



METODE STATISTIKA

PRAKTIKUM - WEEK 8
REGRESI LINEAR SEDERHANA

Dibuat oleh:

Asisten Praktikum Metode Statistika 2023/2024

- Adam Maurizio
- Giovanni Alberta T.
- Najma Attaqiya Alya
- M. Farras Reswara





Regresi Linear Sederhana

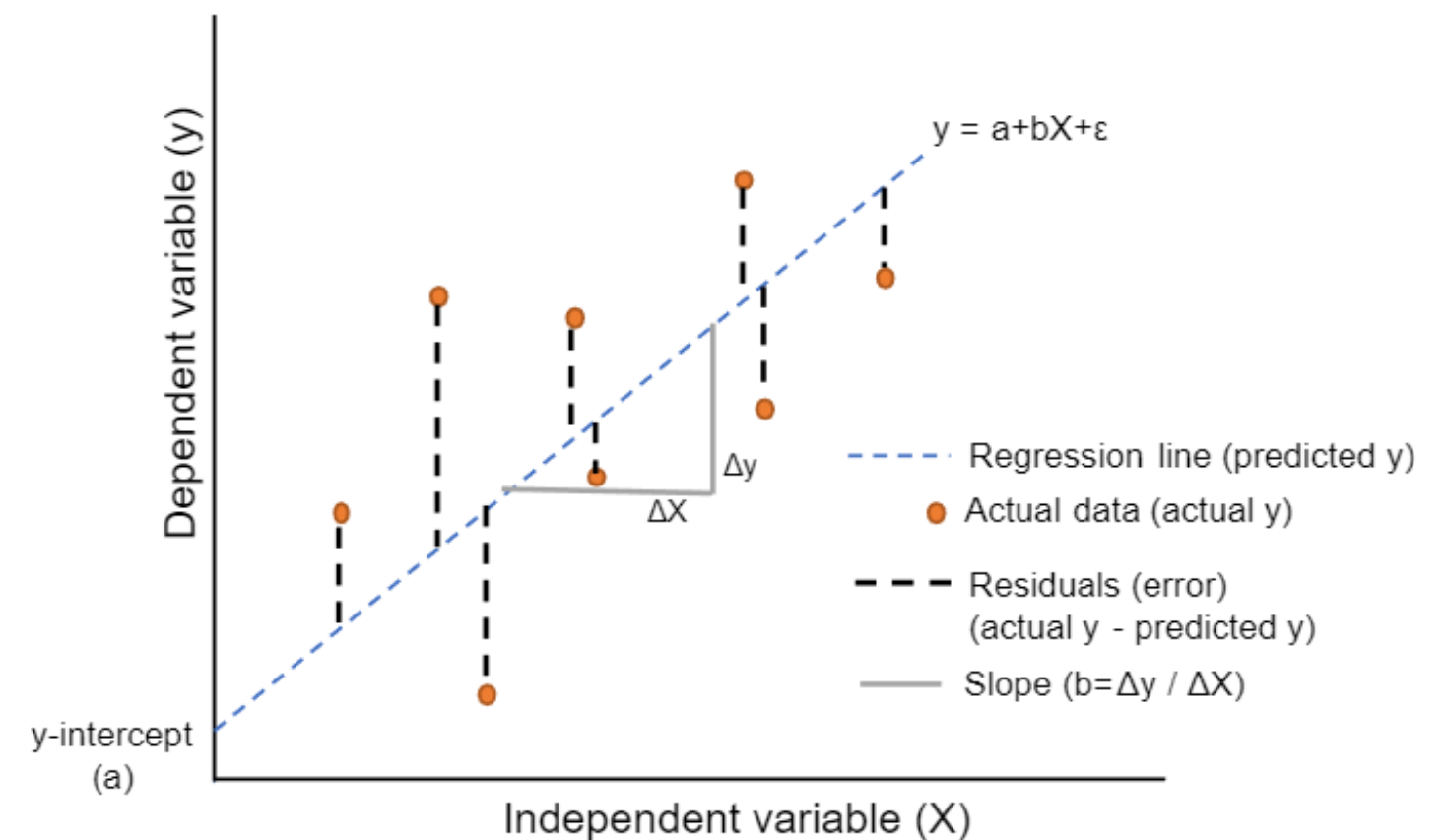
Analisis regresi menyangkut studi tentang **hubungan antara variabel kuantitatif** dengan tujuan mengidentifikasi, memperkirakan, dan memvalidasi hubungan antara variabel.

