

# **MODUL PEMROGRAMAN R**



**Disusun Oleh:**

**TIM DOSEN PRODI SARJANA TERAPAN MANAJEMEN INFORMASI KESEHATAN**

**UNIVERSITAS INDONESIA MAJU**

**PROGRAM STUDI SARJANA TERAPAN MANAJEMEN INFORMASIKESEHATAN**

**FAKULTAS VOKASI**

**UNIVERSITAS INDONESIA MAJU**

**JAKARTA**

**2022**



## **Modul Pemrograman R**

Nama Mahasiswa : \_\_\_\_\_  
NPM : \_\_\_\_\_

Program Studi Sarjana Terapan Manajemen Informasi Kesehatan

Fakultas Vokasi

Universitas Indonesia Maju

2022

## **KATA PENGANTAR**

Buku petunjuk praktikum ini disusun untuk memenuhi kebutuhan mahasiswa sebagai panduan dalam melaksanakan praktikum Pemrograman R, untuk mahasiswa program studi D4 Manajemen Informasi Kesehatan (MIK) UIMA. Dengan adanya buku petunjuk praktikum ini diharapkan akan membantu dan mempermudah mahasiswa dalam memahami dan melaksanakan praktikum Pemrograman R sehingga akan memperoleh hasil yang baik.

Materi yang dipraktikkan merupakan materi yang selaras dengan materi kuliah Pemrograman R. Untuk itu dasar teori yang didapatkan saat kuliah juga akan sangat membantu mahasiswa dalam melaksanakan praktikum ini.

Buku petunjuk ini masih dalam proses penyempurnaan. Insha Allah perbaikan akan terus dilakukan demi kesempurnaan buku petunjuk praktikum ini dan disesuaikan dengan perkembangan ilmu pengetahuan. Semoga buku petunjuk ini dapat dipergunakan sebagaimana mestinya.

Jakarta, September 2022

Penyusun

## DAFTAR ISI

HALAMAN SAMPUL.....	i
KATA PENGANTAR .....	iii
DAFTAR ISI.....	iv
BAB I MENYAJIKAN DATA DALAM GRAFIK .....	1
BAB II DISTRIBUSI SAMPLING .....	33
BAB III KORELASI LINEAR PEARSON.....	58
DAFTAR PUSTAKA .....	66

# BAB I

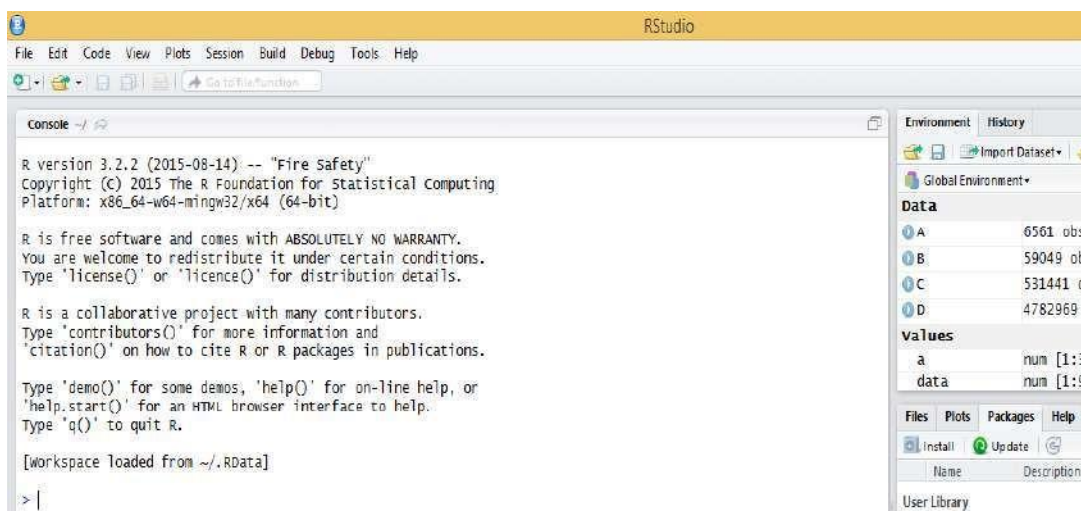
## MENYAJIKAN DATA DALAM GRAFIK

### A. Memulai R

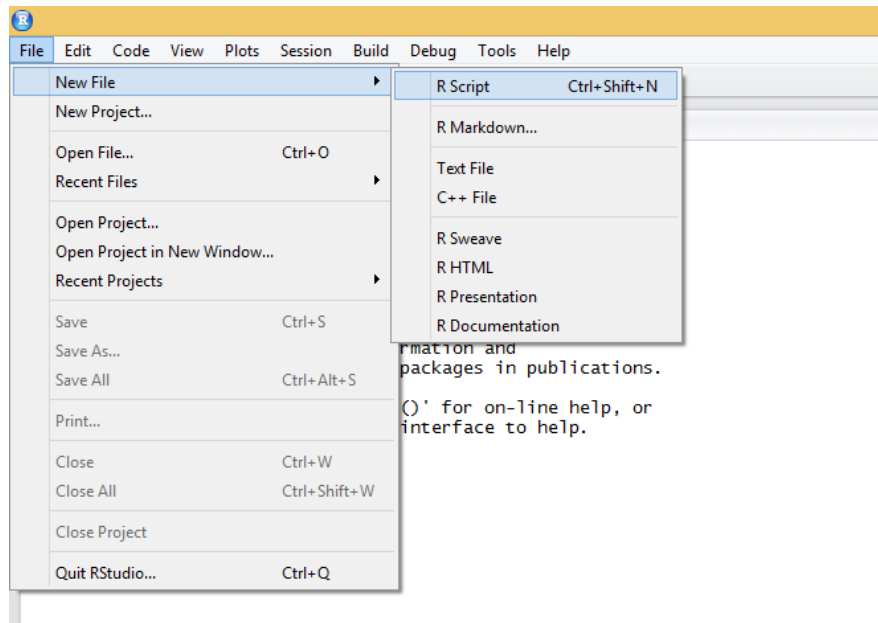
Berikut dipaparkan langkah-langkah untuk masuk ke dalam area kerja R. Aktifkan RStudio terlebih dahulu (Gambar 1.1), sehingga akan muncul tampilan seperti pada Gambar 1.2. Pada Gambar 1.2, pilih *File => New File => R Script* (lihat Gambar 1.3), sehingga muncul tampilan seperti pada Gambar 1.4. Gambar 1.4 merupakan area kerja R, di mana pada pembahasan selanjutnya, kode R akan diinput pada area tersebut. Setelah kode R diinput, selanjutnya kode R tersebut dieksekusi, sehingga muncul *output* berdasarkan eksekusi kode R tersebut.



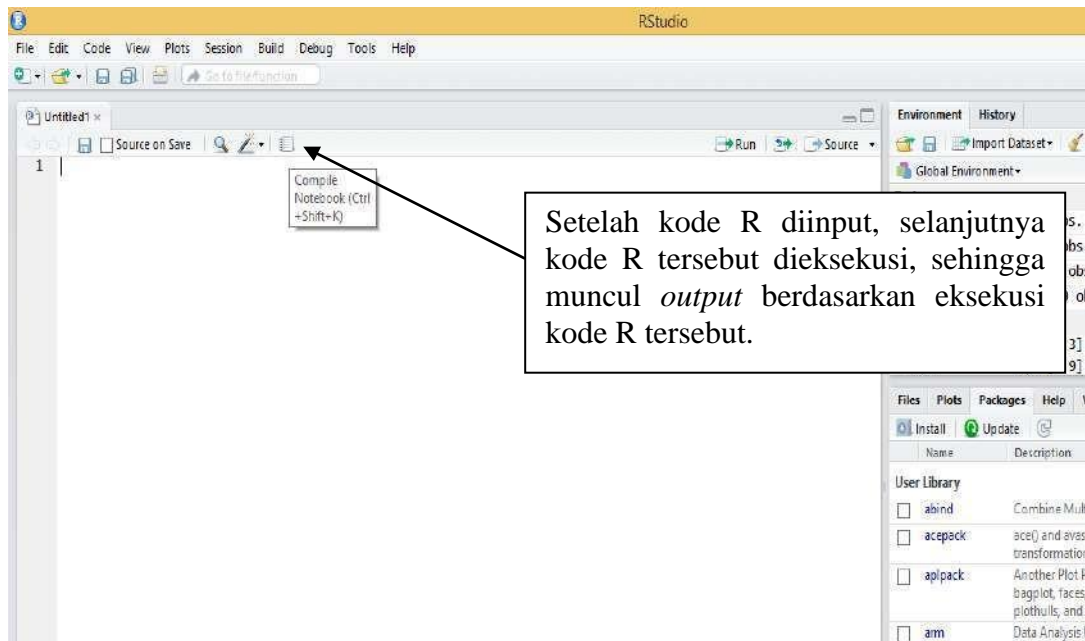
Gambar 1.1



Gambar 1.2



**Gambar 1.3**



**Gambar 1.4**

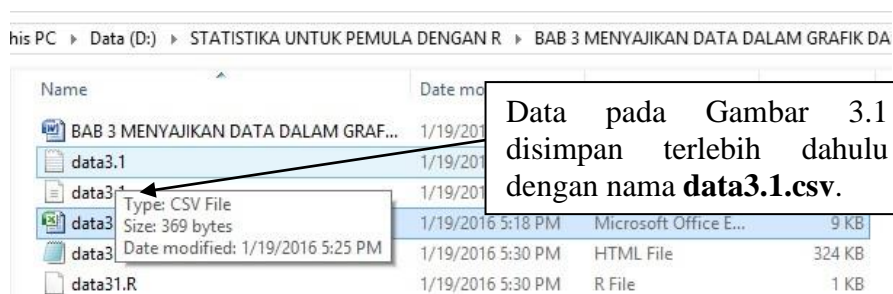
## B. Memplot Data dalam R (Scatter Plot)

Misalkan diberikan data seperti pada Gambar 1.5. Berdasarkan Gambar 1.5, diketahui terdapat 10 responden laki-laki dan 10 responden perempuan. Masing-masing responden disajikan informasi mengenai pendapatan dan pengeluaran per-bulan, dalam jutaan. Sebagai contoh, responden ke-1 adalah laki-laki, dengan pendapatan Rp. 1.000.000, dan pengeluaran Rp. 500.000. Responden ke-20 adalah perempuan, dengan pendapatan Rp. 10.000.000, dan pengeluaran Rp. 8.000.000.

	A	B	C	D
1	jeniskelamin	pendapatan	pengeluaran	
2	laki-laki	1	0.5	
3	laki-laki	2	0.9	
4	laki-laki	3	1.1	
5	laki-laki	4	2	
6	laki-laki	5	4	
7	laki-laki	6	5	
8	laki-laki	7	6.2	
9	laki-laki	8	5.5	
10	laki-laki	9	5.8	
11	laki-laki	10	4.5	
12	perempuan	1	0.7	
13	perempuan	2	1.4	
14	perempuan	3	2.8	
15	perempuan	4	3.3	
16	perempuan	5	4.3	
17	perempuan	6	5	
18	perempuan	7	6.7	
19	perempuan	8	5.9	
20	perempuan	9	5.7	
21	perempuan	10	8	

Gambar 1.5

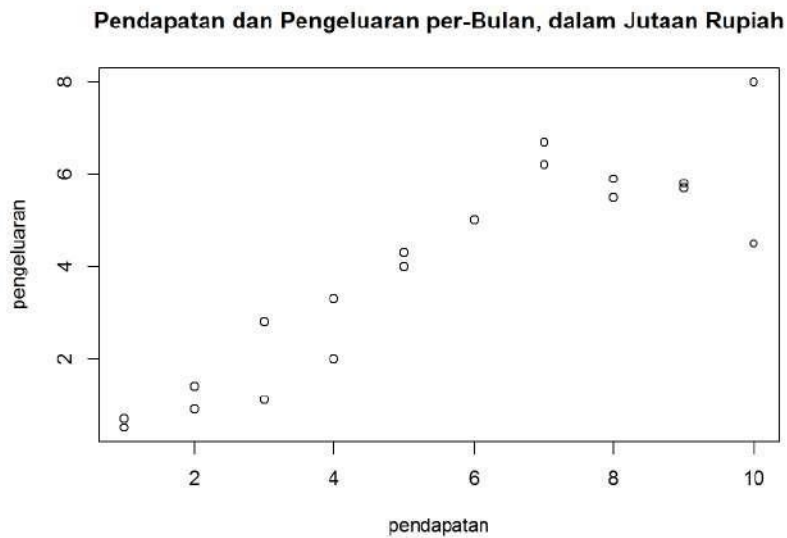
Data pada Gambar 1.5 disimpan terlebih dahulu dengan nama **data3.1.csv** (perhatikan Gambar 1.6).



Gambar 1.6

Data berdasarkan Gambar 1.5 disajikan ke dalam grafik seperti pada Gambar 1.7.

```
plot(simpan[2:3], main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan Rupiah")
```



Gambar 1.7

Kode R untuk menyajikan data pada Gambar 1.5, seperti pada Gambar 1.7, adalah sebagai berikut (Gambar 1.8).

```
1 simpan=read.table("data3.1.csv",header=TRUE, sep=",") #membaca data
2 simpan
3
4 plot(simpan[2:3], main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan Rupiah")
```

Gambar 1.8

Berdasarkan Gambar 1.8, perhatikan kode R berikut (kode R baris pertama).

```
simpan=read.table("data3.1.csv",header=TRUE, sep=",") #membaca data
```

Kode R tersebut (kode R baris pertama) dapat diartikan variabel **simpan** ditugaskan untuk menyimpan data pada variabel **jeniskelamin**, **pendapatan**, dan **pengeluaran** dalam file **data3.1.csv**. Perhatikan kode R berikut (kode R baris kedua).

```
simpan
```

Kode R baris kedua berarti menampilkan nilai yang disimpan dalam variabel **simpan**. Hasilnya seperti pada Gambar 1.9.



```
simpan=read.table("data3.1.csv",header=TRUE, sep=",") #membaca data
simpan
```

```
##   jeniskelamin pendapatan pengeluaran
## 1   laki-laki           1           0.5
## 2   laki-laki           2           0.9
## 3   laki-laki           3           1.1
## 4   laki-laki           4           2.0
## 5   laki-laki           5           4.0
## 6   laki-laki           6           5.0
## 7   laki-laki           7           6.2
## 8   laki-laki           8           5.5
## 9   laki-laki           9           5.8
## 10  laki-laki          10           4.5
## 11  perempuan           1           0.7
## 12  perempuan           2           1.4
## 13  perempuan           3           2.8
## 14  perempuan           4           3.3
## 15  perempuan           5           4.3
## 16  perempuan           6           5.0
## 17  perempuan           7           6.7
## 18  perempuan           8           5.9
## 19  perempuan           9           5.7
## 20  perempuan          10           8.0
```

Gambar 1.9

Kode R pada baris keempat (Gambar 1.10), yakni

**plot(simpan[2:3], main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan Rupiah")**

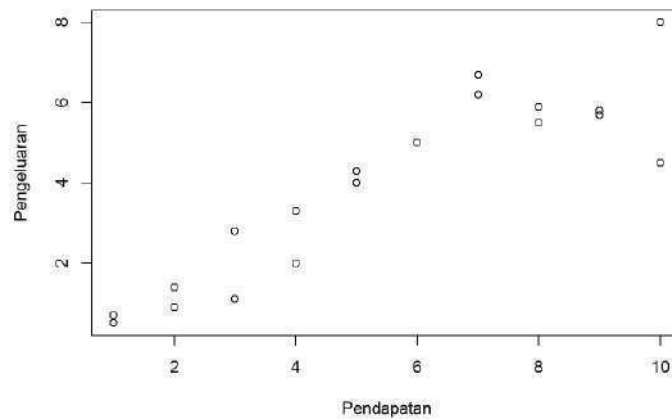
dapat diartikan data pada variabel **pendapatan** (pada kolom 2) dan data pada variabel **pengeluaran** (pada kolom 3), disajikan ke dalam grafik, seperti pada Gambar 3.3. Kode R pada baris keempat mencantumkan **main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan Rupiah"**, yang berguna untuk memberikan judul grafik. Pada Gambar 1.10, kode R pada baris 6 sampai baris 8, apabila dieksekusi, hasilnya seperti pada Gambar 1.11.

```
1  simpan=read.table("data3.1.csv",header=TRUE, sep=",") #membaca data
2  simpan
3
4  plot(simpan[2:3], main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan Rupiah")
5
6  Pendapatan=simpan$pendapatan
7  Pengeluaran=simpan$pengeluaran
8  plot(Pendapatan, Pengeluaran)
```

Gambar 1.10

Pada Gambar 1.10, kode R pada baris keenam, yakni **Pendapatan=simpan\$pendapatan**, berarti variabel **Pendapatan** ditugaskan untuk menyimpan data pada variabel **pendapatan**, dalam variabel **simpan**. Kode R pada baris ketujuh, yakni **Pengeluaran=simpan\$pengeluaran**, berarti variabel **Pengeluaran** ditugaskan untuk menyimpan data pada variabel **pengeluaran**, dalam variabel **simpan**. Kode R pada baris kedelapan, yakni **plot(Pendapatan, Pengeluaran)**, berarti memplot data ke dalam grafik, dengan variabel

**Pendapatan** sebagai sumbu horizontal, dan variabel **Pengeluaran** sebagai sumbu vertikal. Hasilnya seperti pada Gambar 1.11.



**Gambar 1.11**

Pada Gambar 1.12, kode R pada baris 10 sampai baris 14, apabila dieksekusi, hasilnya seperti pada Gambar 1.13.

```

1  simpan=read.table("data3.1.csv",header=TRUE, sep=",") #membaca data
2  simpan
3
4  plot(simpan[2:3], main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan Rupiah")
5
6  Pendapatan=simpan$pendapatan
7  Pengeluaran=simpan$pengeluaran
8  plot(Pendapatan, Pengeluaran)
9
10 library(ggplot2)
11 PENDAPATAN=simpan$pendapatan
12 PENGELUARAN=simpan$pengeluaran
13 qplot(PENDAPATAN, PENGELUARAN, main="Pendapatan dan Pengeluaran per-Bulan, dalam jutaan",
14       xlab="Pendapatan per-Bulan", ylab="Pengeluaran per-Bulan")

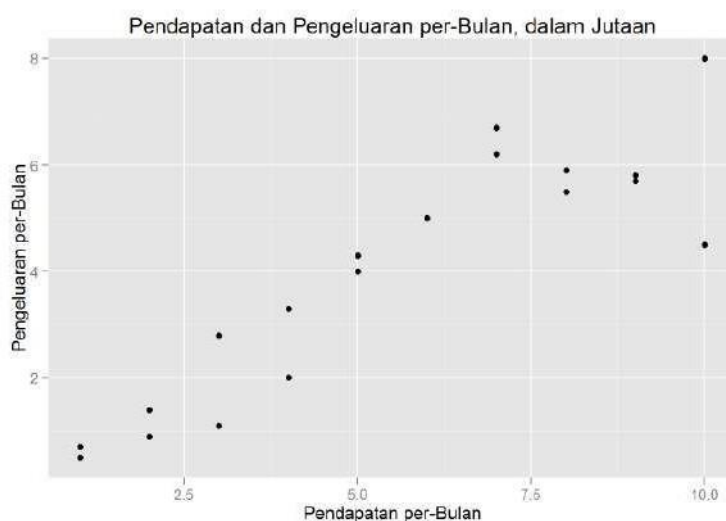
```

**Gambar 1.12**

```

PENDAPATAN=simpan$pendapatan
PENGELUARAN=simpan$pengeluaran
qplot(PENDAPATAN, PENGELUARAN, main="Pendapatan dan Pengeluaran per-Bulan, dalam jutaan",
      xlab="Pendapatan per-Bulan", ylab="Pengeluaran per-Bulan")

```



**Gambar 1.13**

Pada Gambar 1.12, kode R pada baris kesepuluh, yakni `library(ggplot2)`, berarti mengaktifkan *package ggplot2*. Pengaktifkan *package ggplot2* bertujuan untuk menggunakan fungsi `qplot()`. Kode R pada baris kesebelas, yakni `PENDAPATAN=simpan$pendapatan`, berarti variabel `PENDAPATAN` ditugaskan untuk menyimpan data pada variabel `pendapatan`, dalam variabel `simpan`. Kode R pada baris kedubelas, yakni `PENGELUARAN=simpan$pengeluaran`, berarti variabel `PENGELUARAN` ditugaskan untuk menyimpan data pada variabel `pendapatan`, dalam variabel `simpan`. Kode R pada baris ketigabelas dan keempatbelas, yakni `qplot(PENDAPATAN, PENGELUARAN, main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan", xlab="Pendapatan per-Bulan", ylab="Pengeluaran per-Bulan")`, berarti memplot data ke dalam grafik. Hasilnya seperti pada Gambar 1.13.

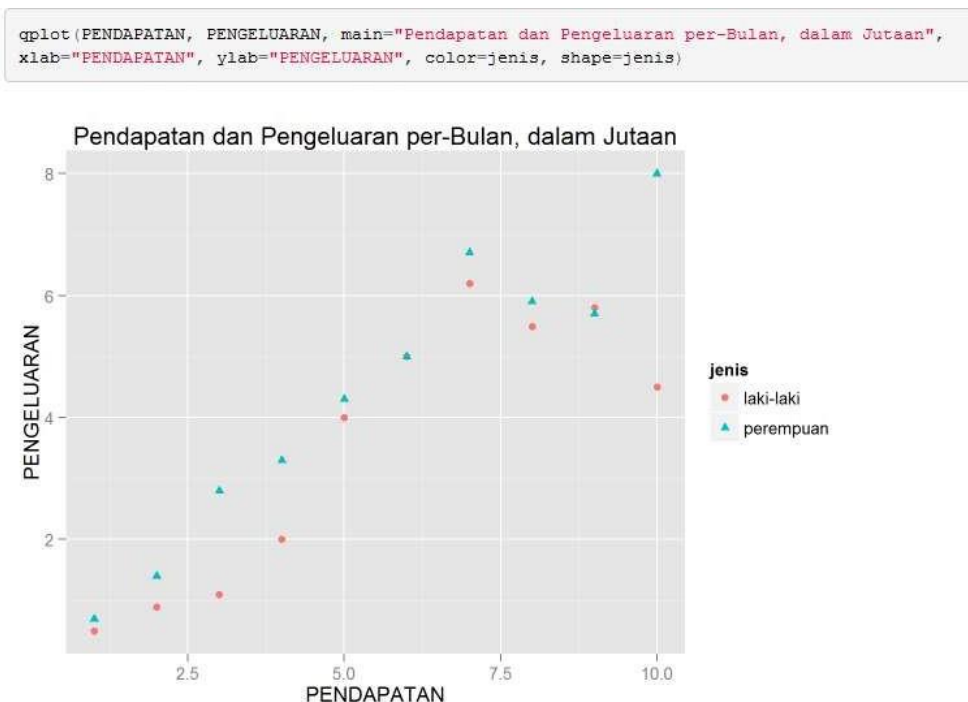
Pada Gambar 1.14, kode R pada baris 21 sampai baris 22, apabila dieksekusi, hasilnya seperti pada Gambar 1.15.

```

15
16 library(ggplot2)
17 jenis=simpan$jeniskelamin
18 qplot(PENDAPATAN, PENGELUARAN, main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan",
19 xlab="Pendapatan", ylab="Pengeluaran", color=jenis)
20
21 qplot(PENDAPATAN, PENGELUARAN, main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan",
22 xlab="PENDAPATAN", ylab="PENGELUARAN", color=jenis, shape=jenis)

```

**Gambar 1.14**



**Gambar 1.15**

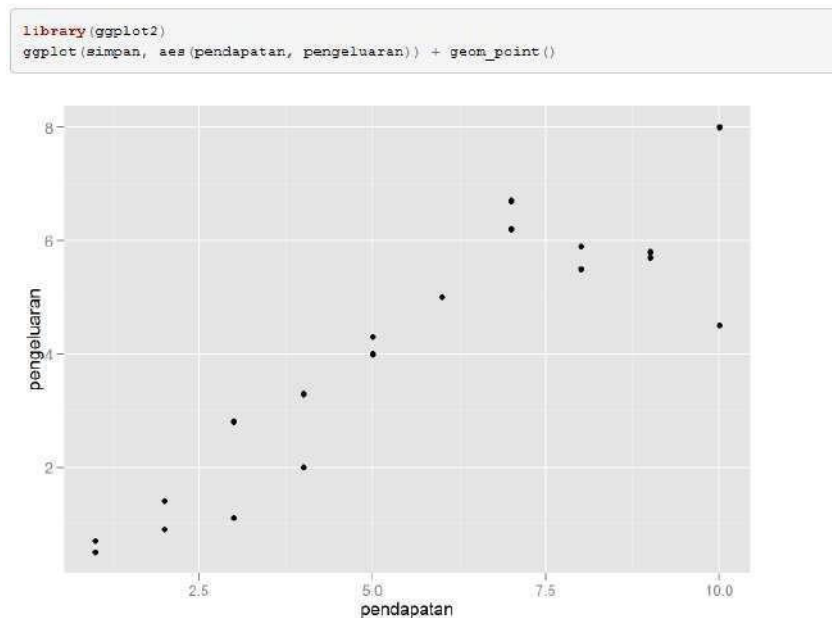
Pada Gambar 1.16, kode R pada baris 24 sampai baris 25, apabila dieksekusi, hasilnya seperti pada Gambar 1.17.

```

19 xlab="Pendapatan", ylab="Pengeluaran", color=jenis)
20
21 qplot(PENDAPATAN, PENGELUARAN, main="Pendapatan dan Pengeluaran per-Bulan, dalam Jutaan",
22 xlab="PENDAPATAN", ylab="PENGELUARAN", color=jenis, shape=jenis)
23
24 library(ggplot2)
25 ggplot(simpan, aes(pendapatan, pengeluaran)) + geom_point()
26

```

Gambar 1.16



Gambar 1.17

Ketik kode R seperti pada Gambar 1.18, dan amati hasil eksekusi dari kode R tersebut.

