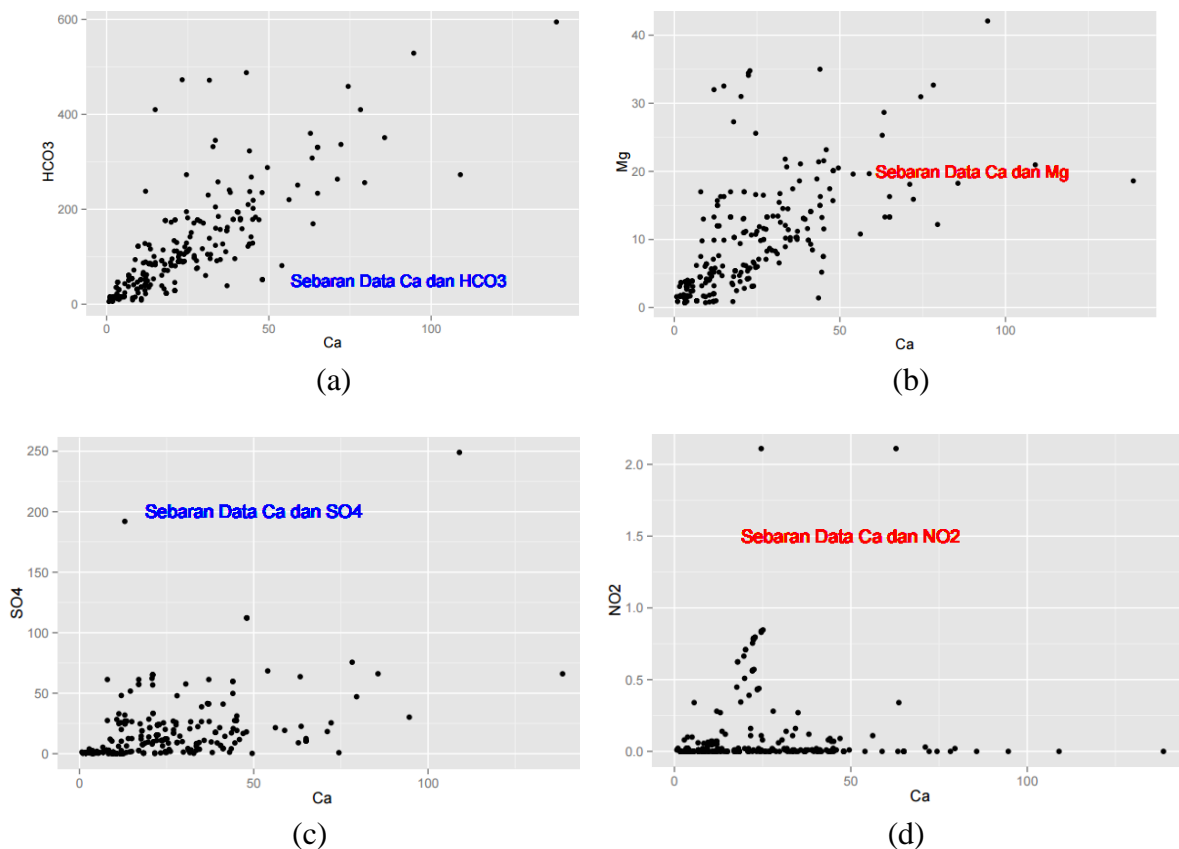


BAB 11

KORELASI LINEAR PEARSON

Analisis Korelasi (Hubungan) Linear dengan Grafik

Berikut disajikan grafik dari sebaran data antara Ca v/s HCO₃, Ca v/s Mg, Ca v/s SO₄, dan Ca v/s NO₂.



