

BAB III

PENYAJIAN DATA

Capaian Pembelajaran

- Mahasiswa mampu menyajikan data dengan metode yang tepat sehingga data mudah dipahami.
- Mahasiswa mampu memberikan analisa keputusan awal dari sebuah permasalahan berdasarkan deskriptif statistika

Materi

Data statistika tidak hanya cukup dikumpulkan dan diolah, tetapi juga perlu disajikan dalam bentuk yang mudah dibaca dan dimengerti oleh pengambil keputusan. Penyajian data dalam bentuk tabel atau grafik merupakan cara untuk meringkas informasi dari sekumpulan data, mendapatkan pola, serta memahami apa yang terjadi di dalam data. Terdapat dua cara untuk menyajikan data, yaitu:

- a. *Tabel* merupakan kumpulan angka-angka yang disusun menurut kategori - kategori.
- b. *Grafik* merupakan gambar-gambar data secara visual dari data yang biasanya berasal dari tabel-tabel yang telah dibuat.

3.1 Penyajian dalam Tabel

Penyajian data menggunakan tabel dilakukan untuk meringkas informasi dari sekumpulan data. Salah satu caranya adalah menggunakan tabel distribusi frekuensi untuk menampilkan distribusi dari satu variabel numerik. Pada variabel kategorik, tabel distribusi frekuensi akan menampilkan kategori dan jumlah, proporsi, atau persentase dari setiap kategori. Proporsi dan persentase juga disebut sebagai frekuensi relative yang merangkum distribusi dari variabel kategorik secara numerik (Franklin, Klingenberg, & Agresti, 2017).

Contoh aplikasi tabel frekuensi dalam R adalah sebagai berikut.

```
# creating a dataframe
data_table <- data.table(col1 = sample(6:9, 9, replace = TRUE),
                        col2 = letters[1 : 3],
                        col3 = c(1, 4, 1, 2, 2, 2, 1, 2, 2))
print ("Original DataFrame")
print (data_table)
```

Diperoleh data sebagai berikut:

	col1	col2	col3
1:	7	a	1
2:	7	b	4
3:	7	c	1
4:	8	a	2
5:	8	b	2
6:	7	c	2
7:	9	a	1
8:	9	b	2
9:	6	c	2

Frekuensi dari data pada kolom ke-1 (col 1) adalah sebagai berikut

```
> freq <- table(data_table$col1)
> print ("Modified Frequency Table")
[1] "Modified Frequency Table"
> print (freq)
 6 7 8 9
1 4 2 2
```

Berdasarkan output di atas diketahui bahwa jumlah angka terbanyak pada col 1 adalah angka 7, yaitu sebanyak empat angka. Sedangkan jumlah angka terkecil pada col 1 adalah angka 6, yaitu sebanyak satu angka.

Frekuensi kumulatif dapat dihitung dengan cara berikut

```
> print ("Cumulative Frequency Table")
[1] "Cumulative Frequency Table"
> cumsum <- cumsum(freq)
> print (cumsum)
 6 7 8 9
1 5 7 9
```

Frekuensi kumulatif menunjukkan frekuensi data terakhir merupakan jumlahan dari banyaknya data sebelumnya. Oleh karena itu, frekuensi data terakhir dihitung dari $1 + 4 + 2 + 2 = 9$.

Frekuensi relatif (proporsi) dapat dihitung dengan cara berikut.

```
> print ("Relative Frequency Table")
[1] "Relative Frequency Table"
> prob <- prop.table(freq)
> print (prob)

      6      7      8      9
0.1111111 0.4444444 0.2222222 0.2222222
```

Frekuensi relative atau proporsi menunjukkan persentase dari banyaknya suatu data dibagi dengan banyak semua data. Oleh karena itu, proporsi angka 6 dalam data adalah sebesar $1/9 = 0,111$, proporsi angka 7 adalah $4/9 = 0,444$, begitu seterusnya untuk angka 8 dan 9.

