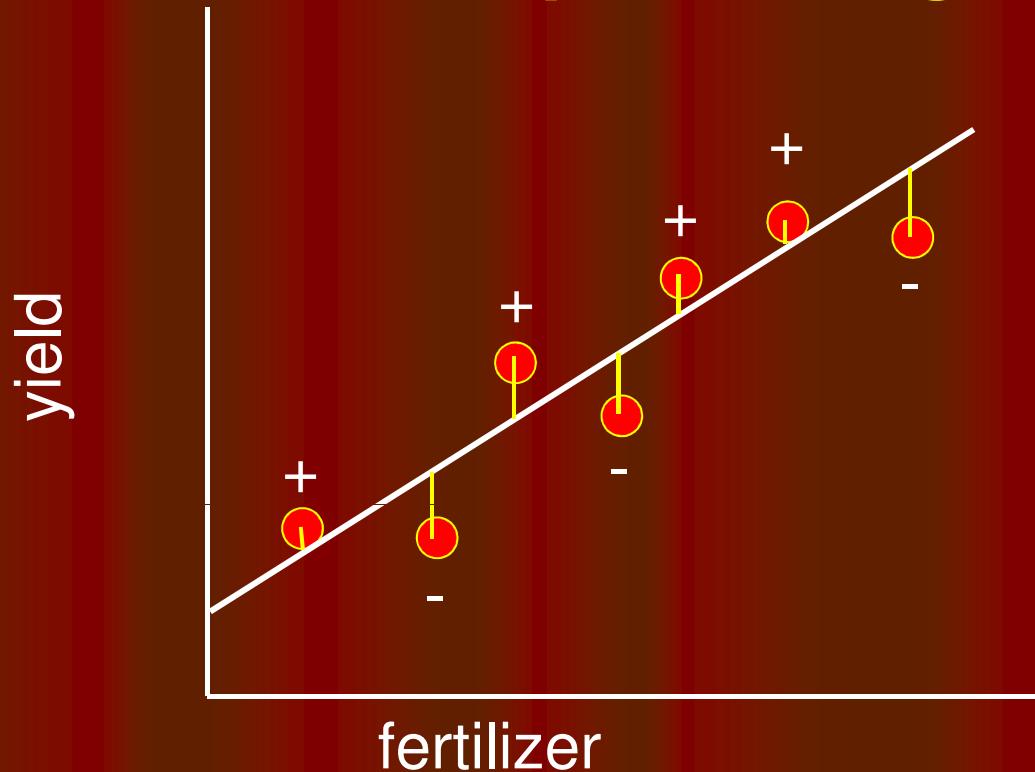
- Since all data is not exactly on the straight line, we use the equation $Y = b_0 + b_1X + \epsilon$
- Where ϵ is error or **residual**

Error (Galat regresi)



The distance from each point to the regression line is known as the residual.