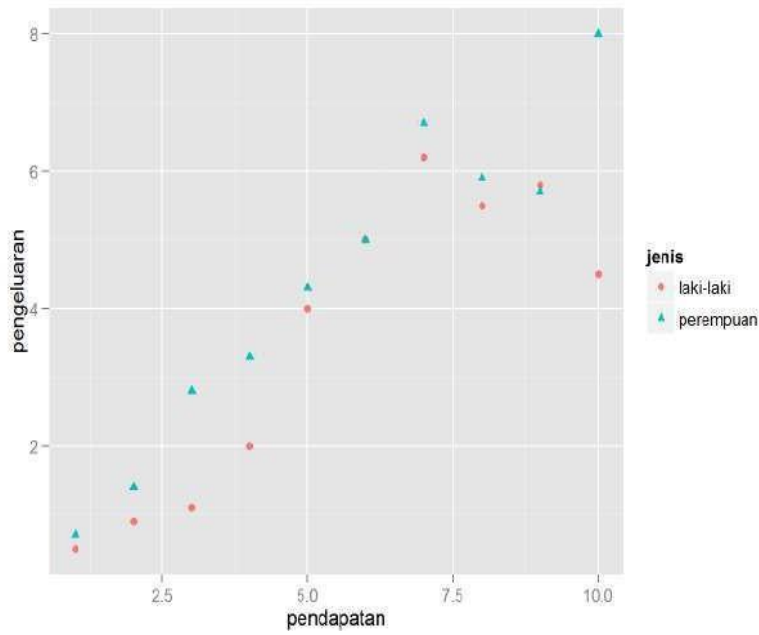
```

22 xlab= PENDAPATAN , ylab= PENGELUARAN , color=jenis, shape=jenis)
23
24 library(ggplot2)
25 ggplot(simpan, aes(pendapatan, pengeluaran)) + geom_point()
26
27
28 ggplot(simpan, aes(pendapatan, pengeluaran)) + geom_point(aes(color = jenis, shape = jenis))
29
30 grafik <- ggplot(simpan, aes(pendapatan, pengeluaran)) + geom_point(aes(color = jenis, shape = jenis))
31 grafik + scale_colour_manual(values = c("blue", "orange"))
32
33 grafik + scale_shape_manual(values = c(16, 5))
34
35 grafik + scale_colour_manual(values = c("blue", "orange")) + scale_shape_manual(values = c(5, 5))
36
37 grafik + facet_grid(.~ jeniskelamin)
38
39 grafik + facet_grid(. ~ jeniskelamin) + scale_colour_manual(values = c("blue", "orange"))
40
41 grafik + geom_vline(xintercept = 2.5)
42
43 grafik + geom_vline(xintercept = 2.5) + geom_vline(xintercept = 5)
44
45 grafik + geom_vline(xintercept = 1:5)
46
47 grafik + geom_vline(xintercept = c(2.5, 5, 7.5))
48
49 grafik + geom_vline(xintercept = c(2.5, 5, 7.5), colour="green", linetype = "longdash")
50
51 grafik + geom_vline(xintercept = c(2.5, 5, 7.5), colour="green", linetype = "longdash") +
52 geom_hline(yintercept = c(2, 4, 6), colour="red", linetype = "longdash")
53

```

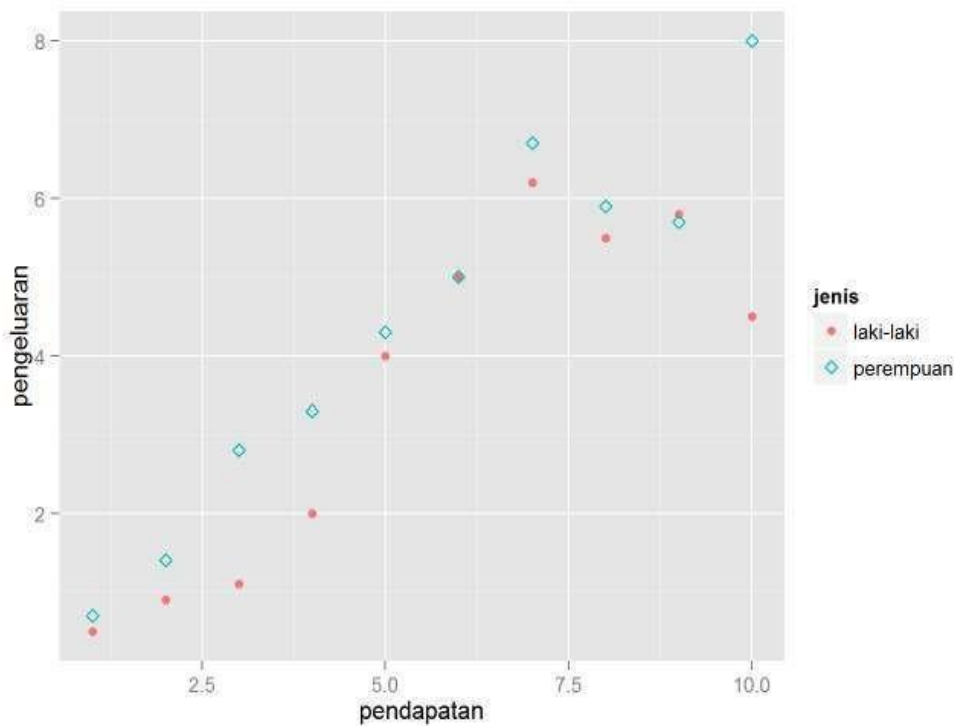
Gambar 1.18

```
ggplot(simpan, aes(pendapatan, pengeluaran)) + geom_point(aes(color = jenis, shape = jenis))
```



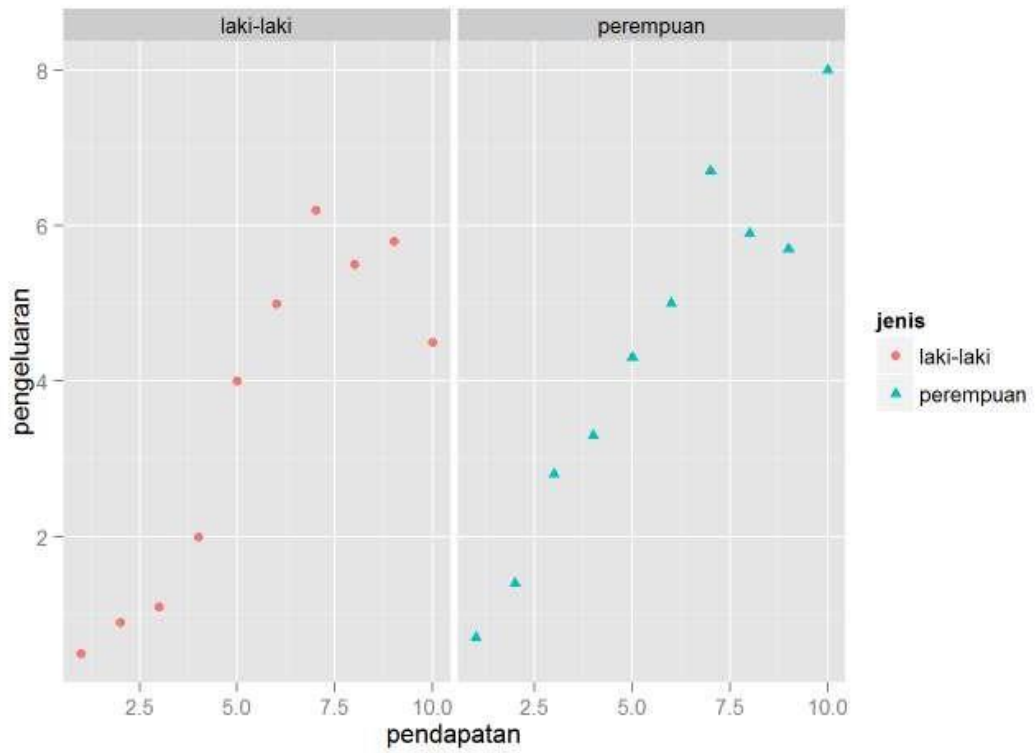
Gambar 1.19

```
grafik + scale_shape_manual(values = c(16, 5))
```



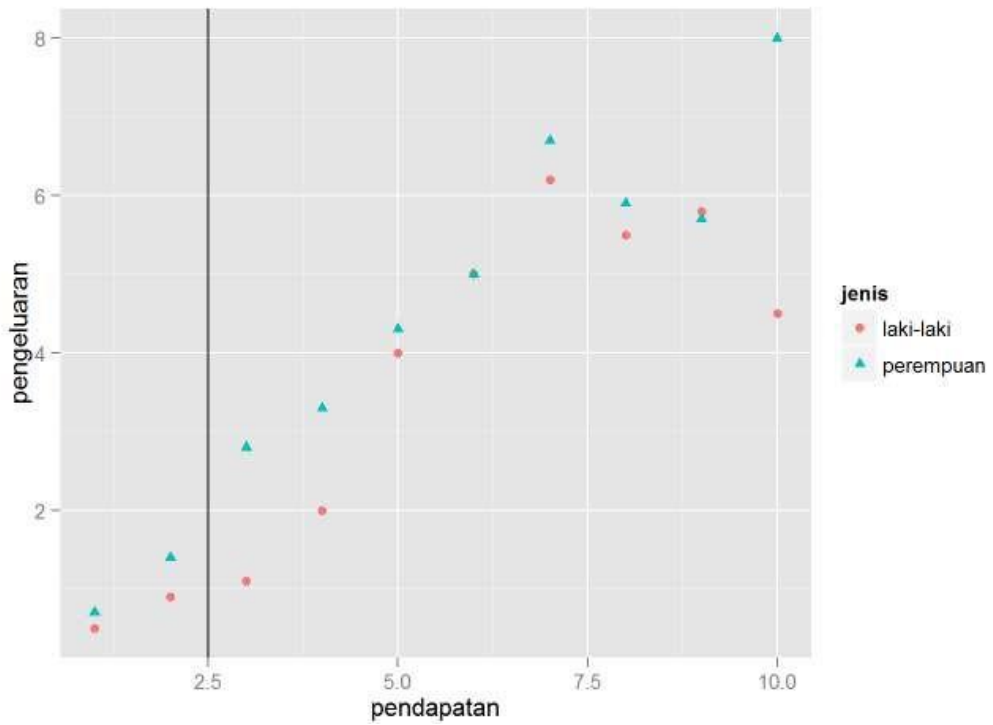
Gambar 1.20

```
grafik + facet_grid(~ jeniskelamin)
```



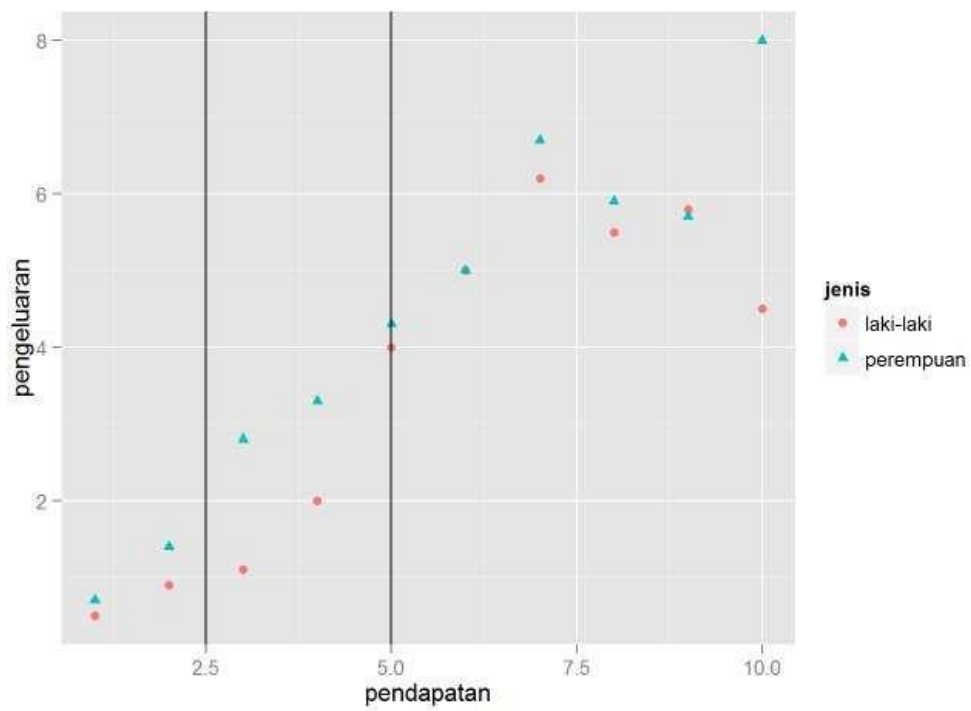
Gambar 1.21

```
grafik + geom_vline(xintercept = 2.5)
```



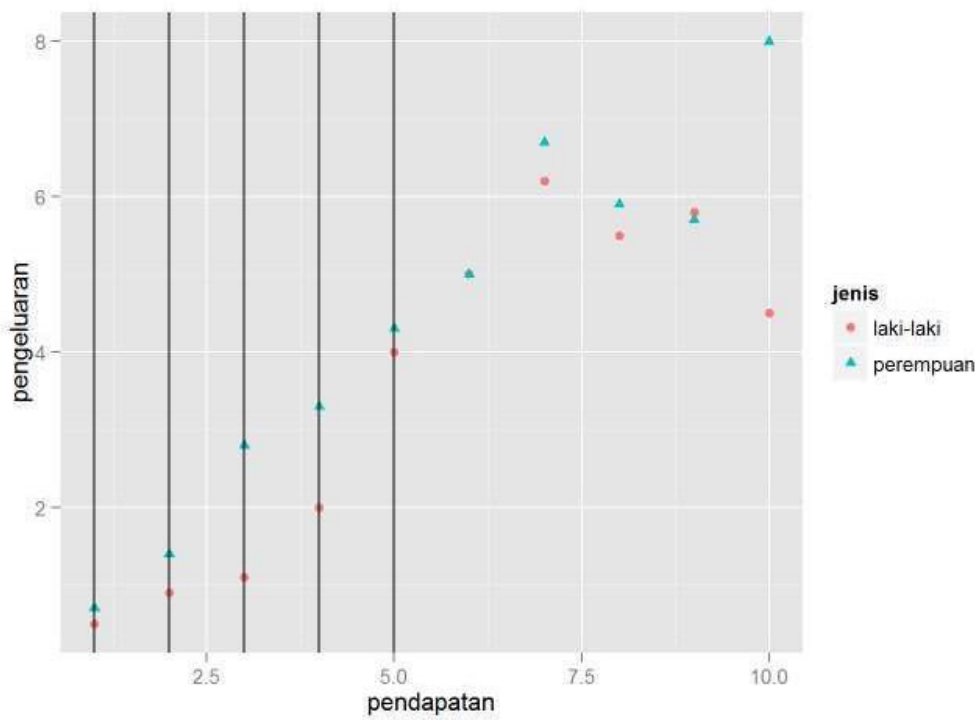
Gambar 1.22

```
grafik + geom_vline(xintercept = 2.5) + geom_vline(xintercept = 5)
```



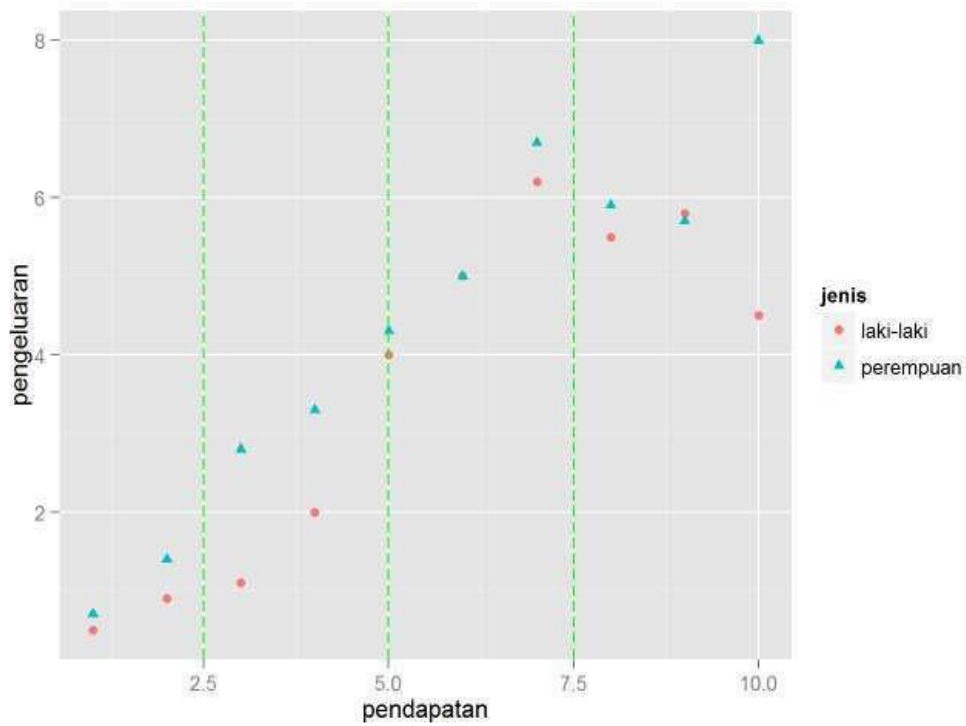
Gambar 1.23

```
grafik + geom_vline(xintercept = 1:5)
```



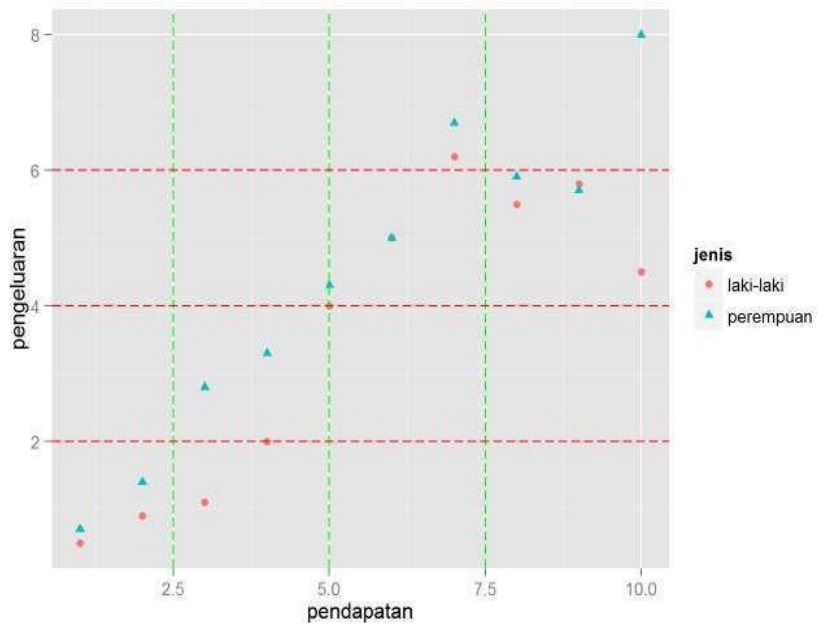
Gambar 1.24

```
grafik + geom_vline(xintercept = c(2.5, 5, 7.5), colour="green", linetype = "longdash")
```



Gambar 1.25

```
grafik + geom_vline(xintercept = c(2.5, 5, 7.5), colour="green", linetype = "longdash") +  
geom_hline(yintercept = c(2, 4, 6), colour="red", linetype = "longdash")
```

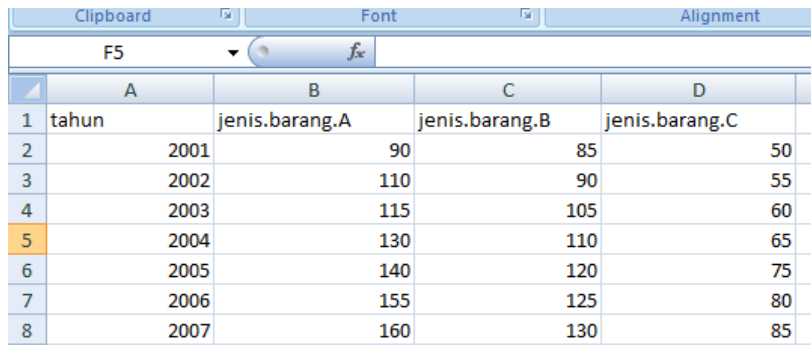


Gambar 1.26



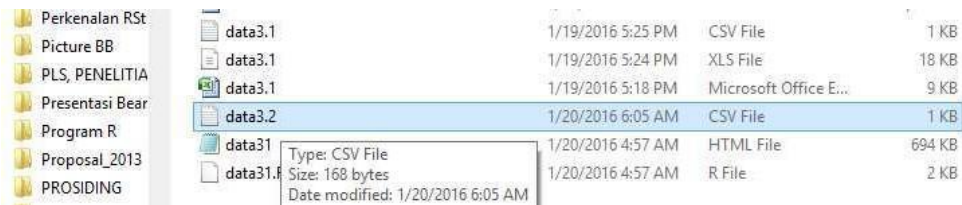
### C. Menyajikan Data dengan Grafik Garis

Misalkan diberikan data seperti pada Gambar 1.27. Gambar 1.27 menyajikan hasil penjualan barang A, B, dan C, selama kurun waktu 2001-2007. Data pada Gambar 1.27 disimpan terlebih dahulu dengan nama **data3.2.csv** (perhatikan Gambar 1.28).



