

Praktikum 1

STK331–Analisis Regresi

Pengenalan Analisis Regresi

Aturan Praktikum

Asisten:

Nabil Bintang Prayoga

Farik Firsteady Haristiyanto

Jadwal Praktikum:

Selasa, 13:00 – 15:00

Asynchronous/Synchronous

Aturan:

- Kehadiran praktikum 100%
- Usahakan hadir tepat waktu
- Presensi IPB Mobile & TTD Basah
- Proporsi Nilai Praktikum: **(10%)**
 - **Kuis**
 - **Tugas**
 - **Keaktifan**
- Kuis praktikum dilaksanakan 1x per sesi (UTS/UAS)

Pengenalan R

- R merupakan serangkaian fasilitas *software* yang terintegrasikan untuk memanipulasi data, perhitungan, dan tampilan grafik.
- *Open source*
- Analisis statistika di R berupa serangkaian *steps* dengan hasil yang disimpan dalam suatu objek

Objek Data dalam R

- Vektor
- Matriks
- Array
- Factor
- List
- Data Frame

Pengenalan R

OBJEK DATA

■ Vektor

Vektor merupakan tipe sederhana dari objek dan setiap elemen memiliki mode yang sama.

Contoh:

```
a <- c(2, 4)
```

```
b <- c("laki-laki", "perempuan")
```

Perintah dalam vektor:

rep (pengulangan)

```
rep(a,2) → 2 4 2 4
```

```
rep(a, each=2) → 2 2 4 4
```

seq (membuat suatu urutan)

```
seq(from=3, to=1) → 3 2 1
```

```
seq(from=1, to=2, by=0.5) → 1 1.5 2
```

is.vector (memeriksa suatu objek termasuk dalam vektor)

as.vector (merubah objek menjadi vector)

operator (+, -, /, ^, %, dll)

Pengenalan R

OBJEK DATA

Akses elemen ke i dalam suatu vektor

Contoh:

```
angka<-c(2,3,4,1)
```

```
name(angka)<-c("a", "b", "c", "d")
```

Untuk mengakses 2 dapat menggunakan syntax angka["a"] atau angka[1]

```
angka[2:3]
```

```
angka[c(1,4)]
```

```
angka[angka<2]
```

```
angka[-2]
```

Pengenalan R

OBJEK DATA

■ Matriks

Matriks umumnya dibuat melalui vektor

Contoh:

```
Data<-matrix(c(1,2,1, 100,102,103), nrow = 2, ncol = 3, byrow = TRUE,  
             dimnames = list(c("row1", "row2"), c("C.1", "C.2", "C.3")))
```

	C.1	C.2	C.3
row1	1	2	1
row2	100	102	103

Perintah dalam vektor:

dim (mengetahui ukuran matriks)

rbind (menggabungkan terhadap baris)

```
a<-c(2,3,4)
```

```
b<-c(1,0,1)
```

```
rbind(a,b)
```

	[,1]	[,2]	[,3]
a	2	3	4
b	1	0	1

cbind (menggabungkan terhadap kolom)

```
a<-c(2,3,4)
```

```
b<-c(1,0,1)
```

```
rbind(a,b)
```

	a	b
[1,]	2	1
[2,]	3	0
[3,]	4	1

Pengenalan R

OBJEK DATA

Akses elemen dalam suatu matriks

Contoh:

```
Data<-matrix(c(1,2,1, 100,102,103), nrow = 2, ncol = 3, byrow = FALSE,  
             dimnames = list(c("row1", "row2"), c("C.1", "C.2", "C.3")))
```

Data[1,3]

Data[,1]

Data[-1,-2]

	C.1	C.2	C.3
row1	1	1	102
row2	2	100	103

Pengenalan R

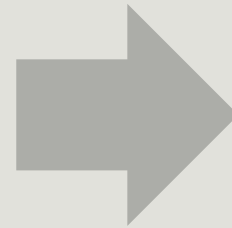
OBJEK DATA

■ Faktor

Faktor merupakan bentuk khusus dari vektor dengan elemen kategorik dan memiliki level.