dalam analisis regresi, dilakukan **pemodelan hubungan linear** antara satu atau lebih variabel independen (X) dengan variabel dependen (Y)

Persamaan regresi linear sederhana

$$y = a + bX + \epsilon$$

Intercept slope of the regression line (unbiased estimate) error term (residuals)





CONTOH SOAL 1



Contoh Soal 1

Dalam satu tahap pengembangan obat baru untuk alergi, sebuah percobaan dilakukan untuk mempelajari bagaimana dosis obat yang berbeda mempengaruhi durasi penyembuhan gejala alergi. Setiap pasien menerima dosis obat tertentu dan diminta untuk melaporkan kembali segera setelah pengaruh/efek obat hilang.

TABLE 1 Dosage x (in Milligrams) and the Number of Hours of Relief y from Allergy for Ten Patients

Dosage x	Duration of Relief y
3	9
3	5
4	12
5	9
6	14
6	16
7	22
8	18
8	24
9	22



Contoh soal 1

```
library(readxl)
data1 <- read_excel("Contoh Soal.xlsx", sheet = "Contoh soal 1") #Import dataset
head(data1) #Menampilkan data

X = data1$Dosage
Y = data1$Duration_of_relief
#Regresi
reg <- lm(Y ~ X, data = data1)
#Summary regresi
summary(reg)
## Membuat scatter plot atau diagram pencar
plot(X,Y)
## Membuat garis regresi
abline(lm(Y~X))
```



Contoh soal 1

Output:

```
> summary(reg)
```

Call:

```
lm(formula = Y ~ X, data = data1)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.6333	-2.0128	-0.3741	2.0428	3.8851

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.0709	2.7509	-0.389	0.707219
X	2.7408	0.4411	6.214	0.000255 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.821 on 8 degrees of freedom

Multiple R-squared: 0.8284, Adjusted R-squared: 0.8069

F-statistic: 38.62 on 1 and 8 DF, p-value: 0.0002555



Contoh soal 1

Manual:

```
slope <- function(X, Y){  
  mean_x <- mean(X)  
  mean_y <- mean(Y)  
  sxy <- sum((X - mean_x)*(Y-mean_y))  
  sxx <- sum((X - mean_x)^2)  
  b1 <- sxy / sxx  
  return(b1)  
}
```

```
intercept <- function(X, Y, b1){  
  b0 <- mean(Y) - (b1 * mean(X))  
  return(b0)  
}
```

