Regresi Linear

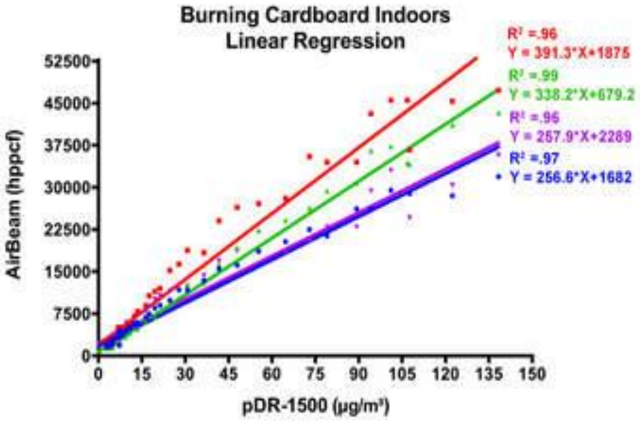
twitter: @kh_notodiputro

email: khairil@apps.ipb.ac.id

Prof. Dr. Ir. Khairil Anwar Notodiputro, MS

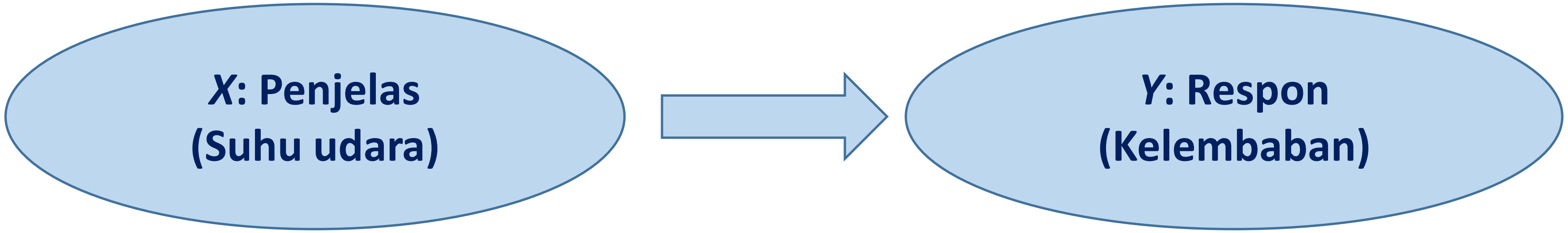
Guru Besar Tetap pada Fakultas Matematika dan Ilmu Pengetahuan Alam

IPB University



Hubungan Dua Kejadian

- Dua kejadian sering kali berhubungan
- Misal: suhu udara vs kelembaban; harga vs permintaan; pupuk vs hasil tanaman; tinggi vs bobot
- Jika dua kejadian itu terukur (skala kategorik maupun kontinu) maka metode statistika dapat digunakan
- Misal dua kejadian itu adalah *X* dan *Y* dan keduanya berskala kontinu (misal tinggi badan dan bobot badan) maka kita sebut *X* = penjelas dan *Y* = respon



- Dalam praktek sering tdk mudah menentukan mana yang berperan sbg penjelas dan mana respon

Email vs FB

- Hubungan penggunaan *email* vs *FB*

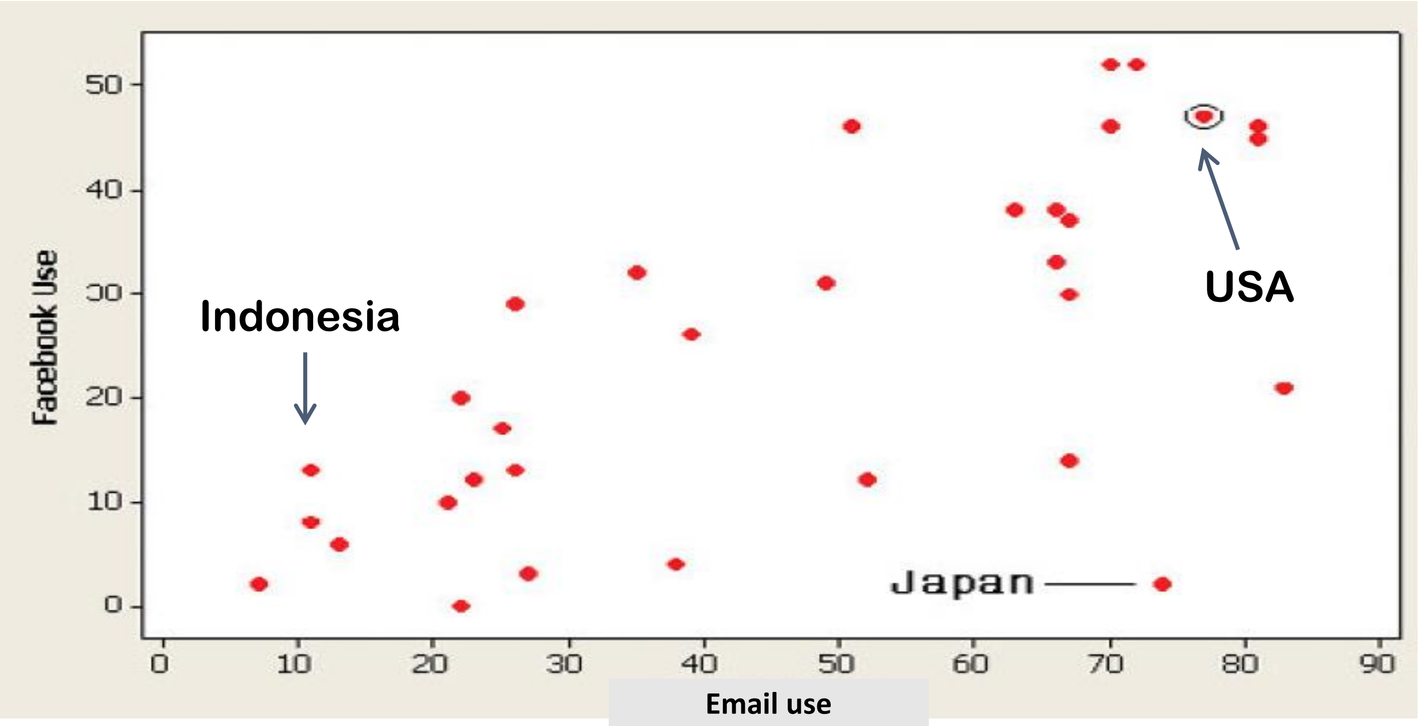
Country	Internet Penetration	Facebook Penetration
Argentina	49.40%	30.53%
Australia	80.60%	46.01%
Belgium	67.30%	36.98%
Brazil	37.76%	4.39%
Canada	72.30%	52.08%
Chile	50.90%	46.14%
China	22.40%	0.05%
Colombia	38.80%	25.90%
Egypt	12.90%	5.68%
France	65.70%	32.91%
Germany	67.00%	14.07%
Hong Kong	69.50%	52.33%
India	7.10%	1.52%
Indonesia	10.50%	13.49%
Italy	48.80%	30.62%
Japan	73.80%	2.00%
Malaysia	62.80%	37.77%
Mexico	24.90%	16.80%
Netherlands	82.90%	20.54%

(Continued)

- Siapa yang punya *FB*?
- Lebih sering *FB* atau email?
- Apakah ada hubungan antara penggunaan email dengan *FB*?
- Apa keunggulan *FB* dibanding email?
- Sekarang perhatikan data internet dan *FB* di dunia

Email vs FB

- Data pengguna email dan FB dalam diagram

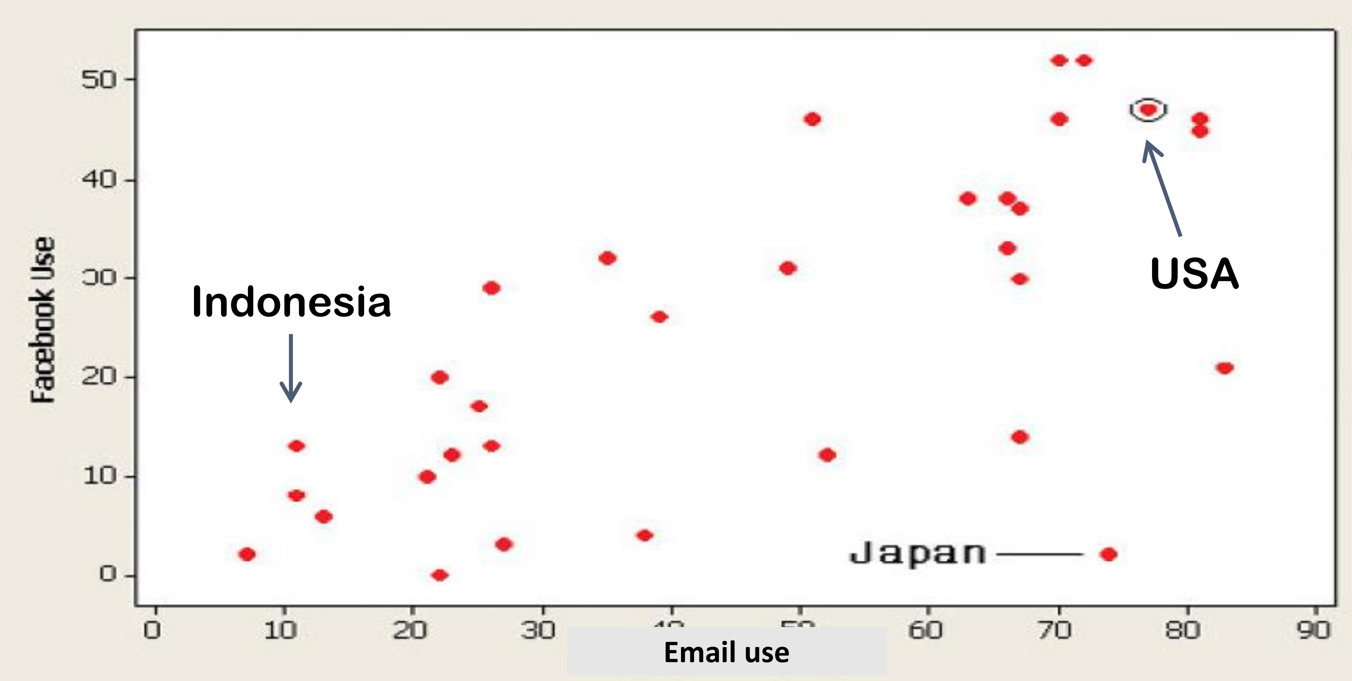


Apa yang terjadi di: USA, Jepang, Indonesia?