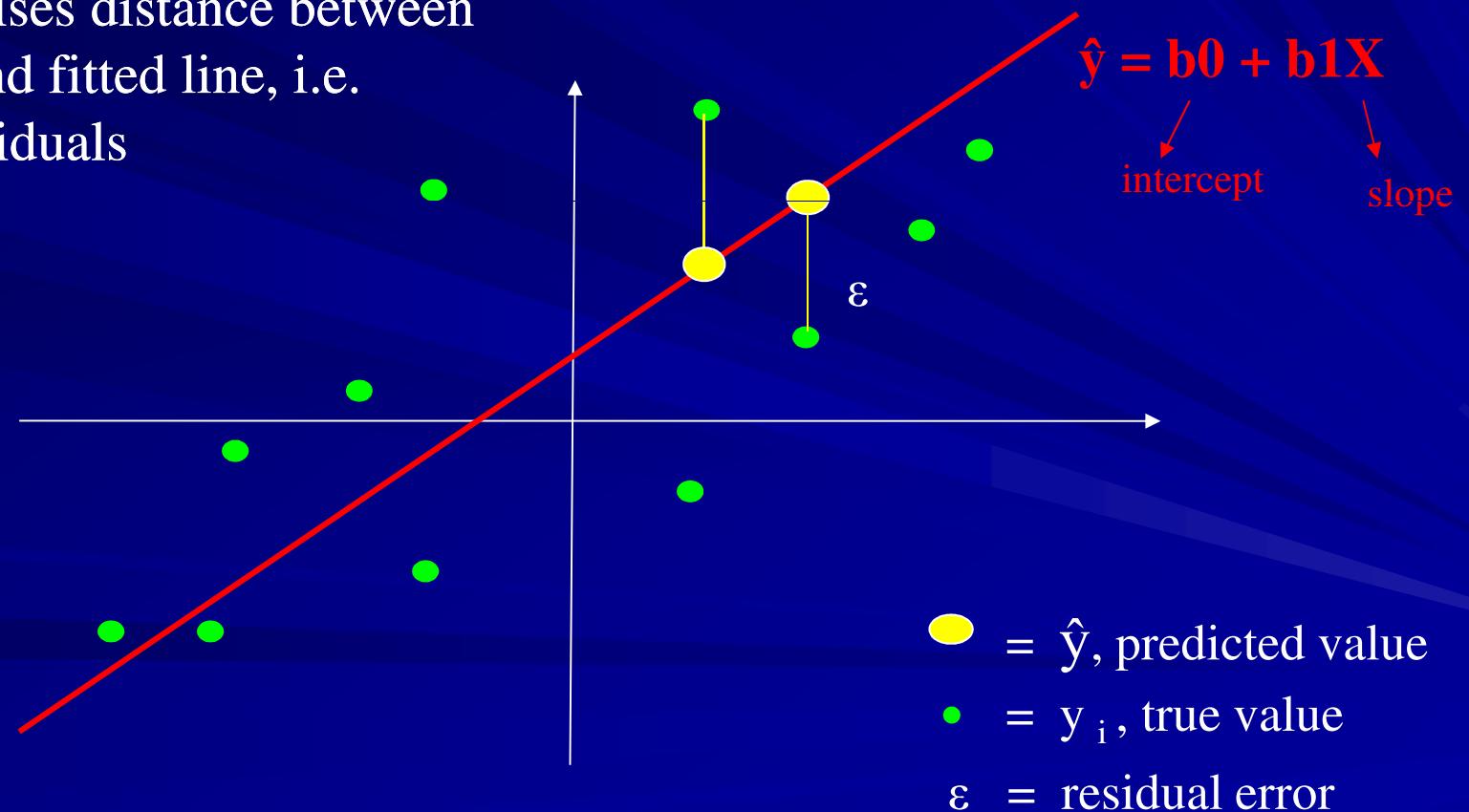
Error (Galat regresi)



- Since some of these deviations are positive and some negative, we use the sum of their squares (just like with SD and variance).
- Known as **least squares method**.

Best-fit Line

- Aim of linear regression is to fit a straight line, $\hat{y} = b_0 + b_1X$, to data that gives best prediction of y for any value of x
- This will be the line that minimises distance between data and fitted line, i.e. the residuals