Perintah dalam vektor:

factor



```
x <- c("apple", "banana", "apple", "orange")
factor_x <- factor(x, levels = c("apple", "banana", "orange"), labels = c("A", "B", "C"))
print(factor_x)

[1] A B A C
Levels: A B C
```

Konversi & mengatur level dan label

as.factor

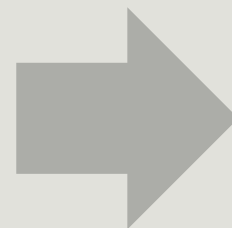


```
x <- c("apple", "banana", "apple", "orange")
factor_x <- as.factor(x)
print(factor_x)

[1] apple banana apple orange
Levels: apple banana orange
```

Konversi vektor menjadi faktor

is.factor



```
# Membuat vektor karakter
colors <- c("red", "green", "blue")

# Memeriksa apakah objek adalah faktor
is_colors_factor <- is.factor(colors)
print(is_colors_factor) # Output: FALSE
```

Mengecek data apakah faktor

Pengenalan R

OBJEK DATA

- List

List merupakan struktur data yang dapat menyimpan berbagai jenis data dalam satu wadah

```
> list(a,dt,Data)
[[1]]
[1] 2 3 4

[[2]]
[1] "b" "a" "a"

[[3]]
      c.1 c.2 c.3
row1    1  1 102
row2    2 100 103
```

a = Vektor

dt = Faktor

Data = Matriks

Pengenalan R

OBJEK DATA

- **Data Frame**

Merupakan bentuk khusus dari list untuk menyimpan data dari berbagai tipe dalam bentuk matriks.
Perintah yang umum digunakan dalam data frame:

`data.frame()` → Membuat data frame baru dari vektor atau list

`as.data.frame()` → Mengkonversi objek menjadi data frame

`read.table()` → Membaca data dari file teks

Pengenalan R

OBJEK DATA

- Data Frame

Contoh:

```
dtf<-data.frame(Mahasiswa=c("Amir","Budi", "Citra"), JK=c("L","L","P"))
```

```
dtf <- rbind(dtf, data.frame(Mahasiswa="Intan", JK="P"))
```

```
  Mahasiswa JK
1      Amir  L
2       Budi  L
3      Citra  P
```

```
  Mahasiswa JK
1      Amir  L
2       Budi  L
3      Citra  P
4      Intan  P
```

colnames(dtf) → untuk mengetahui nama-nama kolom

```
[1] "Mahasiswa" "JK"
```

length(dtf) atau dim(dtf) → untuk mengetahui jumlah kolom atau dimensi

```
[1] 2
```

```
[1] 4 2
```

Pengenalan R

OBJEK DATA


■ Data Frame

`dtf[2]` atau `dtf["JK"]` atau `dtf$JK` → untuk mengakses semua elemen JK

`dtf[1,]` → untuk mengakses baris ke-1 semua kolom

`dtf[2,1]<-"Bambang"`

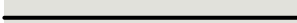
	Mahasiswa	JK
1	Amir	L
2	Budi	L
3	Citra	P



	Mahasiswa	JK
1	Amir	L
2	Bambang	L
3	Citra	P

`names(dtf)[1]<="Nama"` → mengganti nama kolom ke-1

	Mahasiswa	JK
1	Amir	L
2	Bambang	L
3	Citra	P



	Nama	JK
1	Amir	L
2	Bambang	L
3	Citra	P

Pengenalan R

INPUT DATA DALAM R

- **File .txt**

`data<-read.table("E:/...../namadata.txt", header=TRUE)`

Jika pemisah antar kolom (*delimiter*) bukan spasi maka dapat menggunakan syntax berikut

`data<- read.delim("E:/...../namadata.txt", header = TRUE, sep = ";")`

- **File .csv**

`data <- read.csv("E:/...../namadata.txt", header = TRUE)`

Jika menggunakan delimiter ";" maka menggunakan `read.csv2`

- **File .xlsx**

`library(readxl)`

`data<- read_excel(path = "data/namadata.xlsx", col_names = TRUE)`

Korelasi dan Regresi

KORELASI	REGRESI
Mengukur kekuatan atau derajat hubungan antara dua peubah	Menduga rata-rata satu peubah atas dasar nilai yang tetap dari peubah-peubah lain
Dua peubah ini diperlakukan secara simetri	Ada asimetri bagaimana peubah tak bebas dan peubah bebas diperlakukan

Koefisien Korelasi Populasi

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\{\text{var}(X) \text{var}(Y)\}}} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y}$$

