BAB 4

UKURAN PENYEBARAN DATA

Pada Bab sebelumnya kita telah mempelajari beberapa ukuran pemusatan data, yaitu ukuran yang memberikan informasi tentang bagaimana data-data ini mengumpul atau memusat. Pada bagian Bab ini kita akan mempelajari ukuran penyebaran data. Ukuran ini digunakan untuk mengetahui variasi atau dispersi data, yaitu derajat penyebaran data terhadap rata-rata. Diantara ukuran penyebaran yang sering digunakan adalah range, rata-rata deviasi, range semi-interkuartil, range percentil 10-90, dan standar deviasi.

4.1 Range

Range atau jangkauan suatu kelompok data didefinisikan sebagai selisih antara nilai terbesar dan nilai terkecil, yaitu

$$\text{range} := \text{nilai terbesar} - \text{nilai terkecil}.$$

Contoh 4.1.1. Range atau jangkauan kumpulan data 2, 3, 3, 5, 5, 5, 8, 10, 12 adalah 12- 2 = 10. Di sini 12 adalah nilai terbesar dan 2 adalah nilai terkecil.

Untuk data dalam bentuk tabel distribusi frekuensi, range dihitung dengan dua pendekatan berikut, yaitu:

a. selisih antara tanda kelas pada interval kelas teratas dan tanda kelas pada inteval kelas terendah.

b. selih antara batas atas kelas pada interval kelas teratas dan batas bawah pada inteval kelas terendah.

Contoh 4.1.2. Diperhatikan tabel distribusi berikut.

Kelas	Tanda Kelas (X)	Frekuensi (f)
60-62	61	5
63-65	64	18
66-68	67	42
69-71	70	27
72-74	73	8
		$\sum f = 100$

Range dihitung berdasarkan kedua pendekatan ini adalah:

a. Metoda 1: range = 73 - 61 = 12.

b. Metoda 2: range = 74.5 - 59.5 = 16.

4.2 Rata-rata deviasi

Pada Bab III telah didefinisikan deviasi atau penyimpangan suatu data terhadap rata-rata. Jumlah semua deviasi data terhadap rata-rata bernilai nol. Pada bagian ini simpangan data terhadap rata-rata diambil sebagai nilai mutlaknya, yaitu $|X_j - \bar{X}|$. Jadi simpangan mutlaknya tidak pernah negatif. Selanjutnya, rata-rata deviasi didefinisikan sebagai

$$\text{rata-rata deviasi (RD)} := \frac{1}{N}\sum_{j=1}^{n} |X_j - \bar{X}|.$$

Contoh 4.2.1. Diketahui data 2, 3, 6, 8, 11. Maka rata-rata deviasinya dihitung sebagi berikut:

$$\overline{x} = \frac{1}{5}(2 + 3 + 6 + 8 + 11) = 6.$$

Selanjutnya,

$$\text{RD} = \frac{|2-6| + |3-6| + |6-6| + |8-6| + |11-6|}{5} = \frac{4+3+0+2+5}{5} = 2.8.$$

Bila pada kumpulan data tersebut terdapat f_1 data bernilai X_1, sebanyak f_2 bernilai X_2, dan seterusnya terdapat f_k bernilai X_K maka rata-rata deviasi dapat dihitung dengan cara berikut:

$$RD = \frac{1}{N}\sum_{j=1}^{K} f_j \cdot |X_j - \bar{X}|.$$

4.3 Deviasi kuartil

Deviasi kuartil atau rentang semi-interkuartil didefinisikan sebagai

$$DK := \frac{Q_3 - Q1}{2}$$

dimana Q_1 dan Q_2 masing-masing kuarti ke 1 dan kuartil ke 3. Ada juga ukuran penyebaran yang disebut rentang persentil 10-90, yaitu

$$\text{RP10-90} := P_{90} - P_{10}.$$

4.4 Deviasi standar dan variansi

Inilah kedua ukuran penyebaran data yang paling sering digunakan dan paling jelas interpretasinya. Deviasi standar atau simpakan baku didefinisikan sebagai

$$S := \sqrt{\frac{\sum_{j=1}^{n}(X_i - \overline{X})^2}{n}}.$$

Kadangkala, simpangan baku didefinisikan dengan menggunakan pembagi $n-1$ bukan n seperti diatas, yaitu:

$$S := \sqrt{\frac{\sum_{j=1}^{n}(X_i - \overline{X})^2}{n-1}}.$$

Formula kedua digunakan untuk sampel, sdengakan formula sebelumnya digunakan untuk populasi atau sampel yang berukuran cukup besar. Penjelasan teoritis mengenai kedua formula ini akan dibahas pada statistika matematika. Kuadrat dari simpangan baku disebut variansi. Jadi variansi didefinisikan sebagai:

$$\text{var} := S^2 = \frac{\sum_{j=1}^{n}(X_i - \overline{X})^2}{n}.$$

Argumen yang sama untuk pembaginya menggunakan $n-1$. Dengan mudah dapat dibuktikan bahwa

$$S^2 = \overline{X^2} - (\overline{X})^2.$$

4.5 Sifat-sifat deviasi standar

1. Jumlah deviasi kuadrat terhadap bilangan a didefinisikan sebagai

$$\sum_{j=1}^{n}(X_i - a)^2.$$

   Jumlah ini akan menjadi minimum jika diambil $a := \overline{X}$. Jadi deviasi standar $s := \sqrt{\frac{\sum_{j=1}^{n}(X_i - \overline{X})^2}{n}}$ merupakan akar dari rata-rata jumlah kuadrat deviasi (simpangan) minimum.