- Apa penjelas (*X*)nya? Apa respon (*Y*)nya?
- Bolehkah dibalik posisi penjelas dan respon itu?
- Spt apa bentuk hubungan X dengan Y disini?

Email vs FB

- **Jika dianggap hubungan email dan FB linear bgm?**
- **Ingat : persamaan garis lurus (linear) → *Y = a + bX***



Lazimnya

***a*: intercept, yaitu titik potong garis dengan sumbu tegak**

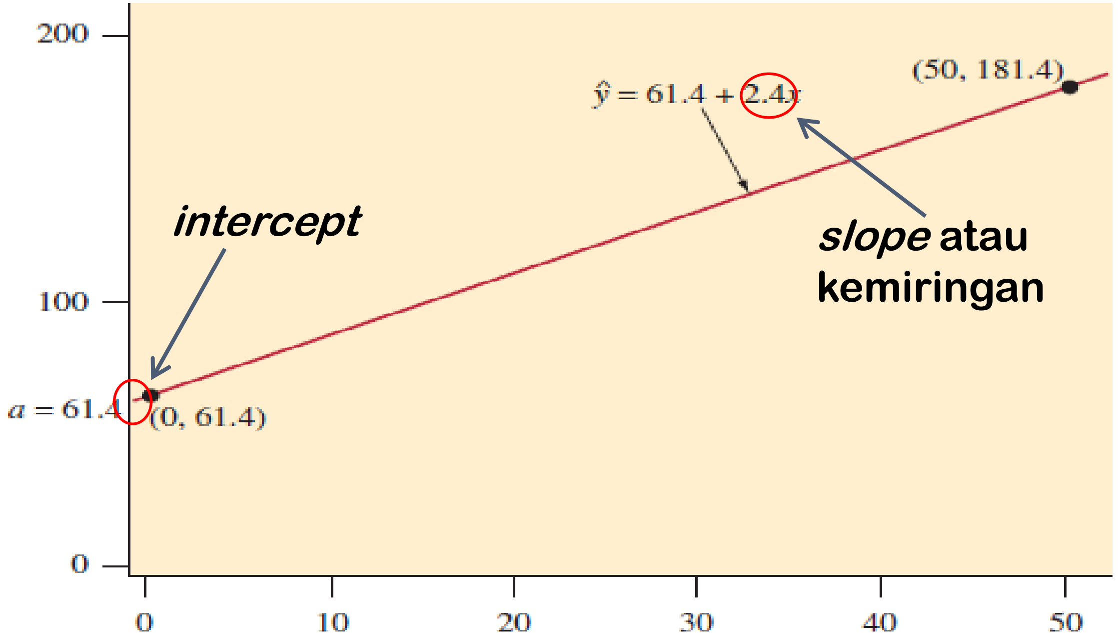
b*: kemiringan garis atau *slope

- **Karena hubungan keduanya tdk sempurna maka kita gunakan model :**

$$FB_i = \alpha + \beta Email_i + \varepsilon_i \qquad (i = 1, 2, \ldots, n)$$

Hubungan Linear

- Perhatikan gambar persamaan garis $y = 61.4 + 2.4\,x$



Data kita tidak sempurna, data kita mengandung galat, shg digunakan

MODEL $Y_i = \alpha + \beta X + \varepsilon_i \; (i=1,2,\ldots,n)$

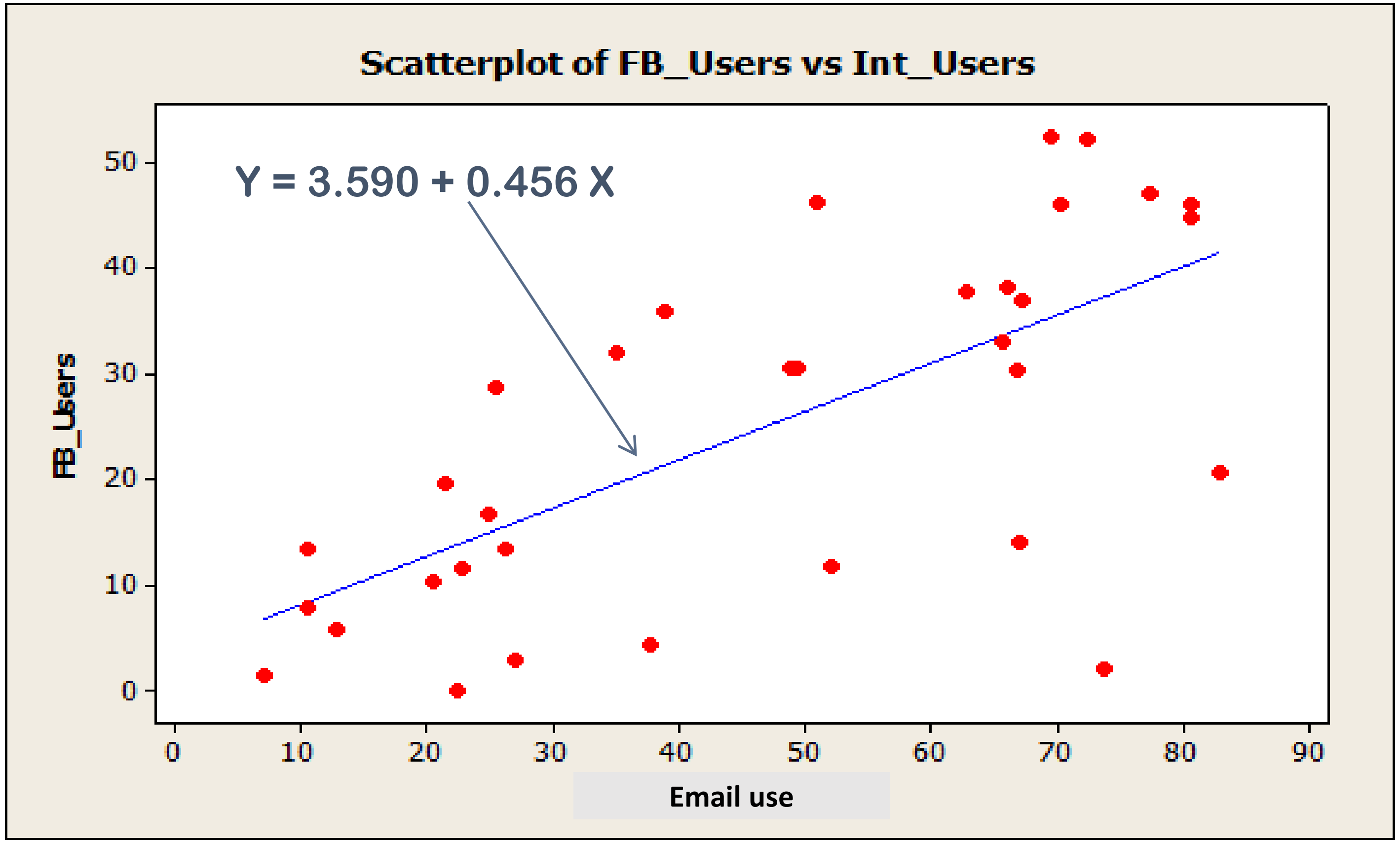
- Data kita tidak sempurna, mengandung galat, shg digunakan

MODEL $Y_i = \alpha + \beta X + \varepsilon_i \; (i=1,2,\ldots,n)$

- Perhatikan bhw ε_i adalah galat pengamatan ke-*i*

Menduga Garis Regresi

- **Untuk data pengguna internet tadi kita bisa minta MINITAB menghitung berapa dugaan (*estimate*) dari α dan β pada model tadi sbb:**



Bagaimana MINITAB bisa memperoleh dugaan
***α = 3.590* dan**
β = 0.456??

Metode Kuadrat Terkecil (*Least Squares*)

Ilustrasi

Andaikan dari suatu survei diperoleh persamaan regresi antara pendidikan (*X*) dalam tahun dgn pendapatan (*Y*) dalam dollar per tahun.

Persamaannya: $\hat{Y}$ = *-20,000 + 4000x.*

Maknanya, semakin lama pendidikan seseorang semakin tinggi pendapatannya.

