

# Linier Regression

Statistik (MAM 4137)

Ledhyane I. Harlyan

# TIK (TUJUAN INSTRUKSIONAL KHUSUS)

- Mahasiswa mampu melakukan analisis regresi sederhana dengan menggunakan metode kuadrat galat terkecil

# History

*“Peramalan”*

*“Meramalkan suatu peubah tak bebas (Y) lewat satu/lebih peubah bebas (X)”* → **Persamaan REGRESI**

Contoh: - Hubungan nilai UTS dengan skor intelegensia

- Hubungan jumlah pendapatan dengan tingkat kepuasan
- Menduga kecerahan air dari konsentrasi klorofil
- Hubungan hasil tangkapan per unit effort dengan effort, dll.

## ***HUBUNGAN FUNGSI***

..”konsentrasi klorofil dapat dihubungkan dgn kecerahan air..”

→ Jika konsentrasi klorofil=C; kecerahan air=D

→  $C = f(D)$  artinya: D digunakan sebagai indikator C

Manakah peubah/variabel bebas?? Manakah peubah tak bebas?

**“D” atau “C”**

# *X dan Y dalam populasi*

X (peubah bebas) dan Y (peubah tak bebas)  $\rightarrow$  adalah anggota dari populasi

Notasi dlm populasi :  $\{x_i, y_i\}; i = 1, 2, 3 \dots n \rightarrow (x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$

- Jika data tiap anggota populasi diplotkan / disebarakan (*Diagram Pencar*)  
Analisa relasi antar variabel adalah dengan membuat diagram pencar (*scatter diagram*) yang menggambarkan titik-titik plot dari data yang diperoleh.
- Titik-titik akan “mengikuti “ garis lurus  $\rightarrow$  dua peubah (X & Y) berhubungan secara linear (*Garis Regresi Linear*)