Cara menghitung b_0 , b_1

b_0 = intersep,
 b_1 = koefisien regresi



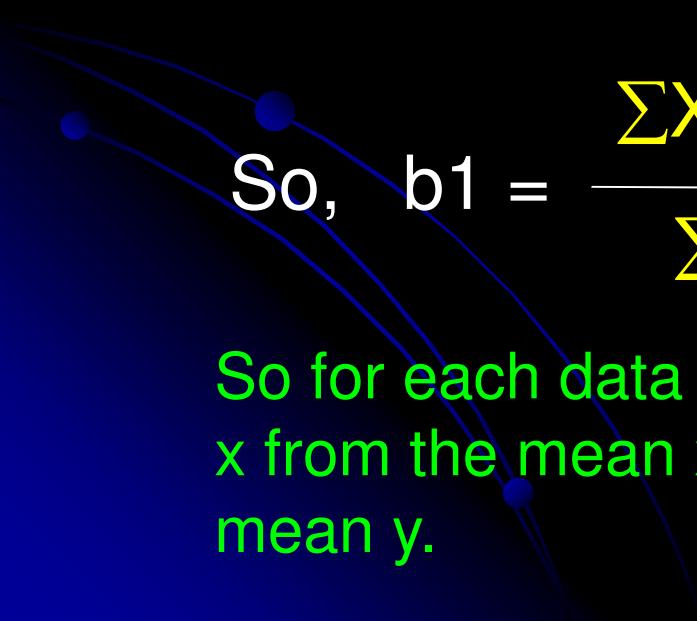
Menghitung Koefisien Regresi b1

The regression coefficient (slope) is calculated by

$$b_1 = \frac{\sum xy}{\sum x^2}$$

Where $\sum xy = \sum (x_i - \bar{x})(y_i - \bar{y})$
 $= \sum XY - \{(\sum X)(\sum Y)\}/n$

And $\sum x^2 = \sum (x_i - \bar{x})^2$
 $= \sum X^2 - (\sum X)^2/n$

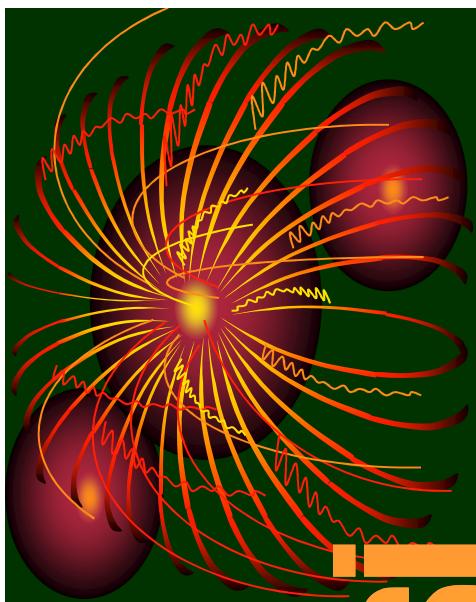


So, $b_1 = \frac{\sum XY - \{(\sum X)(\sum Y)\}/n}{\sum X^2 - (\sum X)^2/n}$

So for each data point (x, y) multiply the deviation of x from the mean \bar{x} by the deviation of y from the mean \bar{y} .

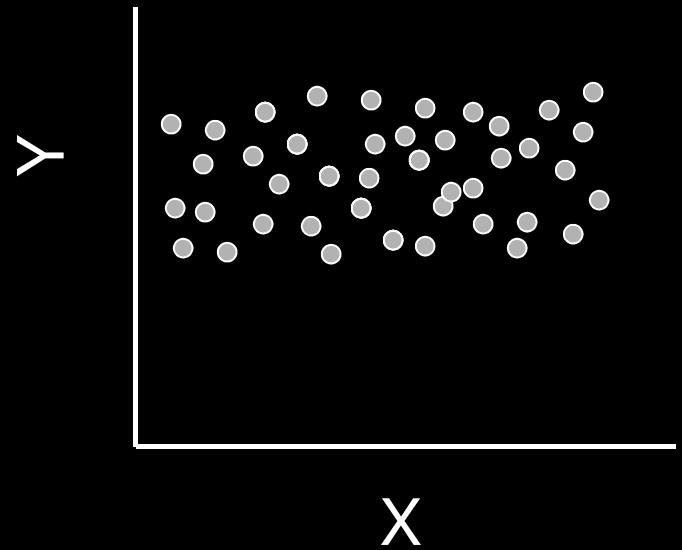
Menghitung intersep (b_0)

- It can also be shown that the point (X mean, Y mean) always lies on the best fit line.
 - $Y = b_0 + b_1X$
 - Substituting, we get $\bar{Y} = b_0 + b_1\bar{X}$
- Therefore, to calculate intercept, a rearrange:
$$b_0 = \bar{Y} - b_1\bar{X}$$
- Now, we have formulas to calculate b_0 and b_1