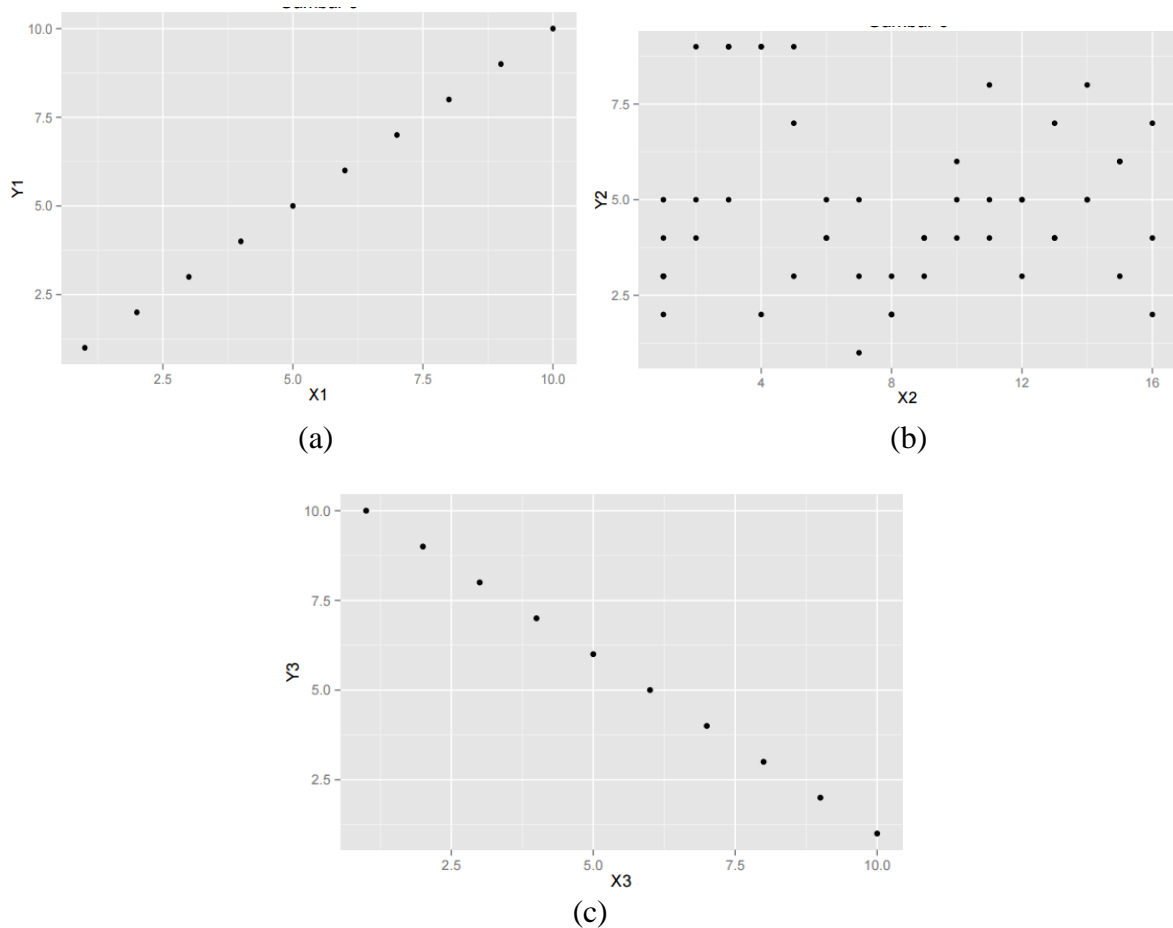
Gambar 11.1

Berdasarkan sebaran data dari Gambar 11.1(a) hingga Gambar 11.1(d), gambar manakah yang kira-kira memiliki sebaran data paling linear? Gambar manakah yang kira-kira memiliki sebaran data paling tidak linear? Pada pembahasan selanjutnya akan diperkenalkan suatu nilai yang dapat mengukur seberapa linear sebaran data untuk dua variabel.