# Model Regresi (pangkat 1)

## Model dugaan

$$\hat{y} = a + bx$$

$\hat{y}$  = nilai ramalan hasil dari analisis regresi

a = intercept/perpotongan sumbu tegak

b = slope/kemiringan

**Note:**  $\hat{y} \neq y$  !!

$y$  = nilai pengamatan sesungguhnya



$$\hat{C} = a + b D$$

$\hat{C}$  = peubah tak bebas

D = peubah bebas

a = intercept

b = slope

## Model observasi

$$\dot{y} = \alpha + \beta x$$

$\dot{y}$  = nilai rata-rata observasi

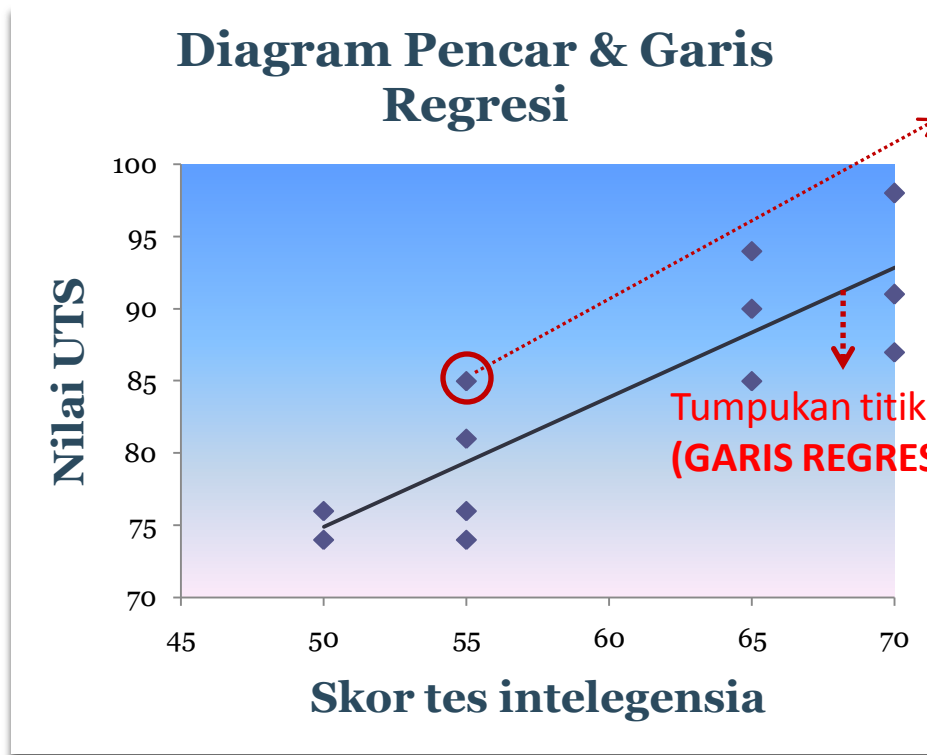
$\alpha$  = intercepts

$\beta$  = slope/kemiringan

$\epsilon_i$  = galat/sisa

$$y_i = \hat{y} + \epsilon_i$$

# Diagram pencar & Garis Regresi



Nilai pengamatan  
sesungguhnya

Tumpukan titik-titik ramalan  $\rightarrow \hat{y} = a + bx$   
(GARIS REGRESI)

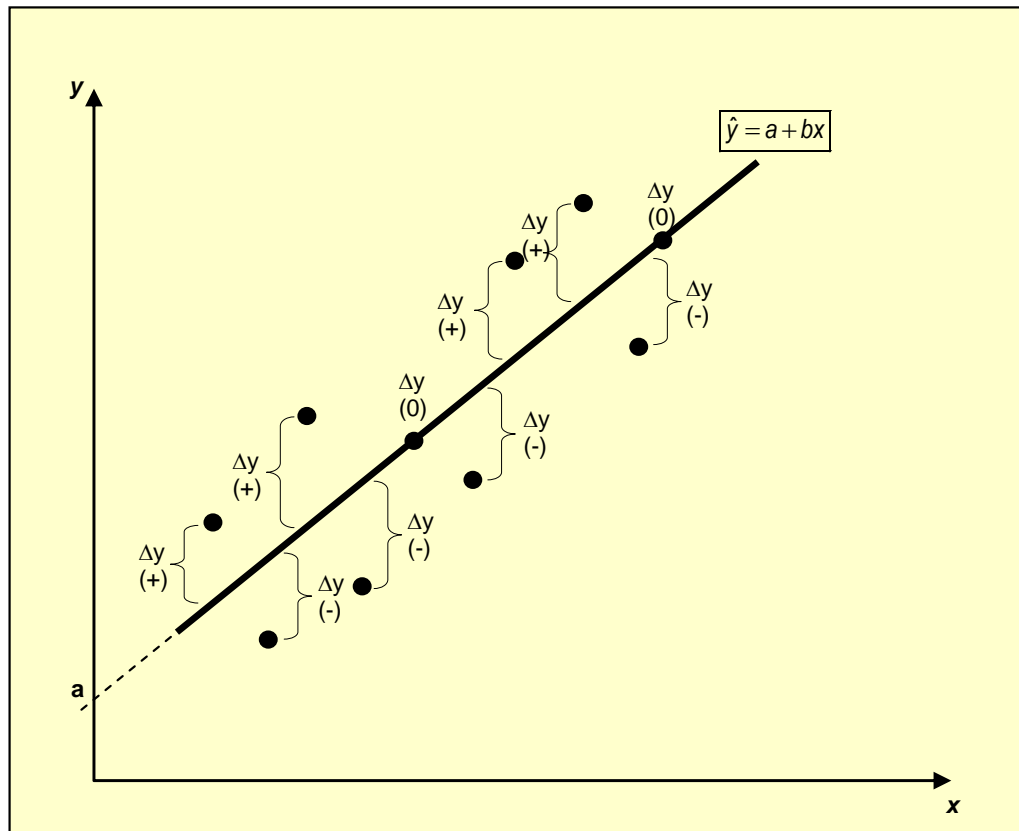
Tumpukan titik2 ramalan/garis regresi  
 $\rightarrow$  digunakan utk peramalan

Misal:

-skor tes 60, maka nilai UTS=83.86

-83.86 adalah nilai harapan bagi mahasiswa  
yg memiliki skor tes 60

# Garis linier pada diagram pencar



Gambar Garis regresi linier pada diagram pencar

# Asumsi Penggunaan Regresi

- $\epsilon_i \sim N(0, \delta^2)$
- $\epsilon_i$  bebas satu sama lain
- Setiap nilai  $x$  mempunyai sebaran bagi nilai  $y$
- $x$  bersifat non measurement error



# Metode Jumlah Kuadrat Galat Terkecil

(Least Squares Method)

merupakan salah satu kriteria yang memenuhi, agar apabila kuadrat dari kesalahan itu dijumlahkan akan se minimum mungkin.

$$b = \frac{\sum XY - \frac{1}{n} \sum X \sum Y}{\sum X^2 - \frac{1}{n} (\sum X)^2}$$

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

# Metode Jumlah Kuadrat Galat Terkecil

Tabel Sidik Ragam / Analysis of Varians (ANOVA) →  
melihat kesesuaian model regresi

Sumber Keragaman	db	JK	KT	Fhit	Ftab	Sig.F
Regresi	1	JKR	KTR	KTR/KTS	dbr,dbs	
Galat	n-2	JKS	KTS	-	-	
Total	n-1	JKT	-	-	-	

$$JKT = \sum Y^2 - \frac{1}{n} (\sum Y)^2$$

$$JKR = b \left( \sum XY - \frac{1}{n} \sum X \sum Y \right)$$

# Kaidah Penarikan Kesimpulan

Hipotesis:

H0: Model regresi dapat dipercaya

H1: Model regresi tidak dapat dipercaya

## Membandingkan hasil F hit dan F tab

- $F_{hit} > F_{tab} 5\%$   
→ Tolak  $H_0$ , terima  $H_1$  artinya model regresi dapat dipercaya dengan selang kepercayaan 95%
- $F_{hit} > F_{tab} 1\%$   
→ Tolak  $H_0$ , terima  $H_1$  artinya model regresi sangat dapat dipercaya dengan selang kepercayaan 99%
- $F_{hit} < F_{tab} 5\%$   
→ Gagal tolak  $H_0$  artinya model regresi tidak dapat dipercaya

# Uji Lanjutan

R<sup>2</sup> (koefisien determinasi)

→ JKR x 100%

JKT

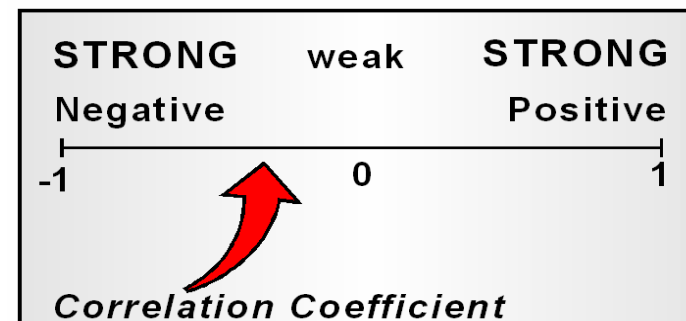
“berapa persen keragaman nilai Y dapat dijelaskan hubungan linearnya dgn X?”

r (koefisien korelasi)

$$r = \sqrt{R^2}$$

→ Nilai rentang:  $-1 \leq r \leq 1$

“seberapa kuat hubungan antara dua peubah (bebas & tak bebas)? “



# Contoh soal

Jika diketahui bahwa lama perendaman (X) akan mempengaruhi kadar protein umpan (gr/100 gr umpan) (Y), maka berikut ini akan dibuktikan bahwa X mempengaruhi Y!