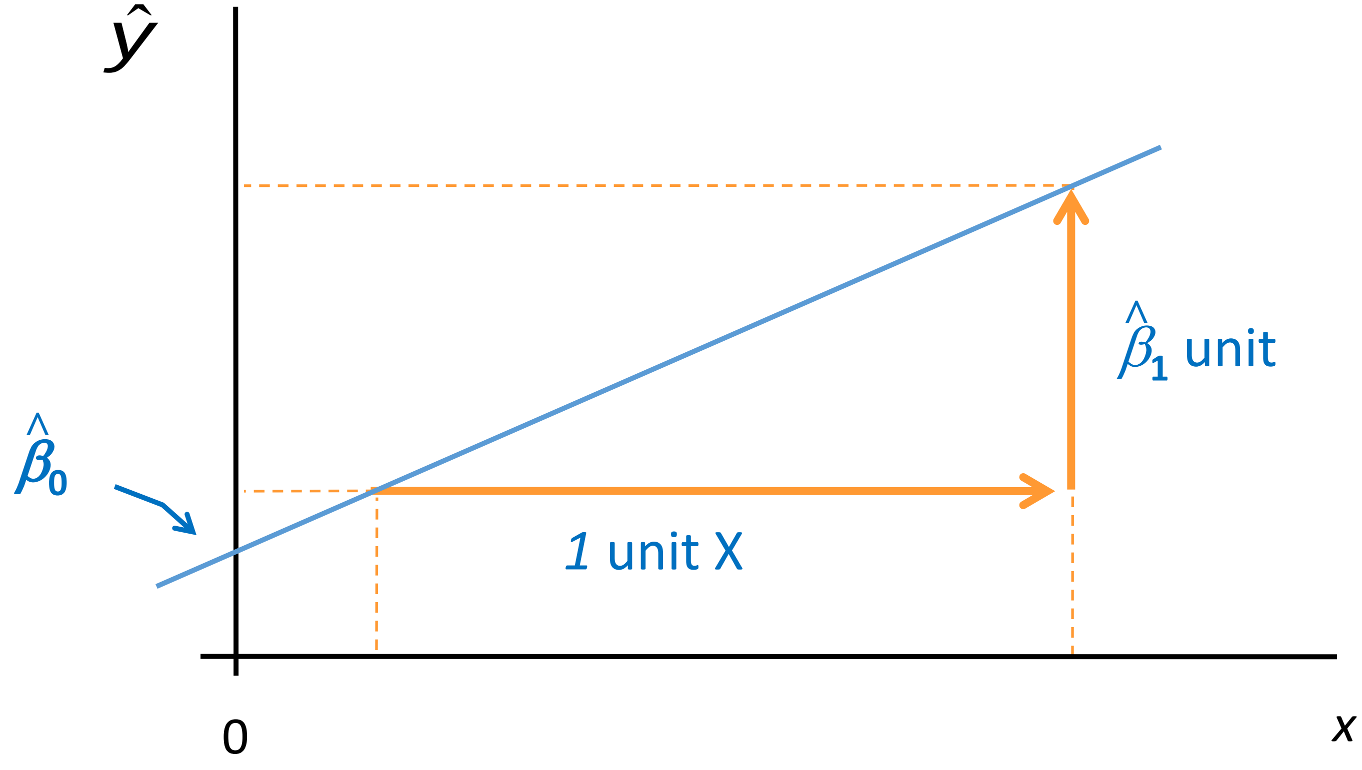
Misal pula bahwa setiap pendapatan pada tingkat pendidikan tertentu menyebar normal dengan sipangan baku σ = *13,000.*

$$\hat{Y} = -20{,}000 + 4000x$$

- *x = 12* → $\hat{Y}$ *= -20,000 + 4000(12)*, shg $\hat{Y}$ *= 28,000,* σ *= 13,000*
- *x = 16* → $\hat{Y}$ *= -20,000 + 4000(16)*, shg $\hat{Y}$ *= 44,000,* σ *= 13,000*

Makna Koefisien Regresi

$$\hat{y} = \hat{E}(Y \mid X) = \hat{\beta}_0 + \hat{\beta}_1 x$$



$\hat{\beta}_0$ **= dugaan intersep**

= (nilai $\hat{y}$ pada saat x = 0)

$\hat{\beta}_0$ **tidak selalu bisa ditafsirkan** dalam praktik.

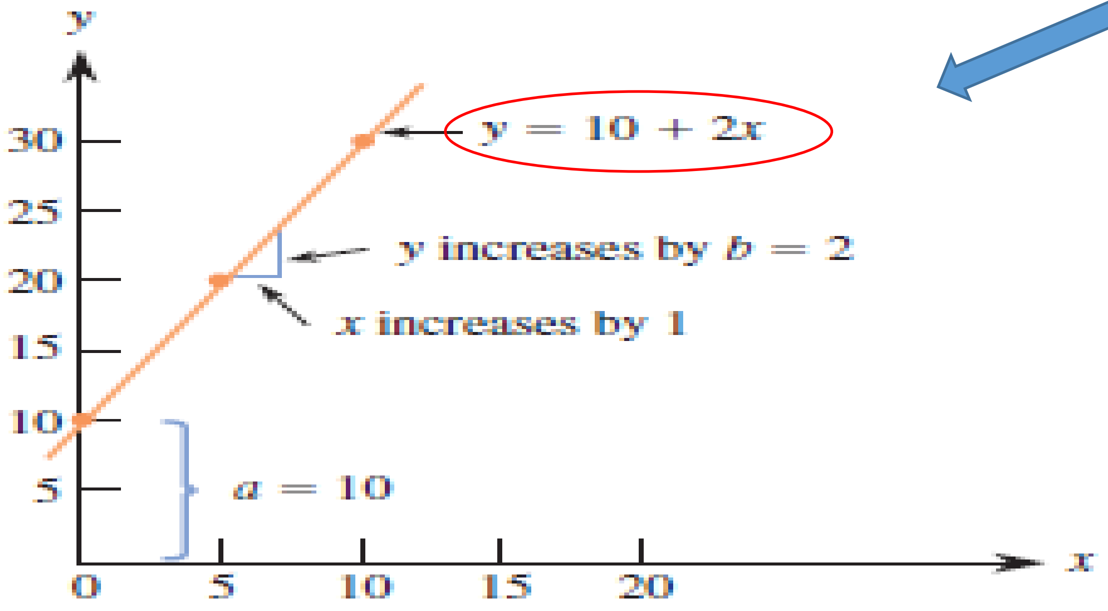
$\hat{\beta}_1$ **= dugaan kemiringan**

= perubahan nilai $\hat{y}$ utk setiap kenaikan 1 unit dari x

$\hat{\beta}_1$ **is always interpretable!**

^

Tanda Koefisien Regresi

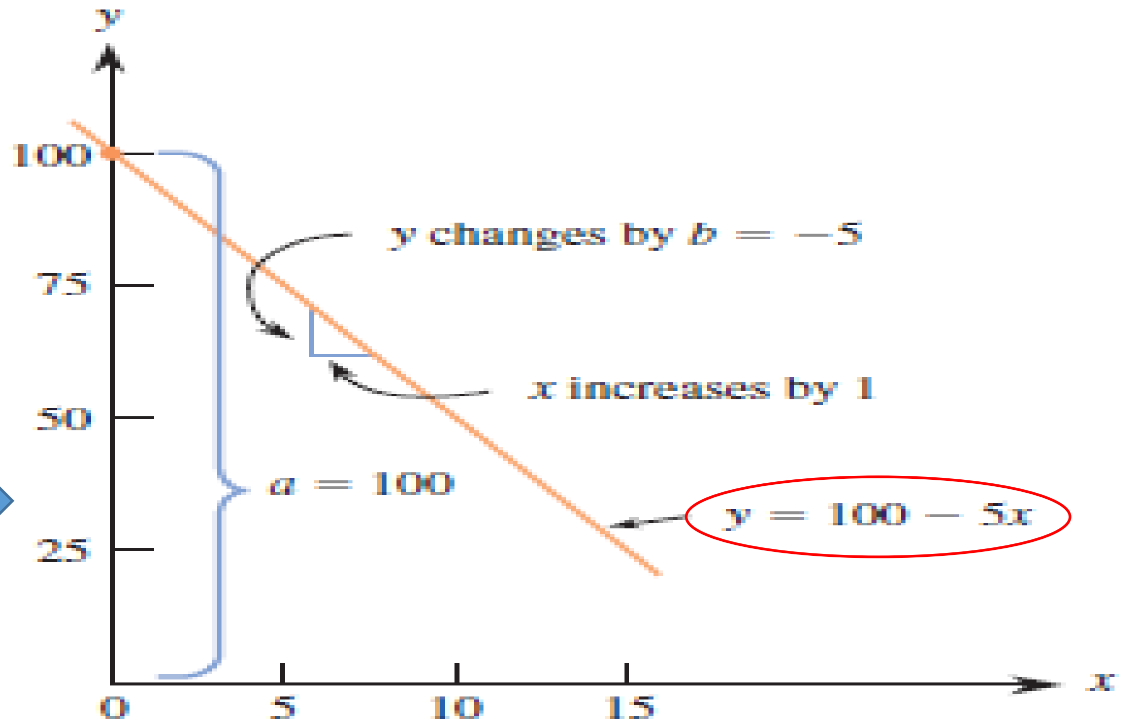


β_1 **positif**

(menaik)

β_1 **negatif**

(menurun)



Model Regresi Linear

Misalkan kita memiliki data dari contoh acaka berupa pasangan (y_1, x_1), …, (y_n, x_n) seperti tabel berikut ini:

$\boldsymbol{y_i}$	y_1	y_2	…	y_n
x_i	x_1	x_2	…	x_n

Biasanya, kita ingin memodelkan hubungan antara respon Y dengan penjelas X menggunakan model linear sbb:

Respon

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \, , \quad i = 1, 2, \ldots, n$$

Intersep

Kemiringan (*slope*)