Koefisien Korelasi Sampel

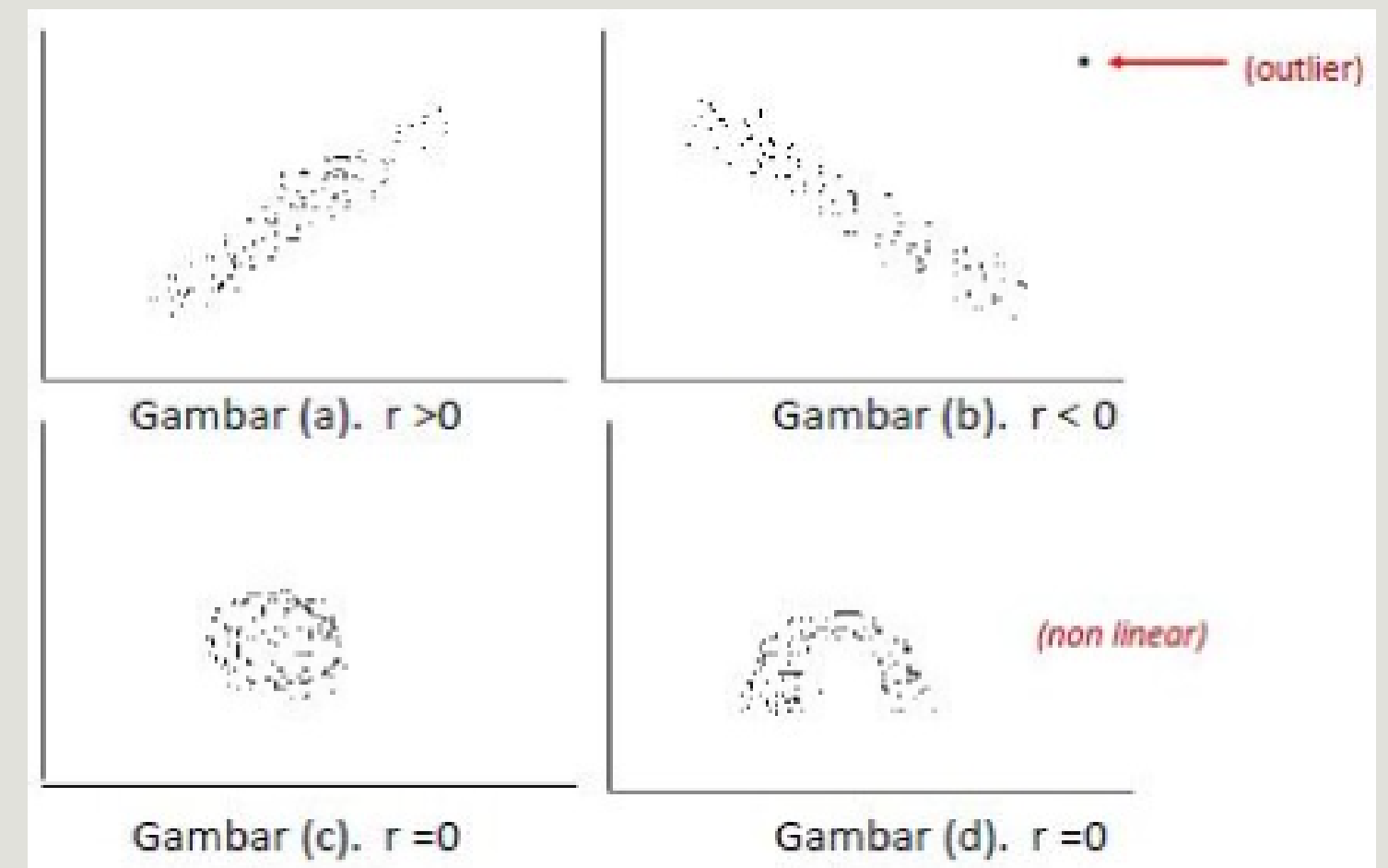
$$r = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}} \\ = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}}$$

Korelasi

- Mengukur keeratan hubungan linier antara dua peubah metrik
- Tidak harus menggambarkan hubungan sebab-akibat
- $-1 \leq r \leq 1$
- r mendekati 1 atau -1, semakin erat hubungan linearnya

Kelemahan Koefisien Korelasi:

- Sangat peka terhadap pencilan (outlier)
- Tidak dapat mendeteksi hubungan non linear