```
b1 <- slope(X, Y)  
b0 <- intercept(X, Y, b1)  
b1  
b0
```

```
> b1  
[1] 2.740831  
> b0  
[1] -1.070905  
> |
```




Contoh soal 1

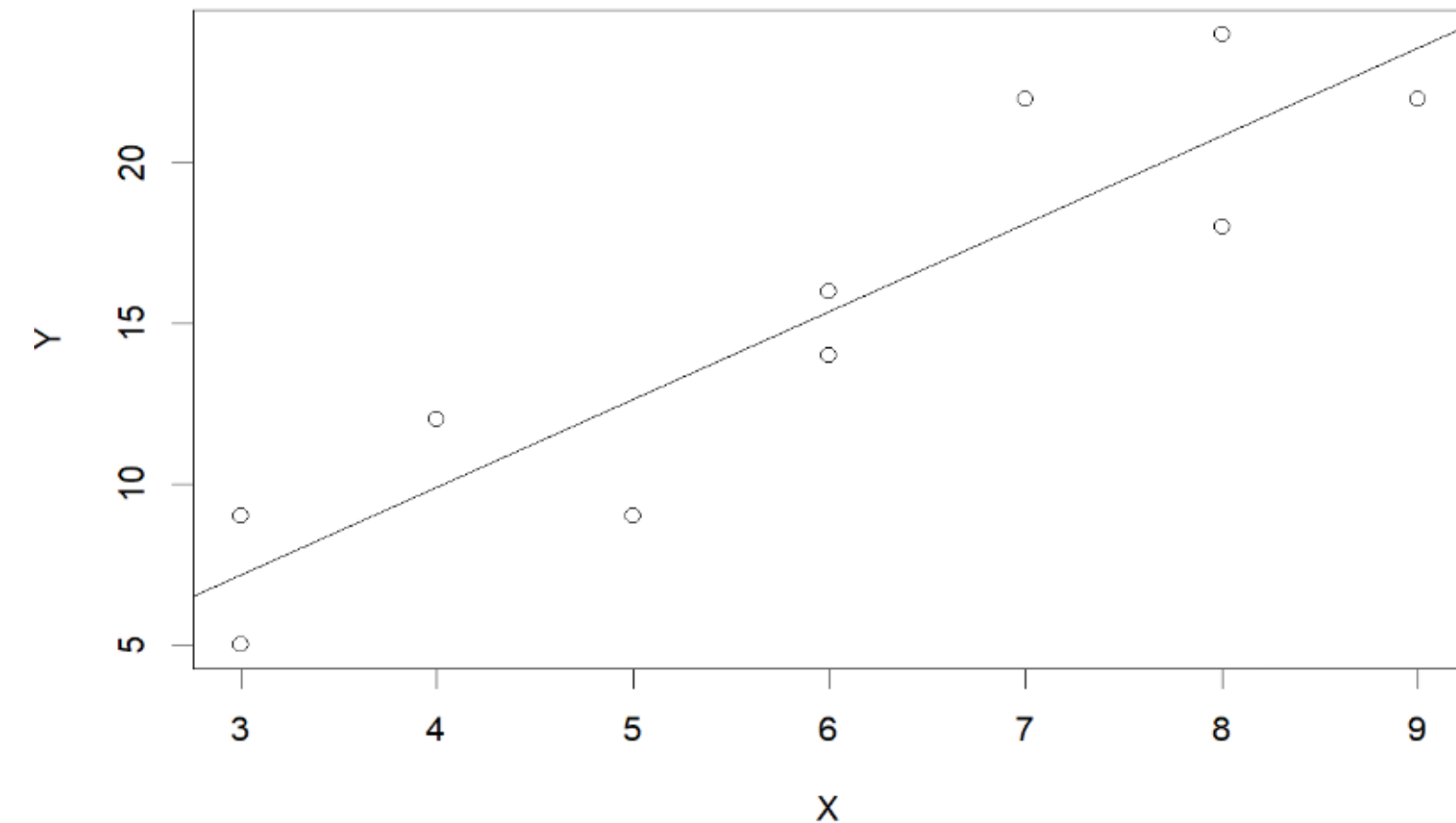
Maka, berdasarkan hasil output diatas, dapat diketahui bahwa:

- **Persamaan regresi** yang diperoleh yakni:

$$\hat{y} = -1,0709 + 2,7408 X_1$$

Berdasarkan persamaan diatas, dapat diketahui bahwa koefisien X pada model regresi adalah **positif**, artinya **terdapat hubungan yang positif** antara dosis obat dan durasi penyembuhan alergi atau semakin tinggi dosis obat yang diberikan, maka semakin baik / cepat penyembuhan

- **Nilai R²** yang dihasilkan yakni sebesar 82,84% yang artinya variabel independen mampu menjelaskan sebesar 82,84% variabel dependen, sedangkan sisanya sebesar 17,16% dipengaruhi oleh faktor lain diluar model. Dapat disimpulkan model regresi yang dibuat sudah sangat baik.
- **Nilai p-value** dari koefisien X memiliki nilai $0,000255 < \alpha$ (5%) -> signifikan, Artinya, dosis obat berpengaruh signifikan terhadap durasi penyembuhan alergi.





CONTOH SOAL 2



Contoh Soal 2