Pada variabel kuantitatif (numerik), tabel distribusi frekuensi digunakan untuk mengelompokkan data interval/rasio dan menghitung banyaknya data dalam satu kelompok/klasifikasi. Langkah-langkah membuat tabel distribusi frekuensi untuk variabel numerik adalah sebagai berikut:

- Hitung sebaran (range) yaitu selisih antara nilai data terbesar dan terkecil, $\text{range} = x_{\max} - x_{\min}$.
- Tentukan banyaknya kelas (k) dengan rumus $k = 1 + 3,3 \log n$.
- Tentukan panjang kelas (p) dengan rumus $p = \text{sebaran} / \text{banyak kelas}$.

Berikut adalah contoh penggunaan tabel distribusi frekuensi untuk variabel numerik.

```
data_frame <- data.frame(coll = c(1,3,5,6,23,6,2,5,7,
                                16,8,9,36,7,12,1,
                                6,4,14,23,19,18,
                                14,2,20,30))

print("Original Data")
print(data_frame)
```

Interval antar kelas ditentukan sebesar 5 dengan perintah berikut

```
# creating intervals between 1 to 30 with a gap of 5 each
interval_table <- table(cut(data_frame$coll, seq(1,30,5)))
print("Data in Intervals")
print(interval_table)
```

Diperoleh output sebagai berikut.

```
> print(interval_table)

(1,6] (6,11] (11,16] (16,21] (21,26]
    9      4      4      3      2
```

Berdasarkan output di atas, diketahui bahwa telah terbentuk kelompok dengan interval sebesar 5. Dari kelompok tersebut, frekuensi terbesar terletak pada interval (1,6] dan frekuensi terkecil berada pada interval (21, 26].

Penyajian data menggunakan grafik disebut juga sebagai visualisasi data, yaitu mengubah informasi atau data ke dalam bentuk gambar atau grafik yang dapat menunjukkan distribusi, pola, dan trend dari data. Tujuan dari visualisasi data adalah untuk menceritakan suatu kejadian (*data stories*) atau menjawab suatu persoalan dengan memilih informasi yang

sesuai dari sekumpulan data untuk menonjolkan cerita tersebut (Pearson, 2018). Penggunaan grafik untuk visualisasi data sangat ditentukan oleh tipe data yang digunakan. Berikut adalah beberapa fungsi grafik yang dapat digunakan dalam perangkat lunak R.

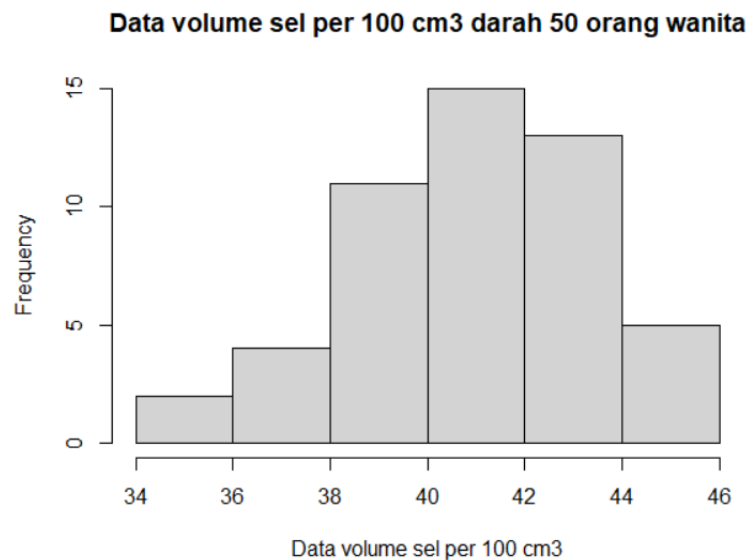
3.2 Penyajian dalam Histogram

Histogram digunakan untuk mengetahui distribusi dari sekumpulan data kuantitatif (numerik). Histogram menggunakan batang untuk menggambarkan frekuensi atau frekuensi relatif dari data. Histogram lebih baik digunakan pada jumlah data yang besar, sedangkan apabila data berjumlah sedikit, maka gunakan dotplot atau stem-and-leaf untuk melihat distribusi dari data (Franklin et al., 2017; Weiss, 2016).