Anova regression

Perhatikan gugusan data pada diagram pencar ini

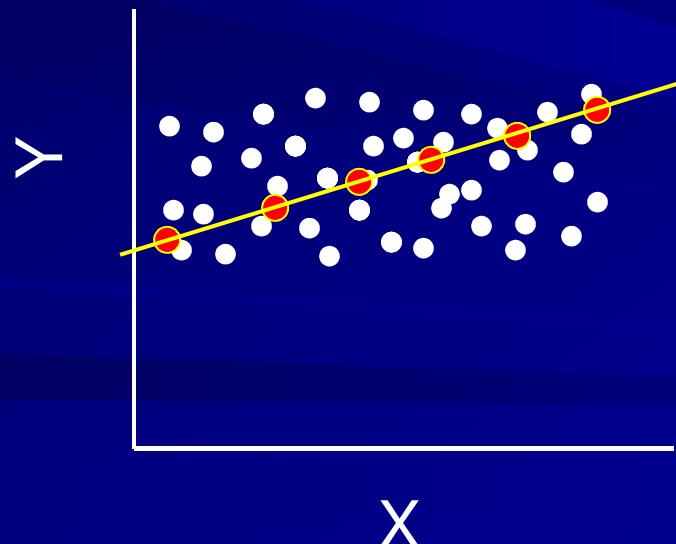


Dari gambar tersebut sebenarnya X tidak mempengaruhi Y, karena garis yang paling tepat Untuk menggambarkan adalah garis lurus, karena kemungkinan nilai $b_1=0$

What is the chance of selecting at random a set of points where the slope (b_1) does not = 0

Uji nyata regresi

- Untuk menguji bentuk hubungan yang digambarkan oleh persamaan regresi → ANOVA testing may be done.
Anova → untuk menguji taraf nyata regresi
- So Ho: $b_1 = 0$ (that X does not affect Y)
HA: $b_1 \neq 0$ (that X does affect Y)



$$F = \frac{KT \text{ Regresi}}{KT \text{ galat}}$$

BENTUK ANOVA REGRESI

<i>SK</i>	<i>Db</i>	<i>JK</i>	<i>KT</i>	<i>Fhit</i>
Regresi	1	$b_1 \sum x_i y_i$	KTr	KTr/KTe
Galat	n-2	JK sisa	KTe	
Total	n-1	$\sum y_i^2$		

$$\text{Dimana, JK total} = \sum y_i^2 = \sum (Y_i - \bar{Y})^2 = \sum Y^2 - (\sum Y)^2/n$$

$$\begin{aligned}\text{Dimana, JK regresi} &= b_1 \sum xy = b_1 \sum (X_i - \bar{X})(Y_i - \bar{Y}) \\ &= b_1 \{ \sum XY - (\sum X)(\sum Y)/n \}\end{aligned}$$

And the rest of the variation is the residuals (source = residual)
JK sisa = JK total – JK Regresi

Koefisien Determinasi

- Koefisien determinasi adalah rasio JK regresi terhadap JK total dan dilambangkan dengan R^2

$$R^2 = \frac{JK_{\text{regresi}}}{JK_{\text{Total}}}$$

- Dapat juga dikatakan sebagai keragaman pada Y yang disebabkan oleh X
- This tells you how much of the variation your regression equation is explaining
- R^2 is not well suited to use as a "goodness of fit" measurement - should not be used to compare regressions with different sum of independent variables or different amounts of replication

helleria content

Cara mencari persamaan regresi

- ◆ Data disamping adalah bobot buah semangka pada berbagai umur buah
- ◆ Carailah persamaan regresi, tabel anova, berikan kesimpulan dan interpretasi datanya
- ◆ Berdasarkan rumus tsb, dapat dibuat tabel dibawah untuk membantu perhitungan

$$b_1 = \frac{\sum XY - \{(\sum X)(\sum Y)\}/n}{\sum X^2 - (\sum X)^2/n}$$

Age (days)	Weight (g)
2	56
4	72
6	99
8	122
10	152
12	170
14	180

Cara perhitungan

No	Age (X)	Weight (Y)	XY	X ²	Y ²
1	2	56			
2	4	72			
3	6	99			
4	8	122			
5	10	152			
6	12	170			
7	14	180			
Sum					
Mean					

Buat tabel untuk menghitung jumlah tiap-tiap komponen. Kemudian isilah tabel tersebut. Hitung perkalian X dengan Y