Salah satu tolak ukur perkembangan suatu negara adalah Human Development Index (HDI). Harapan hidup, melek huruf, pencapaian pendidikan, dan produk domestik bruto per kapita digabungkan menjadi indeks antara 0 dan 1, termasuk 1 sebagai perkembangan tertinggi. Program Pembangunan Perserikatan Bangsa -Bangsa melaporkan nilai - nilai untuk 177 negara. Kami secara acak memilih lima belas negara, di bawah dua puluh lima teratas. Baik IPM dan variabel prediktor x = Penggunaan Internet per 100 orang diperoleh dari laporan mereka. Dapatkan estimasi kuadrat terkecil (least squares) dengan menyesuaikan garis lurus dengan respons penggunaan Internet menggunakan variabel prediktor HDI.

TABLE 8 Human Development Index

Country	Internet/100	HDI
Bahrain	21.3	.866
Poland	26.2	.870
Uruguay	14.3	.852
Bulgaria	20.6	.824
Brazil	19.5	.800
Ukraine	9.7	.788
Dominican Republic	16.9	.799
Moldova	9.6	.708
India	5.5	.619
Madagascar	0.5	.533
Nepal	0.4	.534
Tanzania	0.9	.467
Uganda	1.7	.505
Zambia	2.0	.434
Ethiopia	0.2	.406

Source: Human Development 2007–2008 reports at
UNDP web site <http://hdr.undp.org>



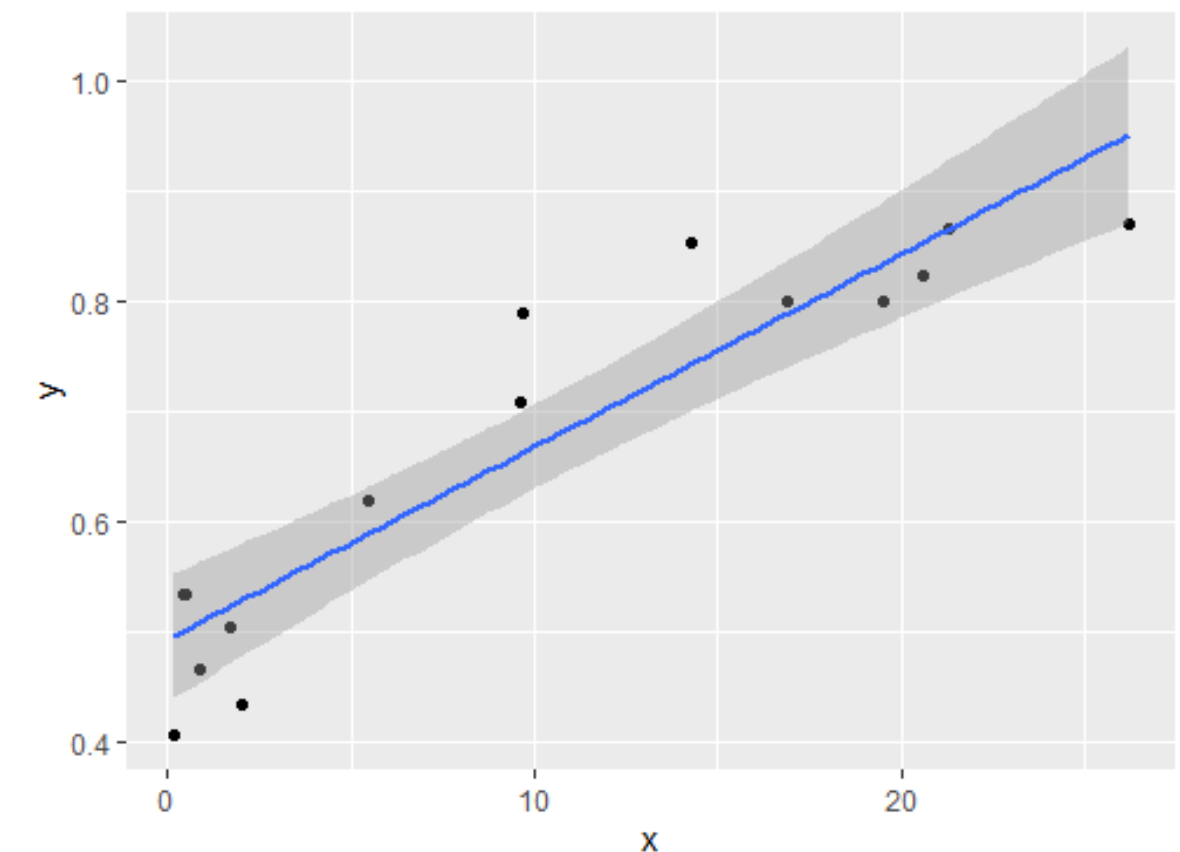
Contoh soal 2