	A	B	C	D
1	tahun	jenis.barang.A	jenis.barang.B	jenis.barang.C
2	2001	90	85	50
3	2002	110	90	55
4	2003	115	105	60
5	2004	130	110	65
6	2005	140	120	75
7	2006	155	125	80
8	2007	160	130	85

Gambar 1.27



Gambar 1.28

Gambar 1.29 sampai dengan Gambar 1.33 merupakan kode R, Eksekusi kode R tersebut, dan amati hasilnya.

```
1  simpan=read.table("data3.2.csv",header=TRUE, sep=",") #membaca data
2  simpan
3
4  Tahun=simpan$tahun
5  Jumlah_A=simpan$jenis.barang.A
6  Jumlah_B=simpan$jenis.barang.B
7  Jumlah_C=simpan$jenis.barang.C
8
9  Jumlah_A
10 Jumlah_B
11 Jumlah_C
12
13 plot(Tahun,Jumlah_A)
14
15 plot(Tahun,Jumlah_A, type="o")
16
17 plot(Tahun,Jumlah_A, type="o", col="blue")
18
19 plot(Tahun,Jumlah_A, type="o", col="green")
20
21 plot(Tahun,Jumlah_A, type="o", col="red")
22 lines(Tahun, Jumlah_B, type="o", col="blue")
23
24 plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(70,180))
25 lines(Tahun, Jumlah_B, type="o", col="blue")
26
27 plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
28 lines(Tahun, Jumlah_B, type="o", col="blue")
29 lines(Tahun, Jumlah_C, type="o", col="green")
30
31 plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
32 lines(Tahun, Jumlah_B, type="o", pch=22, col="blue")
33 lines(Tahun, Jumlah_C, type="o", col="green")
```

Gambar. 1.29

```

34
35 plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
36 lines(Tahun, Jumlah_B, type="o", pch=22, lty=2, col="blue")
37 lines(Tahun, Jumlah_C, type="o", col="green")
38
39 plot(Tahun,Jumlah_A, type="o", pch=22, lty=2, col="red", ylim=c(40,180))
40 lines(Tahun, Jumlah_B, type="o", pch=22, lty=2, col="blue")
41 lines(Tahun, Jumlah_C, pch=22, lty=2, type="o", col="green")
42
43 plot(Tahun,Jumlah_A, type="p", pch=22, lty=2, col="red", ylim=c(40,180))
44 lines(Tahun, Jumlah_B, type="p", pch=22, lty=2, col="blue")
45 lines(Tahun, Jumlah_C, pch=22, lty=2, type="p", col="green")
46
47 plot(Tahun,Jumlah_A, type="o", pch=22, lty=2, col="red", ylim=c(40,180))
48 lines(Tahun, Jumlah_B, type="p", pch=22, lty=2, col="blue")
49 lines(Tahun, Jumlah_C, pch=22, lty=2, type="l", col="green")
50
51 plot.new()
52 plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
53 lines(Tahun, Jumlah_B, type="o", col="blue")
54 lines(Tahun, Jumlah_C, type="o", col="green")
55 title(main="Data Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
56
57 Total = Jumlah_A
58 plot.new()
59 plot(Tahun,Total, type="o", col="red", ylim=c(40,180))
60 lines(Tahun, Jumlah_B, type="o", col="blue")
61 lines(Tahun, Jumlah_C, type="o", col="green")
62 title(main="Data Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
63

```

**Gambar 1.30**

```

64 Total = Jumlah_A
65 plot.new()
66 plot(Tahun,Total, type="o", col="red", ylim=c(40,180))
67 lines(Tahun, Jumlah_B, type="o", col="blue")
68 lines(Tahun, Jumlah_C, type="o", col="green")
69 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), pch=21)
70 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
71
72 Total = Jumlah_A
73 plot.new()
74 plot(Tahun,Total, type="o", col="red", ylim=c(40,180))
75 lines(Tahun, Jumlah_B, type="o", col="blue")
76 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
77 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
78 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
79
80 Total = Jumlah_A
81 plot.new()
82 plot(Tahun,Total, type="o", col="red", ylim=c(40,180))
83 lines(Tahun, Jumlah_B, type="o", col="blue", lty=23)
84 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
85 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
86 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
87
88 Total = Jumlah_A
89 plot.new()
90 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), lty=23)
91 lines(Tahun, Jumlah_B, type="o", col="blue", lty=23)
92 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
93 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
94 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
95

```

**Gambar 1.31**

```

96 Total = Jumlah_A
97 plot.new()
98 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), lty=23)
99 lines(Tahun, Jumlah_B, type="s", col="blue", lty=23)
100 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
101 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
102 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
103
104 Total = Jumlah_A
105 plot.new()
106 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), lty=23)
107 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
108 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
109 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
110 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
111
112 Total = Jumlah_A
113 plot.new()
114 plot(Tahun,Total, type="o", col="red", ylim=c(40,180))
115 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
116 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
117 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
118 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
119
120 Total = Jumlah_A
121 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), xaxt="n")
122 Axis(at=2001:2007, side = 1, labels = c("A","B","C","D","E","F","G"))
123 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
124 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
125 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
126 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
127

```

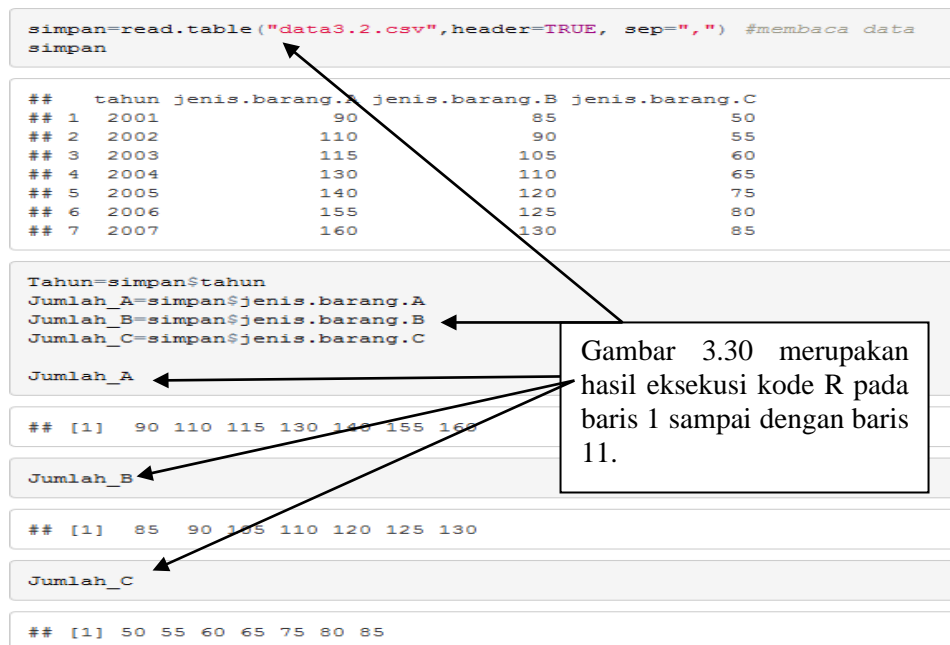
**Gambar 1.32**

```

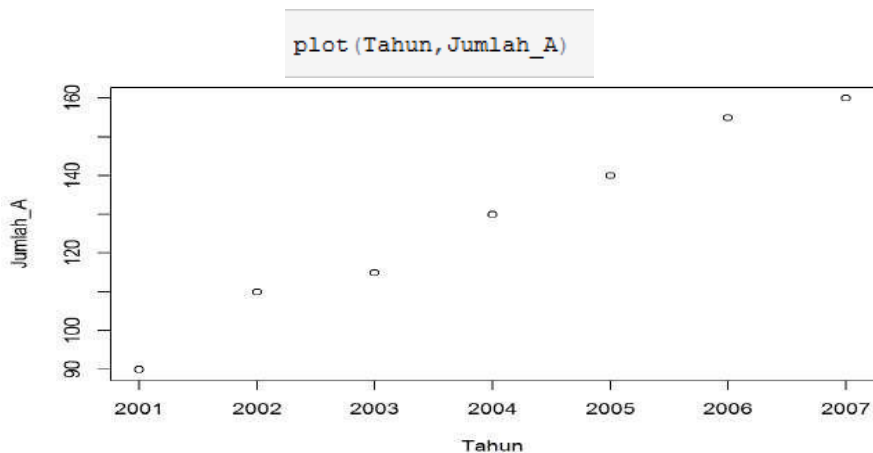
111
112 Total = Jumlah_A
113 plot.new()
114 plot(Tahun,Total, type="o", col="red", ylim=c(40,180))
115 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
116 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
117 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
118 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
119
120 Total = Jumlah_A
121 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), xaxt="n")
122 Axis(at=2001:2007, side = 1, labels = c("A", "B", "C", "D", "E", "F", "G"))
123 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
124 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
125 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
126 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
127
128 Total = Jumlah_A
129 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), xaxt="n")
130 Axis(at=2001:2007, side = 1, labels = c("Tahun 1","Tahun 2","Tahun 3","Tahun 4","Tahun 5","Tahun 6","Tahun 7"))
131 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
132 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
133 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
134 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
135
136 Total = Jumlah_A
137 plot(Tahun,Total, type="o", col="red", ylim=c(40,180), xaxt="n")
138 Axis(at=2001:2007, side = 3, labels = c("Tahun 1","Tahun 2","Tahun 3","Tahun 4","Tahun 5","Tahun 6","Tahun 7"))
139 lines(Tahun, Jumlah_B, type="l", col="blue", lty=23)
140 lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
141 legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"), lty=30)
142 title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
143

```

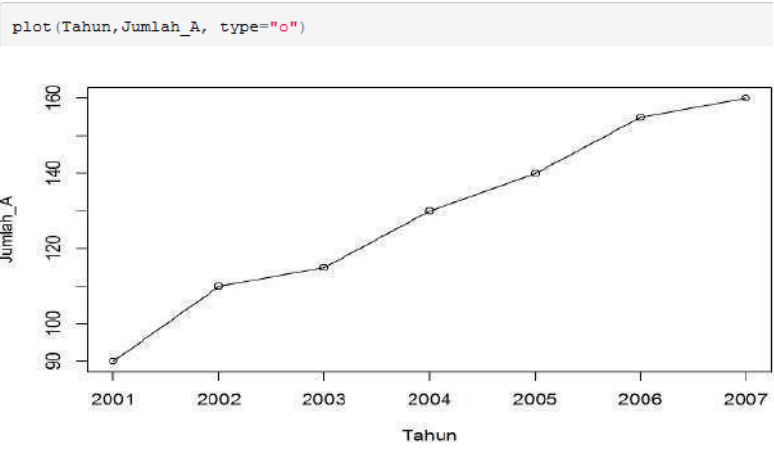
Gambar 1.33



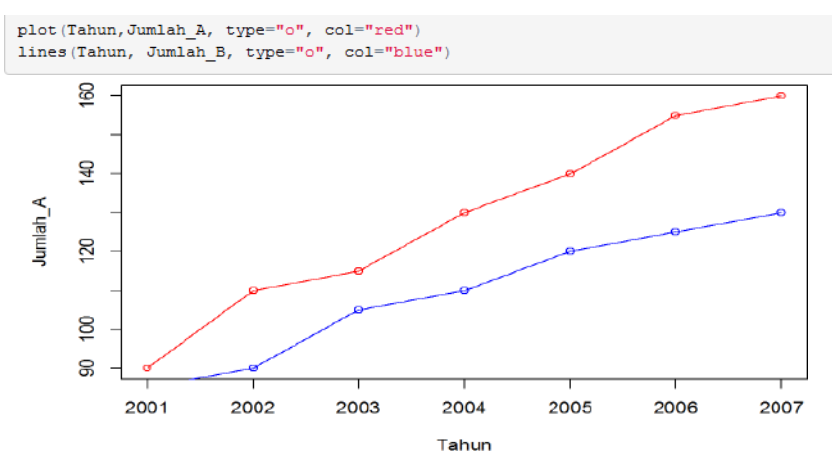
Gambar 1.34



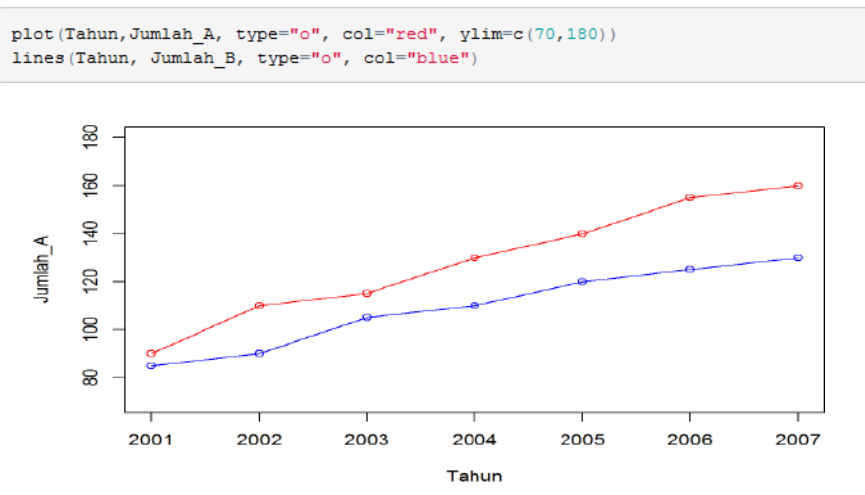
Gambar 1.35



**Gambar 1.36**



**Gambar 1.37**

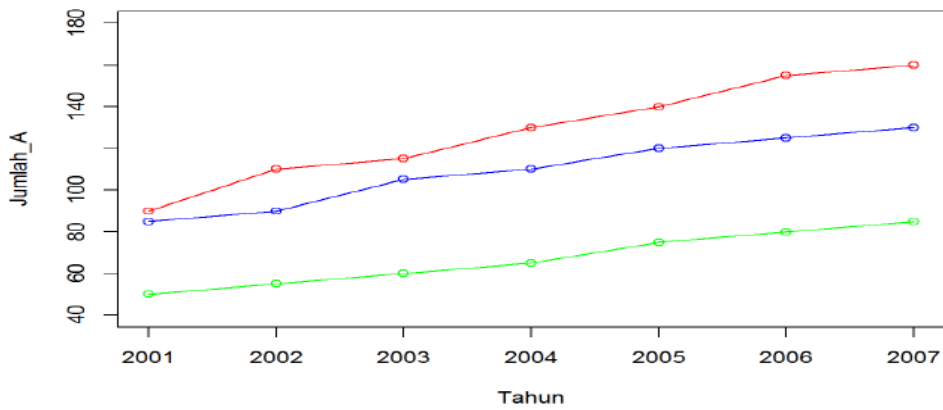


**Gambar 1.38**

```

plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
lines(Tahun, Jumlah_B, type="o", col="blue")
lines(Tahun, Jumlah_C, type="o", col="green")

```

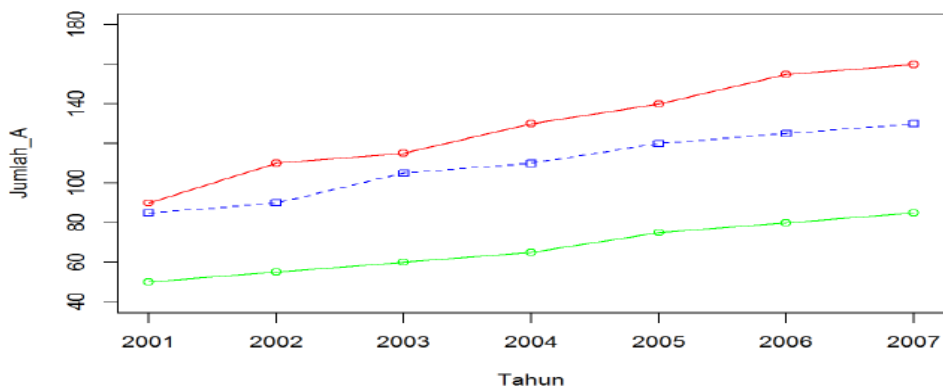


**Gambar 1.38**

```

plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
lines(Tahun, Jumlah_B, type="o", pch=22, lty=2, col="blue")
lines(Tahun, Jumlah_C, type="o", col="green")

```

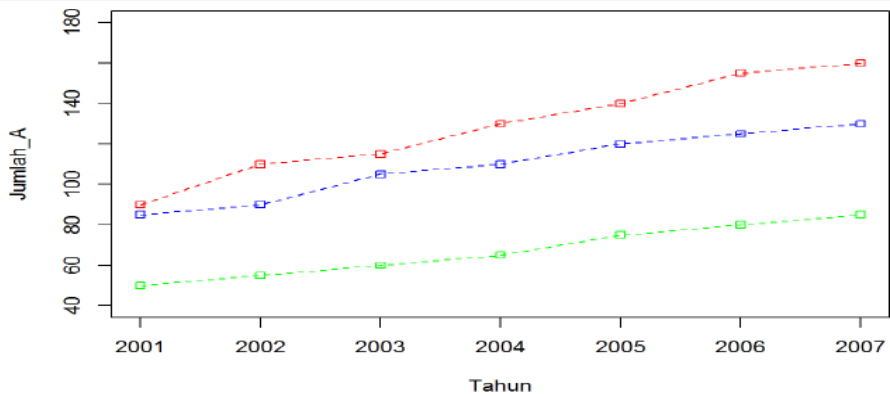


**Gambar 1.39**

```

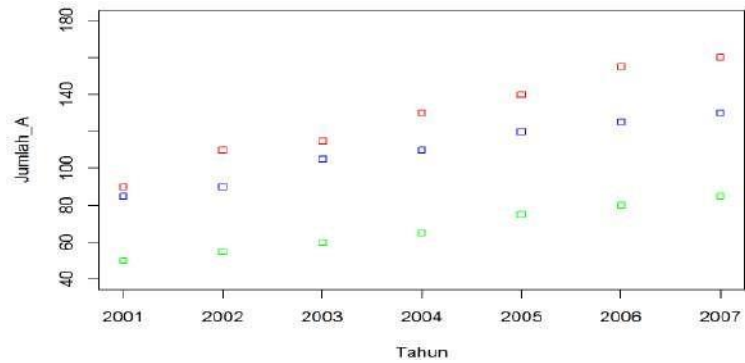
plot(Tahun,Jumlah_A, type="o", pch=22, lty=2, col="red", ylim=c(40,180))
lines(Tahun, Jumlah_B, type="o", pch=22, lty=2, col="blue")
lines(Tahun, Jumlah_C, pch=22, lty=2, type="o", col="green")

```



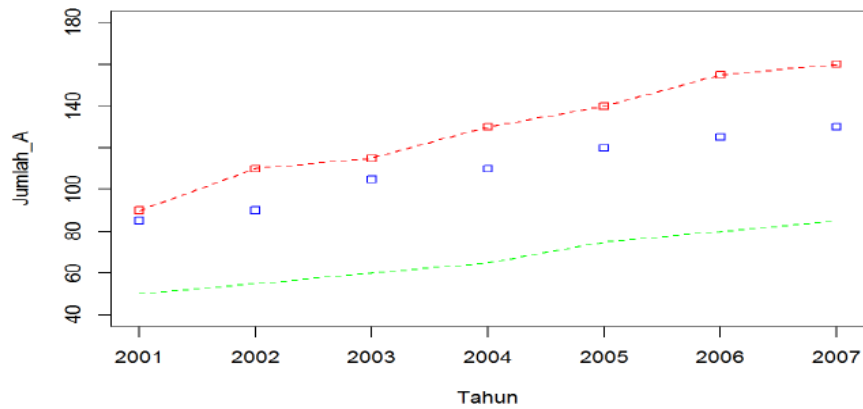
**Gambar 1.40**

```
plot(Tahun,Jumlah_A, type="p", pch=22, lty=2, col="red", ylim=c(40,180))
lines(Tahun, Jumlah_B, type="p", pch=22, lty=2, col="blue")
lines(Tahun, Jumlah_C, pch=22, lty=2, type="p", col="green")
```



**Gambar 1.41**

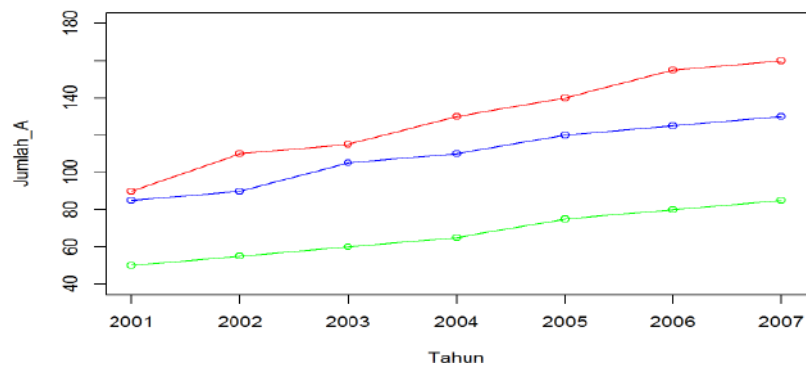
```
plot(Tahun,Jumlah_A, type="o", pch=22, lty=2, col="red", ylim=c(40,180))
lines(Tahun, Jumlah_B, type="p", pch=22, lty=2, col="blue")
lines(Tahun, Jumlah_C, pch=22, lty=2, type="l", col="green")
```



**Gambar 1.42**

```
plot.new()
plot(Tahun,Jumlah_A, type="o", col="red", ylim=c(40,180))
lines(Tahun, Jumlah_B, type="o", col="blue")
lines(Tahun, Jumlah_C, type="o", col="green")
title(main="Data Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)
```

**Data Penjualan Barang A, B, C, dari Tahun 2001-2007**

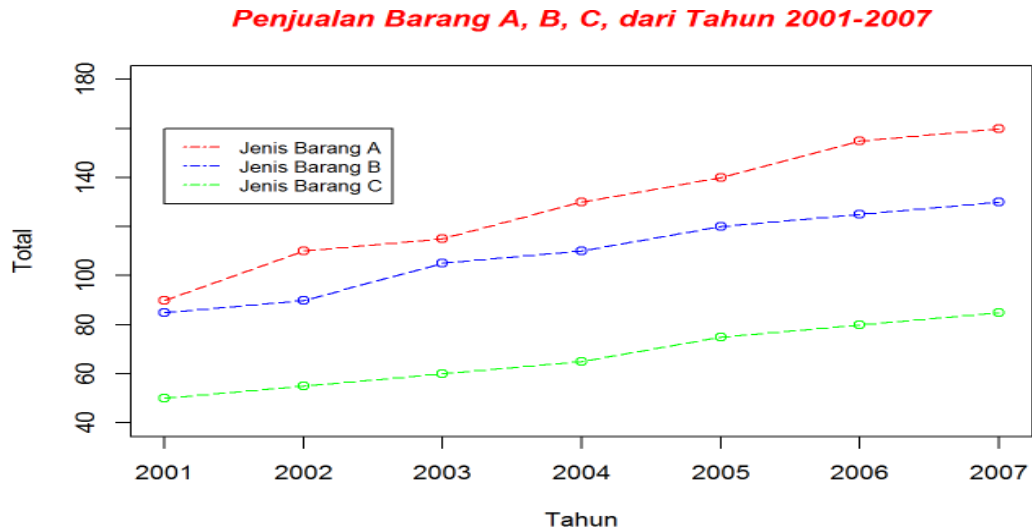


**Gambar 1.43**

```

Total = Jumlah_A
plot.new()
plot(Tahun,Total, type="o", col="red", ylim=c(40,180), lty=23)
lines(Tahun, Jumlah_B, type="o", col="blue", lty=23)
lines(Tahun, Jumlah_C, type="o", col="green", lty=23)
legend(2001,160,c("Jenis Barang A", "Jenis Barang B", "Jenis Barang C"), cex=0.8, col=c("red","blue","green"),
lty=30)
title(main="Penjualan Barang A, B, C, dari Tahun 2001-2007", col.main="red", font.main=4)

```



**Gambar 1.44**

#### D. Menyajikan Data dengan Grafik Batang (Bagian Pertama)

Misalkan diberikan data seperti pada Gambar 1.45. Gambar 1.45 menyajikan hasil penjualan barang A, selama kurun waktu 2001-2007. Data pada Gambar 1.45 disimpan terlebih dahulu dengan nama **data3.3.csv** (perhatikan Gambar 1.46).