Koefisien Korelasi Linear Pearson

Koefisien korelasi Pearson (dalam hal ini korelasi linear) merupakan suatu nilai yang dapat mengukur seberapa erat hubungan linear yang terjadi di antara dua variabel. Nilai dari koefisien korelasi Pearson berkisar dari -1 sampai 1. Nilai koefisien korelasi Pearson yang semakin mendekati 1 atau -1 menandakan terjadi hubungan linear yang kuat antara dua variabel, sementara jika mendekati 0 menandakan terjadi hubungan linear yang lemah antara dua variabel (mungkin bisa didekati dengan hubungan non-linear, alternatif dari hubungan

linear). Hubungan linear yang terjadi dapat bersifat positif, yakni ditandai dengan nilai koefisien korelasi Pearson yang bernilai positif, atau dapat bersifat negatif, ditandai dengan nilai koefisien korelasi Pearson yang bernilai negatif. Perhatikan Gambar 11.2(a) hingga Gambar 11.2(c).



Gambar 11.2

Gambar 11.2(a) menunjukkan terjadinya hubungan linear positif yang **sempurna** antara X1 dan Y1 (apabila nilai koefisien korelasi Pearson dihitung, maka akan bernilai 1). **Hubungan positif berarti sebaran data cenderung menyebar dari kiri bawah ke kanan atas.** Sebaran data pada Gambar 11.2(b) cenderung acak (tidak beraturan), sehingga hubungan linear yang terjadi antara X2 dan Y2 lemah. Apabila nilai koefisien korelasi Pearson dihitung, maka akan bernilai mendekati 0. Gambar 11.2(c) menunjukkan terjadinya hubungan linear negatif yang **sempurna** antara X3 dan Y3 (apabila nilai koefisien korelasi Pearson dihitung, maka akan bernilai -1). **Hubungan negatif berarti sebaran data cenderung menyebar dari kiri atas ke kanan bawah.**

Menyajikan Grafik Sebaran Data dan Menghitung Koefisien Korelasi Linear Pearson dengan R

Misalkan diberikan data seperti pada Gambar 11.3 dengan nama *file* **contohdata.csv**, dan Gambar 11.4 dengan nama *file* **contohdata2.csv**. Gambar 11.6 disajikan kode R. Apabila kode R pada Gambar 11.6 dieksekusi, hasilnya seperti pada Gambar 11.7 dan Gambar 11.8. Berdasarkan Gambar 11.8, diketahui nilai koefisien korelasi linear Pearson antara X1 dan Y1

adalah 1. Hal ini berarti sebaran data bersifat positif dan linear sempurna (positif berarti sebaran data cenderung bergerak dari kiri bawah ke kanan atas). Perhatikan Gambar 11.2(a). Berdasarkan Gambar 11.8, diketahui nilai koefisien korelasi linear Pearson antara X2 dan Y2 adalah -0,0255. Perhatikan bahwa nilai koefisien korelasi Pearson -0,0255 mendekati 0. Hal ini berarti hubungan linear yang terjadi antara X2 dan Y2 lemah (perhatikan bahwa data menyebar cenderung acak, Gambar 11.2(b)). Berdasarkan Gambar 11.8, diketahui nilai koefisien korelasi Pearson antara X3 dan Y3 adalah -1. Hal ini berarti sebaran data bersifat negatif dan linear sempurna (negatif berarti sebaran data cenderung bergerak dari kiri atas ke kanan bawah). Perhatikan Gambar 11.2(c).