Cara perhitungan

No	Age (X)	Weight (Y)	XY	X ²	Y ²
1	2	56	112		
2	4	72	288		
3	6	99	594		
4	8	122	976		
5	10	152	1520		
6	12	170	2040		
7	14	180	2520		
Sum					
Mean					

Hitung kuadrat dari X

Cara perhitungan

No	Age (X)	Weight (Y)	XY	X ²	Y ²
1	2	56	112	4	
2	4	72	288	16	
3	6	99	594	36	
4	8	122	976	64	
5	10	152	1520	100	
6	12	170	2040	144	
7	14	180	2520	196	
Sum					
Mean					

Hitunglah kuadrat dari Y

Cara perhitungan

No	Age (X)	Weight (Y)	XY	X ²	Y ²
1	2	56	112	4	3136
2	4	72	288	16	5184
3	6	99	594	36	9801
4	8	122	976	64	14884
5	10	152	1520	100	23104
6	12	170	2040	144	28900
7	14	180	2520	196	32400
Sum					
Mean					

Hitunglah jumlah masing-masing komponen

Cara perhitungan

No	Age (X)	Weight (Y)	XY	X ²	Y ²
1	2	56	112	4	3136
2	4	72	288	16	5184
3	6	99	594	36	9801
4	8	122	976	64	14884
5	10	152	1520	100	23104
6	12	170	2040	144	28900
7	14	180	2520	196	32400
Sum	56	851	8050	560	117409
Mean					

Hitung nilai rata-rata X dan Y → untuk menghitung intersep

Tabel lengkap, kemudian masukkan ke rumus

No	Age (X)	Weight (Y)	XY	X ²	Y ²
1	2	56	112	4	3136
2	4	72	288	16	5184
3	6	99	594	36	9801
4	8	122	976	64	14884
5	10	152	1520	100	23104
6	12	170	2040	144	28900
7	14	180	2520	196	32400
Sum	56	851	8050	560	117409
Mean	8	121,57			

..... untuk menghitung koefisien regresi dan intersep

Masukkan rumus

$$b_1 = \frac{\sum XY - \{(\sum X)(\sum Y)\}/n}{\sum X^2 - (\sum X)^2/n} = \frac{8050 - \{(56)(851)\}/7}{560 - (56)^2/7} = 11,089$$

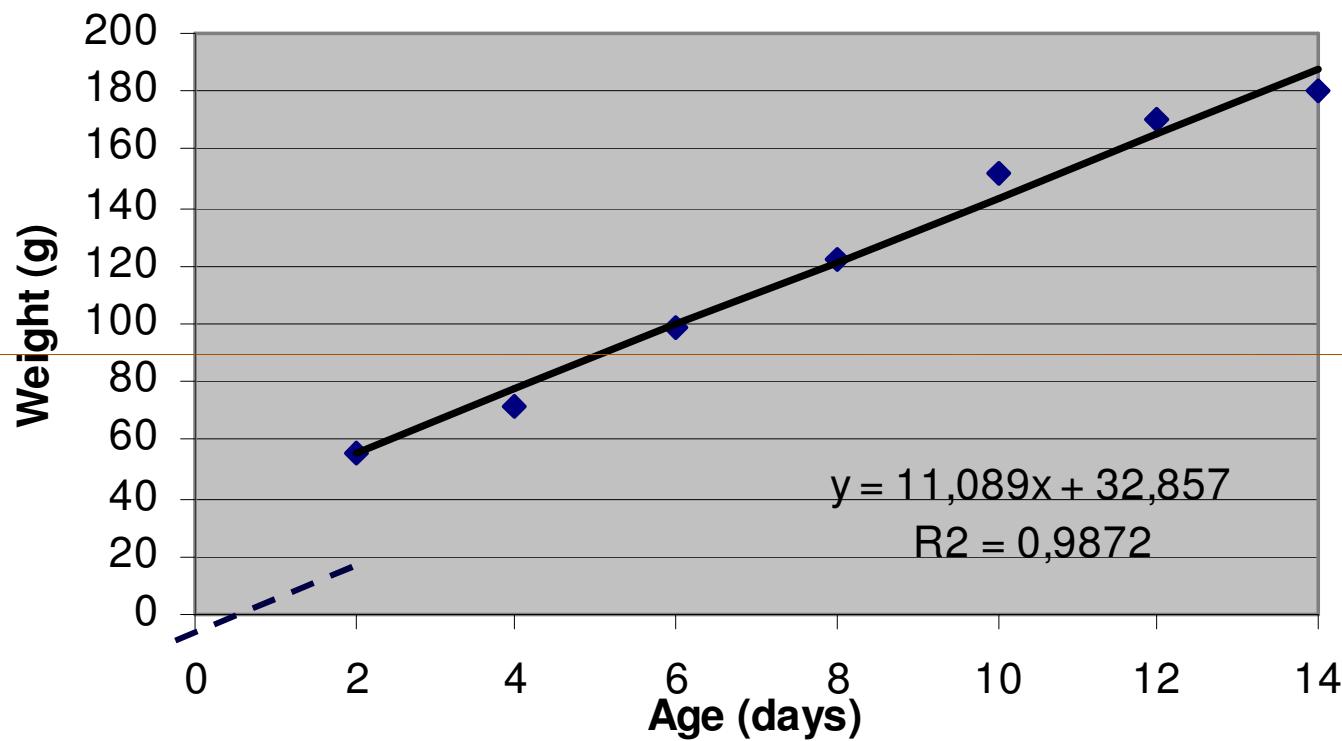
$$\text{Hitung nilai intersep } b_0 = \bar{Y} - b_1 \bar{X} = 121,57 - (11,089) (8) \\ = 32,857$$