Penjelas

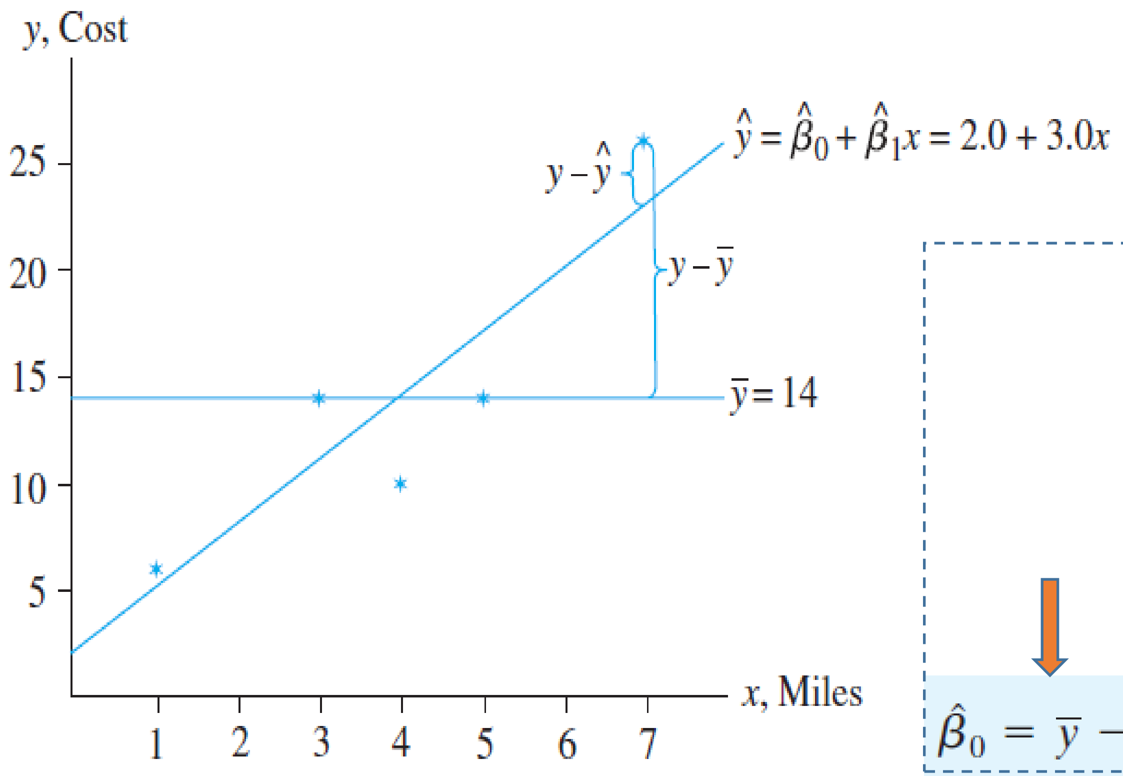
Galat

Y_i ~ Normal $((\beta_0 + \beta_1 X_i), \sigma^2)$

Metode Kuadrat Terkecil
(*Least Square Method*)

Perhatikan data biaya dan jarak sbb:

Cost y_i (in thousands of dollars):	6.0	14.0	10.0	14.0	26.0
Mileage x_i (in miles):	1.0	3.0	4.0	5.0	7.0



least-squares method

chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize

$$\sum_i (y_i - \hat{y}_i)^2 = \sum_i [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \sum_i (x_i - \bar{x})^2$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Diskusi Dulu.....

Sesi 1… beres!!!

Sesi 2…

Metode Kuadrat Terkecil

(*Least Square Method*)

Bagaimana metode kuadrat terkecil bisa menemukan $\hat{\beta}_0$ dan $\hat{\beta}_1$?

Model : $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n,$

Galat: $\varepsilon_i = y_i - \beta_0 - \beta_1 x_i, \quad i = 1, 2, \ldots, n$

JK galat: $S(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

Meminimumkan *S* thd β_0 dan β_1 melalui:

$$\frac{\partial S}{\partial \beta_0} = -2\sum(y_i - \beta_0 - \beta_1 x_i) = 0 \rightarrow \beta_0 = ??? \Longrightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\frac{\partial S}{\partial \beta_1} = -2\sum(y_i - \beta_0 - \beta_1 x_i)\, x_i = 0 \rightarrow \beta_1 = ??? \Longrightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Tidak memerlukan iterasi

Metode Kuadrat Terkecil

(*Least Square Method*)

Sifat-sifat $\hat{\beta}_0$ dan $\hat{\beta}_1$

1

- TAKBIAS *(unbiased)*
- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \rightarrow E(\hat{\beta}_1) = E(\frac{S_{xy}}{S_{xx}}) = \frac{1}{S_{xx}} E(S_{xy}) =$???
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \rightarrow E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1\bar{x}) = E(\bar{y}) - \bar{x}\, E(\hat{\beta}_1) =$???

Mengapa takbias *(unbiased)*?

- $E(\frac{\partial S}{\partial \beta_0})=0$ dan $E(\frac{\partial S}{\partial \beta_1})=0 \rightarrow E(\hat{\beta}_0) = \beta_0$ dan $E\hat{\beta}_1) = \beta_1$
- Jadi statistik $\hat{\beta}_0$ dan $\hat{\beta}_1$ tidak berbias thd parameter β_0 dan β_1

Metode Kuadrat Terkecil
(*Least Square Method*)

Sifat-sifat $\hat{\beta}_0$ dan $\hat{\beta}_1$

- Var($\hat{\beta}_0$) = Var ($\bar{y}$ - $\hat{\beta}_1\bar{x}$) $\Rightarrow$ $$Var(\hat{\beta}_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right]$$
- Var($\hat{\beta}_1$) = Var ($\frac{S_{xy}}{S_{xx}}$) $\Rightarrow$ $$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

➢ s^2 atau $\hat{\sigma}^2$ perlu diduga dari data sbb:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2} = \frac{\text{SSE}}{n-2}$$

$$\text{s.e.}(\hat{\beta}_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$$

$$\text{s.e.}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x})^2}}$$

Pengujian Parameter Regresi

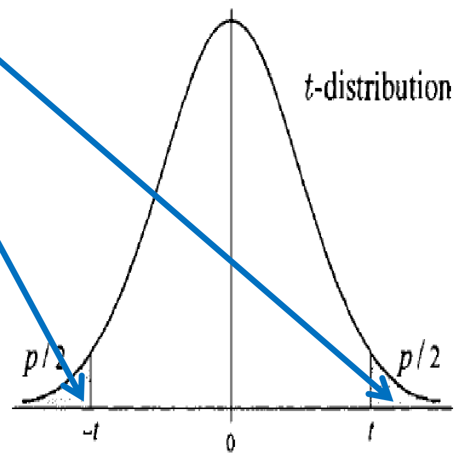
- Jika Y_1, Y_2, …, Y_n adalah contoh acak dari populasi normal → $Y_i \sim Normal\ (\mu,\sigma^2)$ maka sebaran dari $\hat{\beta}_0$ dan $\hat{\beta}_1$ adalah
  - $\hat{\beta}_0 \sim Normal\ (\beta_0, var\ (\hat{\beta}_0))$ dan $\hat{\beta}_0$ dan $\hat{\beta}_1 \sim Normal\ (\beta_1, var\ (\hat{\beta}_1))$

- Pengujian hipotesis H_0: $\boldsymbol{\beta_0 = 0}$ **vs** $\boldsymbol{H_1: \beta_0 \neq 0}$ didasarkan pada statistik $t_{hit} = \frac{\hat{\beta}_0}{s_{\hat{\beta}_0}}$

- Pengujian hipotesis H_0: $\boldsymbol{\beta_1 = 0}$ **vs** $\boldsymbol{H_1: \beta_1 \neq 0}$ didasarkan pada statistik $t_{hit} = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$

- Derajat bebas dari uji-t adalah (n-2) → kita menolak H_0 tersebut jika $|t_{hit}| > t_{\alpha/2(n-2)}$ untuk uji dua arah. Untuk uji satu arah perlu disesuaikan.

$|t_1| \geq t_{(n-2,\alpha/2)}$



Selang Kepercayaan Parameter

- Dengan mengetahui bhw sebaran $\hat{\beta}_0$ dan $\hat{\beta}_1$ adalah normal maka SK (1-α)x100% dari $\boldsymbol{\beta_0}$:

$$\hat{\beta}_0 \pm t_{(n-2,\alpha/2)} \times \mathrm{s.e.}(\hat{\beta}_0)$$

- Begitu juga SK (1-α)x100% dari $\boldsymbol{\beta_1}$:

$$\hat{\beta}_1 \pm t_{(n-2,\alpha/2)} \times \mathrm{s.e.}(\hat{\beta}_1)$$

- Ingat kembali apa interpretasi dari selang kepercayaan ini: apakah SK ini adalah ukurang peluang? Apakah SK ini konsep jangka pendek? Apa hubungan SK dgn uji hipotesis? SIAPA YG MASIH INGAT???

Ilustrasi

- Perhatikan data durasi reparasi komputer dan banyaknya unit yang bisa direparasi:

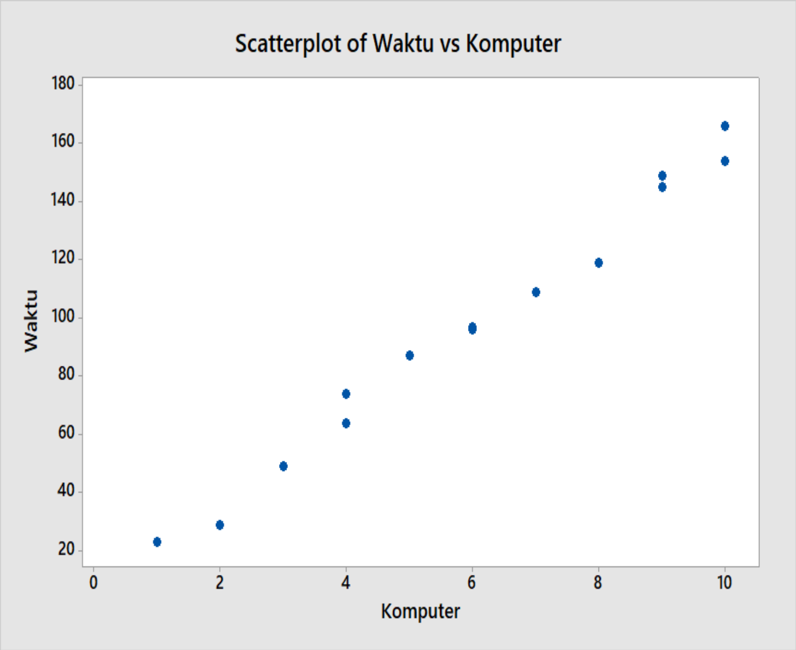
Row	Minutes	Units	Row	Minutes	Units
1	23	1	8	97	6
2	29	2	9	109	7
3	49	3	10	119	8
4	64	4	11	149	9
5	74	4	12	145	9
6	87	5	13	154	10
7	96	6	14	166	10