contohdata.csv

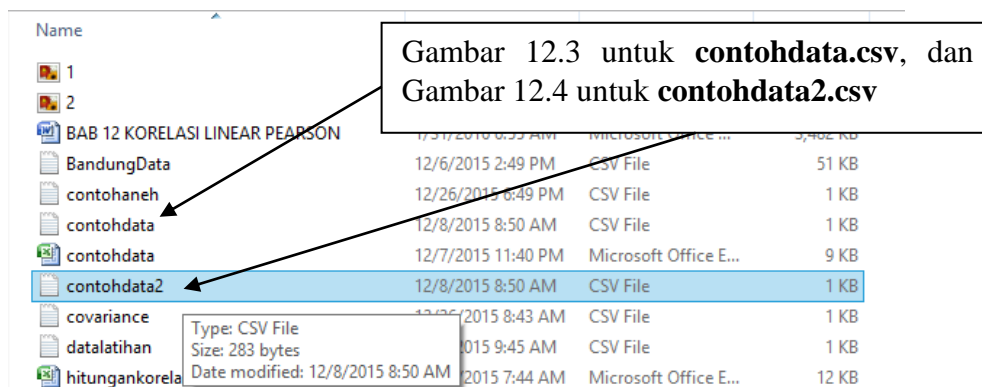
	A	B	C	D	E
1	X1	Y1	X3	Y3	
2	1	1	1	10	
3	2	2	2	9	
4	3	3	3	8	
5	4	4	4	7	
6	5	5	5	6	
7	6	6	6	5	
8	7	7	7	4	
9	8	8	8	3	
10	9	9	9	2	
11	10	10	10	1	
12					

Gambar 11.3

contohdata2.csv

	A	B	C	D	E
1	X2	Y2			
2	1	3			
3	1	4			
4	2	5			
5	3	9			
6	4	2			
7	5	3			
8	6	4			
9	7	1			
10	8	2			
11	9	4			
12	10	6			
13	11	5			
14	12	3			
15	13	4			
16	14	5			
17	15	3			
18	16	2			
19	1	3			
20	1	5			
21	2	9			
22	3	9			
23	4	9			
24	5	7			
25	6	4			

Gambar 11.4



Gambar 11.5

```

1 data1=read.csv("contohdata.csv")
2 data2=read.csv("contohdata2.csv")
3 data1
4 data2
5
6 library(ggplot2)
7 ggplot(data1, aes(x1, y1)) + geom_point() + ggtitle("Hasilnya seperti pada Gambar 12.2(a)")
8 ggplot(data2, aes(x2, y2)) + geom_point() + ggtitle("Hasilnya seperti pada Gambar 12.2(b)")
9 ggplot(data1, aes(x3, y3)) + geom_point() + ggtitle("Hasilnya seperti pada Gambar 12.2(c)")
10
11 cor(data1$x1, data1$y1, method = "pearson")
12 cor(data2$x2, data2$y2, method = "pearson")
13 cor(data1$x3, data1$y3, method = "pearson")
14

```

Gambar 11.6



Gambar 11.7

```
cor(data1$X1, data1$Y1, method = "pearson")
```

```
## [1] 1
```

```
cor(data2$X2, data2$Y2, method = "pearson")
```

```
## [1] -0.02557102
```

```
cor(data1$X3, data1$Y3, method = "pearson")
```

```
## [1] -1
```

Gambar 11.8

Berdasarkan Gambar 11.6, secara umum, perintah untuk menghitung koefisien korelasi linear Pearson dalam R sebagai berikut.

```
cor(variabel1, variabel2, method = "pearson")
```

Berdasarkan Gambar 11.6, *package* **ggplot2** diaktifkan (kode R pada baris 6), dengan maksud untuk menggunakan fungsi **ggplot**. Fungsi **ggplot** bertujuan untuk menyajikan grafik sebaran data.

Menyajikan Grafik Sebaran Data dalam R (Bagian 2)

Grafik dari sebaran data antara Ca v/s HCO₃, Ca v/s Mg, Ca v/s SO₄, dan Ca v/s NO₂, seperti pada Gambar 11.1, akan disajikan kembali, seperti pada Gambar 11.10. Kode R disajikan pada Gambar 11.9.

```
1 data=read.csv("BandungData.csv")
2
3 Ca=data$Ca
4 HCO3=data$HCO3
5 Mg=data$Mg
6 SO4=data$SO4
7 NO2=data$NO2
8
9 library(ggplot2)
10 df <- data.frame(Ca, HCO3, Mg, SO4, NO2)
11 library(reshape)
12 df.melted <- melt(df, id = "Ca")
13 ggplot(data = df.melted, aes(Ca, y = value)) + geom_point(color="firebrick") + facet_grid(variable ~ ., scales='free_y')
14
```