Clipboard		Font	
B4		fx 115	
	A	B	
1	tahun	jenis.barang.A	
2	2001	90	
3	2002	110	
4	2003	115	
5	2004	130	
6	2005	140	
7	2006	155	
8	2007	160	

**Gambar 1.45**

data3.1	1/19/2016 5:24 PM	XLS File	18 KB
data3.1	1/19/2016 5:18 PM	Microsoft Office E...	9 KB
data3.2	1/20/2016 6:14 AM	CSV File	1 KB
data3.3	1/20/2016 8:11 AM	CSV File	1 KB
data31	1/20/2016 4:57 AM	HTML File	694 KB
data3	1/20/2016 4:57 AM	R File	2 KB
datati	1/20/2016 7:28 AM	HTML File	731 KB
datatigadua.R	1/20/2016 7:28 AM	R File	7 KB

**Gambar 1.46**

Gambar 1.47 merupakan kode R. Eksekusi dan amati hasilnya.

```

1  simpan=read.table("data3.3.csv",header=TRUE, sep=",") #membaca data
2  simpan
3
4  Tahun=simpan$Tahun
5  Jumlah_A=simpan$jenis.barang.A
6  barplot(Jumlah_A,Tahun)
7
8  barplot(Jumlah_A,Tahun, main="Penjualan Barang Jenis A dari Tahun 2001-2007", xlab="Tahun",
9          ylab="Jumlah Barang yang Terjual", names.arg=c("2001","2002","2003","2004","2005","2006","2007"))
10
11 barplot(Jumlah_A,Tahun, main="Penjualan Barang Jenis A dari Tahun 2001-2007", xlab="Tahun",
12         ylab="Jumlah Barang yang Terjual", names.arg=c("2001","2002","2003","2004","2005","2006","2007"))
13
14 barplot(Jumlah_A,Tahun, main="Penjualan Barang Jenis A dari Tahun 2001-2007", xlab="Tahun",
15         ylab="Jumlah Barang yang Terjual", names.arg=c("2001","2002","2003","2004","2005","2006","2007"), border="blue")
16
17 barplot(Jumlah_A,Tahun, main="Penjualan Barang Jenis A dari Tahun 2001-2007", xlab="Tahun",
18         ylab="Jumlah Barang yang Terjual", names.arg=c("2001","2002","2003","2004","2005","2006","2007"), border="red")
19
20 barplot(Jumlah_A,Tahun, main="Penjualan Barang Jenis A dari Tahun 2001-2007", xlab="Tahun",
21         ylab="Jumlah Barang yang Terjual", names.arg=c("2001","2002","2003","2004","2005","2006","2007"),
22         border="green",density=c(10,20,30,40,50,60,70) )
23
24
25 library(ggplot2)
26 ggplot(data=simpan, aes(x=Tahun, y=Jumlah_A)) + geom_bar(stat="identity")
27
28 ggplot(data=simpan, aes(x=Tahun, y=Jumlah_A)) + geom_bar(stat="identity", fill="darkblue")
29
30 ggplot(data=simpan, aes(x=Tahun, y=Jumlah_A)) + geom_bar(stat="identity", fill=heat.colors(7))
31

```

Gambar 1.47

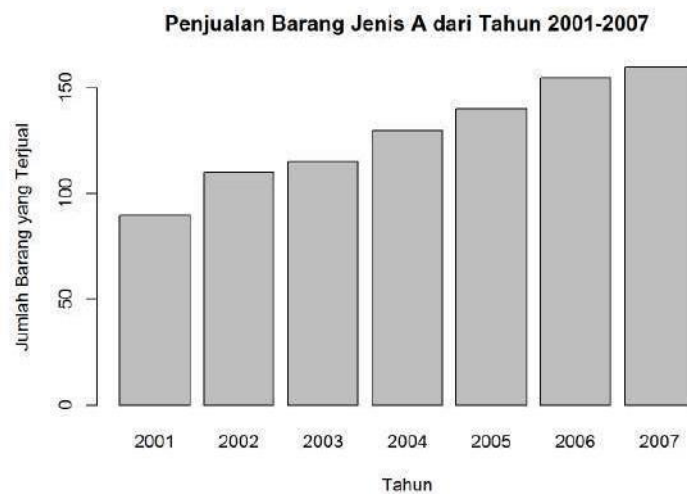
```

simpan=read.table("data3.3.csv",header=TRUE, sep=",") #membaca data
simpan

##  tahun jenis.barang.A
## 1  2001          90
## 2  2002         110
## 3  2003         115
## 4  2004         130
## 5  2005         140
## 6  2006         155
## 7  2007         160

```

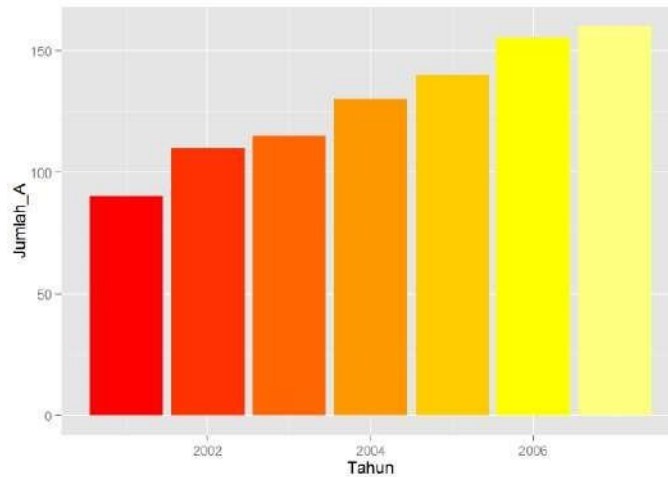
Gambar 1.48



Gambar 1.49



```
ggplot(data=simpan, aes(x=Tahun, y=Jumlah_A)) + geom_bar(stat="identity", fill=heat.colors(7))
```



Gambar 1.50

### E. Menyajikan Data dengan Grafik Batang (Bagian Kedua)

Misalkan diberikan data seperti pada Gambar 1.51. Berdasarkan data pada Gambar 1.51, diketahui responden laki-laki yang memiliki hobi olahraga sebanyak 90 responden, responden laki-laki yang memiliki hobi memasak sebanyak 10 responden, dan seterusnya. Data pada Gambar 1.51 disimpan terlebih dahulu dengan nama **data3.4.csv** (perhatikan Gambar 1.52).

	A	B	C	D
1	Jenis.Kelamin	Hobi	Jumlah	
2	Laki-Laki	Olahraga	90	
3	Laki-Laki	Memasak	10	
4	Perempuan	Olahraga	25	
5	Perempuan	Memasak	75	

Gambar 1.51

File Name	Date Modified	File Type	Size
data3.1	1/19/2016 3:24 PM	ALS File	18 KB
data3.1	1/19/2016 5:18 PM	Microsoft Office E...	9 KB
data3.2	1/20/2016 6:14 AM	CSV File	1 KB
data3.3	1/20/2016 8:54 AM	CSV File	1 KB
data3.4	1/20/2016 8:55 AM	CSV File	1 KB
data31	1/20/2016 4:57 AM	HTML File	694 KB
data31.R	1/20/2016 4:57 AM	R File	2 KB
datatigadua	1/20/2016 7:28 AM	HTML File	731 KB
datatigadua.R	1/20/2016 7:28 AM	R File	7 KB
datatigatiga	1/20/2016 8:30 AM	HTML File	517 KB
datatigatiga.R	1/20/2016 8:30 AM	R File	2 KB

Gambar 1.52

Gambar 1.53 dan Gambar 1.54 merupakan kode R. Eksekusi kode R tersebut dan amati hasilnya.

```
r.Rmd x tes.Rmd x Insurance Data.Rmd x kode R.R x kodeRempatdua.R x data31.R x datatigadua.R x datatigatiga.R x datatigaempat.R* x
1 simpan=read.table("data3.4.csv",header=TRUE, sep=",") #membaca data
2 simpan
3
4 frekuensi=c(90,10,25,75)
5 barplot(t(matrix(frekuensi, ncol=2, byrow=TRUE, dimnames=list(c("Laki-Laki","Perempuan"), c("Olahraga","Memasak")))),
6 main="Hubungan antara Jenis Kelamin dan Hobi", xlab="Jenis kelamin",
7 col=c("darkblue","orange"), beside=TRUE, ylim=c(0,150), legend.text=TRUE,
8 args.legend=list(x="topright"))
9
10
11 frekuensi2=c(2,12,16,6)
12 barplot(frekuensi2, ylim=c(0,20), main="Jumlah Mahasiswa yang Memperoleh Nilai A, B, C, dan D, untuk
13 Matakuliah Matematika 1", names.arg=c("A","B","C","D"), ylab="Jumlah Mahasiswa",
14 xlab="Nilai Mahasiswa", cex.names=0.8, col=c("green","yellow","orange","red") )
15
16 dat = data.frame(
17 jenis_kelamin=factor(c("Laki-Laki","Perempuan"), levels=c("Laki-Laki","Perempuan")), total=c(20,70)
18
19 dat
20
21 library(ggplot2)
22 ggplot(data=dat, aes(x=jenis_kelamin, y=total))+geom_bar(stat="identity")
23
24 ggplot(data=dat, aes(x=jenis_kelamin, y=total, fill=jenis_kelamin))+geom_bar(stat="identity")
25
26 ggplot(data=dat, aes(x=jenis_kelamin, y=total, fill=jenis_kelamin))+geom_bar(stat="identity") + guides(fill=FALSE)
27
28 ggplot(data=dat, aes(x=jenis_kelamin, y=total, fill=jenis_kelamin))+geom_bar(stat="identity") +
29 xlab("Jenis Kelamin") + ylab("Jumlah Mahasiswa") + ggtitle("Universitas XYZ")
30
```

**Gambar 1.53**

```
30 |
31 dat = data.frame( jenis_kelamin=factor(c("Laki-Laki","Laki-Laki","Perempuan","Perempuan")),
32 hobi=factor(c("Olahraga","Memasak","Olahraga","Memasak"),levels=c("Olahraga","Memasak")), total=c(80,20,40,60))
33
34 dat
35
36 ggplot(data=dat, aes(x=hobi, y=total, fill=jenis_kelamin))+geom_bar(stat="identity") +
37 xlab("Hobi Mahasiswa") + ylab("Jumlah Mahasiswa") + ggtitle("universitas XYZ") +
38 geom_text(aes(y=total/1.3, label=total), position="stack" )
39
40 ggplot(data=dat, aes(x=hobi, y=total, fill=jenis_kelamin))+geom_bar(stat="identity",
41 position=position_dodge()) + xlab("Hobi Mahasiswa") + ylab("Jumlah Mahasiswa") +
42 ggtitle("universitas XYZ")
43
44 ggplot(data=dat, aes(x=hobi, y=total, fill=jenis_kelamin))+geom_bar(stat="identity",
45 position=position_dodge()) + xlab("Hobi Mahasiswa") + ylab("Jumlah Mahasiswa") +
46 ggtitle("universitas XYZ") + geom_text(aes(y=total/4, label=total),
47 position=position_dodge(width=1) )
30:1 | (top Level)
```

**Gambar 1.54**

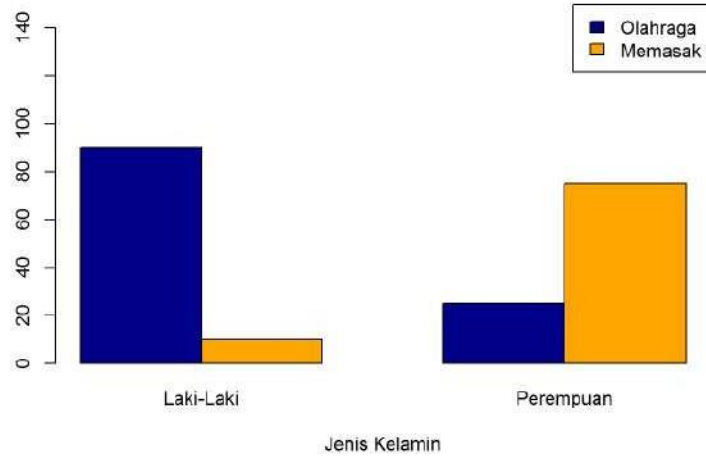
```
simpan=read.table("data3.4.csv",header=TRUE, sep=",") #membaca data
simpan
```

##	Jenis.Kelamin	Hobi	Jumlah
## 1	Laki-Laki	Olahraga	90
## 2	Laki-Laki	Memasak	10
## 3	Perempuan	Olahraga	25
## 4	Perempuan	Memasak	75

**Gambar 1.55**

```
frekuensi=c(90,10,25,75)
barplot(t(matrix(frekuensi, ncol=2, byrow=TRUE, dimnames=list(c("Laki-Laki","Perempuan"), c("Olahraga","Memasak")))),
main="Hubungan antara Jenis Kelamin dan Hobi", xlab="Jenis Kelamin",
col=c("darkblue","orange"), beside=TRUE, ylim=c(0,150), legend.text=TRUE,
args.legend=list(x="topright"))
```

Hubungan antara Jenis Kelamin dan Hobi



Gambar 1.56

```
frekuensi2=c(2,12,16,6)
barplot(frekuensi2, ylim=c(0,20), main="Jumlah Mahasiswa yang Memperoleh Nilai A, B, C, dan D, untuk
Matakuliah Matematika 1", names.arg=c("A","B","C","D"), ylab="Jumlah Mahasiswa",
xlab="Nilai Mahasiswa", cex.names=0.8, col=c("green","yellow","orange","red") )

dat = data.frame(
jenis_kelamin=factor(c("Laki-Laki","Perempuan"), levels=c("Laki-Laki","Perempuan")), total=c(20,70))

dat
```

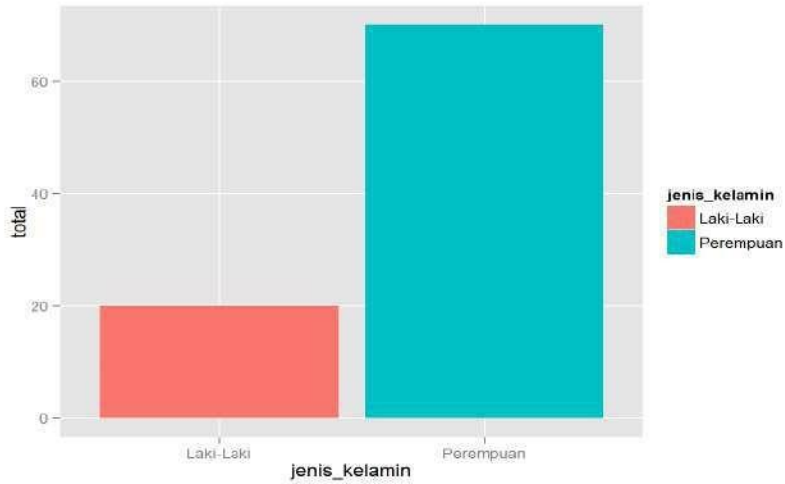
```
## jenis_kelamin total
## 1 Laki-Laki 20
## 2 Perempuan 70
```



Gambar 1.55

```
library(ggplot2)
```

```
ggplot(data=dat, aes(x=jenis_kelamin, y=total, fill=jenis_kelamin))+geom_bar(stat="identity")
```



Gambar 1.56

```
dat = data.frame( jenis_kelamin=factor(c("Laki-Laki", "Laki-Laki", "Perempuan", "Perempuan")),
hobi=factor(c("Olahraga", "Memasak", "Olahraga", "Memasak"), levels=c("Olahraga", "Memasak")), total=c(80, 20, 40, 60))
```

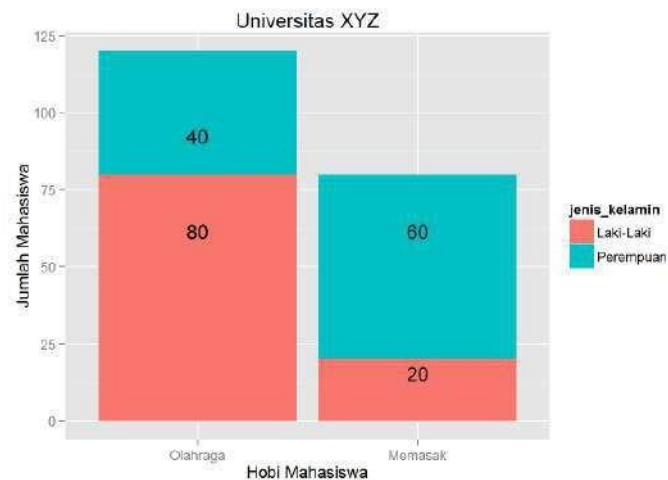
dat

```
## jenis_kelamin hobi total
## 1 Laki-Laki Olahraga 80
## 2 Laki-Laki Memasak 20
## 3 Perempuan Olahraga 40
## 4 Perempuan Memasak 60
```

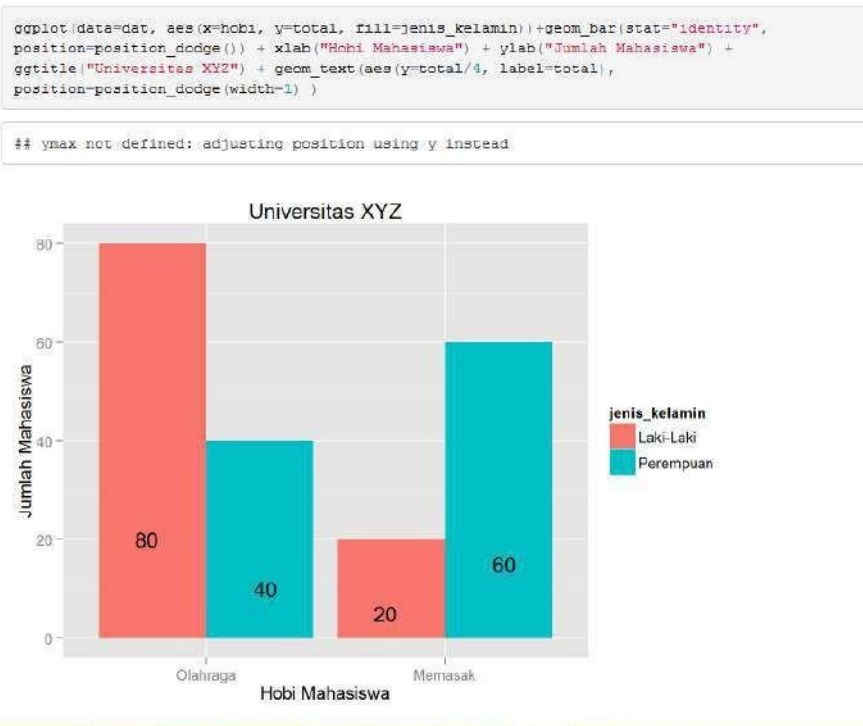
Gambar 1.57

```
ggplot(data=dat, aes(x=hobi, y=total, fill=jenis_kelamin))+geom_bar(stat="identity") +
xlab("Hobi Mahasiswa") + ylab("Jumlah Mahasiswa") + ggtitle("Universitas XYZ") +
geom_text(aes(y=total/1.5, label=total), position="stack")
```

```
## ymax not defined: adjusting position using y instead
```



Gambar 1.58



Gambar 1.59

F. Menyajikan Data dengan Diagram Lingkaran

Misalkan diberikan data seperti pada Gambar 1.60. Berdasarkan Gambar data pada 1.60, diketahui jumlah produk A yang terjual sebanyak 12 unit, jumlah produk B yang terjual sebanyak 5 unit, dan seterusnya. Data pada Gambar 1.60 disimpan terlebih dahulu dengan nama **data3.5.csv** (perhatikan Gambar 1.61).

	A	B	C
1	Produk	Jumlah	
2	A	12	
3	B	5	
4	C	8	
5	D	20	
6			

Gambar 1.60

File Name	Date/Time	Type	Size
data3.1	1/19/2016 5:18 PM	Microsoft Office E...	9 KB
data3.2	1/20/2016 6:14 AM	CSV File	1 KB
data3.3	1/20/2016 8:54 AM	CSV File	1 KB
data3.4	1/20/2016 8:58 AM	CSV File	1 KB
data3.5	1/20/2016 10:10 AM	CSV File	1 KB
data3	1/20/2016 4:57 AM	HTML File	694 KB
data3	1/20/2016 4:57 AM	R File	2 KB

Gambar 1.61

Gambar 1.62 dan Gambar 1.63 merupakan kode R. Eksekusi kode R tersebut, dan amati hasilnya.

```

1  simpan=read.table("data3.5.csv",header=TRUE, sep=",") #membaca data
2  simpan
3
4  pie(simpan$Jumlah,labels=simpan$Produk, main="Data Penjualan Produk A, B, C, dan D")
5
6  pie(simpan$Jumlah,labels=simpan$Produk, main="Data Penjualan Produk A, B, C, dan D", col=heat.colors(4) )
7
8  pie(simpan$Jumlah,labels=simpan$Jumlah, main="Data Penjualan Produk A, B, C, dan D", col=heat.colors(4))
9  colors=heat.colors(4)
10 legend(1,0.5, c("Produk A","Produk B","Produk C", "Produk D"), cex=0.8, fill=colors )
11
12 pie(simpan$Jumlah,labels=simpan$Jumlah, main="Data Penjualan Produk A, B, C, dan D",
13 col=c("darkblue","orange","yellow","red"))
14 colors=c("darkblue","orange","yellow","red")
15 legend(1,0.5, c("Produk A","Produk B","Produk C", "Produk D"), cex=0.8, fill=colors )
16
17 Persen=round(simpan$Jumlah/sum(simpan$Jumlah)*100,4)
18 Persen=paste(Persen,"%",sep="")
19 pie(simpan$Jumlah,labels=Persen, main="Data Penjualan Produk A, B, C, dan D",
20 col=c("darkblue","orange","yellow","red"))
21 colors=c("darkblue","orange","yellow","red")
22 legend(1,0.5, c("Produk A","Produk B","Produk C", "Produk D"), cex=0.8, fill=colors )
23
24 Persen=round(simpan$Jumlah/sum(simpan$Jumlah)*100,4)
25 Persen=paste(Persen,"%",sep="")
26 pie(simpan$Jumlah,labels=Persen, main="Data Penjualan Produk A, B, C, dan D",
27 col=c("darkblue","orange","yellow","red"))
28 colors=c("darkblue","orange","yellow","red")
29 legend(1,0.5, c("Produk A","Produk B","Produk C", "Produk D"), cex=0.8, fill=colors )
30

```

**Gambar 1.62**

```

30
31 Jumlah=simpan$Jumlah
32 Produk=simpan$Produk
33 library(ggplot2)
34 pie = ggplot(simpan, aes(x="", y=Jumlah, fill=Produk))+geom_bar(width=1,stat="identity")+coord_polar("y",start=0)
35
36 pie
37
38 library(ggplot2)
39 library(grid)
40 library(gridExtra)
41
42 blank_theme = theme(
43 axis.title.x=element_blank(),
44 axis.title.y=element_blank(),
45 axis.text.x = element_blank(),
46 axis.text.y = element_blank(),
47 panel.border = element_blank(),
48 panel.grid=element_blank(),
49 axis.ticks= element_blank(),
50 plot.title=element_text(size=14, face="bold")
51 )
52
53 library(scales)
54 pie + blank_theme + geom_text(aes(y=Jumlah/4 + c(0,cumsum(Jumlah)[-length(Jumlah)]), label=Jumlah), size=5)
55 pie + blank_theme + geom_text(aes(y=Jumlah/4 + c(0,cumsum(Jumlah)[-length(Jumlah)]), label=Jumlah), size=5) +
56 scale_fill_manual(values=c(heat.colors(4)))
57
58 Persen=round(simpan$Jumlah/sum(simpan$Jumlah)*100,2)
59 Persen=paste(Persen,"%",sep="")
60 pie + blank_theme + geom_text(aes(y=Jumlah/4 + c(0,cumsum(Jumlah)[-length(Jumlah)]),
61 label=Persen ), size=5) +
62 scale_fill_manual(values=c(heat.colors(4)))
63

```

**Gambar 1.63**

```

simpan=read.table("data3.5.csv",header=TRUE, sep=",") #membaca data
simpan

```

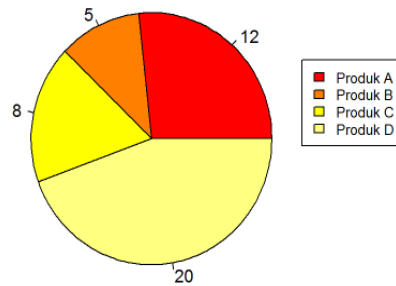
```

##   Produk Jumlah
## 1     A      12
## 2     B       5
## 3     C       8
## 4     D      20

```

```
pie(simpan$Jumlah,labels=simpan$Jumlah, main="Data Penjualan Produk A, B, C, dan D", col=heat.colors(4))
colors=heat.colors(4)
legend(1,0.5, c("Produk A","Produk B","Produk C", "Produk D"), cex=0.8, fill=colors )
```

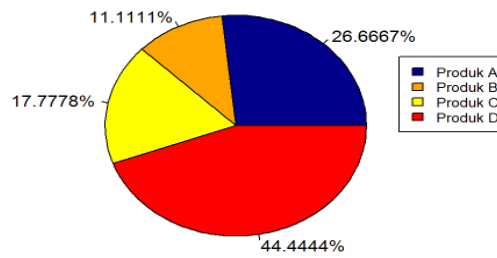
Data Penjualan Produk A, B, C, dan D



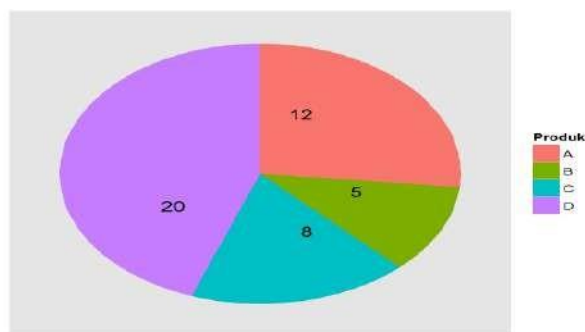
Gambar 1.64

```
Persen=round(simpan$Jumlah/sum(simpan$Jumlah)*100,4)
Persen=paste(Persen,"%",sep="")
pie(simpan$Jumlah,labels=Persen, main="Data Penjualan Produk A, B, C, dan D",
col=c("darkblue","orange","yellow","red"))
colors=c("darkblue","orange","yellow","red")
legend(1,0.5, c("Produk A","Produk B","Produk C", "Produk D"), cex=0.8, fill=colors )
```

Data Penjualan Produk A, B, C, dan D

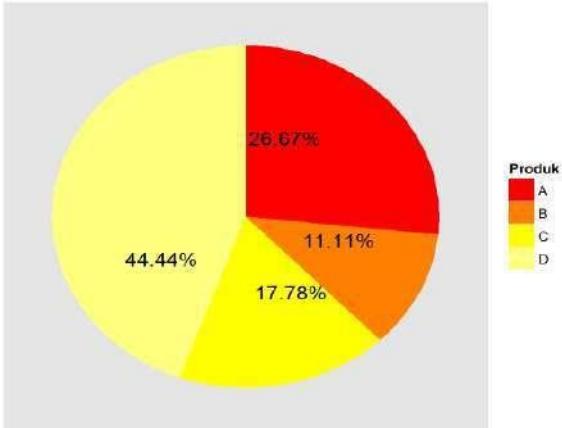


Gambar 1.65



Gambar 1.66

```
Persen=round(simpan$Jumlah/sum(simpan$Jumlah)*100,2)
Persen=paste(Persen,"%",sep="")
pie + blank_theme + geom_text(aes(y=Jumlah/4 + c(0,cumeum(Jumlah)[-length(Jumlah)]),
label=Persen), size=5) +
scale_fill_manual(values=c(heat.colors(4)))
```



Gambar 1.67

**G. Menyajikan Data dengan Histogram**

Misalkan diberikan data mengenai skor IQ seperti pada Gambar 1.68. Berdasarkan data pada Gambar 1.68, jumlah pengamatan sebanyak 77. Data pada Gambar 1.68 disimpan terlebih dahulu dengan nama **IQ.csv** (perhatikan Gambar 1.69). Gambar 1.70 dan Gambar 1.71 disajikan kode R. Eksekusi kode R tersebut, dan amati hasilnya.