$$\sum X = 975$$

$$\sum X^2 = 192175$$

$$\sum XY = 8259$$

$$\sum Y = 110.7$$

$$\sum Y^2 = 2730.95$$

$$\bar{X} = 139.2857$$

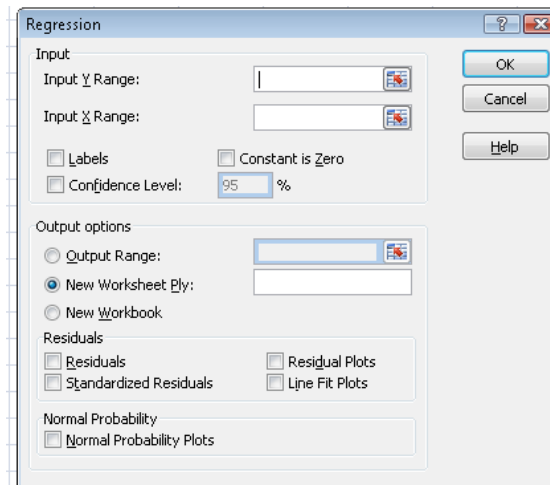
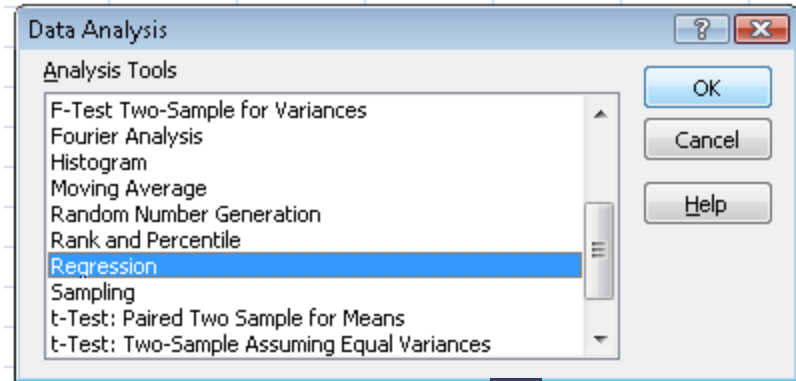
$$\bar{Y} = 15.81$$

$$n = 7$$

# Pengerjaan dengan Ms. Excell

NO	SECCHI DISH (cm) /X	CHLOROPH YLL (TE/F)/ Y
1	45	28,0
2	250	3,2
3	130	14,7
4	270	0,5
5	65	20,4
6	35	35,0
7	180	8,9

- Masuk ke Ms.excell
- Buka data analysis



# Metode Kuadrat Terkecil (Excell)

SUMMARY OUTPUT					
<b>Regression Statistics</b>					
Multiple R	0.963158517				
R Square	0.927674329				
Adjusted R Square	0.913209194				
Standard Error	3.765673261				
Observations	7				
<b>ANOVA</b>					
	<b>df</b>	<b>SS</b>	<b>MS</b>	<b>F</b>	<b>Significance F</b>
Regression	1	909.4070959	909.4070959	64.13175	0.000490536
Residual	5	70.90147554	14.18029511		
Total	6	980.3085714			
	<b>Coefficients</b>	<b>Standard Error</b>	<b>t Stat</b>	<b>P-value</b>	<b>Lower 95%</b>
Intercept	33.50544222	2.627922413	12.74978365	5.28E-05	26.7501526
X Variable 1	-0.127013431	0.015860363	-8.008229944	0.000491	-0.167783792

## INTERPRETASI

Setiap penambahan 1 cm secchi disk, maka konsentrasi klorofil akan berkurang sebesar 0.127 TE/F

# Assignment!!

- Sebuah penelitian dilakukan oleh seorang pedagang eceran untuk menentukan hubungan antara biaya pemasangan iklan (Rp. dalam ribuan) per minggu dan hasil penjualannya.

Data yang diperoleh adalah sebagai berikut :

Biaya iklan (Rp. dalam ribuan)	40	20	25	20	30	50	40	20	50	40	25	50
Penjualan (Rp. dalam ribuan)	385	400	395	365	475	440	490	420	560	525	480	510

Tentukan :

1. Persamaan regresinya ?
2. Perkirakan besar penjualan per minggu bila biaya iklan sebesar Rp. 35.000,- ?
3. Koefisien korelasinya (  $r$  ) ?