Dengan demikian persamaan regresi adalah

$$Y = 32,857 + 11,089 X$$

Dapat juga dituliskan $y = 11,089x + 32,857$. Kemudian Gambarlah garis regresinya →....

Garis Regresi Linier



Gambar garis regresinya

Hitung anova regresi

$$JK \text{ Total} = \sum Y^2 - (\sum Y)^2/n = 117409 - (851)^2/7 = 13951,71$$

$$\begin{aligned} JK \text{ regresi} &= b_1 \{ \sum XY - (\sum X)(\sum Y)/n \} \\ &= 11,089 \times \{ 8050 - (56)(851)/7 \} \\ &= 13.772,89 \end{aligned}$$

$$JK \text{ sisa} = JK \text{ total} - JK \text{ regresi} = 178,82$$

Setelah ketemu, masukkan dalam tabel anova regresi

Tabel anova

Sumber keragaman	Db	JK	KT	F hit
Regresi	1	13.772,89	13.772,89	385,15**
Residu (sisa)	5	178,82	35,76	
Total	6	13.951,71		

Koefisien Determinasi $R^2 = JK \text{ Regresi}/JK \text{ Total}$

$$= 0,9872$$

Kesimpulan dan interpretasi

- Hubungan antara umur buah dengan bobot dapat dituliskan dengan persamaan regresi linier $Y = 32,857 + 11,089X$, artinya setiap penambahan umur satu hari, bobot buah akan bertambah 11,089 gram.
- Dari tabel anova dapat disimpulkan bahwa terdapat hubungan regresi yang sangat nyata antara umur buah dengan bobot buah berdasarkan regresi linier.
- Dari koefisien determinasi dapat disimpulkan bahwa sebanyak 98,72% keragaman berat buah disebabkan oleh pengaruh umur buah

Interpretasi :
perhatikan persamaan ini

$$Y = 32,857 + 11,089X$$

- ◆ What weight would you expect for a age of 5 days?
 - $Y = 32,875 + (11,089)(5) = 88,32 \text{ g}$
- ◆ Pada umur berapa bobot buah akan mencapai 100 g?
 - $100 = 32,875 + 11,089 X$
 - Maka $X = (100 - 32,875) / 11,089 = 6,053 \text{ hari}$
- ◆ Dari rumus tsb, dapat diduga bobot buah dari berbagai umur buah dan sebaliknya. Tentunya sampai batas tertentu saja

Bahan
Diskusi



- Untuk analisis regresi, carilah data independen dan dependen, cari persamaan regresinya, tunjukkan gambarnya, hitunglah anovanya, hitung koefisien determinasinya, dan berikan kesimpulannya