Row	A	B
1	IQ	
2	111	
3	111	
4	111	
5	111	
6	111	
7	111	
8	111	
9	110	
10	110	
11	110	
12	110	
13	110	
14	110	
15	110	
16	112	
17	112	
18	112	
19	112	
20	112	
21	112	
22	112	
23	113	
24	113	
25	113	
26	113	
27	113	
28	114	
29	114	
30	114	
31	114	
32	115	
33	115	
34	115	
35	116	
36	116	
37	117	
38	90	
39	91	
40	92	
41	92	
42	93	
43	93	
44	93	
45	94	
46	94	
47	94	
48	94	
49	95	
50	95	
51	95	
52	95	
53	95	
54	95	
55	96	
56	96	
57	96	
58	96	
59	97	
60	97	
61	97	
62	97	
63	97	
64	98	
65	98	
66	98	
67	101	
68	101	
69	101	
70	102	
71	102	
72	103	
73	104	
74	103	
75	102	
76	108	
77	109	
78	118	
79		
80		

Gambar 1.68



datatigadua.R	1/20/2016 7:28 AM	R File	7 KB
datatigaempat	1/20/2016 9:14 AM	HTML File	519 KB
datatigaempat.R	1/20/2016 10:43 AM	R File	3 KB
datatigalima	1/20/2016 10:28 AM	HTML File	498 KB
datatigalima.R	1/20/2016 10:28 AM	R File	3 KB
datatigatiga	1/20/2016 8:30 AM	HTML File	517 KB
datatigatiga.R	1/20/2016 8:30 AM	R File	2 KB
IQ	9/23/2015 7:52 PM	CSV File	1 KB

Type: CSV File  
Size: 360 bytes  
Date modified: 9/23/2015 7:52 PM

**Gambar 1.69**

```

1  simpan=read.csv("IQ.csv", header=TRUE)
2  simpan
3
4  simpan_skor_IQ=simpan$IQ
5  hist(simpan_skor_IQ)
6
7  hist(simpan_skor_IQ, col="lightblue")
8
9  hist(simpan_skor_IQ, col="darkblue", ylim=c(0,40), main="Contoh Histogram", ylab="Frekuensi")
10
11 hist(simpan_skor_IQ, col="orange", ylim=c(0,40), main="Contoh Histogram", ylab="Frekuensi", breaks=c(90,100,110,120) )
12
13 hist(simpan_skor_IQ, col=heat.colors(6), ylim=c(0,30), main="Contoh Histogram", ylab="Frekuensi",
14 breaks=c(90,95,100,105,110,115,120), xlim=c(90,125) )
15
16 hist(simpan_skor_IQ, col=heat.colors(6), ylim=c(0,30), main="Contoh Histogram", ylab="Frekuensi",
17 breaks=c(90,93,96,99,102,105,108,111,114,117,120), xlim=c(90,125) )
18
19 hist(simpan_skor_IQ, breaks=6, col=heat.colors(6), ylim=c(0,30), main="Contoh Histogram",
20 ylab="Frekuensi", xlim=c(90,125) )
21
22 hist(simpan_skor_IQ, breaks=c(90,117,120), ylim=c(0,50), xlim=c(90,125), main="Contoh Histogram",
23 col=heat.colors(2) )
24
25 hist(simpan_skor_IQ, breaks=c(90,117,120), ylim=c(0,80), xlim=c(90,125), main="Contoh Histogram",
26 col=heat.colors(2), freq=TRUE )
27
28 hist(simpan_skor_IQ, breaks=c(90,92,97,117,120), ylim=c(0,80), xlim=c(90,125), main="Contoh Histogram",
29 col=heat.colors(4), freq=TRUE )
30

```

**Gambar 1.70**

```

34
35 ggplot(data=simpan, aes(IQ)) + geom_histogram(breaks=c(90,95,100,105,110,115,120), col="darkblue",
36 fill=heat.colors(6) )
37
38 ggplot(data=simpan, aes(IQ)) + geom_histogram(breaks=c(90,95,100,105,110,115,120), col="red",
39 aes(fill=.count..)) + labs(title="Contoh Histogram") + labs(x="IQ", y="Jumlah") +
40 xlim(c(90,125)) + ylim(c(0,20)) + scale_fill_gradient("count", low="green", high="red")
41
42 library(ggplot2)
43
44 ggplot(data=simpan, aes(IQ)) + geom_histogram(breaks=c(90,93,96,99,102,105,111,114,115,120),
45 col="darkblue", fill=heat.colors(9), aes(fill=.count..)) + labs(title="Contoh Histogram") +
46 labs(x="IQ", y="Jumlah") + xlim(c(90,125)) + ylim(c(0,20)) + scale_fill_gradient("count",
47 low=heat.colors(9), high=heat.colors(9))
48
49

```

**Gambar 1.71**

```

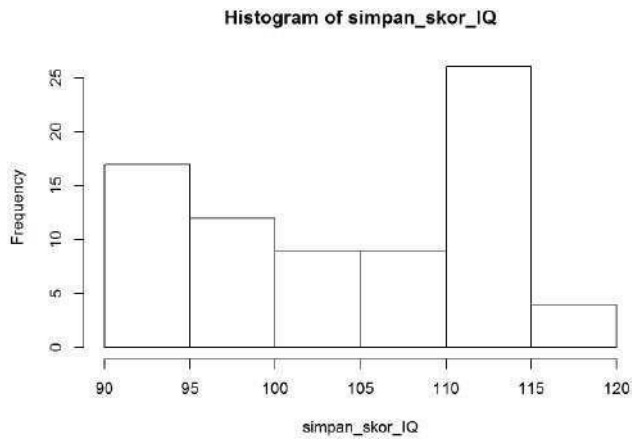
simpan=read.csv("IQ.csv", header=TRUE)
simpan

##      IQ
## 1  111
## 2  111
## 3  111
## 4  111
## 5  111
## 6  111
## 7  111
## 8  110
## 9  110
## 10 110
## 11 110

```

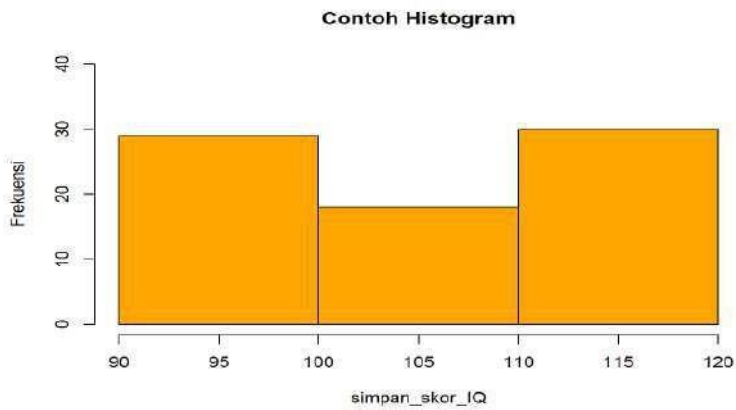
**Gambar 1.72**

```
simpan_skor_IQ=simpan$IQ
hist(simpan_skor_IQ)
```



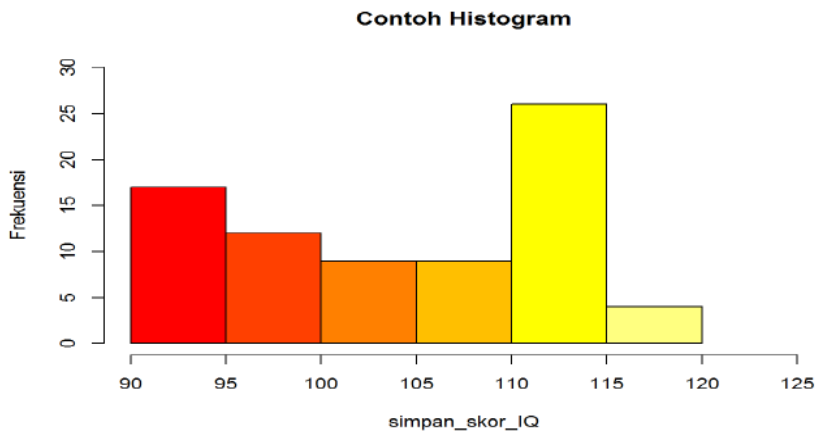
**Gambar 1.73**

```
hist(simpan_skor_IQ, col="orange", ylim=c(0,40), main="Contoh Histogram", ylab="Frekuensi", breaks=c(90,100,110,120) )
```



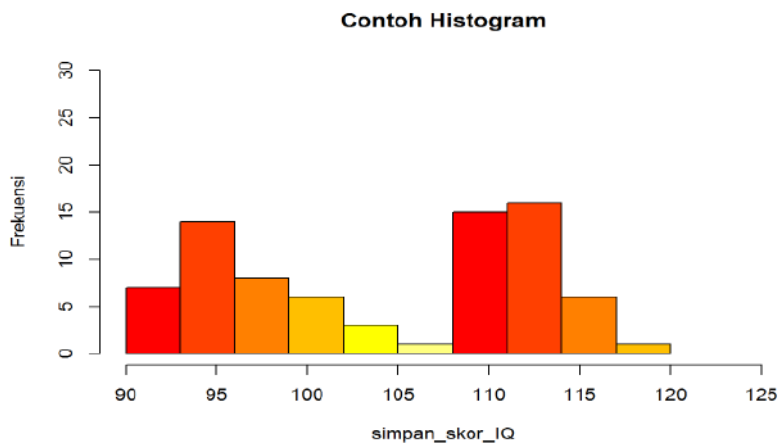
**Gambar 1.74**

```
hist(simpan_skor_IQ, col=heat.colors(6), ylim=c(0,30), main="Contoh Histogram", ylab="Frekuensi", breaks=c(90,95,100,105,110,115,120), xlim=c(90,125) )
```



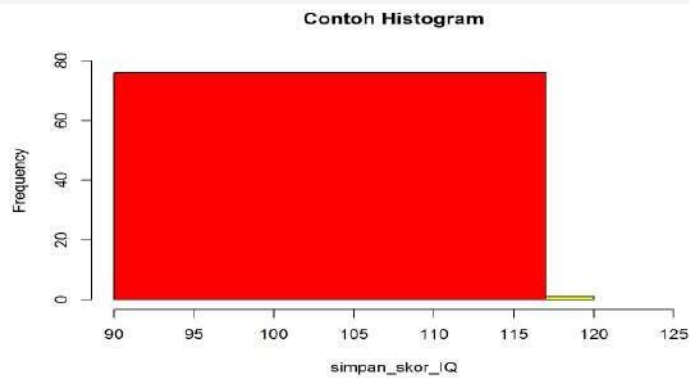
**Gambar 1.75**

```
hist(simpan_skor_IQ, col=heat.colors(6), ylim=c(0,30), main="Contoh Histogram", ylab="Frekuensi",
breaks=c(90,93,96,99,102,105,108,111,114,117,120), xlim=c(90,125) )
```



**Gambar 1.76**

```
hist(simpan_skor_IQ, breaks=c(90,117,120), ylim=c(0,80), xlim=c(90,125), main="Contoh Histogram",
col=heat.colors(2), freq=TRUE )
```

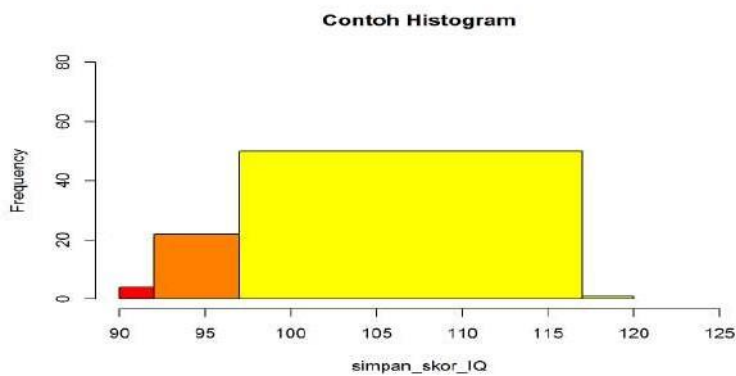


**Gambar 1.77**

```
hist(simpan_skor_IQ, breaks=c(90,92,97,117,120), ylim=c(0,80), xlim=c(90,125), main="Contoh Histogram",
col=heat.colors(4), freq=TRUE )
```

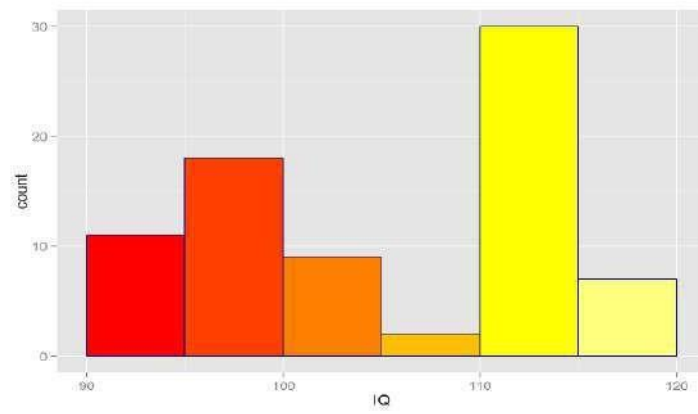
```
## Warning in plot.histogram(x, freq = freq1, col = col, border = border,
## angle = angle, : the AREAS in the plot are wrong -- rather use 'freq =
## FALSE'
```

```
library(ggplot2)
```



**Gambar 1.78**

```
ggplot(data=simpan, aes(IQ)) + geom_histogram(breaks=c(90, 95, 100, 105, 110, 115, 120), col="darkblue", fill="heat.colors(6) )
```



**Gambar 1.79**

## BAB II DISTRIBUSI SAMPLING

### A. Distribusi Populasi (Population Distribution)

Distribusi populasi dapat diartikan sebagai distribusi probabilitas dari data populasi. Andaikan dalam suatu kelas hanya terdiri lima mahasiswa jurusan matematika. Berikut disajikan nilai ujian matakuliah kalkulus dari lima mahasiswa tersebut.

70, 75, 80, 80, 90

Andaikan  $X$  menyatakan nilai ujian matakuliah kalkulus dan  $P(X = x)$  atau  $f(x)$  menyatakan probabilitas dari suatu nilai ujian matakuliah kalkulus. Berikut disajikan distribusi probabilitas dari data populasi nilai ujian matakuliah kalkulus (Tabel 2.1).

**Tabel 2.1 Distribusi Probabilitas dari Data Populasi Nilai Ujian Kalkulus**

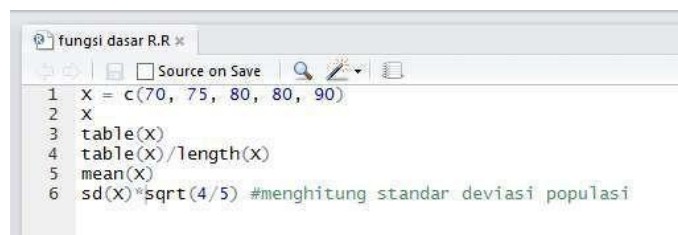
$X$	$P(X = x)$
70	0.2
75	0.2
80	0.4
90	0.2
$\sum P(X = x) = \sum f(x) = 1$	

Nilai rata-rata dan standar deviasi berdasarkan data pada Tabel 2.1 dihitung sebagai berikut.

$$\mu = \frac{70 + 75 + 80 + 80 + 90}{5} = 79$$

$$\sigma = \sqrt{\frac{(70 - 79)^2 + (75 - 79)^2 + \dots + (90 - 79)^2}{5}} = 6,633$$

Perhatikan bahwa  $\mu$  dan  $\sigma$  merupakan nilai-nilai parameter populasi. Parameter dapat diartikan sebagai suatu nilai atau ukuran yang dihitung berdasarkan populasi. Gambar 2.1 dan Gambar 2.2 merupakan ilustrasi dalam R.



```
fungsi dasar R.R x
Source on Save
1 x = c(70, 75, 80, 80, 90)
2 x
3 table(x)
4 table(x)/length(x)
5 mean(x)
6 sd(x)*sqrt(4/5) #menghitung standar deviasi populasi
```

**Gambar 2.1**

```

X = c(70, 75, 80, 80, 90)
X

## [1] 70 75 80 80 90

table(X)

## X
## 70 75 80 90
## 1 1 2 1

table(X)/length(X)

## X
## 70 75 80 90
## 0.2 0.2 0.4 0.2

mean(X)

## [1] 79

sd(X)*sqrt(4/5) #menghitung standar deviasi populasi

## [1] 6.63325

```

Untuk menampilkan distribusi probabilitas.

Penambahan `sqrt(4/5)` dengan maksud untuk menghitung standar deviasi populasi. Jika `sqrt(4/5)` dihilangkan, berarti menghitung standar deviasi sampel (bukan populasi).

Gambar 2.2

### ***Distribusi Sampling Rata-Rata Sampel $\bar{X}$ (Sampling Distribution of $\bar{X}$ )***

Berbeda dengan statistika deskriptif yang rangkaian pengerjaannya meliputi mengorganisasi (*organizing*), menampilkan (*displaying*), dan menjelaskan data dengan menggunakan tabel, grafik, serta ukuran-ukuran seperti rata-rata, median, serta modus, pada statistika inferensi sampai pada tahap pengambilan keputusan atau prediksi mengenai populasi berdasarkan sampel yang diteliti. Konsep mengenai **distribusi sampling** memberikan teori yang penting untuk membuat prosedur-prosedur statistik inferensi. Daniel (2005:129) menyatakan sebagai berikut.

*“Sampling distributions serve two purposes: (1) they allow us to answer probability questions about sample statistics, and (2) they provide the necessary theory for making statistical inference procedures valid”.*

Nilai dari parameter suatu populasi bersifat konstan. Dalam hal ini, untuk setiap data populasi hanya memiliki satu nilai rata-rata populasi  $\mu$ . Namun hal ini belum tentu berlaku untuk rata-rata sampel  $\bar{X}$  Sampel-sampel yang ditarik dari populasi yang sama dan dengan ukuran yang sama dapat menghasilkan nilai rata-rata sampel yang berbeda-beda. Jadi, nilai rata-rata sampel bergantung pada nilai-nilai yang berada dalam sampel tersebut. **Oleh karena itu, rata-rata sampel  $\bar{X}$  merupakan variabel acak (*random variable*).** Sebagaimana pada variabel acak, **maka rata-rata sampel  $\bar{X}$  memiliki distribusi probabilitas.** Distribusi probabilitas  $\bar{X}$  sering disebut dengan istilah **distribusi sampling dari  $\bar{X}$**  Ukuran-ukuran statistik lainnya seperti median, modus, dan standar deviasi juga memiliki distribusi sampling (Mann dan Lacke, 2011:302).

Pada pembahasan sebelumnya mengenai “Distribusi Probabilitas”, diketahui data populasi sebagai berikut.

70, 75, 80, 80, 90

Andaikan masing-masing nilai diberi kode huruf sebagai berikut.

V = 70, W = 75, X = 80, Y = 80, dan Z = 90

Maka, V, W, X, Y, dan Z merupakan kode-kode huruf yang menyatakan kelima nilai ujian matakuliah kalkulus. Kemudian misalkan akan diambil sampel yang terdiri tiga nilai tanpa pengembalian (*without replacement*). Maka banyaknya kemungkinan sampel yang terambil sebagai berikut.

$$C_3^5 = \frac{5!}{(5-3)!3!} = \frac{5.4.3.2.1}{(2.1)(3.2.1)} = 10 \text{ kemungkinan sampel}$$

VWX, VWY, VWZ, VXY, VXZ, VYZ, WXY, WXZ, WYZ, XYZ

**Tabel 2.3 Sampel-Sampel yang Mungkin Terambil beserta Nilai Rata-Rata**

Sampel	Nilai-Nilai dalam Sampel			$\bar{X}$
VWX	70	75	80	75
VWY	70	75	80	75
VWZ	70	75	90	78.33
VXY	70	80	80	76.67
VXZ	70	80	90	80
VYZ	70	80	90	80
WXY	75	80	80	78.33
WXZ	75	80	90	81.67
WYZ	75	80	90	81.67
XYZ	80	80	90	83.33

Perhatikan bahwa terdapat 10 kemungkinan sampel. Sampel VWX berarti mengandung nilai 70, 75, dan 80, sampel WYZ berarti mengandung nilai 75, 80, dan 90, dan seterusnya. Tabel 2.3 menyajikan sampel-sampel yang mungkin terambil beserta penghitungan nilai rata-rata. Berdasarkan Tabel 2.3, selanjutnya dibentuk tabel distribusi frekuensi dan frekuensi relatif berdasarkan nilai rata-rata sampel (Tabel 2.4). Tabel 2.5 menyajikan distribusi sampling dari rata-rata sampel  $\bar{X}$  berdasarkan data pada Tabel 2.3.