```
# membuat data frame
df <- data.frame(
  inet =
c(21.3,26.2,14.3,20.6,19.5,9.7,16.9,9.6,5.
5,0.5,0.4,0.9,1.7,2,0.2),
  hdi =
c(0.866,0.870,0.852,0.824,0.8,0.788,
0.799,0.708,0.619,0.533,0.534,0.467,
0.505,0.434,0.406))
df

# inisiasi variabel
x = df$inet
y = df$hdi
```

inet	hdi
21.3	0.866
26.2	0.870
14.3	0.852
20.6	0.824
19.5	0.800
9.7	0.788
16.9	0.799
9.6	0.708
5.5	0.619
0.5	0.533
0.4	0.534
0.9	0.467
1.7	0.505
2.0	0.434
0.2	0.406

```
# plot and fit a linear model
library(ggplot2)
ggplot(df, aes(x, y)) +
  geom_point() +
  geom_smooth(method = "lm")
```





Contoh soal 2

```
# manual
slope <- function(x, y){
  mean_x <- mean(x)
  mean_y <- mean(y)
  sxy <- sum((x - mean_x)*(y-mean_y))
  sxx <- sum((x - mean_x)^2)
  b1 <- sxy / sxx
  return(b1)
}
```

```
intercept <- function(x, y, b1){
  b0 <- mean(y) - (b1 * mean(x))
  return(b0)
}
```

```
b1 <- slope(x, y)
b0 <- intercept(x, y, b1)
# pembuatan model
model = lm(y~x, data=df)
summary(model)
```

```
Call:
lm(formula = y ~ x, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.094255 -0.037803  0.001059  0.032284  0.125419

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.493365   0.026542  18.588 9.55e-11 ***
x             0.017445   0.001993   8.752 8.26e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06828 on 13 degrees of freedom
Multiple R-squared:  0.8549,    Adjusted R-squared:  0.8437 
F-statistic: 76.59 on 1 and 13 DF,  p-value: 8.255e-07
```



Contoh soal 2