Histogram dalam R dapat ditampilkan menggunakan perintah `hist()` sebagai berikut.

```
#Histogram
hist(selDarah,
      main="Data volume sel per 100 cm3 darah 50 orang wanita",
      xlab="Data volume sel per 100 cm3")
```

Histogram yang dihasilkan adalah sebagai berikut.



Gambar 1. Histogram

Berdasarkan histogram pada Gambar 3, dapat disimpulkan bahwa kurva frekuensinya menjulur ke kiri dimana nilai mean paling kecil dibandingkan dengan nilai median dan modus. Dan berarti bahwa data tidak berdistribusi normal. Namun, untuk melihat kenormalan data dengan histogram belum bisa dijadikan suatu keputusan yang valid untuk menyatakan bahwa data telah mengikuti distribusi normal atau tidak. Agar lebih meyakinkan, perlu dilakukan uji kenormalan data yang akan dipelajari dalam statistika lanjutan.

3.3 Penyajian dalam Chart

Grafik batang digunakan untuk menampilkan informasi berupa frekuensi, persentase atau nilai statistik lainnya dari data yang bersifat kategorik. Garis vertikal menunjukkan frekuensi, persentase atau nilai statistik lainnya, sedangkan garis horizontal menunjukkan kategori dengan tubuh batang yang terpisah (Franklin et al., 2017).

Grafik batang dalam R dapat dibuat menggunakan fungsi `barplot()`. Sebagai ilustrasi, data *VADeaths* akan digunakan untuk membuat grafik batang. Data *VADeaths* memuat informasi tentang tingkat kematian per 1000 penduduk di Virginia. Penduduk dikelompokkan berdasarkan kelompok umur (baris) dan kelompok populasi (kolom). Berikut adalah isi dari data *VADeaths*.

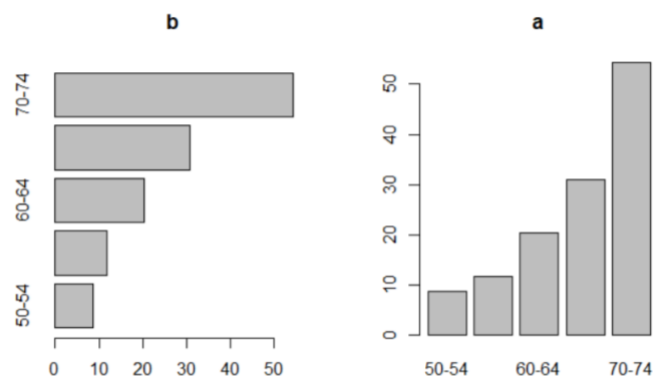
```
> VADeaths
      Rural Male Rural Female Urban Male Urban Female
50-54      11.7      8.7      15.4      8.4
55-59      18.1     11.7     24.3     13.6
60-64      26.9     20.3     37.0     19.3
65-69      41.0     30.9     54.6     35.1
70-74      66.0     54.3     71.1     50.0
```

Beikut adalah *perintah* untuk membuat grafik batang variabel tingkat kematian penduduk perempuan yang tinggal di wilayah pedesaan Virginia (*Rural Male*).

```
#BARPLOT
VADeaths
par(mfrow=c(1,2))
# Horizontal Bar Plot for
barplot(VADeaths[, "Rural Female"], main="b", horiz=TRUE)

# Vertical Bar Plot for
barplot(VADeaths[, "Rural Female"], main="a")
```

Berikut adalah hasil atau output grafik batangnya.