Tabel 2.5 menyajikan distribusi probabilitas dari rata-rata sampel  $\bar{X}$  Sebagai contoh probabilitas untuk memperoleh sampel yang memiliki nilai rata-rata 76,67 sebesar 0,2. Atau dapat dinyatakan sebagai berikut.

$$P(\bar{X}= 81.67) = 0.20$$

**Tabel 2.4 Distribusi Frekuensi dan Frekuensi Relatif Berdasarkan Nilai Rata-Rata Sampel**

$\bar{X}$	Frekuensi	Frekuensi Relatif
75	2	0.2
76.67	1	0.1
78.33	2	0.2
80	2	0.2
81.67	2	0.2
83.33	1	0.1
Jumlah	10	1

**Tabel 2.5 Distribusi Sampling dari  $\bar{X}$  dengan Ukuran Sampel sebanyak 3**

$\bar{X}$	$P(\bar{X} = \bar{x}) = f(\bar{x})$
75	0.2
76.67	0.1
78.33	0.2
80	0.2
81.67	0.2
83.33	0.1
$\sum P(\bar{X} = \bar{x}) = 1$	

Berikut diberikan ilustrasi dalam R.

```

fungsi dasar R.R
Source on Save
Run
1 X = c(70, 75, 80, 80, 90)
2 X
3 library(prob)
4 urnsamples(c(70,75,80,80,90), size = 3, replace = FALSE, ordered = FALSE)
5
6
7 X = c(70, 75, 80, 80, 90)
8 X
9 sampel = combn(X, 3) #pengambilan sampel tanpa pengembalian dan tanpa memperhatikan urutan
10 sampel
11 ratarata = colMeans(sampel)
12 ratarata=format(ratarata, digits=4) #pengaturan desimal
13 prop.table(table(ratarata))
14 table(ratarata)/length(ratarata)
15 barplot(table(ratarata))
  
```

**Gambar 2.3**

Pada Gambar 2.3 mengaktifkan *package* **prob** (kode R baris 3) dengan maksud untuk menggunakan fungsi **urnsamples**.

```

X = c(70, 75, 80, 80, 90)
X

## [1] 70 75 80 80 90

library(prob)
  
```

**Gambar 2.4**



```
urnsamples(c(70,75,80,80,90), size = 3, replace = FALSE, ordered = FALSE)
```

```
##      X1 X2 X3
## 1    70 75 80
## 2    70 75 80
## 3    70 75 90
## 4    70 80 80
## 5    70 80 90
## 6    70 80 90
## 7    75 80 80
## 8    75 80 90
## 9    75 80 90
## 10   80 80 90
```

**Gambar 2.5**

Pada Gambar 2.5, penggalan kode R **replace = FALSE** berarti pengambilan sampel tanpa pengembalian, serta pada penggalan kode R **ordered = FALSE** berarti tanpa memperhatikan urutan.

```
sampel = combn(X, 3) #pengambilan sampel tanpa pengembalian dan tanpa memperhatikan urutan sampel
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]   70   70   70   70   70   70   75   75   75    80
## [2,]   75   75   75   80   80   80   80   80   80    80
## [3,]   80   80   90   80   90   90   80   90   90    90
```

**Gambar 2.6**

Pada Gambar 2.6 menyajikan alternatif kode R (dari yang sebelumnya) untuk menampilkan seluruh kemungkinan sampel yang mungkin terambil. Pada Gambar 2.6 menggunakan fungsi **combn** (*combination*).

```
ratarata = colMeans(sampel)
ratarata=format(ratarata, digits=4) #pengaturan desimal
prop.table(table(ratarata))
```

```
## ratarata
## 75.00 76.67 78.33 80.00 81.67 83.33
## 0.2 0.1 0.2 0.2 0.2 0.1
```

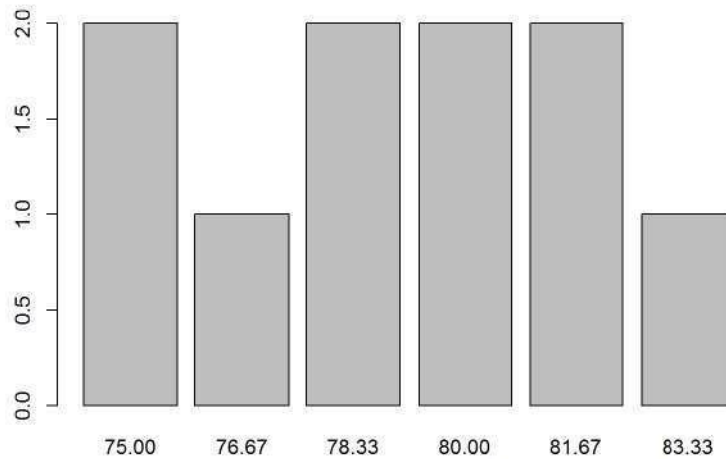
```
table(ratarata)/length(ratarata)
```

```
## ratarata
## 75.00 76.67 78.33 80.00 81.67 83.33
## 0.2 0.1 0.2 0.2 0.2 0.1
```

**Gambar 2.7**

Pada Gambar 2.7 menyajikan distribusi probabilitas dari rata-rata sampel  $\bar{X}$  Pada Gambar 2.8 menyajikan grafik batang yang menyajikan frekuensi dari setiap nilai rata-rata sampel  $\bar{X}$  Berdasarkan Gambar 2.8, nilai rata-rata 75 sebanyak 2, nilai rata-rata 76,67 sebanyak 1, dan seterusnya.

```
barplot(table(ratarata))
```



**Gambar 2.8**

### ***Rata-Rata dari Distribusi Sampling Rata-Rata Sampel $\bar{X}$***

Rata-rata dari distribusi sampling  $\bar{X}$  (*mean of the sampling distribution of  $\bar{X}$*  atau rata-rata dari  $\bar{X}$  dilambangkan dengan  $\mu_{\bar{X}}$  Berdasarkan Tabel 2.3, berikut akan dihitung rata-rata dari distribusi sampling  $\bar{X}$  serta rata-rata populasinya.

$$\mu_{\bar{X}} = \frac{75 + 75 + 78,33 + \dots + 83,33}{10} = 79$$

$$\mu = \frac{70 + 75 + 80 + 80 + 90}{5} = 79$$

Perhatikan bahwa berdasarkan perhitungan diperoleh  $\mu_{\bar{X}} = 79$  dan  $\mu = 79$ . Mann dan Lacke (2011:307) menyatakan sebagai berikut.

*“The mean of the sampling distribution of  $\bar{X}$  is always equal to the mean of the population. Thus,  $\mu_{\bar{X}} = \mu$ ”.*

Rata-rata sampel  $\bar{X}$  disebut juga sebagai *estimator* atau penduga terhadap rata-rata populasi  $\mu$ . Suatu statistik dikatakan sebagai estimator tak-bias atau *unbiased estimator* jika nilai rata-rata dari distribusi sampling statistik tersebut sama dengan nilai parameter tertentu. Perhatikan bahwa statistik rata-rata sampel  $\bar{X}$  merupakan estimator tak-bias dari parameter rata-rata populasi ( $\mu$ ), karena nilai rata-rata dari distribusi sampling rata-rata  $\bar{X}$  selalu sama dengan rata-rata populasi, yakni

$$\mu_{\bar{X}} = \mu.$$

Berikut diberikan ilustrasi dalam R.

```

RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
fungsi dasar R.R* *
Source on Save Run
1 X = c(70, 75, 80, 80, 90)
2 X
3 sampel = combn(X, 3) #pengambilan sampel tanpa pengembalian dan tanpa memperhatikan urutan
4 ratarata = colMeans(sampel)
5 ratarata=format(ratarata, digits=4) #pengaturan desimal
6 ratarata
7 mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character
8 ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
9 mode(ratarata)
10 ratarata
11 mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel

```

Gambar 2.9

```

X = c(70, 75, 80, 80, 90)
X
## [1] 70 75 80 80 90

sampel = combn(X, 3) #pengambilan sampel tanpa pengembalian dan tanpa memperhatikan urutan
ratarata = colMeans(sampel)
ratarata=format(ratarata, digits=4) #pengaturan desimal
ratarata
## [1] "75.00" "75.00" "78.33" "76.67" "80.00" "80.00" "78.33" "81.67"
## [9] "81.67" "83.33"

mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character
## [1] "character"

ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
mode(ratarata)
## [1] "numeric"

ratarata
## [1] 75.00 75.00 78.33 76.67 80.00 80.00 78.33 81.67 81.67 83.33

mean(ratarata)
## [1] 79

```

Gambar 2.10

Berdasarkan Gambar 2.9, kode R pada baris 7 bertujuan untuk mengetahui tipe atau jenis data dari variabel **ratarata**. Sementara kode R pada baris 8 bertujuan untuk mengkonversi jenis data variabel **ratarata**, dari *character* menjadi *numeric*. Setelah dikonversi menjadi *numeric*, barulah bisa dihitung nilai rata-rata dari distribusi sampling rata-rata sampel (kode R pada baris 11). Berikut alternatif kode R untuk memperoleh rata-rata dari distribusi sampling rata-rata sampel (perhatikan Gambar 2.11 sampai dengan Gambar 2.13).

```

RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
fungsi dasar R.R* x
Source on Save Run
1 library(prob)
2 sampel=urnsamples(c(70,75,80,80,90), size=3, replace=FALSE, ordered=FALSE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=4) #pengaturan desimal
7 ratarata
8 mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character
9 ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
10 mode(ratarata)
11 ratarata
12 mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel

```

Gambar 2.11

```

sampel=urnsamples(c(70,75,80,80,90), size=3, replace=FALSE, ordered=FALSE)
sampel

##      X1 X2 X3
## 1   70 75 80
## 2   70 75 80
## 3   70 75 90
## 4   70 80 80
## 5   70 80 90
## 6   70 80 90
## 7   75 80 80
## 8   75 80 90
## 9   75 80 90
## 10  80 80 90

ratarata = rowMeans(sampel)
ratarata=format(ratarata, digits=4) #pengaturan desimal
ratarata

## [1] "75.00" "75.00" "78.33" "76.67" "80.00" "80.00" "78.33" "81.67"
## [9] "81.67" "83.33"

mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character

## [1] "character"

ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
mode(ratarata)

## [1] "numeric"

ratarata

## [1] 75.00 75.00 78.33 76.67 80.00 80.00 78.33 81.67 81.67 83.33

mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel

## [1] 79

```

## **Standar Deviasi dari Distribusi Sampling Rata-Rata Sampel $\bar{X}$**

Diketahui pada pembahasan sebelumnya bahwa rata-rata dari distribusi sampling rata-rata  $\bar{X}$  dilambangkan dengan simbol  $\mu_x$ , sedangkan rata-rata populasi dilambangkan dengan simbol  $\mu$ . Standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  dilambangkan dengan simbol  $\sigma_x$  sedangkan standar deviasi populasi dilambangkan dengan simbol  $\sigma$ . Pada pembahasan sebelumnya diketahui bahwa rata-rata dari distribusi sampling rata-rata  $\bar{X}$  sama dengan rata-rata populasi  $\mu$ , yakni

$$\mu_x = \mu.$$

Namun pada standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  tidak sama dengan standar deviasi populasi (kecuali jika  $n = 1$ ). Sebagai contoh untuk kasus  $n = 1$ , misalkan suatu populasi terdiri dari tiga angka, yakni 1, 2, 3. Misalkan dari populasi yang terdiri dari tiga angka tersebut, akan diambil sampel yang terdiri atas satu angka. Maka sampel-sampel yang mungkin adalah

$$1 \quad 2 \quad 3.$$

Diketahui rata-rata dari setiap sampel tersebut adalah

$$1 \quad 2 \quad 3.$$

Maka rata-rata dari distribusi sampling rata-rata  $\bar{X}$  tersebut adalah

$$\mu_x = \frac{1 + 2 + 3}{3} = 2.$$

Sedangkan standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  tersebut adalah

$$\sigma_x = \sqrt{\frac{(1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2}{3}} = 0,8165,$$

yang mana

$$\sigma_x = \sigma \text{ (ketika } n = 1\text{)}.$$

Mann dan Lacke (2011:307) menyatakan rumus

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

berlaku ketika paling tidak memenuhi salah satu dari kriteria sebagai berikut.

- ⇒ Jumlah elemen dalam populasi berhingga (*finite*) dan pengambilan elemen untuk sampel dari suatu populasi dengan pengembalian (*with replacement*).
- ⇒ Jumlah elemen dalam populasi tak berhingga (*infinite*) dan pengambilan elemen untuk sampel dari suatu populasi tanpa pengembalian (*without replacement*).

Namun kriteria-kriteria tersebut dapat diganti ketika ukuran sampel kecil (*sample size is small*) dalam perbandingannya terhadap ukuran populasi (*in comparison to the population size*). Ukuran sampel dapat dipandang (*is considered*) kecil dalam perbandingannya terhadap ukuran populasi ketika ukuran sampel lebih kecil atau sama dengan 5% dari ukuran populasi, yakni

$$\frac{n}{N} \leq 0,05,$$

dengan  $n$  merupakan ukuran sampel dan  $N$  ukuran populasi. Namun ketika tidak terpenuhi, maka penghitungan  $\sigma_x$  dihitung dengan rumus

$$\sigma_x = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

di mana

$$\sqrt{\frac{N-n}{N-1}}$$

merupakan faktor koreksi populasi berhingga (Mann dan Lacke, 2011:307).

Berikut diberikan contoh kasus untuk perhitungan standar deviasi dari distribusi sampling  $\bar{X}$  dengan rumus  $\sigma_x = \frac{\sigma}{\sqrt{n}}$  Misalkan suatu populasi terdiri dari tiga angka, yakni 1, 2, 3.

Misalkan dari populasi yang terdiri dari tiga angka tersebut, akan diambil sampel yang terdiri atas dua angka dengan pengembalian (*with replacement*). Maka sampel-sampel yang mungkin adalah sebagai berikut.

(1,1)	(1,2)	(1,3)
(2,1)	(2,2)	(2,3)
(3,1)	(3,2)	(3,3)

Perhatikan bahwa karena jumlah elemen dalam populasi berhingga, yakni tiga, dan pengambilan elemen sampel dengan pengembalian, maka standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  dihitung dengan rumus sebagai berikut.

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

Hasil perhitungan rata-rata untuk setiap sampel sebagai berikut.

1	1,5	2
1,5	2	2,5
2	2,5	3

Maka rata-rata dari distribusi sampling rata-rata  $\bar{X}$  tersebut adalah

$$\mu_x = \frac{1 + 1,5 + 2 + 1,5 + 2 + 2,5 + 2 + 2,5 + 3}{9} = \frac{18}{9} = 2.$$

Berikut perhitungan standar deviasi dari distribusi sampling rata-rata  $\bar{X}$

$$\begin{array}{ccc} (1 - 2)^2 & (1,5 - 2)^2 & (2 - 2)^2 \\ (1,5 - 2)^2 & (2 - 2)^2 & (2,5 - 2)^2 \\ (2 - 2)^2 & (2,5 - 2)^2 & (3 - 2)^2 \end{array}$$

Maka diperoleh hasil sebagai berikut.

$$\begin{array}{ccc} 1 & 0,25 & 0 \\ 0,25 & 0 & 0,25 \\ 0 & 0,25 & 1 \end{array}$$

Sehingga

$$\sigma_x = \sqrt{\frac{(1 - 2)^2 + (1,5 - 2)^2 + (2 - 2)^2 + \dots + (3 - 2)^2}{9}}$$

$$\sigma_x = \sqrt{\frac{1 + 0,25 + 0 + 0,25 + 0 + 0,25 + 0 + 0,25 + 1}{9}}$$

$$\sigma_x = \sqrt{\frac{3}{9}} = \sqrt{0,33333333} = 0,57735$$

Perhatikan bahwa berdasarkan perhitungan sebelumnya diperoleh

$$\mu_x = 2$$

$$\sigma_x = 0,57735.$$

Diketahui

$$\mu = \frac{1 + 2 + 3}{3} = 2$$

$$\sigma = \sqrt{\frac{(1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2}{3}} = \sqrt{0,66666666} = 0,81649658.$$

Perhatikan bahwa

$$\sigma_x \neq \sigma,$$

namun

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$0,57735 = \frac{0,81649658}{\sqrt{2}}$$

$$0,57735 = 0,57735.$$

Berikut diberikan ilustrasi dalam R.

```

fungsi dasar R,R x
Source on Save
Run
1 library(prob)
2 sampel=urnsamples(c(1,2,3), size=2, replace=TRUE, ordered=TRUE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=4) #pengaturan desimal
7 ratarata
8 mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character
9 ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
10 mode(ratarata)
11 ratarata
12 mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel
13 #kode R pada baris 14 untuk menghitung standar deviasi dari distribusi samling rata-rata sampel
14 sd(ratarata)*sqrt((length(ratarata)-1)/(length(ratarata)))

```

Gambar 2.13

```

sampel=urnsamples(c(1,2,3), size=2, replace=TRUE, ordered=TRUE)
sampel

##      X1 X2
## 1  1  1
## 2  2  1
## 3  3  1
## 4  1  2
## 5  2  2
## 6  3  2
## 7  1  3
## 8  2  3
## 9  3  3

ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
ratarata=format(ratarata, digits=4) #pengaturan desimal
ratarata

## [1] "1.0" "1.5" "2.0" "1.5" "2.0" "2.5" "2.0" "2.5" "3.0"

mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character

## [1] "character"

ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
mode(ratarata)

## [1] "numeric"

ratarata

## [1] 1.0 1.5 2.0 1.5 2.0 2.5 2.0 2.5 3.0

mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel

## [1] 2

#kode R pada baris 14 untuk menghitung standar deviasi dari distribusi samling rata-rata sampel
sd(ratarata)*sqrt((length(ratarata)-1)/(length(ratarata)))

## [1] 0.5773503

```

Gambar 2.14



Berikut diberikan contoh kasus untuk perhitungan standar deviasi dari distribusi sampling  $\bar{X}$  dengan rumus  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ . Misalkan suatu populasi terdiri dari tiga angka, yakni 1, 2, 3.

Misalkan dari populasi yang terdiri dari tiga angka tersebut, akan diambil sampel yang terdiri atas dua angka tanpa pengembalian (*without replacement*). Maka sampel-sampel yang mungkin adalah

$$(1,2) \quad (1,3) \quad (2,3)$$

Perhatikan bahwa karena jumlah elemen dalam populasi berhingga, yakni tiga, namun pengambilan elemen sampel tanpa pengembalian, maka standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  dihitung dengan rumus sebagai berikut.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Diketahui rata-rata dari setiap sampel tersebut adalah

$$1,5 \quad 2 \quad 2,5,$$

sehingga rata-rata dari distribusi sampling rata-rata ( $\bar{X}$ ) tersebut adalah

$$\mu_{\bar{X}} = \frac{1,5 + 2 + 2,5}{3} = \frac{6}{3} = 2.$$

Standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  tersebut adalah

$$\sigma_{\bar{X}} = \sqrt{\frac{(1,5 - 2)^2 + (2 - 2)^2 + (2,5 - 2)^2}{3}}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{0,25 + 0 + 0,25}{3}}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{0,5}{3}} = \sqrt{0,16666667} = 0,408248.$$

Perhatikan bahwa berdasarkan perhitungan sebelumnya diperoleh

$$\mu_{\bar{X}} = 2$$

$$\sigma_{\bar{X}} = 0,408248.$$

Diketahui

$$\mu = \frac{1 + 2 + 3}{3} = 2$$

$$\sigma = \sqrt{\frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3}} = \sqrt{0,6666666} = 0,81649658.$$

Perhatikan bahwa

$$\sigma_x \neq \sigma.$$

Namun

$$\begin{aligned}\sigma_x &= \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \\ 0,408248 &= \frac{0,81649658}{\sqrt{2}} \sqrt{\frac{3-2}{3-1}} \\ 0,408248 &= \frac{0,81649658}{\sqrt{2}} \sqrt{\frac{1}{2}} \\ 0,408248 &= \frac{0,81649658}{2} \\ 0,408248 &= 0,408248\end{aligned}$$

Beberapa hal penting mengenai distribusi sampling rata-rata  $\bar{X}$  yakni:

- ⇒ Nilai standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  lebih kecil dibandingkan nilai standar deviasi populasi, yakni  $\sigma_x < \sigma$  ketika  $n$  lebih besar dari 1. Hal ini terlihat jelas dari rumus

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

Sebagai contoh misalkan  $\sigma = 20$  dan  $n = 4$ , maka

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{4}} = 10.$$

Perhatikan bahwa

$$\sigma_x < \sigma$$

$$10 < 20.$$

- ⇒ Nilai dari standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  akan semakin mengecil ketika ukuran sampel  $n$  semakin besar.

$$\text{ketika } n \uparrow \text{ maka } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \downarrow$$

Sebagai contoh misalkan  $\sigma = 20$  dan  $n = 4$ , maka

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{4}} = 10.$$

Untuk  $n = 20$  maka

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{20}} = 4,4721.$$

Untuk  $n = 50$  maka

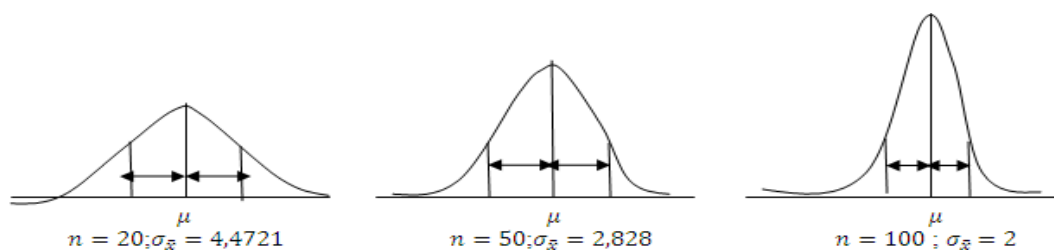
$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{50}} = 2,828.$$

Untuk  $n = 100$  maka

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2.$$

Perhatikan bahwa nilai  $\sigma_x$  semakin mengecil ketika ukuran sampel  $n$  semakin besar. Suatu statistik dikatakan estimator konsisten jika nilai standar deviasi dari distribusi sampling statistik tersebut semakin mengecil ketika ukuran sampel  $n$  semakin besar, sehingga statistik rata-rata  $\bar{X}$  merupakan estimator konsisten dari parameter rata-rata  $\mu$  (Mann dan Lacke, 2011:307)

Standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  merupakan suatu nilai yang mengukur pencaran atau sebaran dari rata-rata sampel dari distribusi sampling rata-rata  $\bar{X}$  terhadap rata-rata populasinya  $\mu$ . Semakin kecil nilai standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  maka rata-rata sampel dari distribusi sampling rata-rata  $\bar{X}$  semakin mengumpul atau lebih dekat terhadap rata-rata populasinya  $\mu$ . Pada pembahasan sebelumnya, diketahui untuk untuk  $n = 20$  diperoleh  $\sigma_x = 4,4721$ , untuk  $n = 50$  diperoleh  $\sigma_x = 2,828$ , dan untuk  $n = 100$  diperoleh  $\sigma_x = 2$ . Perhatikan ilustrasi gambar berikut ini (Gambar 5.15).



**Gambar 2.15**

Berikut diberikan ilustrasi dalam R (perhatikan Gambar 2.16 dan Gambar 2.17).

```

fungsi dasar R.R x
1 library(prob)
2 sampel=urnsamples(c(1,2,3), size=2, replace=FALSE, ordered=FALSE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=4) #pengaturan desimal
7 ratarata
8 mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character
9 ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
10 mode(ratarata)
11 ratarata
12 mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel
13 #kode R pada baris 14 untuk menghitung standar deviasi dari distribusi samling rata-rata sampel
14 sd(ratarata)*sqrt((length(ratarata)-1)/(length(ratarata)))

```

Gambar 2.16

```

sampel=urnsamples(c(1,2,3), size=2, replace=FALSE, ordered=FALSE)
sampel

## X1 X2
## 1 1 2
## 2 1 3
## 3 2 3

ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
ratarata=format(ratarata, digits=4) #pengaturan desimal
ratarata

## [1] "1.5" "2.0" "2.5"

mode(ratarata) #perhatikan bahwa tipe data rata-rata adalah character

## [1] "character"

ratarata=as.numeric(ratarata) #mengkonversi tipe data rata-rata, dari character menjadi numeric
mode(ratarata)

## [1] "numeric"

ratarata

## [1] 1.5 2.0 2.5

mean(ratarata) #menghitung rata-rata dari distribusi sampling rata-rata sampel

## [1] 2

#kode R pada baris 14 untuk menghitung standar deviasi dari distribusi samling rata-rata sampel
sd(ratarata)*sqrt((length(ratarata)-1)/(length(ratarata)))

## [1] 0.4082483

```

Gambar 2.17

### ***Bentuk Distribusi Sampling dari Rata-Rata Sampel $\bar{X}$***

Mann dan Lacke (2011:310) menyatakan bentuk distribusi sampling dari rata-rata  $\bar{X}$  berkenaan (*relates*) atas dua hal, yakni:

- ⇒ Sampel yang ditarik dari populasi yang berdistribusi normal.
- ⇒ Sampel yang ditarik dari populasi yang tidak berdistribusi normal.