Korelasi

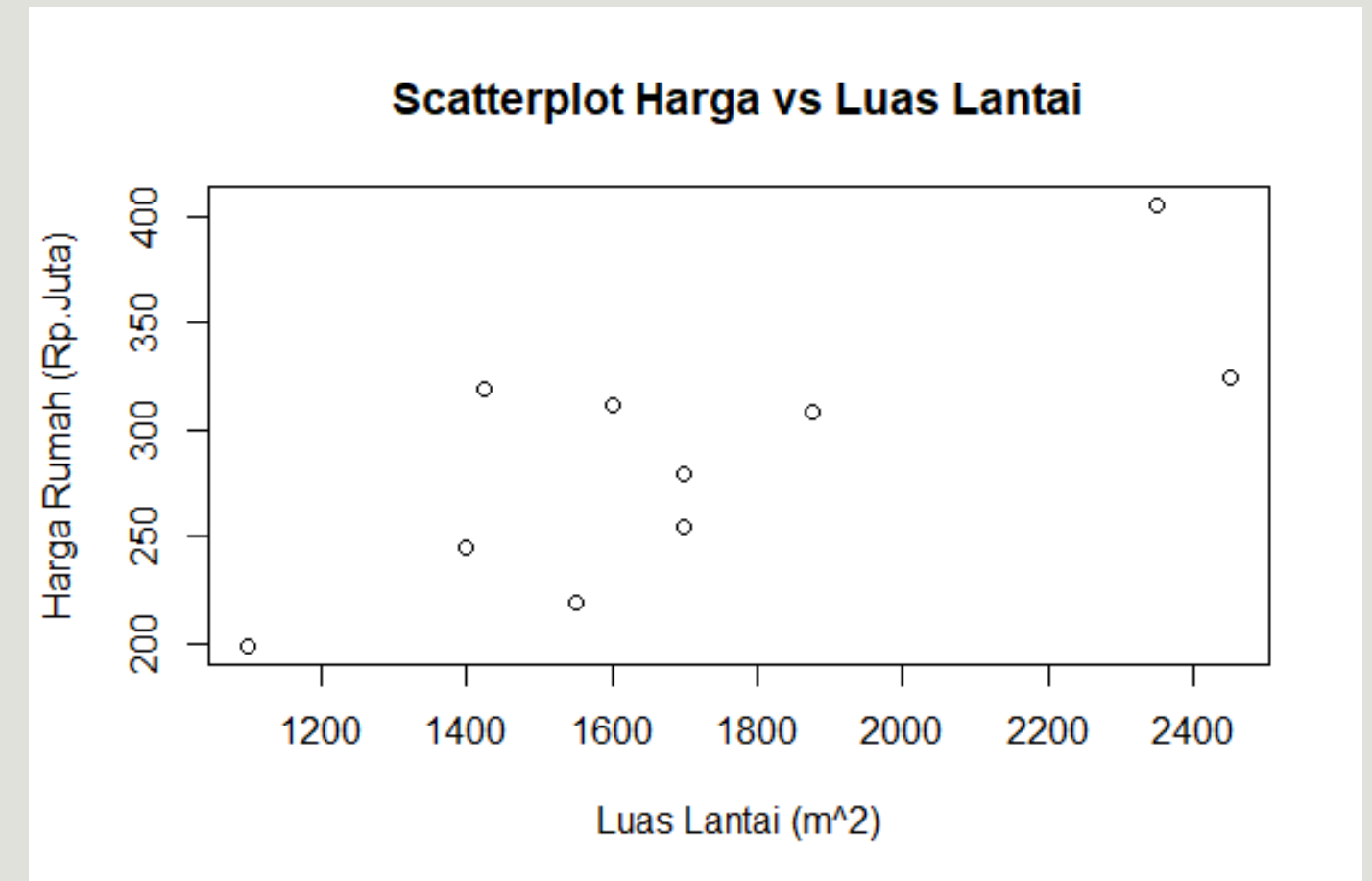
Harga Rumah (Rp.juta) (y)	Luas Lantai (m2) (x)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

#Input data

```
harga <- c(245, 312, 279, 308, 199, 219,  
           405, 324, 319, 255)  
luas <- c(1400, 1600, 1700, 1875, 1100, 1550,  
          2350, 2450, 1425, 1700)
```

#Membuat Scatterplot

```
plot (x = luas, y = harga,  
      main = "Scatterplot Harga vs Luas Lantai",  
      xlab = "Luas Lantai (m^2)",  
      ylab = "Harga Rumah (Rp.Juta)")
```



Korelasi

#Koefisien Korelasi

```
cor(luas, harga, method = "pearson")
```

```
> cor(luas, harga, method = "pearson")  
[1] 0.7621137
```