- Model regresi → $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \ldots, 14$

- y_i = durasi, x_i = banyaknya unit, β_0 *dan* β_1 adalah koefisien, sedangkan ε_i= galat

Ilustrasi

- Plot data →
- Hub linear positif



$$\hat{\beta}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{1768}{114} = 15.509$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 97.21 - 15.509 \times 6 = 4.162$$

$$\widehat{\text{Minutes}} = 4.162 + 15.509 \cdot \text{Units}$$

Testing H_0: $\beta_1 = 12$ **Mengapa 12???**

$$t_1 = \frac{\hat{\beta}_1 - 12}{\text{s.e.}(\hat{\beta}_1)} = \frac{15.509 - 12}{0.505} = 6.948$$

$$t_{(12, 0.025)} = 2.18$$

H_0 ditolak

- Jumlah $(y_i - \bar{y})^2$ = 27768.36
- Jumlah $(x_i - \bar{x})^2$ = 114
- Jumlah $(y_i - \bar{y})(x_i - \bar{x})$ = 1768.00

Selang kepercayaan 95% bagi $\boldsymbol{\beta_1}$:

$$15.509 \pm 2.18 \times 0.505 = (14.408, 16.610)$$

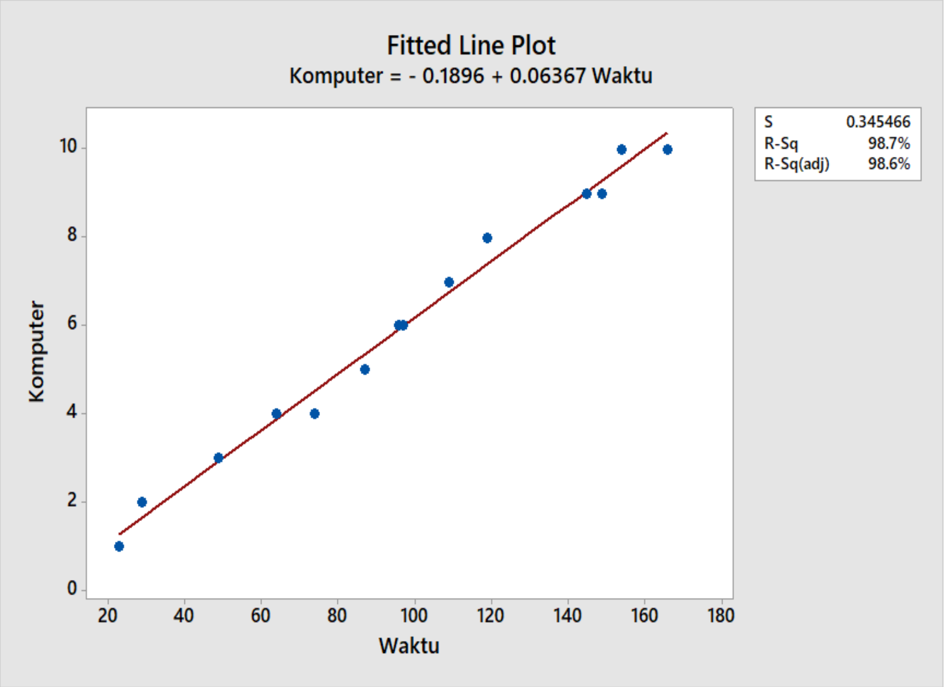
$t_{0.025\ (12)}$

Jika Waktu dan Unit dibalik?

Jika model dibalik menjadi:

$$X = \beta_0 + \beta_1 Y + \varepsilon$$

$$\text{Komputer} = \beta_0 + \beta_1 \text{Waktu} + \varepsilon$$



The **regression equation** is
Komputer = - 0.1896 + 0.06367 Waktu

S = 0.345466   R-Sq = 98.7%   R-Sq(adj) = 98.6%

Analysis of Variance

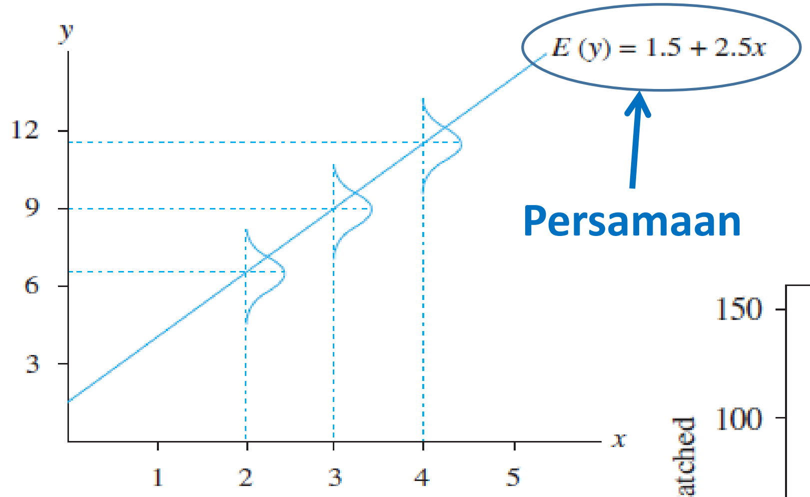
Source	DF	SS	MS	F	P-value
Regression	1	112.568	112.568	943.20	0.000
Error	12	1.432	0.119		
Total	13	114.000			

- Tidak masalah X dan Y dipertukarkan, selama ada maksud yang jelas dalam praktek.
- Hasilnya akan berubah dan interpretasinya otomatis akan berubah pula

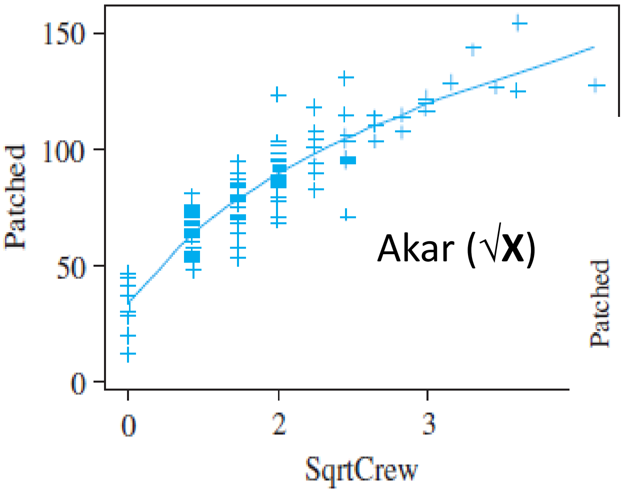
Asumsi Model Regresi

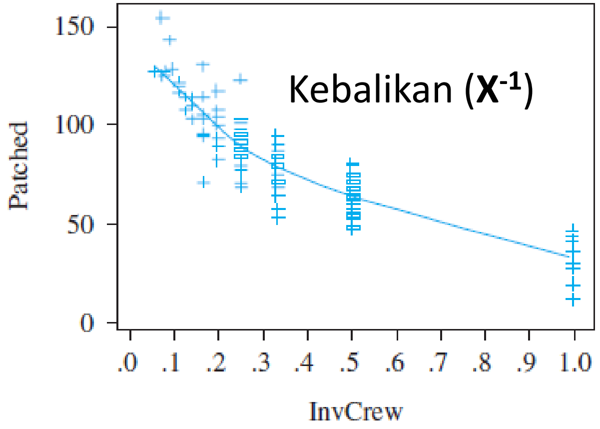
Formal assumptions of regression analysis:

1. The relation is, in fact, linear, so that the errors all have expected value zero: $E(\varepsilon_i) = 0$ for all i.
2. The errors all have the same variance: $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for all i.
3. The errors are independent of each other.
4. The errors are all normally distributed; ε_i is normally distributed for all i.

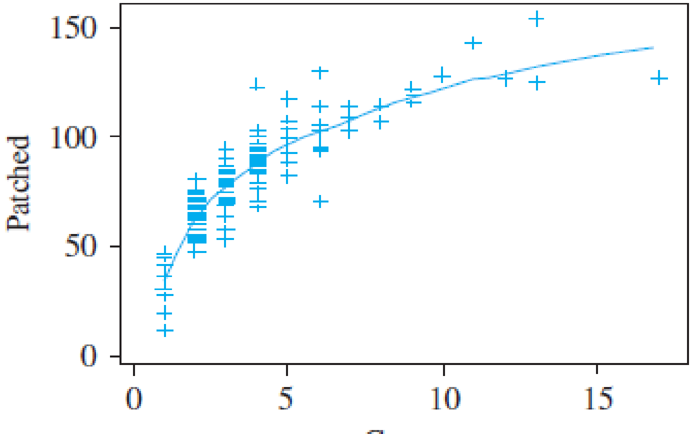


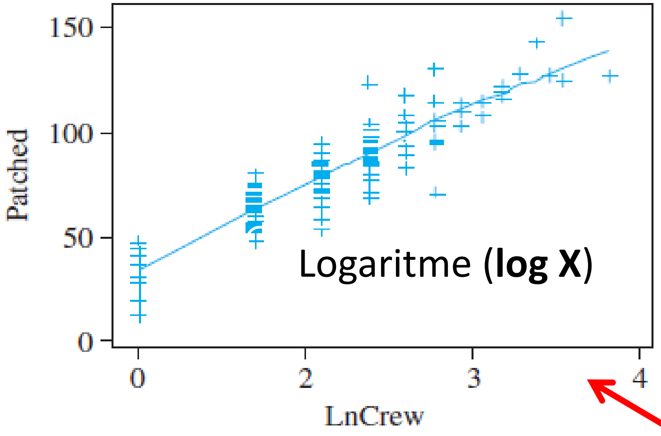
Persamaan





Hubungan taklinear





Terbaik

Korelasi

(*Correlation*)

- Korelasi merupakan ukuran keeratan linear
- Dalam model regresi, koefisien β_1 (*slope*) menjelaskan apakah hubungan X dengan Y positif ataukah negatif sehingga menentukan kecenderungan (*trend*) perubahan dari Y jika X berubah.
- Tetapi koefisien β_1 tidak menjelaskan seberapa kuat hubungan antara X dengan Y → KORELASI
- Korelasi (lengkapnya: *Pearson's Correlation*) dihitung sbb:

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}} \Longrightarrow r = \hat{\beta}_1\left(\frac{s_x}{s_y}\right)$$

- Koefisien $\hat{\beta}_1$ mrp penduga *slope* regresi, sedangkan S_x dan S_y simpangan baku dari X dan Y.