```
# estimasi parameter  
model$coefficients
```

```
# Rsquared  
summary(model)$r.squared
```

```
# predict hdi when internet 5.5  
predict(model, data.frame(x = 5.5))
```

```
> # estimasi parameter  
> model$coefficients  
(Intercept)          x  
  0.49336548  0.01744486  
>  
> # Rsquared  
> summary(model)$r.squared  
[1] 0.8548993  
>  
> # predict hdi when internet 5.5  
> predict(model, data.frame(x = 5.5))  
          1  
0.5893122
```

INTERPRETASI?



What is residuals?

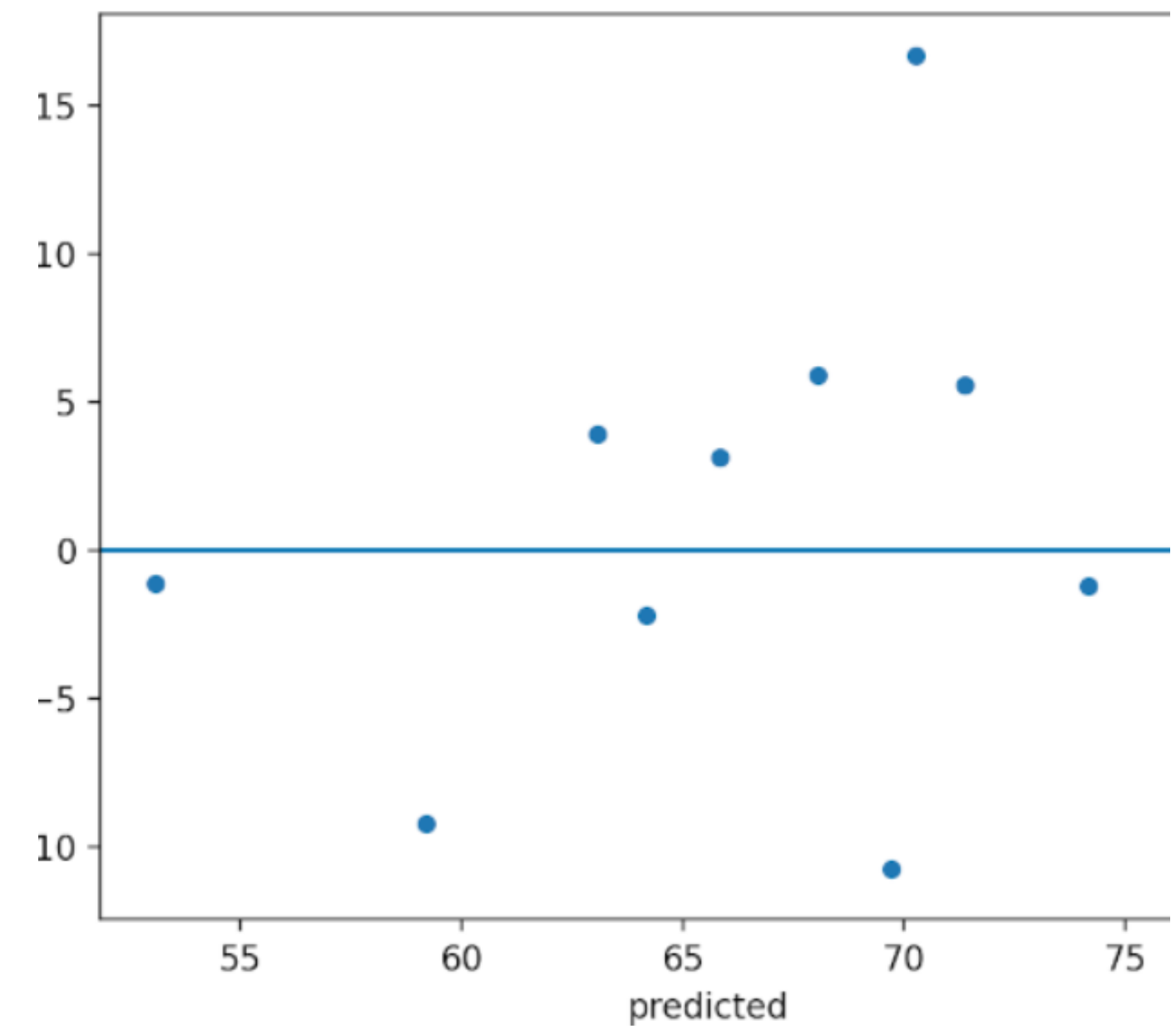
Persamaan regresi linear sederhana

$$y = a + bX + \epsilon$$

\downarrow Intercept

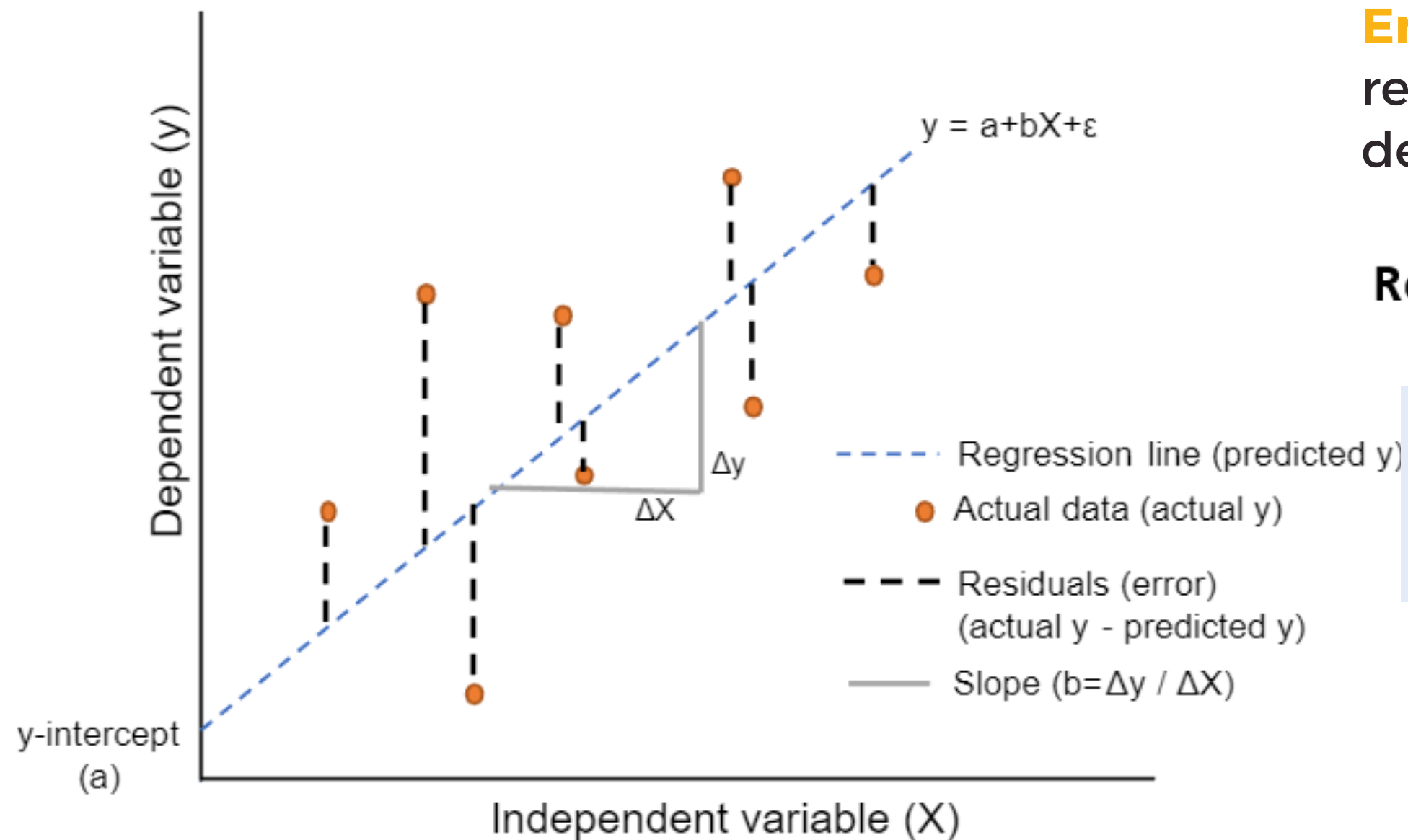
\downarrow slope of the regression line (unbiased estimate)

\rightarrow error term (residuals)





What is residuals?



Error term (ϵ) dalam model regresi disebut residual, yaitu selisih antara nilai aktual y dengan nilai prediksi y (garis regresi).

Residual = actual y (y_i) - predicted y (\hat{y}_i)

Residuals

$$\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad i = 1, \dots, n$$



How to calculate residuals?