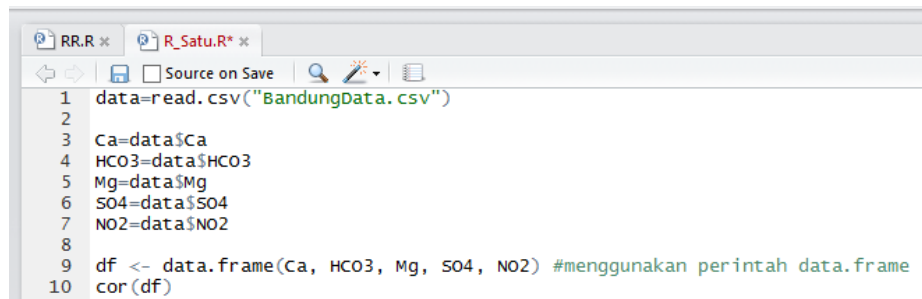
Gambar 11.9



Gambar 11.10

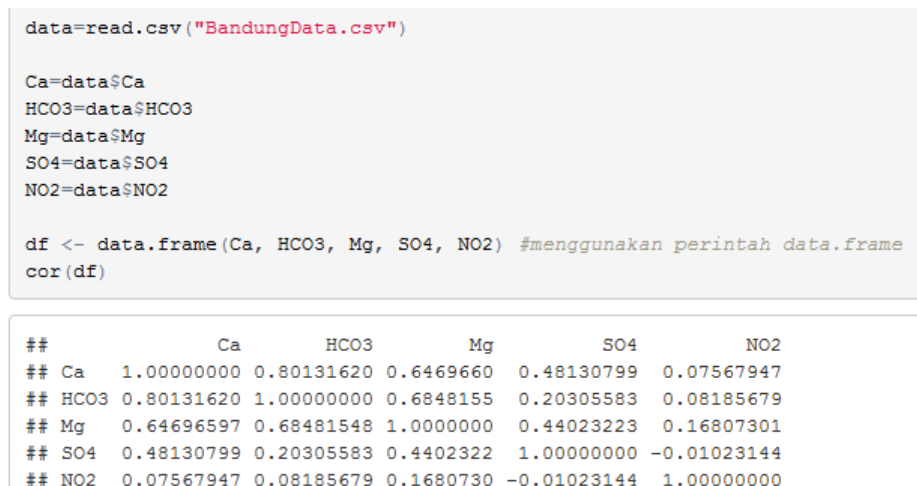
Menghitung Koefisien Korelasi Linear Pearson secara Sekaligus dengan R

Pada pemaparan sebelumnya, penghitungan nilai koefisien korelasi linear Pearson dilakukan secara satu persatu (Gambar 11.8). Dalam R, perhitungan nilai koefisien korelasi linear Pearson dapat dilakukan secara sekaligus dengan menggunakan perintah **data.frame** terlebih dahulu. Perhatikan ilustrasi berikut.



```
1 data=read.csv("BandungData.csv")
2
3 Ca=data$Ca
4 HCO3=data$HCO3
5 Mg=data$Mg
6 SO4=data$SO4
7 NO2=data$NO2
8
9 df <- data.frame(Ca, HCO3, Mg, SO4, NO2) #menggunakan perintah data.frame
10 cor(df)
```

Gambar 11.11



```
data=read.csv("BandungData.csv")

Ca=data$Ca
HCO3=data$HCO3
Mg=data$Mg
SO4=data$SO4
NO2=data$NO2

df <- data.frame(Ca, HCO3, Mg, SO4, NO2) #menggunakan perintah data.frame
cor(df)
```