Negative



Positive



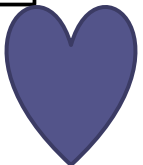
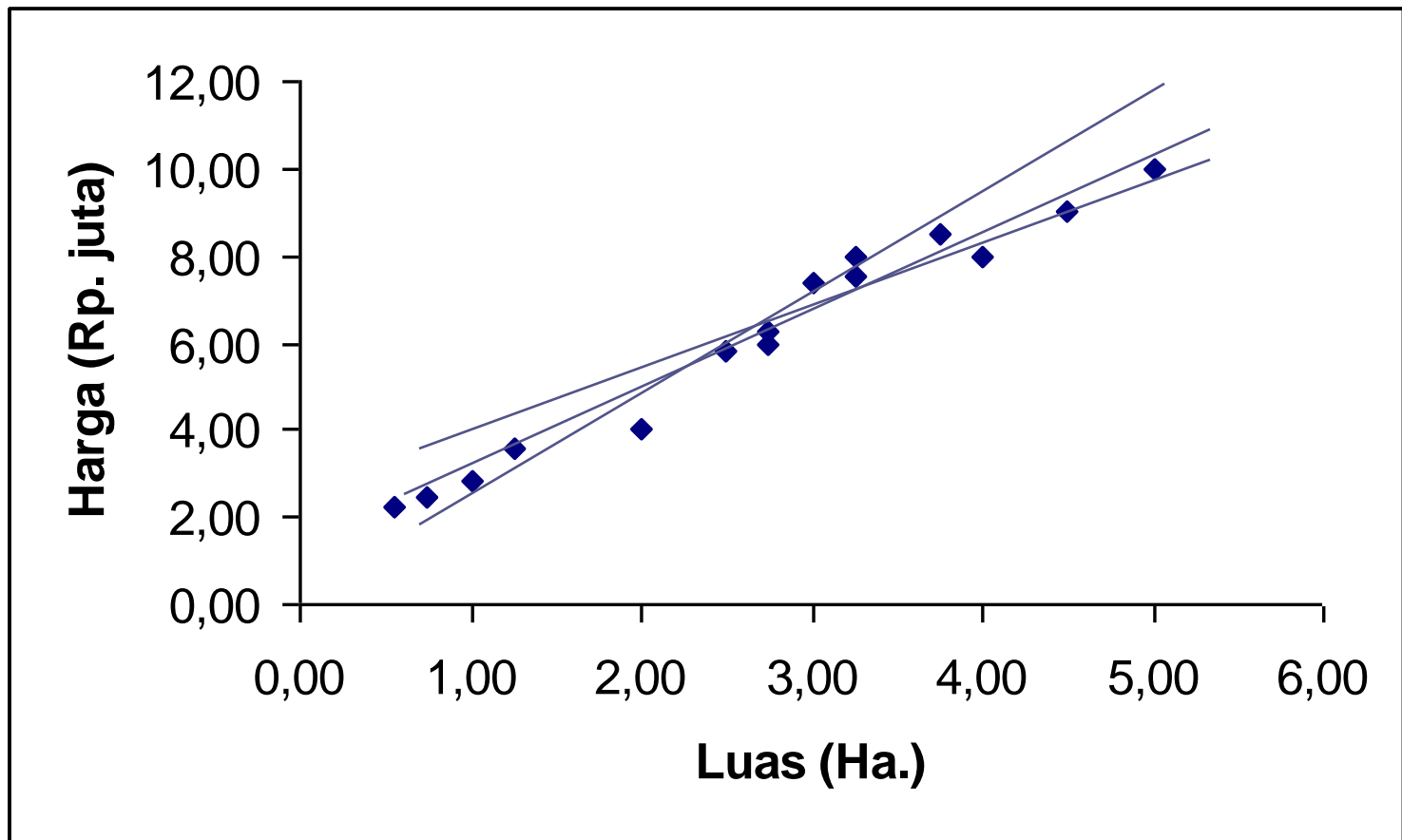
Zero



# Model Dugaan

Mana pendekatan yang baik ?