Gambar 2. Grafik batang

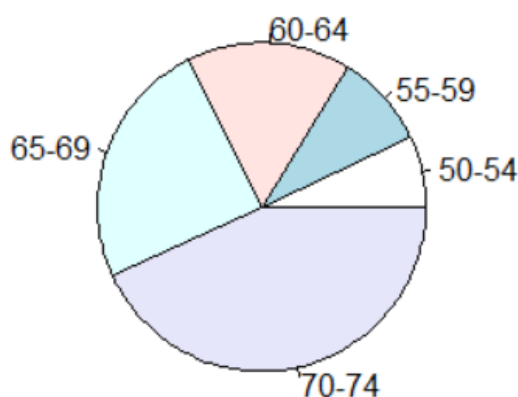
Berdasarkan grafik batang pada Gambar 1, diketahui bahwa tingkat kematian penduduk perempuan yang tinggal di wilayah pedesaan Virginia berada pada kelompok umur 70 – 74 tahun, sedangkan tingkat kematian terendah penduduk perempuan yang tinggal di wilayah pedesaan Virginia berada pada kelompok umur 50 – 54 tahun.

Grafik Pie (*Pie Chart*) merupakan grafik yang hanya digunakan untuk data numerik yang bersifat non-negatif. Setiap bagian lingkaran menunjukkan persentase dari setiap kategori. Perintah yang digunakan untuk membuat grafik pie dalam R adalah `pie()`.

Sebagai ilustrasi, data *VADeaths* akan digunakan untuk membuat garfik pie pada variabel tingkat kematian penduduk perempuan yang tinggal di wilayah pedesaan Virginia sebagai berikut.

```
#PIE CHART
pie(VADeaths[, "Rural Female"])
```

Berikut adalah output grafik pie yang dihasilkan.



Fungsi `plot()` merupakan fungsi grafik yang paling umum untuk membuat plot atau grafik di R. Format dasarnya adalah `plot(x, y, type="p")`. Beberapa tipe plot yang dapat digunakan adalah sebagai berikut.

Tabel 1. Tipe-Tipe Plot dalam R

Tipe	Keterangan
<i>p</i>	Membuat plot titik atau scatterplot. Nilai ini merupakan default pada fungsi <code>plot()</code> .
<i>l</i>	Plot garis
<i>b</i>	Plot titik yang terhubung dengan garis
<i>o</i>	Plot titik yang ditimpa oleh garis
<i>h</i>	Plot garis vertikal dari titik ke garis $y=0$
<i>s</i>	Fungsi tangga
<i>n</i>	Tidak membuat grafik plot sama sekali, kecuali plot dari axis. Dapat digunakan untuk mengatur tampilan suatu plot utama yang diikuti oleh sekelompok plot tambahan

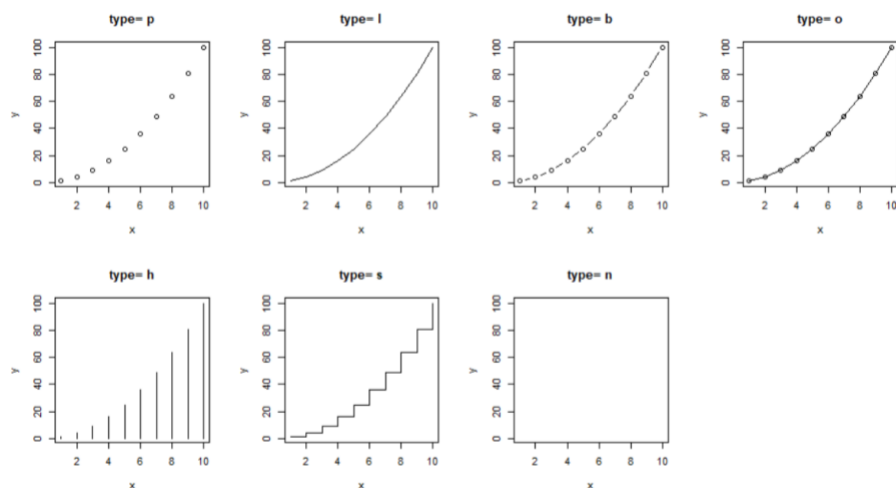
Contoh:

```
# membuat vektor data
x <- c(1:10); y <- x^2

# membagi jendela grafik menjadi 2 baris dan 4 kolom
par(mfrow=c(2,4))

# loop
type <- c("p","l","b","o","h","s","n")
for (i in type){
  plot(x,y, type= i,
       main= paste("type=", i))
}
```

