

BAB IV

STATISTIKA DESKRIPTIF: UKURAN PEMUSATAN DAN PENYEBARAN DATA

Capaian Pembelajaran

- Mahasiswa memahami definisi mean, median, dan modus.
- Mahasiswa mampu mendapatkan nilai rata-rata, nilai tengah (median), dan nilai modus.
- Mahasiswa mengerti perbedaan penggunaan dan latar belakang dari ketiga nilai tersebut
- Mahasiswa mampu menghitung besarnya keragaman data.

Materi

Statistik merupakan karakteristik dari sampel atau catatan angka-angka yang memiliki makna yang disajikan dalam bentuk tabel, grafik, diagram dan lain sebagainya. Statistika deskriptif dilakukan untuk mendapatkan gambaran atau informasi dari sekumpulan data sampel, namun tidak untuk disimpulkan pada populasi. Statistika deskriptif meliputi dua hal, yaitu ukuran-ukuran numerik dan penyajian data. Pada Bab 4, akan dibahas mengenai ukuran-ukuran numerik yang digunakan untuk mendapatkan informasi, meringkas sekumpulan data, dan menggambarkan kondisi data. Secara umum, terdapat dua ukuran numerik dalam statistika yaitu ukuran pemusatan data dan ukuran penyebaran data. Kedua ukuran numerik ini akan dibahas lebih lanjut pada dua subbagian terpisah.

4.1 Ukuran Pemusatan Data

Ukuran pemusatan data merupakan nilai-nilai yang menggambarkan pusat sekumpulan data. Ukuran pemusatan data terdiri dari rata-rata hitung (mean), median, dan modus. berikut penjelasan dari masing-masing ukuran pemusatan data (Weiss, 2016).

a. Rata-Rata Hitung (Mean)

Rata-rata hitung (mean) merupakan titik berat atau pusat massa (*centroid*) dari suatu data kuantitatif (numerik). Nilai rata-rata hitung sensitif terhadap nilai ekstrim/pencilan (*outlier*), artinya nilai rata-rata dapat berubah drastis jika terdapat nilai pencilan dalam data tersebut. Nilai rata-rata (mean) dapat dihitung dengan formula berikut.

$$\text{Rata-rata hitung populasi} \quad : \quad \mu = \frac{(x_1 + x_2 + \dots + x_N)}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

Rata-rata hitung sampel : $\bar{x} = \frac{(x_1 + x_2 + \dots + x_n)}{n} = \frac{1}{n} \sum_{i=1}^n x_i$

Contoh:

Diberikan data sebagai berikut.

Data I: 1, 4, 6, 9

Nilai rata-ratanya adalah

$$\bar{x}_1 = \frac{(1+4+6+9)}{4} = \frac{20}{4} = 5$$

Data II: 1, 4, 6, 27

Nilai rata-ratanya adalah

$$\bar{x}_2 = \frac{(1+4+6+27)}{4} = \frac{48}{4} = 12$$

Perhatikan perbedaan nilai rata-rata (mean) Data I dan Data II. Hasil perhitungan nilai rata-rata Data II lebih besar diakibatkan oleh adanya pencilan, yaitu nilai 27. Hal ini membuktikan bahwa nilai rata-rata (mean) sangat sensitif terhadap data pencilan.

b. Median

Median merupakan nilai tengah dari sekumpulan data. Median diperoleh dengan mengurutkan data dari nilai yang terkecil sampai nilai terbesar, kemudian membaginya menjadi dua bagian yang sama banyak. Terdapat perbedaan menghitung nilai median berdasarkan banyaknya data (n):

- Jika n **ganjil**, maka mediannya adalah nilai data yang ada ditengah atau nilai data ke $\frac{n+1}{2}$, sehingga median = $Me = x_{\left(\frac{n+1}{2}\right)}$.
- Jika n **genap**, maka mediannya adalah rata-rata dari dua data yang ditengah atau nilai data ke $\frac{n}{2}$, sehingga median = $Me = x_{\left(\frac{n}{2}\right)}$.

c. Modus

Modus merupakan nilai yang paling sering muncul dalam sekelompok data. Modus menunjukkan nilai data yang memiliki frekuensi terbesar. Sekelompok data bisa memiliki lebih dari satu modus atau tidak memiliki modus sama sekali. Misalkan diberi data sebagai berikut.