	Ca	HCO3	Mg	SO4	NO2
Ca	1.00000000	0.80131620	0.6469660	0.48130799	0.07567947
HCO3	0.80131620	1.00000000	0.6848155	0.20305583	0.08185679
Mg	0.64696597	0.68481548	1.00000000	0.44023223	0.16807301
SO4	0.48130799	0.20305583	0.4402322	1.00000000	-0.01023144
NO2	0.07567947	0.08185679	0.1680730	-0.01023144	1.00000000

Gambar 11.12

Berdasarkan Gambar 11.12, nilai koefisien korelasi linear Pearson antara Ca dan HCO3 adalah 0,80131620, nilai koefisien korelasi linear Pearson antara Ca dan Mg adalah 0,6469660, nilai koefisien korelasi linear Pearson antara Ca dan SO4 adalah 0,48130799, dan seterusnya. Di antara variabel HCO3, Mg, SO4, dan NO2, variabel HCO3 yang memiliki keeratan linear yang paling tinggi terhadap variabel Ca, yakni bernilai 0,80131620.

Contoh Perhitungan Koefisien Korelasi Linear Pearson dan Penyelesaian dalam R

Misalkan diberikan data seperti pada Tabel 11.1. Berdasarkan data pada Tabel 11.1, berikut rumus untuk menghitung nilai koefisien korelasi linear Pearson (r).

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}}$$

Tabel 11.1

X	Y
10	3,01
12	3,15
9	2,9
10	3,1
8	2,7
11	3,25
15	3,6
17	3,7
16	3,65
10	3,15

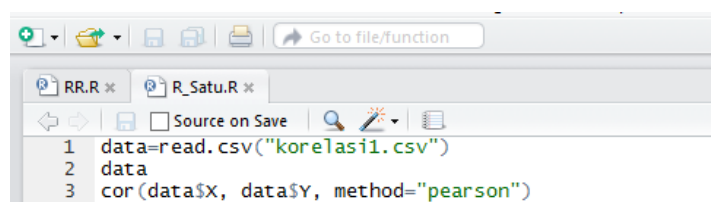
Tabel 11.2

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
10	3,01	-1,8	-0,211	3,24	0,044521	0,3798
12	3,15	0,2	-0,071	0,04	0,005041	-0,0142
9	2,9	-2,8	-0,321	7,84	0,103041	0,8988
10	3,1	-1,8	-0,121	3,24	0,014641	0,2178
8	2,7	-3,8	-0,521	14,44	0,271441	1,9798
11	3,25	-0,8	0,029	0,64	0,000841	-0,0232
15	3,6	3,2	0,379	10,24	0,143641	1,2128
17	3,7	5,2	0,479	27,04	0,229441	2,4908
16	3,65	4,2	0,429	17,64	0,184041	1,8018
10	3,15	-1,8	-0,071	3,24	0,005041	0,1278
Jumlah	118	32,21	0	87,6	1,00169	9,072
Rata-Rata	11,8	3,221				

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}} = \frac{9,072}{\sqrt{87,6} \sqrt{1,00169}} = \frac{9,072}{(9,359487165)(1,000844643)}$$

$$r = 0,968465868$$

Berdasarkan perhitungan secara manual, diperoleh nilai koefisien korelasi linear Pearson $r = 0,968465868$. Berikut hasil perhitungan nilai koefisien korelasi linear Pearson berdasarkan R.



```

1 data=read.csv("korelas1.csv")
2 data
3 cor(data$X, data$Y, method="pearson")

```

Gambar 11.13

```
data=read.csv("korelasi1.csv")
data

##      X      Y
## 1  10  3.01
## 2  12  3.15
## 3   9  2.90
## 4  10  3.10
## 5   8  2.70
## 6  11  3.25
## 7  15  3.60
## 8  17  3.70
## 9  16  3.65
## 10 10  3.15

cor(data$X, data$Y, method="pearson")

## [1] 0.9684659
```