Ilustrasi

- Hasil penelitian untuk mengukur korelasi penggunaan internet dengan 3 peubah lainnya sbb:

Unemployment rate: Total percentage of labor force unemployed

GDP: Gross domestic product, per capita, in thousands of U.S. dollars (a measure of a nation's economic development)

CO_2: Carbon dioxide emissions, per capita (a measure of air pollution)

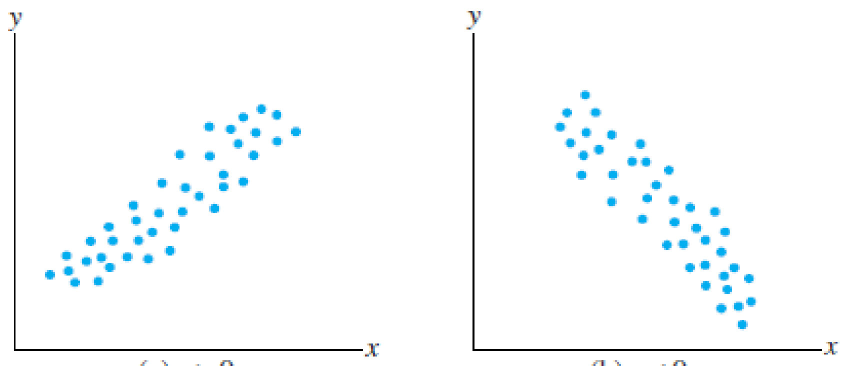
	Internet users (per 100 people)
Unemployment rate	0.238
GDP per capita	0.938
Carbon dioxide emissions	0.569

- Peubah apa yang paling erat hubungannya dengan penggunaan internet? → GDP
- Peubah apa yang paling tidak erat hubungannya dengan penggunaan internet? → *Unemployment*
- Apa kesimpulan yang bisa diambil dari data ini? → Semakin tinggi GDP (makmur?) semakin banyak orang menggunakan internet.

Sifat Korelasi

- Korelasi HANYA mengukur keeratan linear→ hubungan taklinear tidak dapat diukur oleh korelasi.

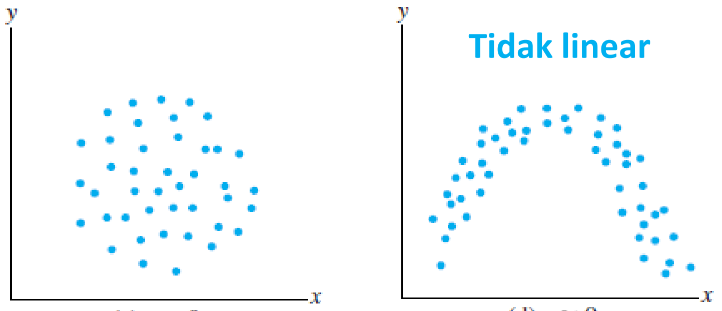
- The correlation r has the same sign as the slope b. Thus, $r > 0$ when the points in the scatterplot have an upward trend and $r < 0$ when the points have a downward trend.
- The correlation r always falls between -1 and $+1$, that is, $-1 \le r \le +1$.
- The larger the absolute value of r, the stronger the linear association, with $r = \pm 1$ when the data points all fall exactly on the regression line.



(a) $r > 0$ (b) $r < 0$

Ada korelasi

Tidak ada korelasi



(c) $r \approx 0$ (d) $r \approx 0$

Dalam regresi → ada koefisien determinasi (r^2) atau *R-square*, yaitu ukuran kekuatan prediksi dari model.

Jika tidak ada hubungan antara X dan Y maka korelasinya akan sama dengan nol. **Tapi tidak berlaku sebaliknya.**

- Parameter korelasi adalah ρ (*rho*), sedangkan statistiknya adalah r.

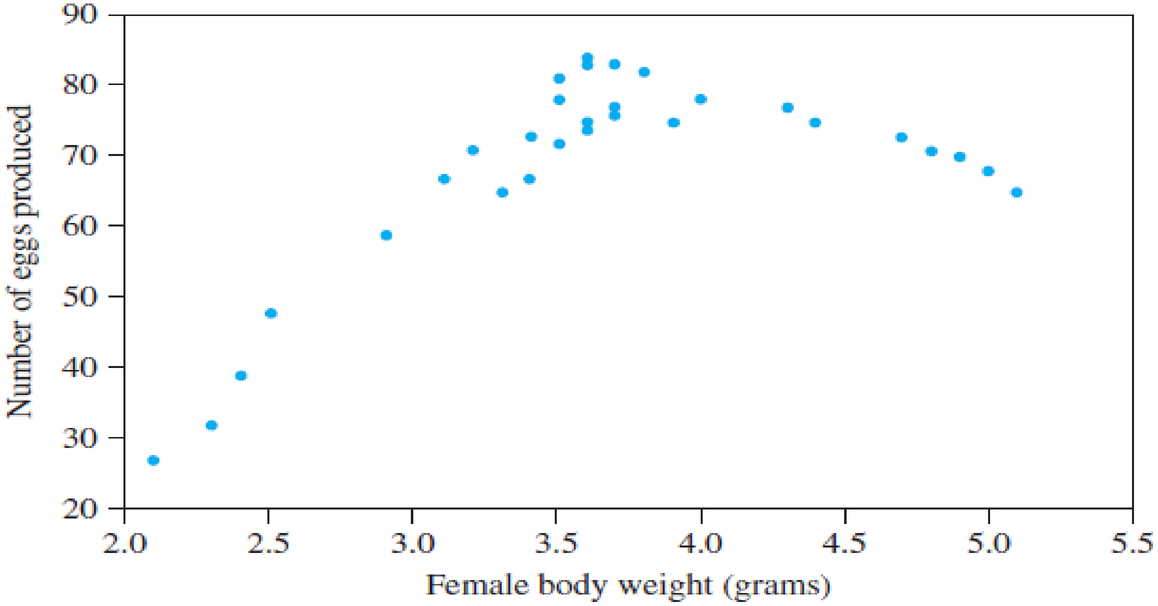
Diskusi Dulu.....

Sesi 2… beres!!! → Sesi 3…

Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)

In a study of the reproductive success of grasshoppers, an entomologist collected a sample of 30 female grasshoppers. She recorded the number of mature eggs produced and the body weight of each of the females (in grams). The data are given here:

Number of eggs(y)	Weight of female(x)	Number of eggs(y)	Weight of female(x)
27	2.1	75	3.6
32	2.3	84	3.6
39	2.4	77	3.7
48	2.5	83	3.7
59	2.9	76	3.7
67	3.1	82	3.8
71	3.2	75	3.9
65	3.3	78	4.0
73	3.4	77	4.3
67	3.4	75	4.4
78	3.5	73	4.7
72	3.5	71	4.8
81	3.5	70	4.9
74	3.6	68	5.0
83	3.6	65	5.1



Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)

Model:

No_eggs = β_0 + β_1 Weight + ϵ

Hypotheses:

Case 1: $H_0\colon \rho_{yx} \le 0$ vs. $H_a\colon \rho_{yx} > 0$

Case 2: $H_0\colon \rho_{yx} \ge 0$ vs. $H_a\colon \rho_{yx} < 0$

Case 3: $H_0\colon \rho_{yx} = 0$ vs. $H_a\colon \rho_{yx} \ne 0$

T.S.: $t = r_{yx}\dfrac{\sqrt{n-2}}{\sqrt{1-r_{yx}^2}}$

R.R.: With $n-2$ df and Type I error probability α,

1. $t > t_\alpha$
2. $t < -t_\alpha$
3. $|t| > t_{\alpha/2}$

Check assumptions and draw conclusions.

$$\sum_{i=1}^{30} x_i = 109.5 \Rightarrow \bar{x} = 3.65, \qquad \sum_{i=1}^{30} y_i = 2065 \Rightarrow \bar{y} = 68.8333$$

$$S_{xx} = \sum_{i=1}^{30}(x_i - \bar{x})^2$$
$$= (2.1 - 3.65)^2 + (2.3 - 3.65)^2 + \cdots + (5.1 - 3.65)^2 = 17.615$$

$$S_{yy} = \sum_{i=1}^{30}(y_i - \bar{y})^2$$
$$= (27 - 68.8333)^2 + (32 - 68.8333)^2 + \cdots + (65 - 68.8333)^2 = 6{,}066.1667$$