#Matriks Korelasi

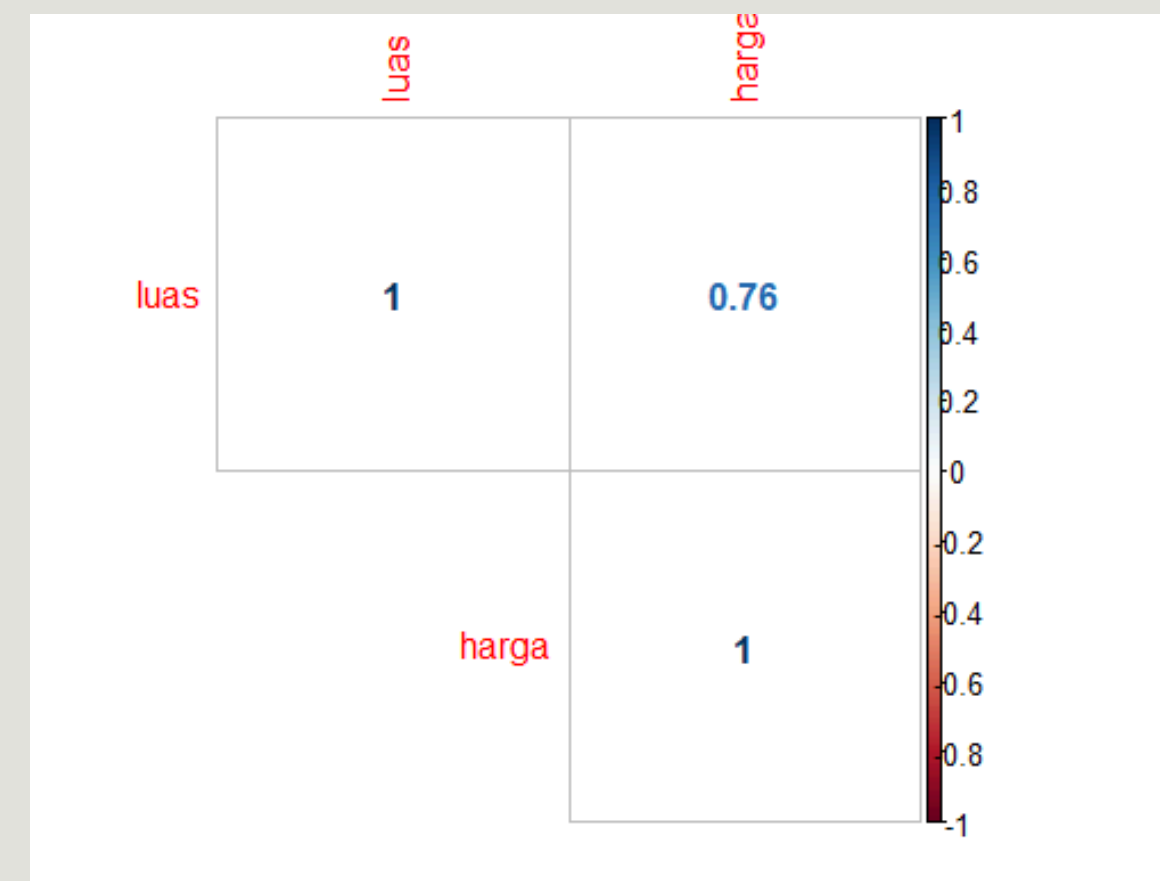
```
df <- data.frame(luas, harga)  
korelasi <- cor(df)  
korelasi
```

	luas	harga
luas	1.0000000	0.7621137
harga	0.7621137	1.0000000

```
#install.packages("corrplot")
```

```
library(corrplot)
```

```
corrplot(corr = korelasi, method = "number", type = "upper")
```



Korelasi

#Uji Signifikansi Koefisien Korelasi

#Hipotesis $H_0: \rho = 0$ vs $H_1: \rho \neq 0$

```
cor.test(x = luas, y = harga, alternative = "two.sided",  
        method = "pearson",  
        exact = NULL, conf.level = 0.95)
```

Pearson's product-moment correlation

```
data: luas and harga  
t = 3.3294, df = 8, p-value = 0.01039  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.2547085 0.9404622  
sample estimates:  
      cor  
0.7621137
```

Nilai p (p-value = 0.01039): Karena nilai $p < 0.05$, kita menolak hipotesis nol dan menyimpulkan bahwa ada korelasi yang signifikan antara Luas Lantai dan Harga Rumah

Hipotesis statistik:

$H_0: \rho = 0$ (tidak ada korelasi)

$H_1: \rho \neq 0$ (ada korelasi)

Statistik uji:

$$Z = \frac{\sqrt{n-3}}{2} \ln \left[\frac{(1+r)(1-\rho)}{(1-r)(1+\rho)} \right]$$

Keputusan:

Tolak H_0 jika $|Z| > \text{critical value } Z_{\alpha/2}$

α	$Z_{\alpha/2}$
1%	2.575
5%	1.960
10%	1.645

Analisis Regresi

Model Deterministik vs Stokhastik

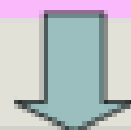
Mis. Diketahui $X=1, Y=3$
 $X=4, Y=9$

$$Y = f(X) = 1 + 2X$$

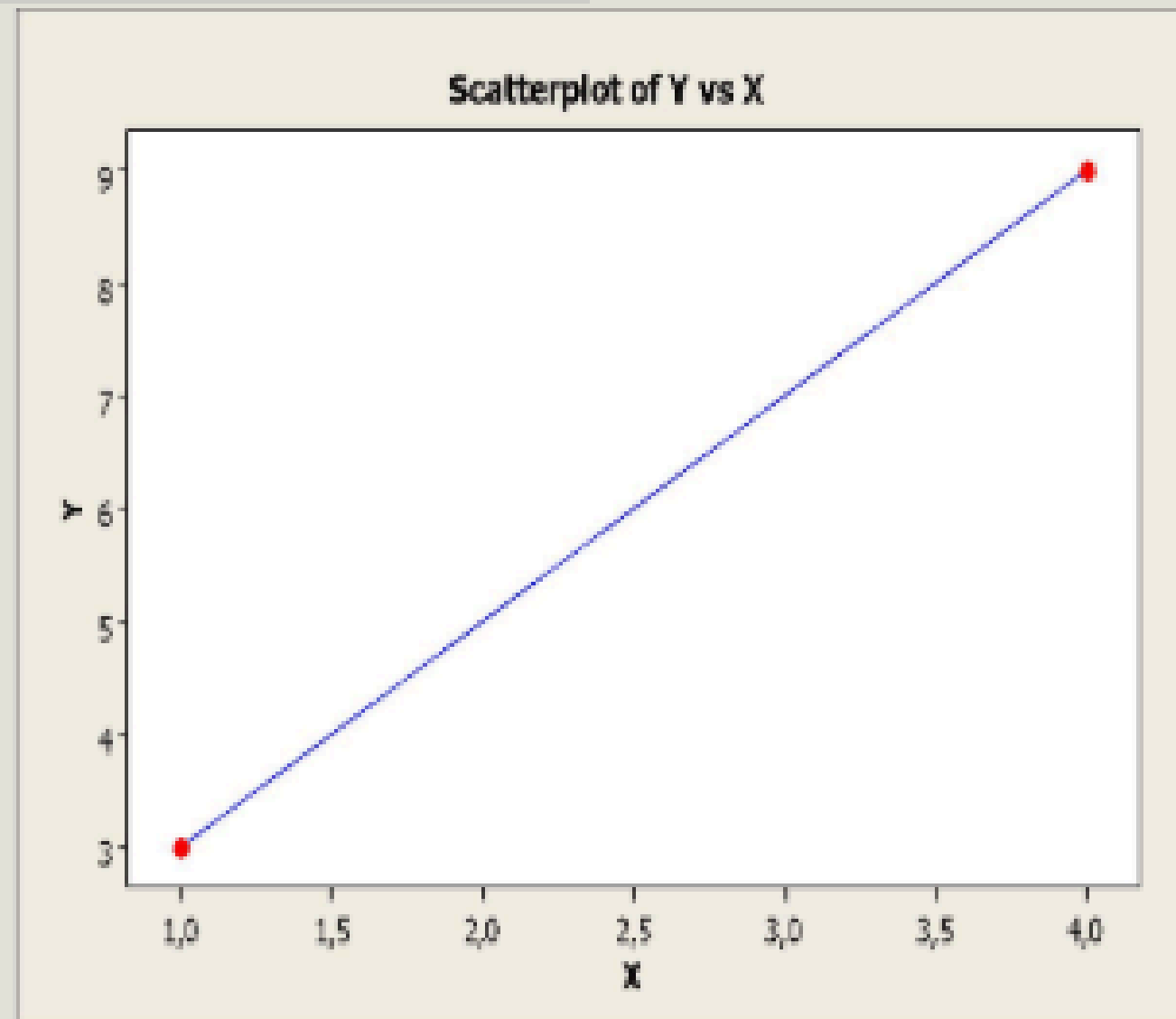
Berapa Y jika $X=2$? \rightarrow pasti
 $Y=5$ Berapa Y jika $X=3$? \rightarrow
pasti $Y=7$