Jika sampel-sampel yang ditarik berasal dari populasi yang berdistribusi normal dengan rata-rata dan standar deviasi masing-masing  $\mu$  dan  $\sigma$ , maka:

- ⇒ Rata-rata distribusi sampling rata-rata  $\bar{X}$  sama dengan rata-rata populasi, yakni

$$\mu_{\bar{X}} = \mu.$$

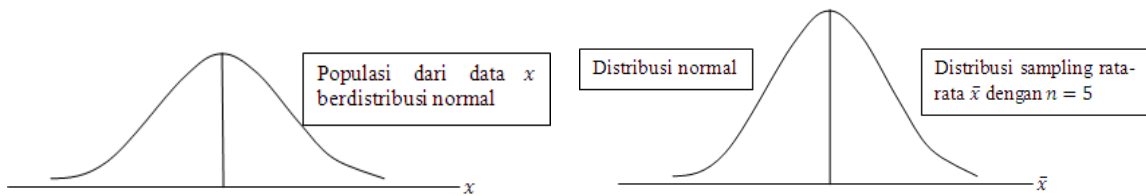
- ⇒ Standar deviasi distribusi sampling rata-rata  $\bar{X}$  sama dengan  $\frac{\sigma}{\sqrt{n}}$  dengan asumsi (*assuming*)  $n/N \leq 0,05$ .
- ⇒ Bentuk dari distribusi sampling rata-rata  $\bar{X}$  berbentuk normal, untuk berapapun ukuran sampel  $n$ .

Jadi, jika sampel-sampel yang ditarik berasal dari populasi yang berdistribusi normal dengan rata-rata adalah  $\mu$  dan standar deviasi adalah  $\sigma$ , maka distribusi sampling dari rata-rata  $\bar{X}$  juga terdistribusi secara normal, dengan rata-rata dan standar deviasi

$$\mu_{\bar{X}} = \mu$$

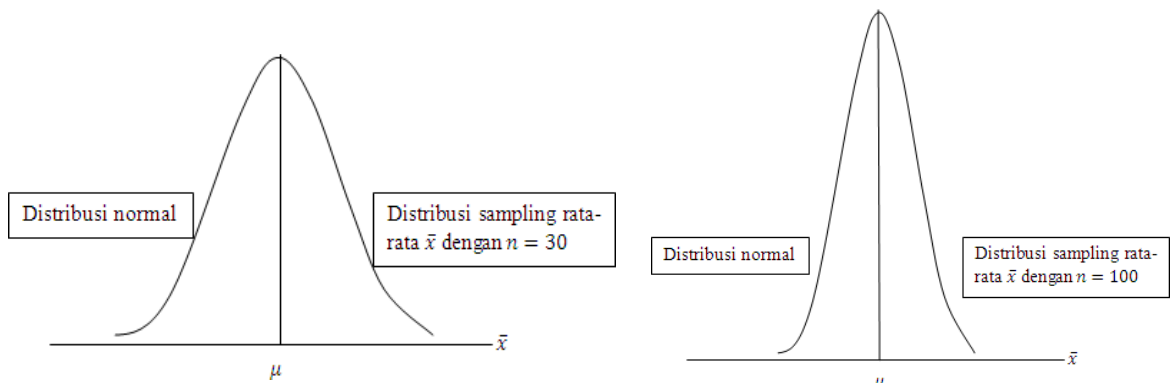
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} ; \frac{n}{N} \leq 0,05.$$

Perhatikan Gambar 5.18 hingga Gambar 5.21.



**Gambar 2.18**

**Gambar 2.19**



**Gambar 2.20**

**Gambar 2.21**

Perhatikan bahwa pada Gambar 2.18 menjelaskan data  $X$  berasal dari populasi berdistribusi normal. Pada Gambar 2.19 merupakan kurva dari distribusi sampling rata-rata  $\bar{X}$  dengan  $n = 5$ . Pada Gambar 2.20 merupakan kurva dari distribusi sampling rata-rata  $\bar{X}$  dengan  $n = 30$ .

Pada Gambar 2.21 merupakan kurva dari distribusi sampling rata-rata  $\bar{X}$  dengan  $n = 100$ . Perhatikan bahwa karena sampel-sampel ditarik dari populasi yang berdistribusi normal, maka kurva dari distribusi sampling rata-rata  $\bar{X}$  membentuk kurva normal (Gambar 2.19 sampai Gambar 2.21). Perhatikan bahwa standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  pada Gambar 2.20 lebih kecil daripada Gambar 2.19, standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  pada Gambar 2.21 lebih kecil daripada Gambar 2.20. Perhatikan bahwa semakin besar ukuran sampel, maka akan semakin kecil nilai standar deviasi dari distribusi sampling rata-rata  $\bar{X}$ . Dalam prakteknya, seringkali populasi yang diteliti tidak berdistribusi normal. Teorema yang sangat penting untuk menyimpulkan bentuk dari distribusi sampling rata-rata  $\bar{X}$  adalah **Teorema Limit Sentral** (*Central Limit Theorem*).

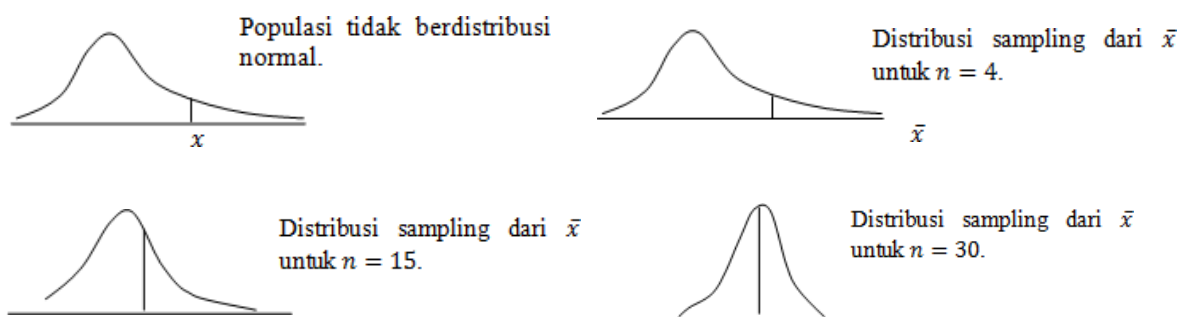
Teorema limit sentral menyatakan bahwa untuk sampel berukuran besar, distribusi sampling rata-rata  $\bar{X}$  akan mendekati normal, tidak peduli apakah sampel-sampel tersebut ditarik dari populasi yang berdistribusi normal atau tidak, dengan rata-rata dan standar deviasi dari distribusi sampling rata-rata  $\bar{X}$  sebagai berikut.

$$\mu_{\bar{X}} = \mu \quad \text{dan} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Ukuran sampel  $n$  dipertimbangkan cukup besar, yakni  $n \geq 30$ . Berdasarkan teorema limit sentral, perlu diperhatikan bahwa, jika populasi tidak berdistribusi normal, bentuk dari distribusi sampling rata-rata  $\bar{X}$  **tidak secara tepat normal, namun mendekati normal**, ketika sampel berukuran besar. Semakin besar ukuran sampel, maka bentuk dari distribusi sampling rata-rata ( $\bar{X}$ ) akan semakin mendekati normal. Berdasarkan teori limit sentral (Mann dan Lacke, 2011:313),

- ⇒ Ketika ukuran sampel  $n \geq 30$ , maka bentuk dari distribusi sampling rata-rata ( $\bar{X}$ ) mendekati normal, tidak peduli apakah sampel-sampel tersebut ditarik dari populasi berdistribusi normal atau tidak.
- ⇒ Rata-rata dari distribusi sampling rata-rata ( $\bar{X}$ ), yakni  $\mu_{\bar{X}}$  sama dengan rata-rata populasi, yakni  $\mu$ .
- ⇒ Standar deviasi dari distribusi sampling rata-rata ( $\bar{X}$ ), yakni  $\sigma_{\bar{X}}$  sama dengan  $\sigma/\sqrt{n}$  dengan syarat  $n/N \leq 0,05$ .

Perhatikan ilustrasi gambar berikut.



**Gambar 2.22**

Berdasarkan Gambar 2.22, populasi tidak berdistribusi normal. Semakin meningkat ukuran sampel, maka distribusi sampling rata-rata  $\bar{X}$  semakin berbentuk distribusi normal. Semakin

meningkat ukuran sampel, semakin kecil nilai standar deviasi dari distribusi sampling rata-rata  $\bar{X}$

## B. Simulasi Distribusi Sampling dalam R (Bagian 1)

Andaikan diberikan data populasi sebagai berikut.

1,2,3,4,5,6,7,8

Dari data populasi tersebut, akan diambil sampel yang terdiri dari 2 angka. Pengambilan sampel dengan pengembalian dan memperhatikan urutan. Dengan menggunakan R, berikut akan ditentukan seluruh kemungkinan sampel yang mungkin terambil, distribusi frekuensi dari rata-rata sampel, distribusi probabilitas dari rata-rata sampel atau distribusi sampling dari rata-rata sampel, dan disajikan secara visual.

```

fungsi dasar R.R x
1 library(prob)
2 sampel=urnsamples(c(1,2,3,4,5,6,7,8), size=2, replace=TRUE, ordered=TRUE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=4) #pengaturan desimal
7 table(ratarata)
8 table(ratarata)/length(ratarata)
9 barplot(table(ratarata))
10 plot(table(ratarata))
  
```

Gambar 2.23

```

sampel=urnsamples(c(1,2,3,4,5,6,7,8), size=2, replace=TRUE, ordered=TRUE)
sampel

##      X1 X2
## 1  1  1
## 2  2  1
## 3  3  1
## 4  4  1
## 5  5  1
## 6  6  1
## 7  7  1
## 8  8  1
## 9  1  2
## 10 2  2
## 11 3  2
## 12 4  2
## 13 5  2
## 14 6  2
## 15 7  2
## 16 8  2
## 17 1  3
## 18 2  3
## 19 3  3
## 20 4  3
## 21 5  3
## 22 6  3
## 23 7  3
## 24 8  3
## 25 1  4
## 26 2  4
## 27 3  4
## 28 4  4
## 29 5  4
## 30 6  4
## 31 7  4
## 32 8  4
## 33 1  5
## 34 2  5
## 35 3  5
## 36 4  5
## 37 5  5
## 38 6  5
## 39 7  5
## 40 8  5
## 41 1  6
## 42 2  6
## 43 3  6
## 44 4  6
## 45 5  6
## 46 6  6
## 47 7  6
## 48 8  6
## 49 1  7
## 50 2  7
## 51 3  7
## 52 4  7
## 53 5  7
## 54 6  7
## 55 7  7
## 56 8  7
## 57 1  8
## 58 2  8
## 59 3  8
## 60 4  8
## 61 5  8
## 62 6  8
## 63 7  8
## 64 8  8
  
```

Gambar 2.24

```
ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
ratarata=format(ratarata, digits=4) #pengaturan desimal
table(ratarata)
```

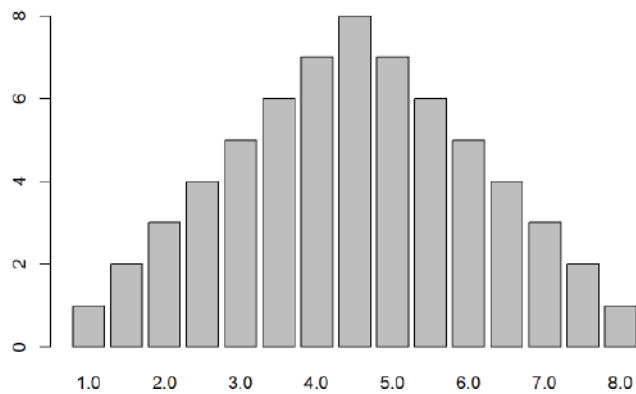
```
## ratarata
## 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0
## 1 2 3 4 5 6 7 8 7 6 5 4 3 2 1
```

```
table(ratarata)/length(ratarata)
```

```
## ratarata
## 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
## 0.015625 0.031250 0.046875 0.062500 0.078125 0.093750 0.109375 0.125000
## 5.0 5.5 6.0 6.5 7.0 7.5 8.0
## 0.109375 0.093750 0.078125 0.062500 0.046875 0.031250 0.015625
```

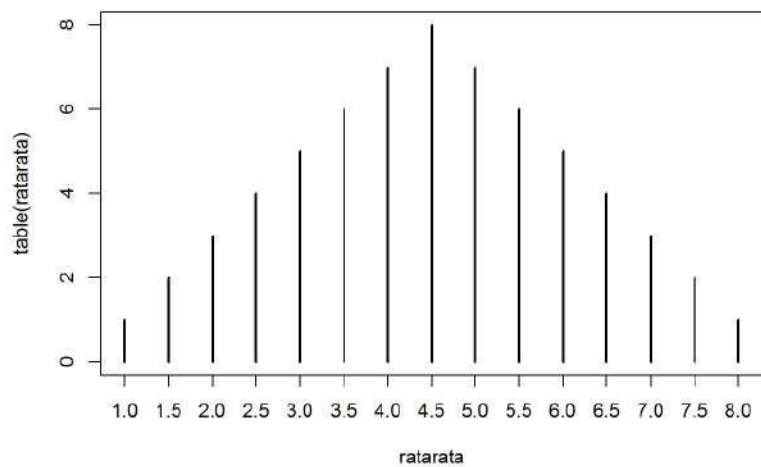
**Gambar 2.25**

```
barplot(table(ratarata))
```



**Gambar 2.26**

```
plot(table(ratarata))
```



**Gambar 2.27**



## C. Simulasi Distribusi Sampling dalam R (Bagian 2)

Andaikan diberikan data populasi sebagai berikut.

1,2,3,4,5,6,7,8

Dari data populasi tersebut, akan diambil sampel yang terdiri dari 3 angka. Pengambilan sampel dengan pengembalian dan memperhatikan urutan. Dengan menggunakan R, berikut akan ditentukan seluruh kemungkinan sampel yang mungkin terambil, distribusi frekuensi dari rata-rata sampel, distribusi probabilitas dari rata-rata sampel atau distribusi sampling dari rata-rata sampel, dan disajikan secara visual.

```
1 library(prob)
2 sampel=urnsamples(c(1,2,3,4,5,6,7,8), size=3, replace=TRUE, ordered=TRUE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=4) #pengaturan desimal
7 table(ratarata)
8 table(ratarata)/length(ratarata)
9 barplot(table(ratarata))
10 plot(table(ratarata))
```

Gambar 2.28

```
sampel=urnsamples(c(1,2,3,4,5,6,7,8), size=3, replace=TRUE, ordered=TRUE)
sampel
```

##	X1	X2	X3
## 1	1	1	1
## 2	2	1	1
## 3	3	1	1
## 4	4	1	1
## 5	5	1	1
## 6	6	1	1
## 7	7	1	1
## 8	8	1	1
## 9	1	2	1
## 10	2	2	1
## 11	3	2	1
## 12	4	2	1
## 13	5	2	1
## 14	6	2	1
## 15	7	2	1
## 16	8	2	1
## 17	1	3	1
## 18	2	3	1
## 19	3	3	1
## 20	4	3	1
## 21	5	3	1
## 22	6	3	1
## 23	7	3	1
## 24	8	3	1
## 25	1	4	1
## 26	2	4	1
## 27	3	4	1
## 28	4	4	1
## 29	5	4	1
## 30	6	4	1
## 31	7	4	1
## 32	8	4	1

##	1	2	3	4	5	6	7	8
## 479	7	4	8					
## 480	8	4	8					
## 481	1	5	8					
## 482	2	5	8					
## 483	3	5	8					
## 484	4	5	8					
## 485	5	5	8					
## 486	6	5	8					
## 487	7	5	8					
## 488	8	5	8					
## 489	1	6	8					
## 490	2	6	8					
## 491	3	6	8					
## 492	4	6	8					
## 493	5	6	8					
## 494	6	6	8					
## 495	7	6	8					
## 496	8	6	8					
## 497	1	7	8					
## 498	2	7	8					
## 499	3	7	8					
## 500	4	7	8					
## 501	5	7	8					
## 502	6	7	8					
## 503	7	7	8					
## 504	8	7	8					
## 505	1	8	8					
## 506	2	8	8					
## 507	3	8	8					
## 508	4	8	8					
## 509	5	8	8					
## 510	6	8	8					
## 511	7	8	8					
## 512	8	8	8					

Gambar 2.29

```
ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
ratarata=format(ratarata, digits=4) #pengaturan desimal
table(ratarata)
```

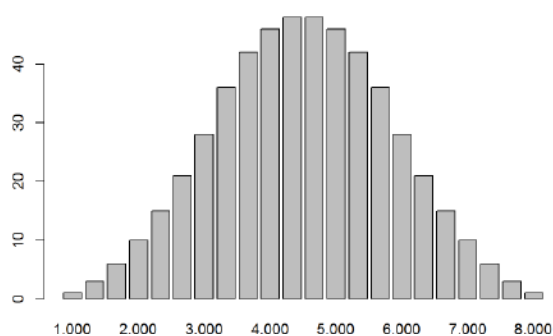
```
## ratarata
## 1.000 1.333 1.667 2.000 2.333 2.667 3.000 3.333 3.667 4.000 4.333 4.667
##      1      3      6     10     15     21     28     36     42     46     48     48
## 5.000 5.333 5.667 6.000 6.333 6.667 7.000 7.333 7.667 8.000
##      46     42     36     28     21     15     10      6      3      1
```

```
table(ratarata)/length(ratarata)
```

```
## ratarata
##      1.000      1.333      1.667      2.000      2.333      2.667
## 0.001953125 0.005859375 0.011718750 0.019531250 0.029296875 0.041015625
##      3.000      3.333      3.667      4.000      4.333      4.667
## 0.054687500 0.070312500 0.082031250 0.089843750 0.093750000 0.093750000
##      5.000      5.333      5.667      6.000      6.333      6.667
## 0.089843750 0.082031250 0.070312500 0.054687500 0.041015625 0.029296875
##      7.000      7.333      7.667      8.000
## 0.019531250 0.011718750 0.005859375 0.001953125
```

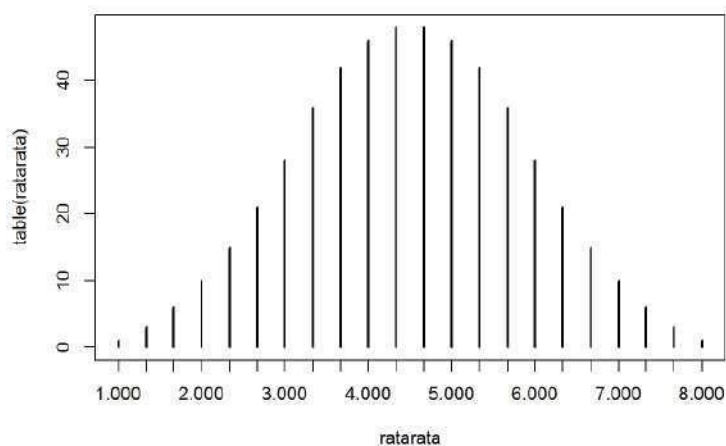
**Gambar 2.30**

```
barplot(table(ratarata))
```



**Gambar 2.31**

```
plot(table(ratarata))
```



**Gambar 2.32**

## D. Simulasi Distribusi Sampling dalam R (Bagian 3)

Andaikan diberikan data populasi sebagai berikut.

1,2,3,4,5,6,7,8

Dari data populasi tersebut, akan diambil sampel yang terdiri dari 4 angka. Pengambilan sampel dengan pengembalian dan memperhatikan urutan. Dengan menggunakan R, berikut akan ditentukan seluruh kemungkinan sampel yang mungkin terambil, distribusi frekuensi dari rata-rata sampel, distribusi probabilitas dari rata-rata sampel atau distribusi sampling dari rata-rata sampel, dan disajikan secara visual.

```
fungsi dasar R.R *
1 library(prob)
2 sampel=urnsamples(c(1,2,3,4,5,6,7,8), size=4, replace=TRUE, ordered=TRUE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=4) #pengaturan desimal
7 table(ratarata)
8 table(ratarata)/length(ratarata)
9 barplot(table(ratarata))
10 plot(table(ratarata))
```

Gambar 2.33

	X1	X2	X3	X4
##				
## 1	1	1	1	1
## 2	2	1	1	1
## 3	3	1	1	1
## 4	4	1	1	1
## 5	5	1	1	1
## 6	6	1	1	1
## 7	7	1	1	1
## 8	8	1	1	1
## 9	1	2	1	1
## 10	2	2	1	1
## 11	3	2	1	1
## 12	4	2	1	1
## 13	5	2	1	1
## 14	6	2	1	1
## 15	7	2	1	1
## 16	8	2	1	1
## 17	1	3	1	1
## 18	2	3	1	1
## 19	3	3	1	1
## 20	4	3	1	1
## 21	5	3	1	1
## 22	6	3	1	1
## 23	7	3	1	1
## 24	8	3	1	1
## 25	1	4	1	1
## 26	2	4	1	1
## 27	3	4	1	1
## 28	4	4	1	1
## 29	5	4	1	1
## 30	6	4	1	1
## 31	7	4	1	1
## 32	8	4	1	1
## 33	1	5	1	1
## 34	2	5	1	1
## 35	3	5	1	1
## 4061	5	4	8	8
## 4062	6	4	8	8
## 4063	7	4	8	8
## 4064	8	4	8	8
## 4065	1	5	8	8
## 4066	2	5	8	8
## 4067	3	5	8	8
## 4068	4	5	8	8
## 4069	5	5	8	8
## 4070	6	5	8	8
## 4071	7	5	8	8
## 4072	8	5	8	8
## 4073	1	6	8	8
## 4074	2	6	8	8
## 4075	3	6	8	8
## 4076	4	6	8	8
## 4077	5	6	8	8
## 4078	6	6	8	8
## 4079	7	6	8	8
## 4080	8	6	8	8
## 4081	1	7	8	8
## 4082	2	7	8	8
## 4083	3	7	8	8
## 4084	4	7	8	8
## 4085	5	7	8	8
## 4086	6	7	8	8
## 4087	7	7	8	8
## 4088	8	7	8	8
## 4089	1	8	8	8
## 4090	2	8	8	8
## 4091	3	8	8	8
## 4092	4	8	8	8
## 4093	5	8	8	8
## 4094	6	8	8	8
## 4095	7	8	8	8
## 4096	8	8	8	8

Gambar 2.34

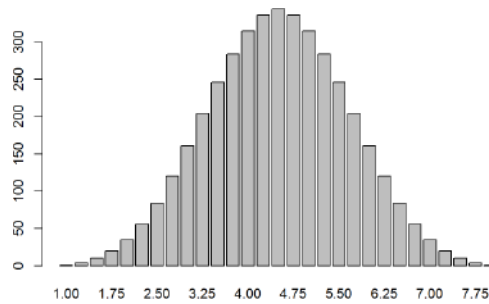
```
ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
ratarata=format(ratarata, digits=4) #pengaturan desimal
table(ratarata)
```

```
## ratarata
## 1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75 3.00 3.25 3.50 3.75 4.00 4.25 4.50
##    1    4   10   20   35   56   84  120  161  204  246  284  315  336  344
## 4.75 5.00 5.25 5.50 5.75 6.00 6.25 6.50 6.75 7.00 7.25 7.50 7.75 8.00
## 336 315 284 246 204 161 120  84  56  35  20  10  4  1
```

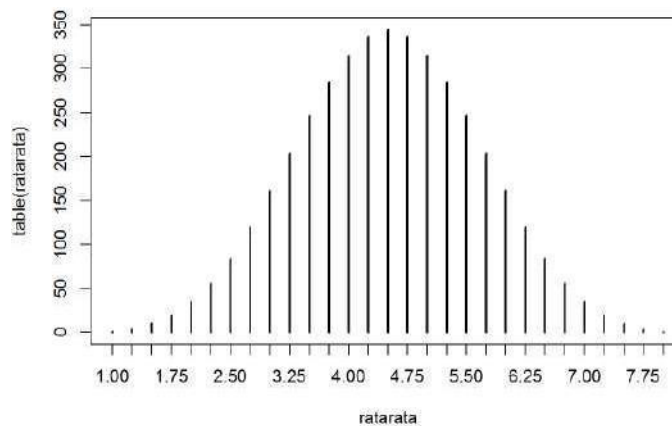
```
table(ratarata)/length(ratarata)
```

```
## ratarata
##      1.00      1.25      1.50      1.75      2.00
## 0.0002441406 0.0009765625 0.0024414062 0.0048828125 0.0085449219
##           2.25           2.50           2.75           3.00           3.25
## 0.0136718750 0.0205078125 0.0292968750 0.0393066406 0.0498046875
##           3.50           3.75           4.00           4.25           4.50
## 0.0600585938 0.0693359375 0.0769042969 0.0820312500 0.0839843750
##           4.75           5.00           5.25           5.50           5.75
## 0.0820312500 0.0769042969 0.0693359375 0.0600585938 0.0498046875
##           6.00           6.25           6.50           6.75           7.00
## 0.0393066406 0.0292968750 0.0205078125 0.0136718750 0.0085449219
##           7.25           7.50           7.75           8.00
## 0.0048828125 0.0024414062 0.0009765625 0.0002441406
```

**Gambar 2.35**



**Gambar 2.36**



**Gambar 2.37**

## E. Simulasi Distribusi Sampling dalam R (Bagian 4)

Andaikan diberikan data populasi sebagai berikut.