Berikut adalah hasilnya:



3.4 Penyajian dalam *Stem-and-Leaf*

Jika kumpulan data relatif kecil, diagram *stem-and-leaf* (diagram batang-daun) berguna untuk melihat bentuk distribusi dan nilainya. Angka di sebelah kiri adalah batangnya (stem), yang digunakan untuk mengelompokkan nilai. Sedangkan angka di sebelah kanan adalah digitnya (leaf), yang menunjukkan skor individu dalam setiap kelompok.

Secara manual, grafik stem-and-leaf dibentuk dengan mengurutkan data dari nilai terkecil ke nilai terbesar. Kemudian, letakkan stem pada satu kolom, mulai dari angka terkecil. Pada sisi kanan garis vertikal (leaf), letakkan angka desimal mulai dari angka terkecil. Berikut contoh penggunaannya.

Input data pada Tabel 1 ke section *Rscript* atau *Rconsole* dengan perintah berikut:

```
selDarah <- c(35.2, 39.3, 40.6, 41.8, 43.1,
              35.3, 39.4, 40.8, 41.8, 43.4,
              36.5, 39.5, 40.8, 42.0, 43.5,
              37.0, 39.8, 40.9, 42.1, 43.7,
              37.0, 39.8, 41.1, 42.2, 44.0,
```

```
37.7, 40.0, 41.2, 42.1, 44.2,  
38.3, 40.0, 41.3, 42.3, 44.6,  
38.5, 40.3, 41.4, 42.5, 44.9,  
38.7, 40.4, 41.4, 42.5, 45.1,  
39.2, 40.6, 42.6, 42.6, 45.9)
```

Grafik stem-and-leaf dapat dibentuk menggunakan fungsi `stem()` sebagai berikut.

```
> stem(selDarah, scale=2)
```

```
The decimal point is at  
the |
```

```
35 | 23  
36 | 5  
37 | 007  
38 | 357  
39 | 234588  
40 | 003466889  
41 | 1234488  
42 | 011235566  
43 | 1457  
44 | 0269  
45 | 19
```

Grafik *stem-and-leaf* di atas menunjukkan bahwa data menyebar dari 35,2 sampai 45,9 tanpa ada *outlier* (pencilan). Data volume sel per 100 cm³ adalah multimodus (mempunyai lebih dari satu modus), yaitu 37,0; 39,8; 40,0; 40,6; 40,8; 41,4; 41,8; 42,1; 42,5 dan 42,6 sebanyak dua kali kemunculan. Nilai Kuartil kedua (median) adalah 41,8. Kurva frekuensinya menjulur ke kiri yang berarti nilai mean paling kecil dibandingkan dengan nilai median dan modus.