Data I	: 1, 2, 4, 4, 6, 7, 7, 7, 7	Modus : 7
Data II	: 1, 1, 5, 5, 6, 7, 9, 9, 10	Modus: 1, 5, 9
Data III	: 1, 1, 1, 3, 3, 3, 8, 8, 8	Modus: Tidak ada

4.2 Ukuran Penyebaran Data

Ukuran pemusatan data tidak cukup untuk mendeskripsikan distribusi data. Dibutuhkan ukuran penyebaran data untuk menunjukkan seberapa besar variasi atau perbedaan nilai-nilai dalam suatu kumpulan data. Berikut adalah beberapa ukuran penyebaran data (Franklin, Klingenberg, & Agresti, 2017).

a. Range

Range adalah nilai yang menunjukkan perbedaan antara nilai terbesar dan nilai terkecil. Range merupakan ukuran penyebaran data yang paling kasar dan sensitif karena hanya bersangkutan dengan nilai data terkecil dan nilai data terbesar.

b. Varians dan Standar Deviasi

Deviasi merupakan ukuran penyebaran yang dihitung berdasarkan selisih dari masing-masing nilai data dari nilai rata-rata hitungnya (mean). Deviasi dari suatu observasi x dari mean \bar{x} adalah $(x - \bar{x})$. Deviasi bernilai positif jika nilai observasi berada diatas nilai mean, sedangkan deviasi bernilai positif jika nilai observasi berada dibawah nilai mean.

Jika varians adalah rata-rata deviasi kuadrat dari nilai rata-rata hitung (mean), maka akar dari varians disebut sebagai standar deviasi. Varians dan standar deviasi dari n observasi adalah sebagai berikut:

	Varians	Standar Deviasi
Populasi	$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$	$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$
Sampel	$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$	$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

Semakin besar nilai standar deviasi s , maka semakin besar variabilitas dari suatu data.

c. Kuartil, Desil, dan Persentil

Cara lain untuk mendeskripsikan suatu distribusi data adalah dengan mengukur posisi data. Range menggunakan posisi data minimum dan maksimum, median menunjukkan posisi tengah data, sedangkan varians dan standar deviasi menunjukkan seberapa jauh posisi dari masing-masing data terhadap nilai mean.

Median adalah salah satu kasus khusus dari ukuran posisi lainnya yang disebut kuartil, desil, persentil. Jika median membagi data menjadi dua bagian, maka kuartil membagi data menjadi empat bagian, desil membagi data menjadi sepuluh bagian, dan persentil membagi data menjadi seratus bagian.

Misalkan diberi data sebagai berikut

Q_1			Q_2			Q_3		
10%	10%	10%	10%	10%	10%	10%	10%	10%
D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9

Kuartil membagi data menjadi empat bagian, sehingga setiap $\frac{1}{4}$ bagian mewakili 25% dari data. Desil membagi data menjadi 10 bagian, sehingga setiap $\frac{1}{10}$ bagian mewakili 10% dari data. Kemudian untuk persentil, setiap $\frac{1}{100}$ mewakili 1% dari data. Kuartil ke-2 (Q_2), Desil ke-5 (D_5), dan Persentil ke 50 (P_{50}) merupakan median dari sekumpulan data.

4.3 Aplikasi Menggunakan R

Statistik deskriptif lebih berkenaan dengan pengumpulan dan peringkasan data, serta penyajian hasil peringkasan tersebut. Data-data statistik yang diperoleh hasil sensus, survei, jajak pendapat atau pengamatan lainnya umumnya masih bersifat acak, mentah dan tidak terorganisir dengan baik. Datadata tersebut harus diringkaskan dengan baik dan teratur, baik dalam bentuk table atau presentasi grafis yang berguna sebagai dasar dalam proses pengambilan keputusan.

Penyajian tabel dan grafis yang digunakan dalam statistik deskriptif dapat berupa: (1) distribusi frekuensi, (2) presentasi grafis seperti histogram, pie chart dan sebagainya. Selain tabel dan grafik, untuk mengetahui deskripsi data diperlukan ukuran yang lebih eksak, yang biasa disebut summary statistics (ringkasan statistik).

Dua ukuran penting yang sering dipakai dalam pengambilan keputusan adalah: (1) mencari central tendency (kecenderungan memusat), seperti mean, median, dan modus, (2) mencari ukuran dispersi, seperti standar deviasi dan varians. Selain *central tendency* dan dispersi ukuran lain yang dipakai adalah Skewness dan Kurtosis yang berfungsi untuk mengetahui kemiringan data (gradien data). Selain itu data yang diambil tergantung dari mana data tersebut diperoleh, hal ini akan berkaitan dengan ukuran statistik, yaitu parameter, dan statistik.

Ukuran statistik adalah bilangan yang diperoleh dari sekumpulan data statistik melalui proses aritmatik tertentu. Dalam analisis data, ukuran statistik ini mengisyaratkan gejala spesifik, misalnya kecenderungan letak di pusat data, pengelompokan data,

penyebaran/variasi/ keseragaman data, atau bentuk lainnya yang dikandung oleh data yang sedang dianalisis.