1,1,2,2,2,2,3,3,3,4,5,6

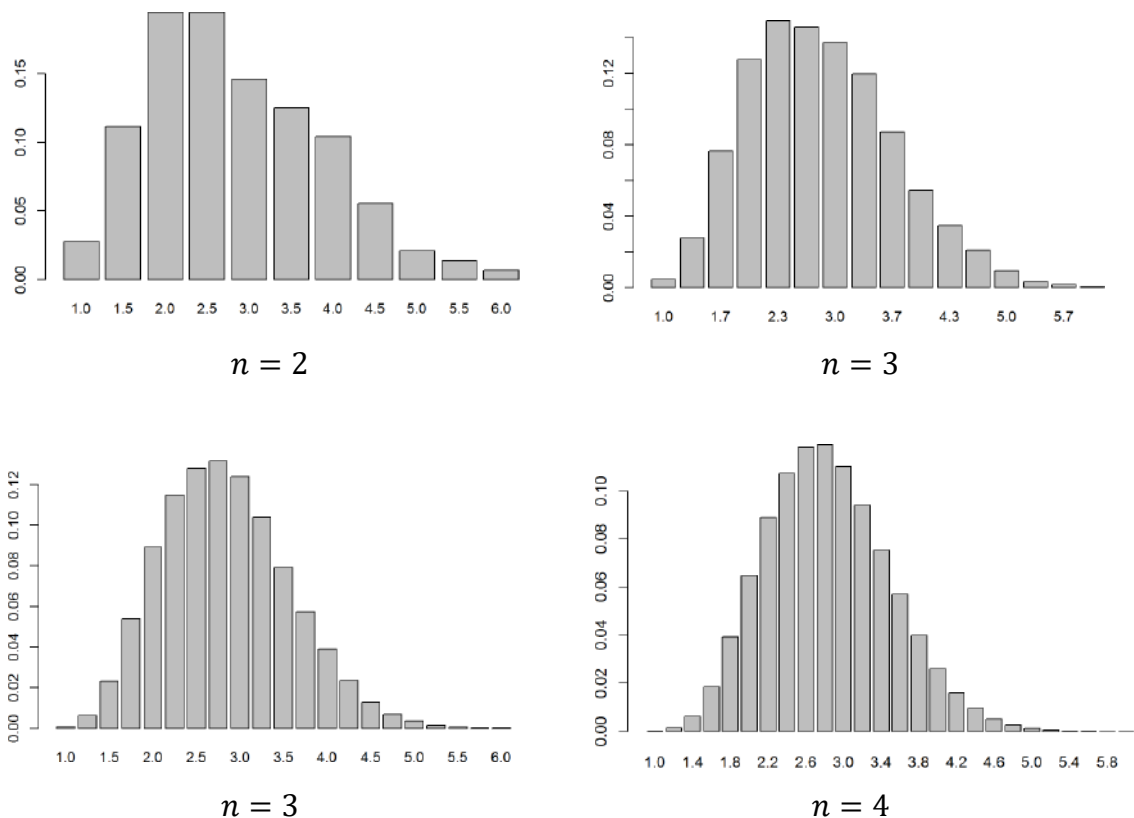
Dari data populasi tersebut, misalkan:

- akan diambil sampel yang terdiri dari 2 angka.
- akan diambil sampel yang terdiri dari 3 angka.
- akan diambil sampel yang terdiri dari 4 angka.
- akan diambil sampel yang terdiri dari 5 angka.

Pengambilan sampel **dengan pengembalian** dan **memperhatikan urutan**. Dengan menggunakan R, berikut akan disajikan secara visual distribusi sampling dari rata-rata sampel.

```
fungsi dasar R.R x
1 library(prob)
2 sampel=urnsamples(c(1,1,2,2,2,2,3,3,3,4,5,6), size=5, replace=TRUE, ordered=TRUE)
3 sampel
4
5 ratarata = rowMeans(sampel) #sebelumnya menggunakan colMeans, sekarang menggunakan rowMeans
6 ratarata=format(ratarata, digits=2) #pengaturan desimal
7 table(ratarata)
8 table(ratarata)/length(ratarata)
9 barplot(table(ratarata) / length(ratarata))
10 plot(table(ratarata) / length(ratarata))
```

Gambar 2.38

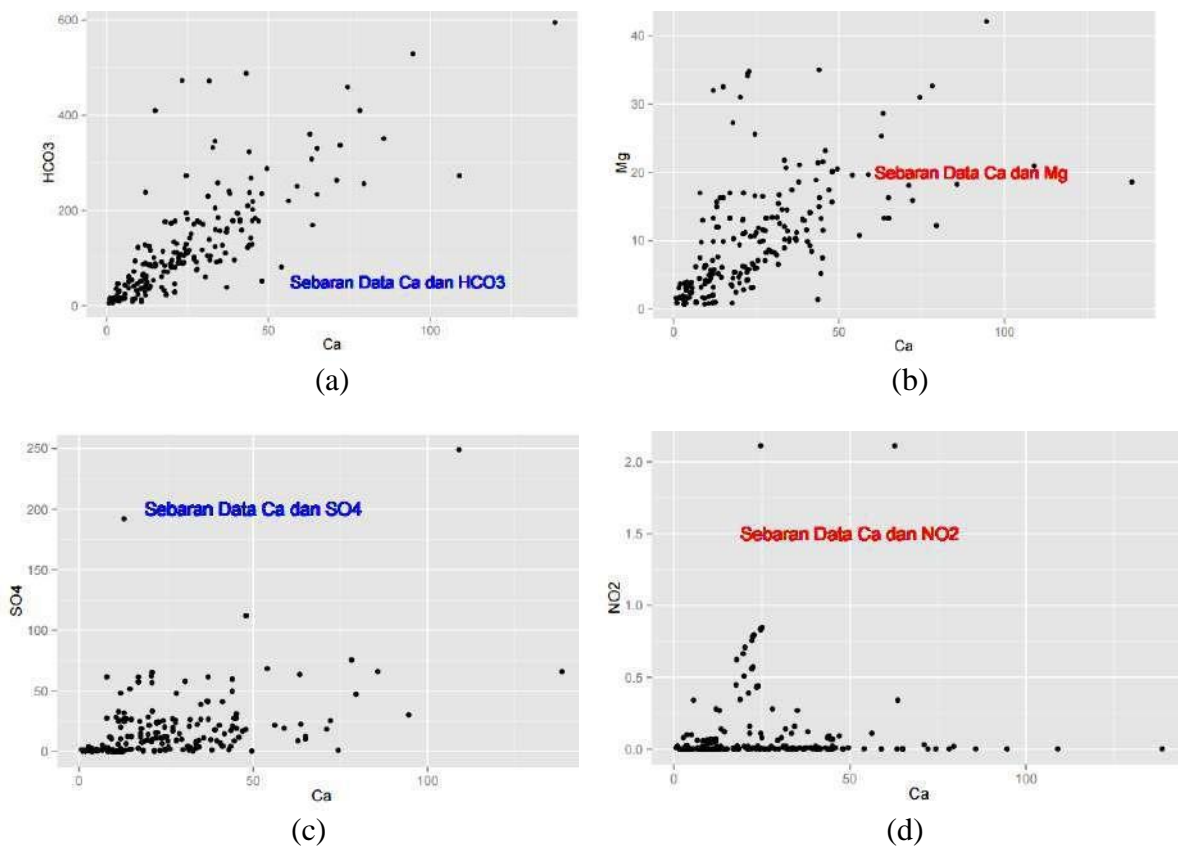


Gambar 2.39

## BAB III KORELASI LINEAR PEARSON

### A. Analisis Korelasi (Hubungan) Linear dengan Grafik

Berikut disajikan grafik dari sebaran data antara Ca v/s HCO<sub>3</sub>, Ca v/s Mg, Ca v/s SO<sub>4</sub>, dan Ca v/s NO<sub>2</sub>.



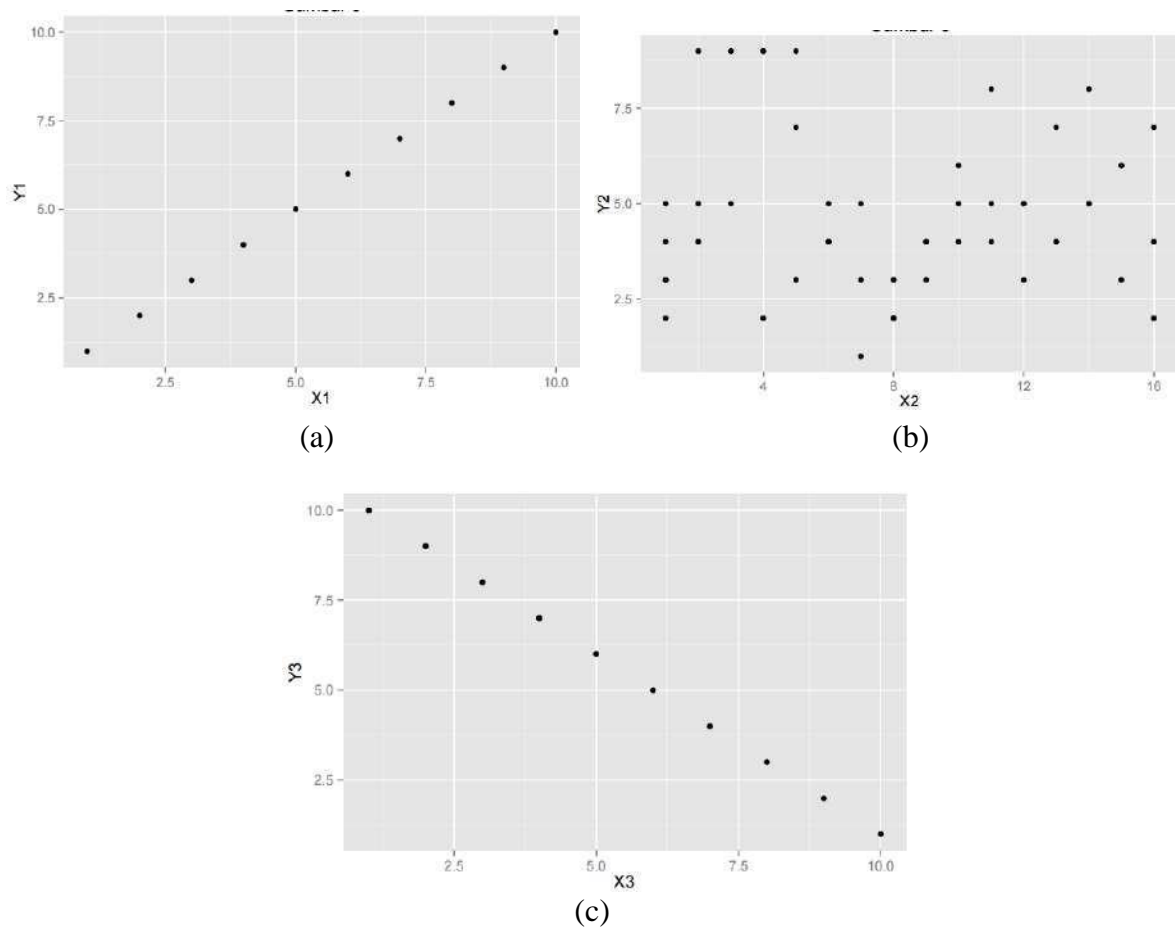
**Gambar 3.1**

Berdasarkan sebaran data dari Gambar 3.1(a) hingga Gambar 3.1(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.

### B. 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 3.2(a) hingga Gambar 3.2(c).



**Gambar 3.2**

Gambar 3.2(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 3.2(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 3.2(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.**

### **C. Menyajikan Grafik Sebaran Data dan Menghitung Koefisien Korelasi Linear Pearson dengan R**

Misalkan diberikan data seperti pada Gambar 3.3 dengan nama *file* **contohdata.csv**, dan Gambar 3.4 dengan nama *file* **contohdata2.csv**. Gambar 3.6 disajikan kode R. Apabila kode R pada Gambar 3.6 dieksekusi, hasilnya seperti pada Gambar 3.7 dan Gambar 3.8. Berdasarkan Gambar 3.8, 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 3.2(a). Berdasarkan Gambar 3.8, 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 3.2(b)). Berdasarkan Gambar 3.8, 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 3.2(c).

**contohdata.csv**

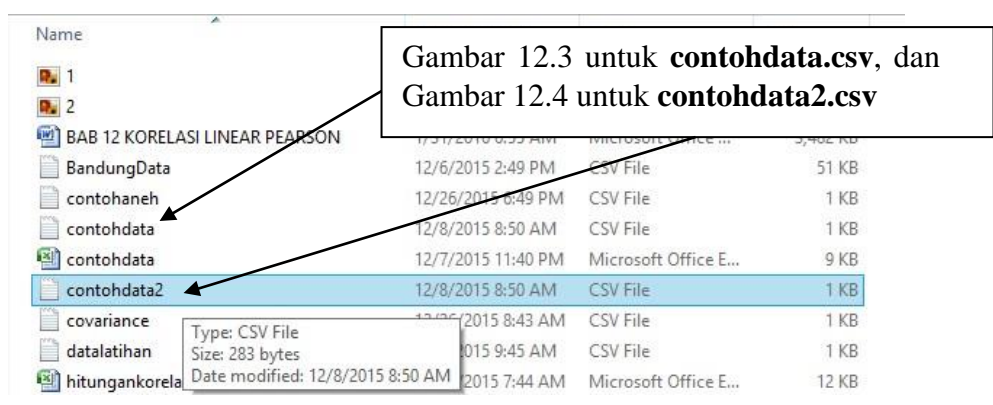
	A	B	C	D	E
1	X1	Y1	X3	Y3	
2	1	1	1	10	
3	2	2	2	9	
4	3	3	3	8	
5	4	4	4	7	
6	5	5	5	6	
7	6	6	6	5	
8	7	7	7	4	
9	8	8	8	3	
10	9	9	9	2	
11	10	10	10	1	
12					

**Gambar 3.3**

**contohdata2.csv**

	A	B	C			
1	X2	Y2		28	9	3
2	1	3		29	10	4
3	1	4		30	11	4
4	2	5		31	12	5
5	3	9		32	13	7
6	4	2		33	14	8
7	5	3		34	15	6
8	6	4		35	16	4
9	7	1		36	1	2
10	8	2		37	1	3
11	9	4		38	2	4
12	10	6		39	3	5
13	11	5		40	4	9
14	12	3		41	5	9
15	13	4		42	6	5
16	14	5		43	7	5
17	15	3		44	8	3
18	16	2		45	9	4
19	1	3		46	10	5
20	1	5		47	11	8
21	2	9		48	12	5
22	3	9		49	13	4
23	4	9		50	14	5
24	5	7		51	15	6
25	6	4		52	16	7

**Gambar 3.4**



**Gambar 3.5**

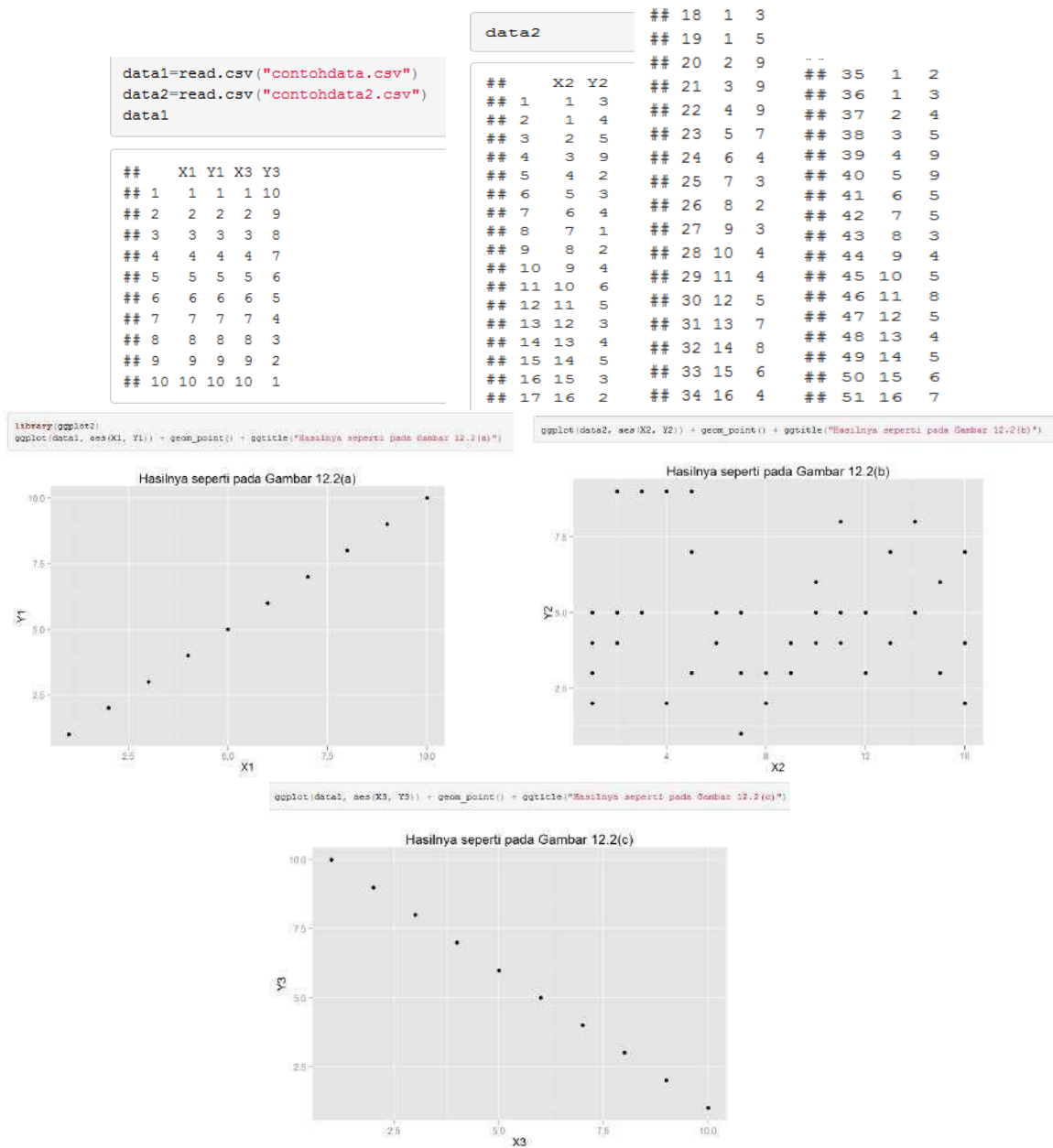
```

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 3.6**





Gambar 3.7

```

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 3.8

Berdasarkan Gambar 3.6, secara umum, perintah untuk menghitung koefisien korelasi linear Pearson dalam R sebagai berikut.

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

Berdasarkan Gambar 3.6, *package ggplot2* diaktifkan (kode R pada baris 6), dengan maksud untuk menggunakan fungsi **ggplot**. Fungsi **ggplot** bertujuan untuk menyajikan grafik sebaran data.

#### D. Menyajikan Grafik Sebaran Data dalam R (Bagian 2)

Grafik dari sebaran data antara Ca v/s HCO<sub>3</sub>, Ca v/s Mg, Ca v/s SO<sub>4</sub>, dan Ca v/s NO<sub>2</sub>, seperti pada Gambar 3.1, akan disajikan kembali, seperti pada Gambar 3.10. Kode R disajikan pada Gambar 3.9.

```
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 3.9



Gambar 3.10

## E. Menghitung Koefisien Korelasi Linear Pearson secara Sekaligus dengan R

Pada pemaparan sebelumnya, penghitungan nilai koefisien korelasi linear Pearson dilakukan secara satu persatu (Gambar 3.8). 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 3.11

```

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	0.64696660	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 3.12

Berdasarkan Gambar 3.12, 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 3.1. Berdasarkan data pada Tabel 3.1, 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 3.1**

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 3.2**

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				
Rata-Rata	11,8	3,221	0	0	87,6	1,00169
						9,072

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

$$r = 0,968465868$$

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("korelasi1.csv")
2 data
3 cor(data$X, data$Y, method="pearson")

```

**Gambar 3.13**

```

data=read.csv("korelasil.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 3.14

**Contoh Perhitungan Covariance dan Penyelesaian dalam R**

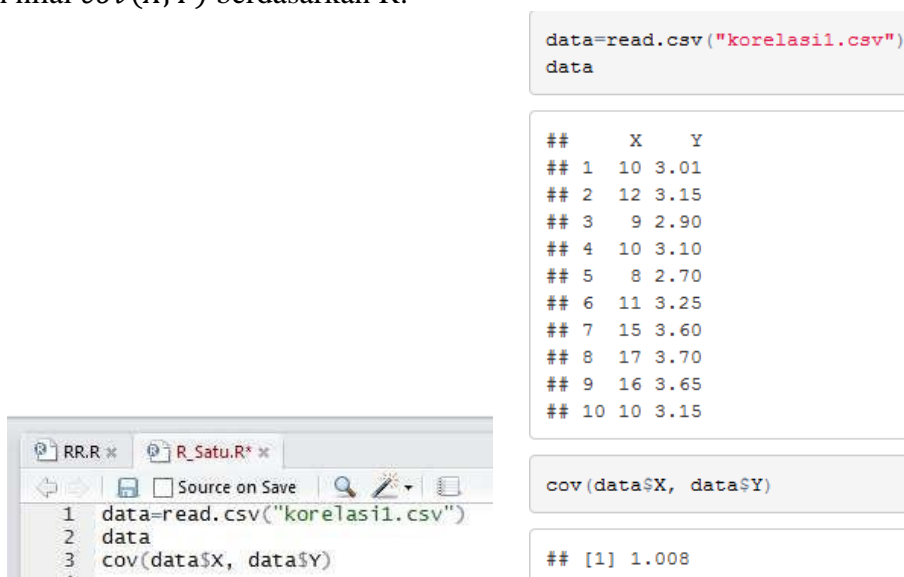
Berdasarkan data pada Tabel 3.1, berikut rumus untuk menghitung *covariance* antara variabel *X* dan variabel *Y* ( $cov(X, Y)$ ).

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

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("korelasil.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 3.15

## DAFTAR PUSTAKA

- Gio, P.U. dan E. Rosmaini, 2015. Belajar Olah Data dengan SPSS, Minitab, R, Microsoft Excel, EViews, LISREL, AMOS, dan SmartPLS. USUpres.
- Hair, J.F Jr., R.E. Anderson, B.J. Babin, dan W.C. Black. 2010. *Multivariate Data Analysis, 7<sup>th</sup> Edition*. Pearson Prentice Hall.
- Johnson, R.A. dan G.K. Bhattacharyya. 2011. *Statistics, Principles and Methods, 6<sup>th</sup> Edition*. John Wiley and Sons, Inc.
- Mann, P. S. dan C.J. Lacke. 2011. *Introductory Statistics, International Student Version, 7<sup>th</sup> Edition*. Asia: John Wiley & Sons, Inc.
- Montgomery, D.C. dan G.C. Runger. 2011. *Applied Statistics and Probability for Engineers, 5<sup>th</sup> Edition*. United States of America: John Wiley & Sons, Inc.