Gambar 11.14

Contoh Perhitungan Covariance dan Penyelesaian dalam R

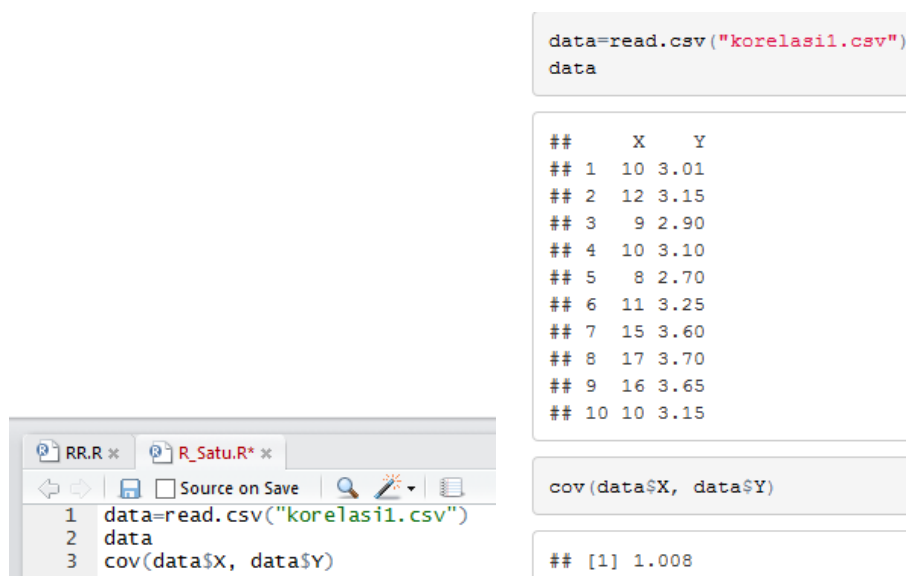
Berdasarkan data pada Tabel 11.1, berikut rumus untuk menghitung *covariance* antara variabel X dan variabel Y ($cov(X, Y)$).

$$cov(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1}$$

Perhatikan bahwa n menyatakan banyaknya data, yakni $n = 10$.

$$cov(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1} = \frac{9,072}{10 - 1} = 1,008$$

Berdasarkan perhitungan secara manual, diperoleh nilai $cov(X, Y) = 1,008$. Berikut hasil perhitungan nilai $cov(X, Y)$ berdasarkan R.



```
data=read.csv("korelasi1.csv")
data

##      X      Y
## 1  10  3.01
## 2  12  3.15
## 3   9  2.90
## 4  10  3.10
## 5   8  2.70
## 6  11  3.25
## 7  15  3.60
## 8  17  3.70
## 9  16  3.65
## 10 10  3.15

cov(data$X, data$Y)

## [1] 1.008
```

Gambar 11.15

Referensi

1. Agresti, A. dan B. Finlay. 2009. *Statistical Methods for the Social Sciences*, 4th Edition. United States of America: Prentice Hall.
2. Field, A. 2009. *Discovering Statistics Using SPSS*, 3rd Edition. London: Sage.
3. Gio, P.U. dan E. Rosmaini, 2015. Belajar Olah Data dengan SPSS, Minitab, R, Microsoft Excel, EViews, LISREL, AMOS, dan SmartPLS. USUpres.
4. Mann, P. S. dan C.J. Lacke. 2011. *Introductory Statistics, International Student Version*, 7th Edition. Asia: John Wiley & Sons, Inc.
5. Montgomery, D. C. dan G. C. Runger. 2011. *Applied Statistics and Probability for Engineers*, 5th Edition. United States of America: John Wiley & Sons, Inc.
6. Ott, R.L. dan M. Longnecker. 2001. *An Introduction to Statistical Methods and Data Analysis*, 5th Edition. United States of America: Duxbury.
7. Smidth, R. K. dan D. H. Sanders. 2000. *Statistics a First Course*, 6th Edition. United States of America: McGraw-Hill Companies.
8. <http://www.statmethods.net/stats/correlations.html>
9. <http://www.r-bloggers.com/correlation-and-linear-regression/>
10. <http://www.r-bloggers.com/pairwise-complete-correlation-considered-dangerous/>