100% kita percaya bahwa nilai $Y=5$
dan $Y=7$ adalah satu-satunya nilai
untuk $X=2$ dan $X=3$

- ☐ Y nilainya pasti, bebas dr eror
- ☐ Y bukan peubah acak



MODEL DETERMINISTIK



Analisis Regresi

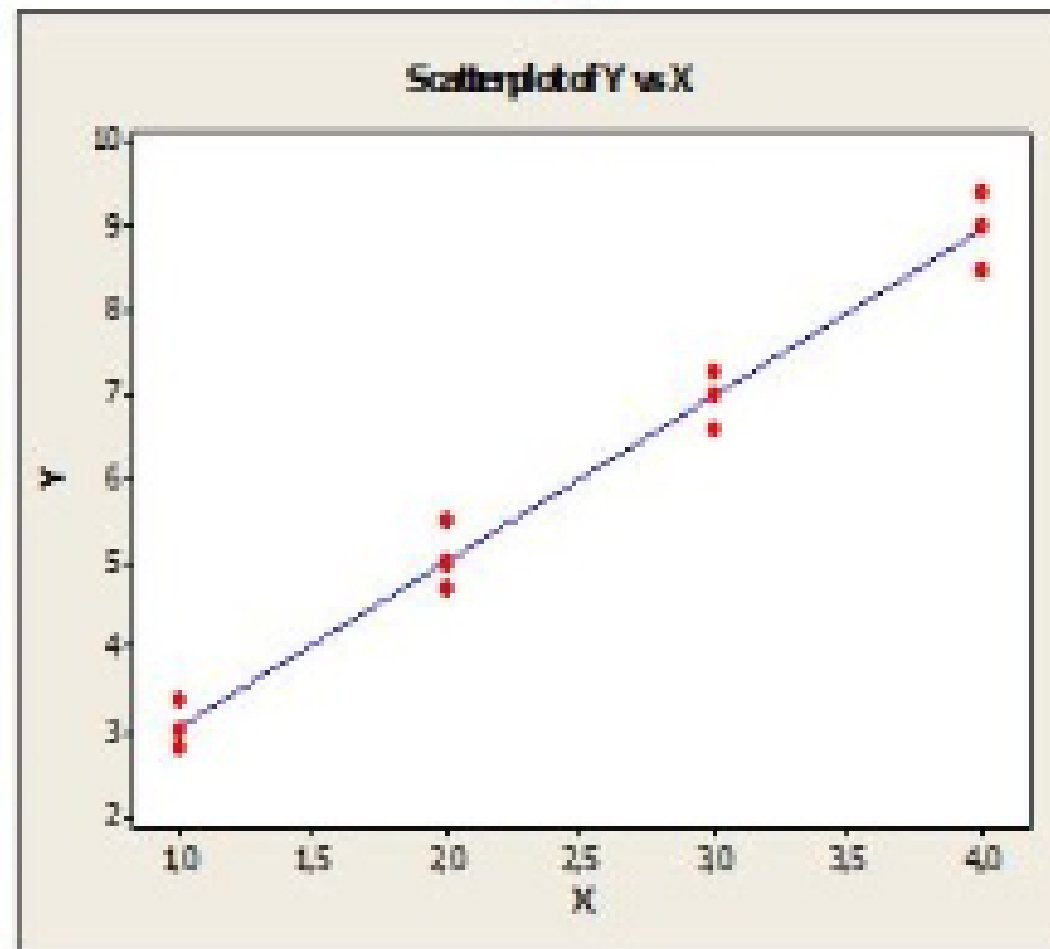
Untuk setiap dosis pupuk yg dicoba-kan, banyaknya padi yang dihasil-kan diamati seba-nyak 3 kali. Hasilnya pd tabel di samping.