2. Untuk data berdistribusi normal, deviasi standar dapat digunakan untuk mengetahui penyebaran data, yaitu:

Gambar 4.1: Jarak 1 deviasi standar dari mean

Gambar 4.2: Jarak 2 deviasi standar dari mean

a. 68.27% data tersebar diantara $\overline{X} - s$ dan $\overline{X} + s$.

b. 95.45% data tersebar diantara $\overline{X} - 2s$ dan $\overline{X} + 2s$.

c. 99.73% data tersebar diantara $\overline{X} - 3s$ dan $\overline{X} + 3s$. Terlihat pada Gam-

Gambar 4.3: Jarak 3 deviasi standar dari mean

bar 4.3 bahwa jika suatu kumpulan data membentuk distribusi normal maka hampir semua data tersebar pada rentang 6 deviasi standar.

3. Bila ada dua kelompok data, masing-masing berukuran n_1 dan n_2 dengan variansi S_1^2 dan S_2^2 dan diasumsikan keduanya mempunyai rata-rata yang

sama maka variansi gabungan kedua kelompok data ini diberikan oleh

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}.$$

4. Teorema Chebyshev mengatakan bahwa untuk $k > 1$, paling tidak ada $(1 - \frac{1}{k^2}) \times 100\%$ data terletak diantara $\overline{X} - ks$ dan $\overline{X} + ks$. Untuk $k = 2$ maka lebih dari$(1 - \frac{1}{4}) \times 100\% = 75\%$ data terletak di dalam interval $[\overline{X} - 2s, \overline{X} + ks]$. Pada teorema ini tidak disyaratkan bahwa data harus berdistribusi normal seperti pada sifat 2 di atas.

Untuk data yang telah tersaji dalam distribusi frekuensi, maka deviasi standarnya dihitung berdasarkan rumus:

$$s := \sqrt{\frac{\sum fX^2}{n} - \left(\frac{\sum fX}{n}\right)^2}$$

dimana X adalah tanda kelas.

Contoh 4.5.1. Diperhatikan data

```
32  113 70  60  84  114  31  58  86  102  113  79  86  24  40  44
42  54  71  25  42  116  68  30  63  121  74   77  77  100 51  31
61  28  26  47  54  74   57  35  77  80   125  105 61  102 45  115
36  52  58  24  24  39   40  95  99  54   35   31  77  29  69  58
32  49 118  44  95  65   71  65  74  122  99.
```

Dengan menggunakan perintah pada excel maka diperoleh beberapa ukuran penyebaran berikut:

```
rata-rata deviasi(AVEDEV)=24.42
deviasi standar(STDEV)=29.2563
variansi sampel(VAR)=855.932
variansi populasi(VARP)=844.52
```

Untuk melihat bagaimana pola data ini menyebar, kita buat tabel distribusi dan histogram dengan ukuran kelas $S \approx 30$ dan $\overline{X} \approx 66$. Diperoleh Tabel dan gambar berikut.

Kelas	Frekuensi (f)
6-36	16
37-66	25
67-96	19
97-126	15
total	75

Gambar 4.4: Histogram dengan ukuran kelas deviasi standar

Diperhatikan banyaknya data yang terletak diantara $\overline{X} - S$ dan $\overline{X} + S$ adalah sebanyak $25 + 19 = 44$ data atau sekitar 58% data. Kemudian, semua data terletak diantara $\overline{X} - 2S$ dan $\overline{X} + 2S$. Fakta ini tidak bertentangan dengan sifat 2 diatas karena data kita tidak persis normal. Sedangkan teorema Chebyshev terpenuhi. Satu lagi informasi yang dapat diketahui melalui deviasi standar adalah semakin kecil deviasi standar semakin banyak data yang mengumpul di sekitar rata-rata.

4.6 Skor Z atau nilai standar

Misalkan kumpulan data X_i mempunyai rata-rata $\overline{X}$ dan deviasi standar S maka Z_i yang didefinisikan sebagai

$$Z_i := \frac{X_i - \overline{X}}{S}$$

disebut **skor** Z atau **nilai standar** dari X_i.

Contoh 4.6.1. Diperhatikan kembali data pada contoh di atas. Nilai standar dari data 70 adalah $\frac{70-65.65}{29.253} = 0.1486$. Semakin dekat nilai data ke nilai rata-ratanya maka semakin dekat ke nol nilai standarnya. Dalam bebebrapa kebutuhan analisis data diperlukan untuk mentransformasi data ke dalam nilai standarnya.

Soal-soal Latihan

1. Untuk data pada contoh 4.5.1, tentukan banyak data yang terletak pada rentang:

   a. $\overline{X} \pm RD$
   b. $\overline{X} \pm DK$

2. Soal latihan 4.4.
3. Soal latihan 4.11.
4. Soal latihan 4.12.
5. Soal latihan 4.14.
6. Soal latihan 4.17.
7. Soal latihan 4.42.
8. Soal latihan 4.57.
9. Soal latihan 4.63.
10. Soal latihan 4.82.