TABLE 3 Computations for the Least Squares Line, SSE, and Residuals Using the Data of Table 1

x	y	x^2	y^2	xy	$\hat{\beta}_0 + \hat{\beta}_1 x$	Residual \hat{e}
3	9	9	81	27	7.15	1.85
3	5	9	25	15	7.15	-2.15
4	12	16	144	48	9.89	2.11
5	9	25	81	45	12.63	-3.63
6	14	36	196	84	15.37	-1.37
6	16	36	256	96	15.37	.63
7	22	49	484	154	18.11	3.89
8	18	64	324	144	20.85	-2.85
8	24	64	576	192	20.85	3.15
9	22	81	484	198	23.59	-1.59

Residual dihitung dengan **mengurangkan nilai aktual titik data** dari **nilai prediksi titik data** tersebut. Nilai prediksi dapat diperoleh dari analisis regresi.

Beberapa residual adalah **positif** dan beberapa **negatif**, dan **sifat kuadrat terkecil** adalah bahwa jumlah **residual selalu nol**.



Contoh soal 1

```
resid(model regresi)
#Plot residual
df1 <- data.frame(fitted(reg), resid(reg))
ggplot(df1, aes(fitted.reg., resid.reg.)) + geom_point(size = 3) + geom_hline(yintercept = 0)
```

Output:

```
> resid(reg)
      1      2      3      4      5      6      7      8
1.8484108 -2.1515892  2.1075795 -3.6332518 -1.3740831  0.6259169  3.8850856 -2.8557457
      9     10
3.1442543 -1.5965770
> |
```



Contoh soal 1