Apabila ukuran statistik ini diperolehnya atas dasar perhitungan yang menyeluruh (*complete enumeration*) atau sensus, maka namanya parameter, sedangkan jika diperolehnya atas dasar perhitungan terhadap data statistik yang ada dalam sampel, ukuran statistik ini disebut statistik. Sebagai ilustrasi perhatikan contoh berikut ini, ada 40 orang karyawan yang sedang diselidiki besarnya upah bulanan dalam ribuan rupiah. Adapun hasilnya adalah sebagai berikut:

Tabel 2. Data upah karyawan

146	156	119	138	147
157	125	140	147	158
126	140	148	161	128
142	149	163	132	142
150	164	135	144	150
165	135	144	152	168
135	145	153	173	136
145	154	176	138	146

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

```
Karyawan<-c(146, 156, 119, 138, 147,  
            157, 125, 140, 147, 158,  
            126, 140, 148, 161, 128,  
            142, 149, 163, 132, 142,  
            150, 164, 135, 144, 150,  
            165, 135, 144, 152, 168,  
            135, 145, 153, 173, 136,  
            145, 154, 176, 138, 146)
```

Kemudian hitung ukuran pemusatan data dengan perintah sebagai berikut:

```
>#rata-rata hitung (mean)  
> mean(Karyawan)  
[1] 146.8  
  
> #median  
> median(Karyawan)  
[1] 146
```

```

> #range
> range(Karyawan)
[1] 119 176
> #varians
> var(Karyawan)
[1] 170.3179

```

Hitung ukuran penyebaran data dengan perintah sebagai berikut:

```

> #standar deviasi
> sd(Karyawan)
[1] 13.05059

> #kuartil
> quantile(Karyawan)
 0%   25%   50%   75%  100%
119.0 138.0 146.0 154.5 176.0

> #desil
> quantile(Karyawan, probs = seq(0.1,0.9,by=0.1))
 10%   20%   30%   40%   50%   60%   70%   80%   90%
131.6 135.8 140.0 144.0 146.0 148.4 152.3 157.2 164.1

> #persentil ke-12, 37, 63, dan 88
> quantile(Karyawan, probs = c(0.12,0.37,0.63,0.88))
 12%   37%   63%   88%
134.04 142.86 149.57 163.32

```

Secara umum, perintah `summary(data)` akan memuat semua informasi numerik statistika deskriptif yang terdiri dari nilai minimum, kuartil ke-1 (Q_1), median(Q_2), mean, kuartil ke-3 (Q_3), dan nilai maksimum. Kelima ukuran ini dikenal sebagai statistik lima serangkai.

```

> summary(Karyawan)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
119.0   138.0   146.0   146.8   154.5   176.0

```

Interpretasi

Pada output yang diberikan dapat dilihat bahwa dari 40 orang karyawan, rata-rata besarnya upah bulanan yang diperoleh adalah Rp 146.800,00 dengan standar deviasi sebesar Rp

13.050,00. Adapun upah minimum dan maksimum yang diperoleh masing-masing adalah Rp 119.000,00 dan Rp 176.000,00 dengan jangkauan Rp 57.000,00.

4.3 LATIHAN

1. Menurut survei yang dilakukan sebuah lembaga penelitian dilaporkan bahwa gaji seorang karyawan pada saat masuk untuk tingkat sarjana berkisar antara Rp.500.000 ribu sampai Rp.800.000,- per bulan. Diambil sampel sebanyak 25 orang dan diperoleh data sebagai berikut (dalam ribuan rupiah)

700	600	725	500	770
750	525	690	770	780
800	575	680	700	650
785	800	580	695	650
650	750	550	750	700

- a. Tentukan mean, median, dan modus dari data tersebut?
 - b. Tentukan range, varians, standar deviasi, Q_3 , D_4 , dan P_{70} dari data tersebut?
 - c. Buatlah summary dari data tersebut.
2. Tingkat aliran air maksimal dari 34 kepala keran air yang berbeda yang dilaporkan dalam salah satu artikel majalah Consumer Reports (Juli 1990) adalah sebagai berikut:

2.9	2.7	1.8	3.4	2.8	2.8
2.7	2.6	2.5	2.0	2.5	2.7
2.6	3.6	2.8	4.7	2.7	2.7
2.2	2.8	2.4	2.5	2.5	2.7
2.5	2.6	2.5	3.1	5.4	2.9
2.8	2.9	4.9			

Dari data di atas, tentukan:

- a. Ukuran pemusatan data (mean, median, maksimum dan minimum)
 - b. Kuartil bawah, kuartil tengah dan kuartil atas
 - c. Ringkasan statistik

DAFTAR PUSTAKA

- Franklin, C. A., Klingenberg, B., & Agresti, A. (2017). *Statistics: The Art and Science of Learning from Data, Global Edition*.
- Weiss, N. A. (2016). Introductory Statistics. In *Angewandte Chemie International Edition*, 6(11), 951–952. Pearson.