

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(coll = 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(col1 = 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$col1, 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.

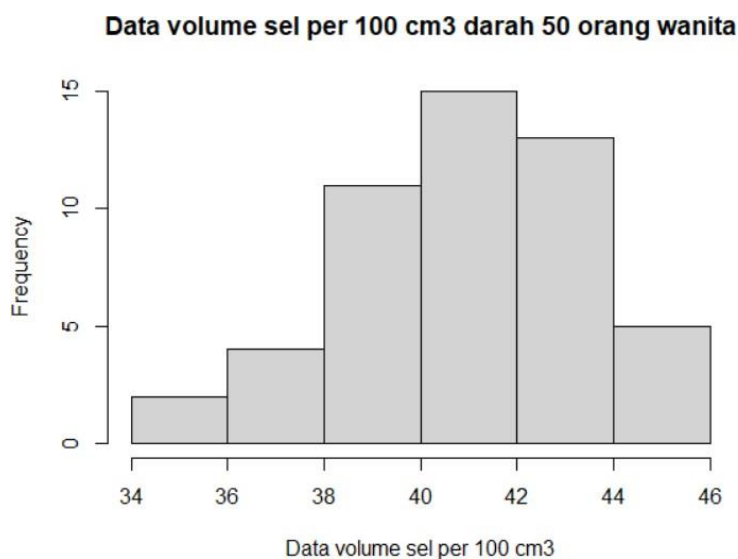
3.2 Penyajian dalam Histogram

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.