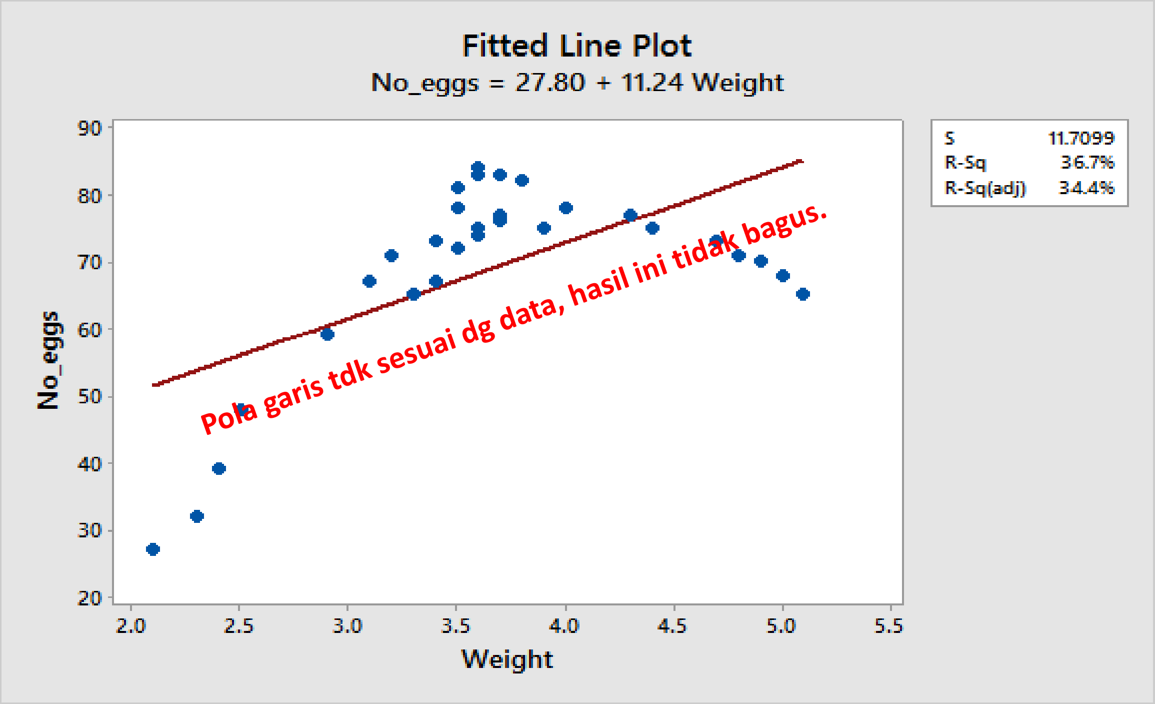
$$S_{xy} = \sum_{i=1}^{30}(x_i - \bar{x})(y_i - \bar{y})$$
$$= (2.1 - 3.65)(27 - 68.8333)$$
$$+ (2.3 - 3.65)(32 - 68.8333) + \cdots + (5.1 - 3.65)(65 - 68.8333) = 198.05$$

$$r_{xy} = \frac{198.05}{\sqrt{(17.615)(6066.1667)}} = 0.606$$

$\hat{\beta}_1$ = 198.05/17.615
= 11.24

$\hat{\beta}_0$ = 68.83-11.24(3.65)
= 27.804

Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)



ID	No_eggs	Weight	RESI1
1	27	2.1	-24.41
2	32	2.3	-21.65
3	39	2.4	-15.78
4	48	2.5	-7.90
5	59	2.9	-1.40
6	67	3.1	4.35
7	71	3.2	7.23
8	65	3.3	0.10
9	73	3.4	6.98
10	67	3.4	0.98
11	78	3.5	10.85
12	72	3.5	4.85
13	81	3.5	13.85
14	74	3.6	5.73
15	83	3.6	14.73
16	75	3.6	6.73
17	84	3.6	15.73
18	77	3.7	7.60
19	83	3.7	13.60
20	76	3.7	6.60
21	82	3.8	11.48
23	78	4.0	5.23
24	77	4.3	0.86
25	75	4.4	-2.27
26	73	4.7	-7.64
27	71	4.8	-10.76
28	70	4.9	-12.89
29	68	5.0	-16.01
30	65	5.1	-20.14

overestimated

Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)

Karena garis dengan data nampak tidak bersesuaian maka hasil regresi ini tidak bagus. → perlu diperbaiki terlebih dahulu.

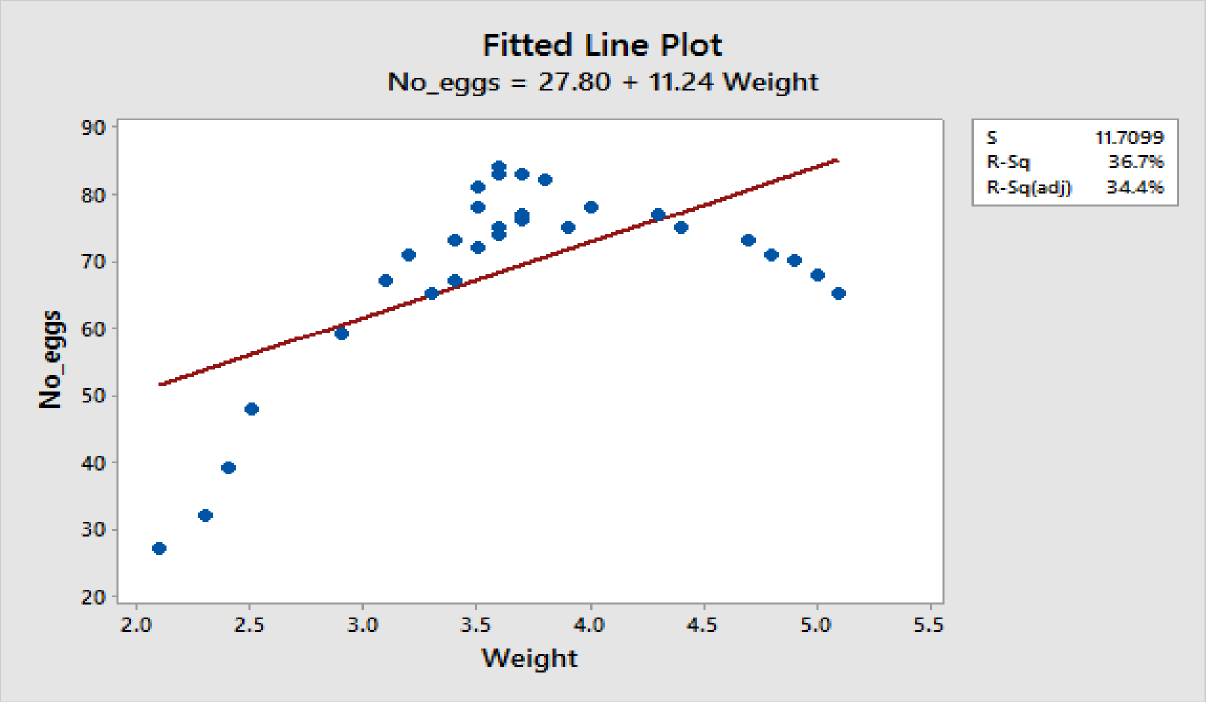
Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)

- **Andai dipaksakan dilakukan uji hipotesis (seharusnya tidak boleh)**
- **Misalkan hipotesis** $H_0: \rho = 0$ vs $H_1: \rho \neq 0$
- **Statistik** $$t = r_{yx}\frac{\sqrt{n-2}}{\sqrt{1-r_{yx}^2}}$$ → $t = 0.606\ (\sqrt{28} \div \sqrt{0.63}) = 4.04$
- **Pada taraf nyata 5% nilai tabel t adalah**

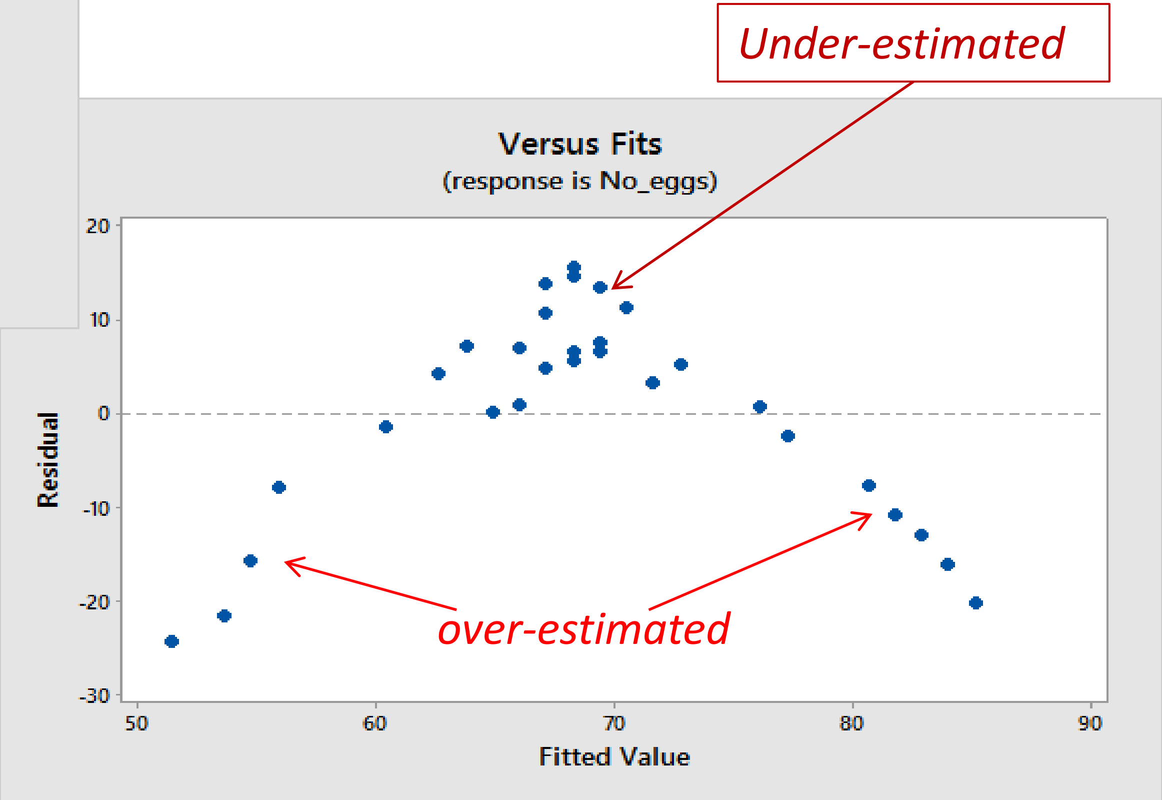
  Tabel → $t_{0.025(28)} = 2.048$

- **Jadi Ho ditolak → ada korelasi antara jumlah telor dengan bobot betina belalang. Semakin tinggi bobot betina belalang semakin banyak telornya**

Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)



Model linear bermasalah dalam kasus ini

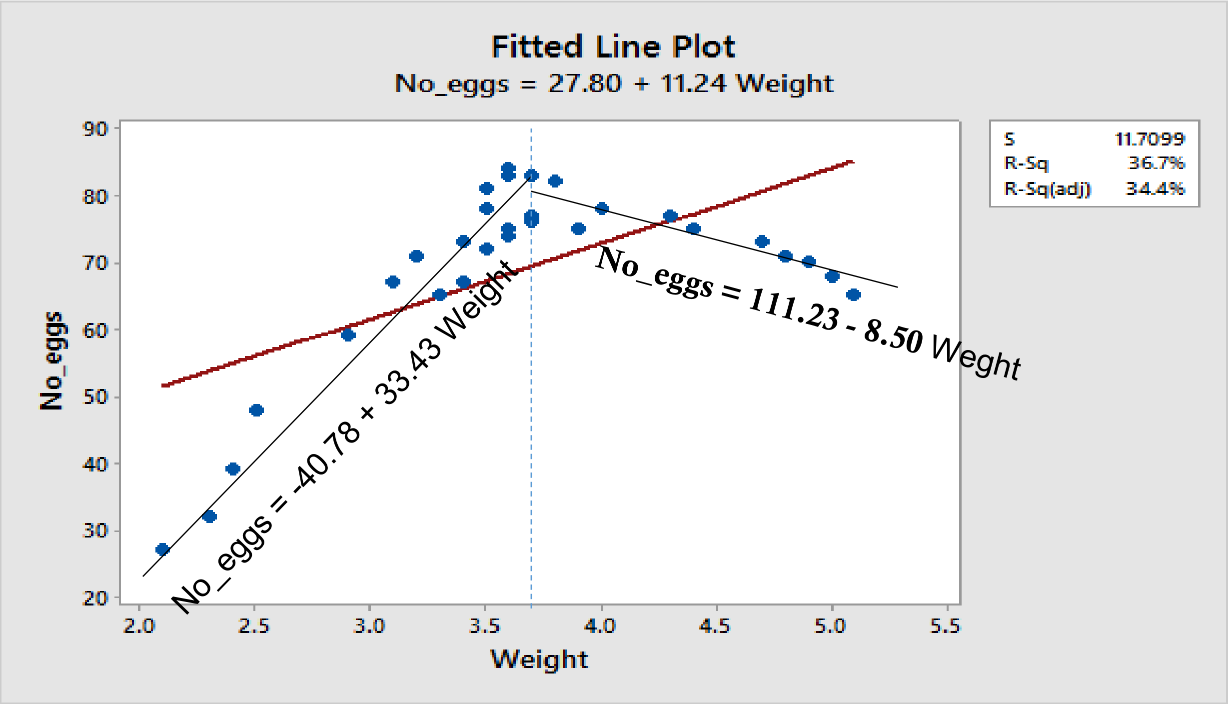


Garis regresi tidak pas dengan data pengamatan.

→ **CARI MODEL LAIN**

Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)

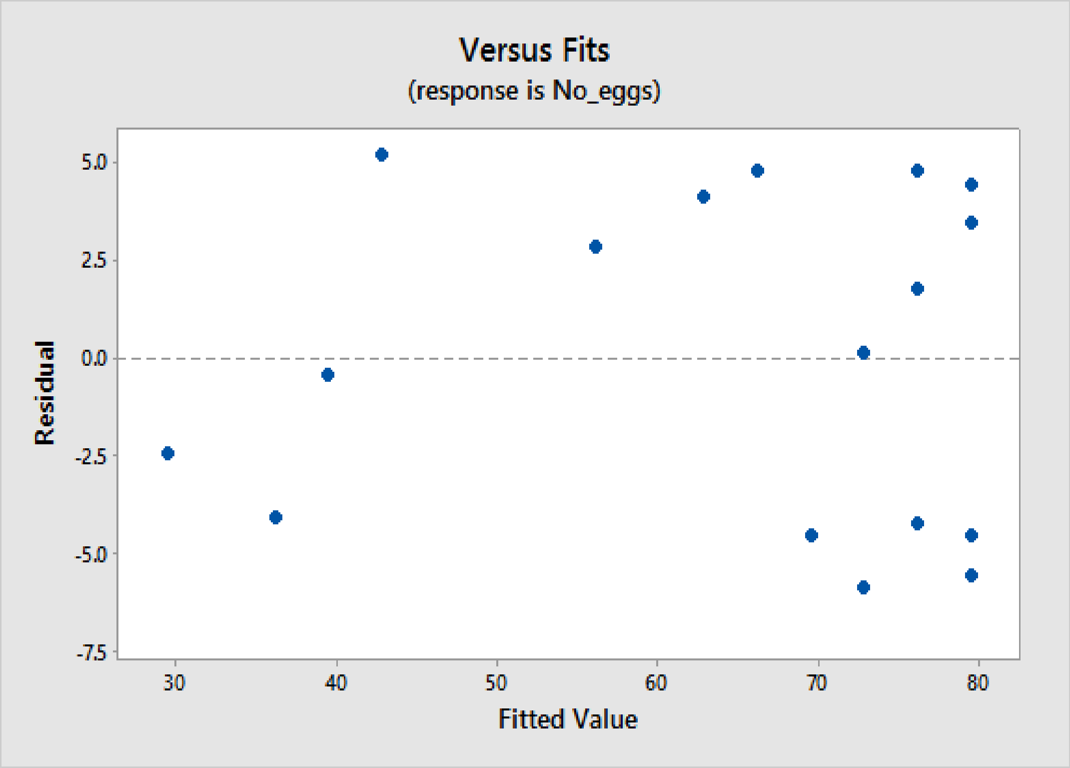
- **Banyak cara untuk mengatasi masalah yang terjadi pada model linear ini.**
- **Salah satu (sekedar ilustrasi) yaitu dengan membagi data menjadi 2 bagian:**
- **Weight < 3.65 dan Weight > 3.65.**
- **Lalu utk tiap bagian dicobakan model linear.**

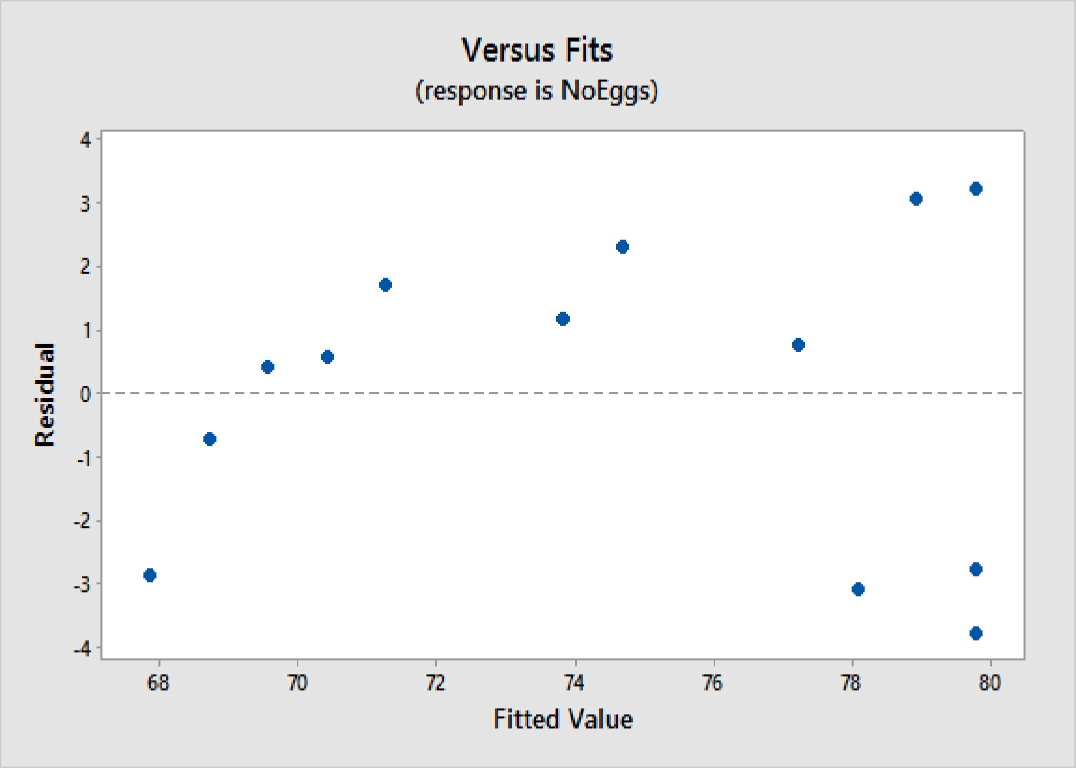


Jadi persamaan regresinya:

$$\widehat{NoEggs} = \begin{cases} -40.78 + 33.43 \text{ Weight, jika Weight} < 3.65 \\ 111.23 - 8.50 \text{ Weight, jika Weight} \geq 3.65 \end{cases}$$

Ilustrasi:
Belalang (Jumlah telur vs Bobot betina)





Garis regresi tidak menampakkan kejadian *under-estimate* dan *over-estimate* secara sistematis,

→ Model sudah relatif pas dengan data.

Thank You

email: khairil@apps.ipb.ac.id

twitter: @kh_notodiputro