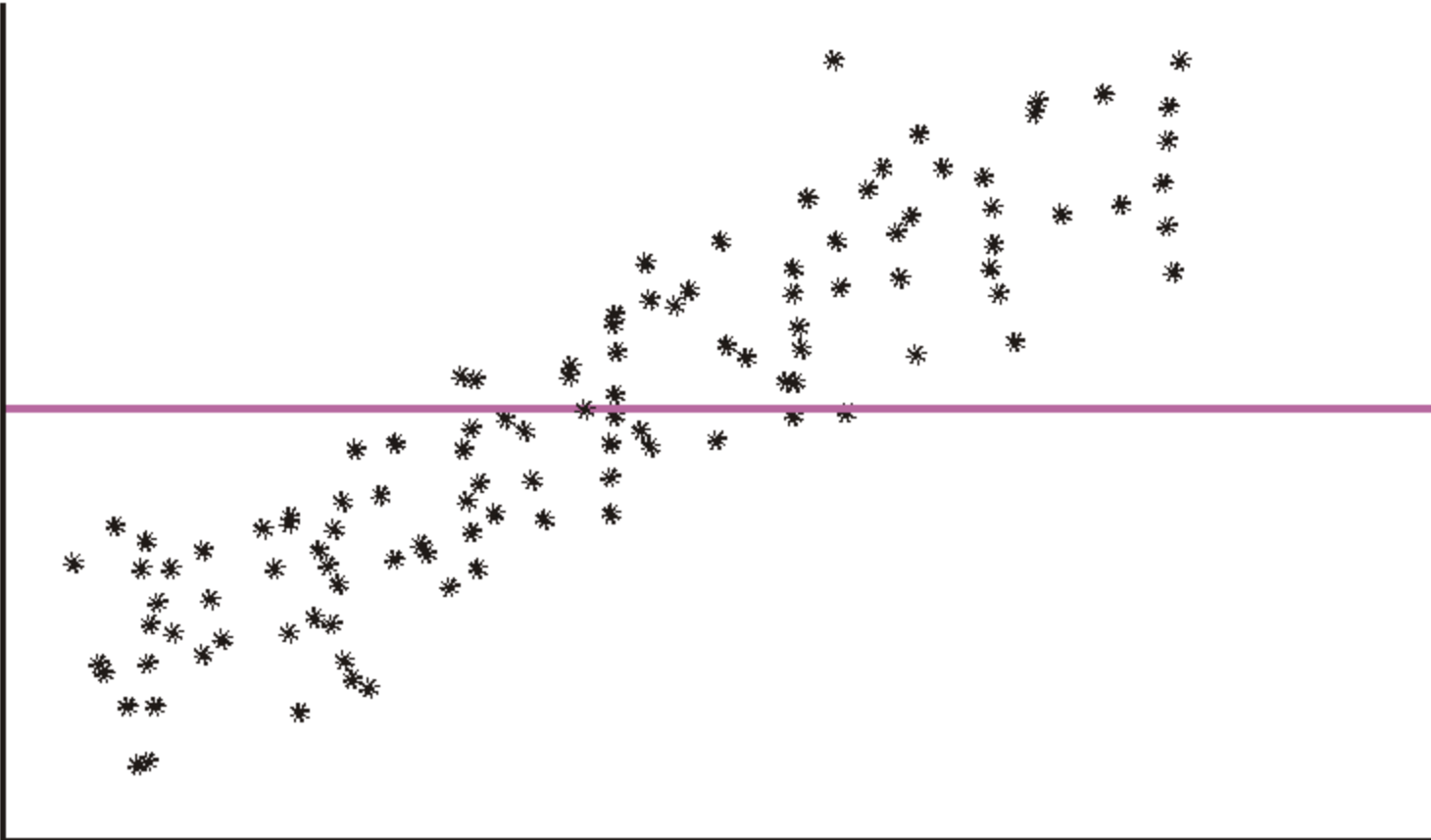
Garis lurus yang sedemikian rupa sehingga melewati seluruh titik (data ) pada diagram pencar → yang mendekati



# Model Observasi

Response (Y)

$\bar{Y}$



Predictor (X)