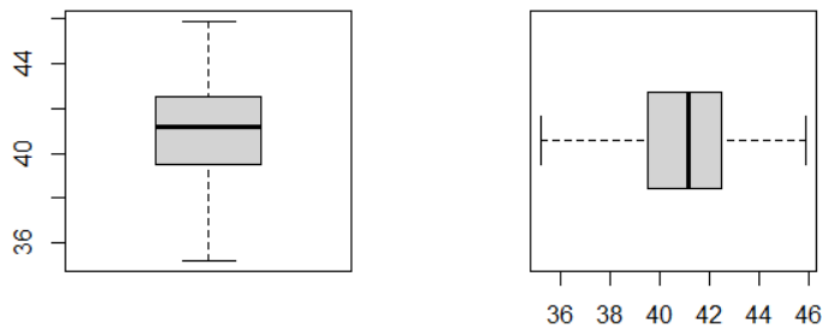
3.5 Penyajian dalam Box Plot

Boxplot merupakan grafik yang dapat memberikan gambaran pusat data, variasi dan distribusi dari sekumpulan data numerik. Untuk membentuk boxplot, perlu diketahui apakah terdapat pengamatan ekstrim (pencilan) pada data, sehingga letaknya akan berada diluar batas nilai minimum dan maksimum data. Namun, jika tidak terdapat pencilan pada data, maka boxplot akan menampilkan data dalam batas nilai minimum dan maksimum saja. Posisi statistik lima serangkai yang terdiri dari nilai minimum, Q_1 , median, Q_3 , dan nilai maksimum dapat divisualisasikan oleh boxplot (Franklin et al., 2017; Weiss, 2016).

Untuk menampilkan boxplot pada R jalankan perintah `boxplot(data, horizontal = FALSE)`. Apabila perintah "`horizontal=TRUE`" digunakan, maka boxplot akan ditampilkan secara horizontal.


```
#Boxplot
par(mfrow=c(1,2))
boxplot(selDarah)
boxplot(selDarah, horizontal=TRUE)
par(mfrow=c(1,1))
```

Berikut adalah boxplot yang dihasilkan.



Gambar 3. Boxplot

Garis tengah pada bagian dalam box di atas menunjukkan letak median dari data. Ujung garis yang paling kanan menunjukkan nilai maksimum sedangkan ujung garis yang paling kiri menunjukkan nilai minimum. Batas paling kiri dari box menunjukkan kuartil pertama (Q1), dan batas paling kanan dari box menunjukkan kuartil ketiga (Q3). Garis perpanjangan dari Q1 ke nilai minimum dan Q3 ke nilai maksimum disebut whisker. Data dianggap simetri jika median berada di tengah kotak dan panjang whisker sama. Boxplot pada Gambar 2-2 mengindikasikan data tidak simetris karena median tidak terletak di tengah kotak dan panjang whisker tidak sama.

3.6 LATIHAN

1. Moore Travel Agency, agen perjalanan berskala nasional, menawarkan tarif khusus pelayaran ke Karibia bagi para warga negara berusia lanjut. Presiden Moore Travel ingin informasi tambahan mengenai usia orang– orang yang mengikuti pelayaran. Sebuah sampel acak berukuran 40 diambil dari mereka yang mengikuti pelayaran tahun lalu, yang usianya seperti berikut ini.

77	18	63	84	38	54	50	59	54	56
36	26	50	34	44	41	58	58	53	51
62	43	52	53	63	62	62	65	61	52
60	60	45	66	83	71	63	58	61	71

Gambarkan hasil penyajian data dalam bentuk boxplot, histogram dan stem and leaf serta interpretasikan hasilnya. Tentukan median dan modus dari data tersebut. Apakah terdapat outliers pada data? Jelaskan!

2. Suatu survei mengenai banyaknya telepon yang diterima sampel pelanggan Southern Phone Company pada minggu lalu, disajikan berikut ini.

52	43	30	38	30	42	12	46
39	37	34	46	32	18	41	5

- Buatlah boxplot, histogram, dan stem and leaf dari data tersebut serta interpretasikan hasilnya.
- Berapakah median dan kuartil dari data tersebut?
- Apakah terdapat outliers pada data? Jelaskan!

DAFTAR PUSTAKA

- Franklin, C. A., Klingenberg, B., & Agresti, A. (2017). *Statistics: The Art and Science of Learning from Data, Global Edition*.
- Pearson, R. K. (2018). *Exploratory Data Analysis Using R*. CRC Press, Taylor & Francis Group.
- Weiss, N. A. (2016). Introductory Statistics. In *Angewandte Chemie International Edition*, 6(11), 951–952. Pearson.