Ternyata hasilnya (Y) tidak pasti.
→ mengandung eror
→ Y Peubah Acak

MODEL STOKHASTIK

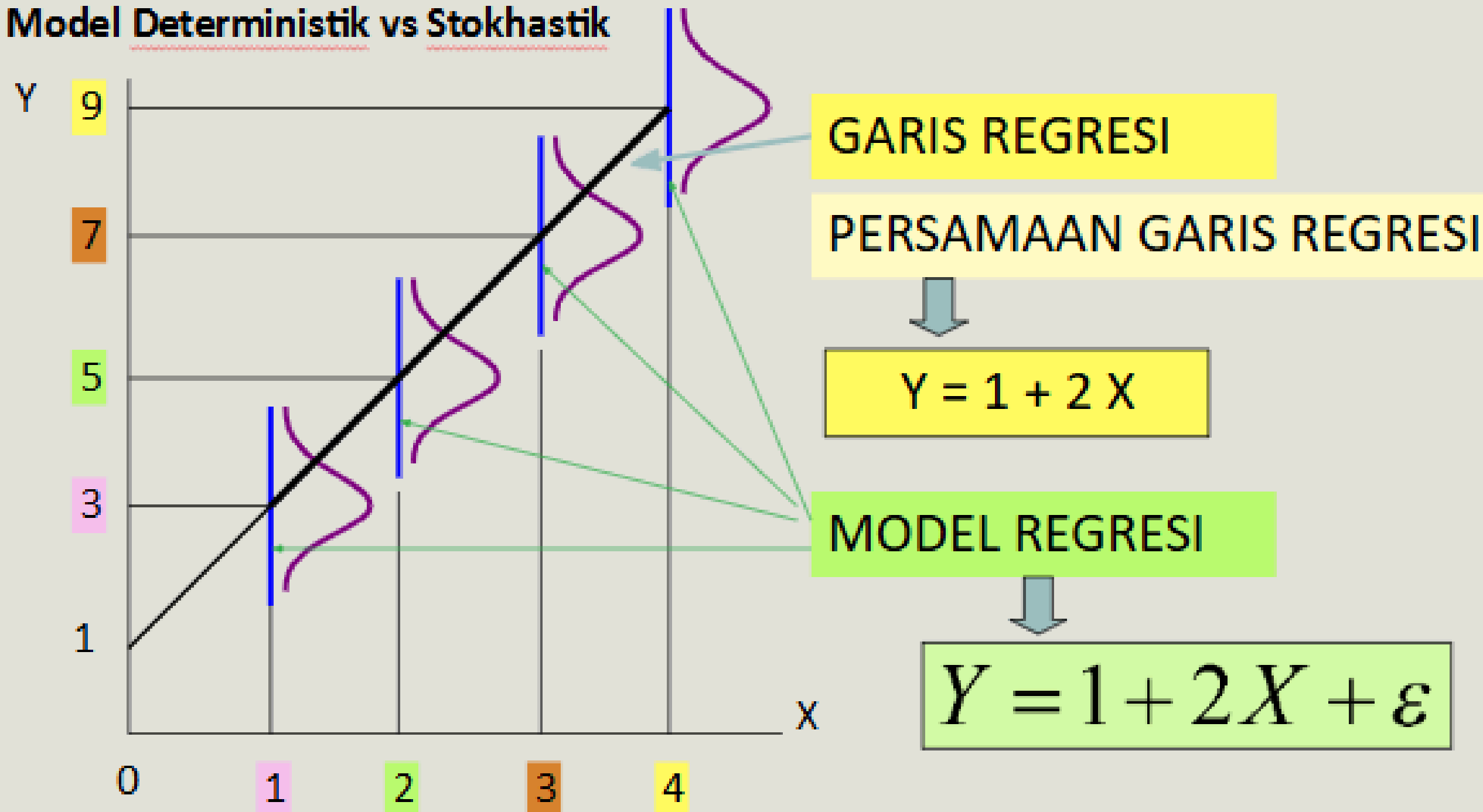
X	Y
1	3
1	3,4
1	2,8
2	5
2	5,4
2	4,7
3	7
3	7,3
3	6,6
4	9
4	9,3
4	8,5

$$Y = 1 + 2x + \varepsilon, \varepsilon = \text{eror}$$
$$y_i = 1 + 2x_i + \varepsilon_i, \varepsilon = \text{eror ke -i}$$



Analisis Regresi

Model Deterministik vs Stokhastik



Latihan

Diduga ada hubungan antara jumlah halaman buku yang digunakan dengan harga jual bukunya. Untuk menjawab dugaan tersebut dipilih sampel dari delapan buku yang dijual di toko tersebut.

Pertanyaan:

1. Buatlah scatterplot dan hitung koefisien korelasinya!
2. Apakah korelasi tersebut signifikan? Jelaskan!

Teknis:

- Kerjakan secara individu
- Disubmit ke <https://ipb.link/tugas-prak-anreg1> dengan format **NIM_Tugas1.pdf**
- **Deadline Jumat (24 Jan 2025), pukul 23.59**

Book	Page	Price (\$)
Intro to History	500	84
Basic Algebra	700	75
Intro to Psyc	800	99
Intro to Sociology	600	72
Bus. Mngt	400	69
Intro to Biology	500	81
Fund. of Jazz	600	63
Princ. of Nursing	800	93