

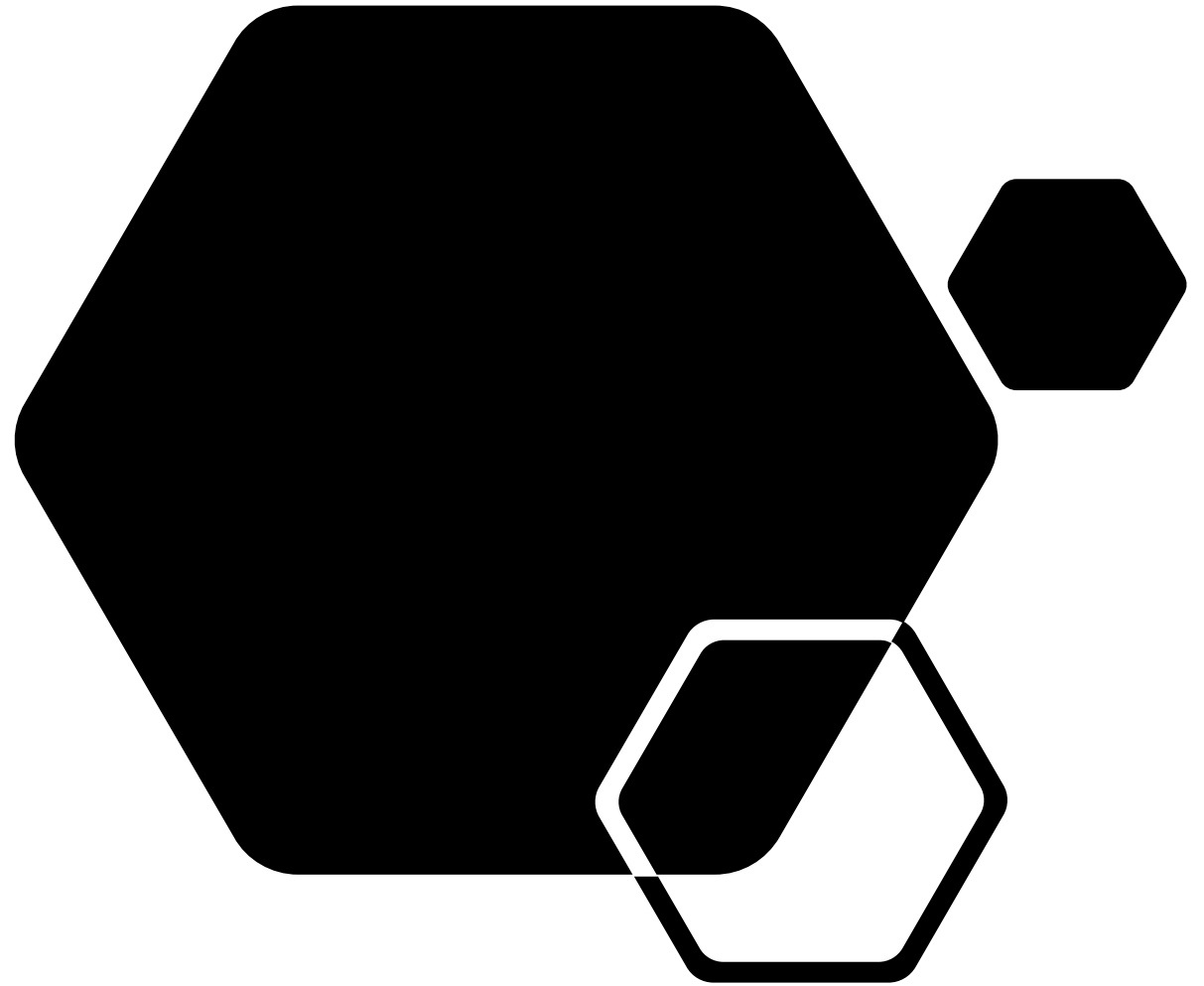
Analisis Data Eksploratif

Pertemuan 3

Persebaran Data

Ringkasan Numerik

Ukuran Persebaran Data



Ukuran Penyebaran Data

Diberikan hasil pengukuran 2 contoh Jus Jeruk (dalam Liter) yang dikemas dalam botol oleh Perusahaan A dan Perusahaan B.

| | | | | | |
|-------|------|------|------|------|------|
| JUS A | 0,97 | 1,00 | 0,94 | 1,03 | 1,06 |
| JUS B | 1,06 | 1,01 | 0,88 | 0,91 | 1,14 |

- Diperoleh rata-rata masing – masing $\bar{x}_A = 1,00$ dan $\bar{x}_B = 1,00$
- Nampak dari table tersebut bahwa isi jus jeruk A “lebih seragam” dibanding jus jeruk B.

Artinya jika harus memilih antara A dan B, mana yang harus dipilih? Kenapa?

Range / Jangkauan/ Wilayah

$$RANGE = X_{max} - X_{min}$$

Contoh :

1. Nilai IQ lima orang anggota keluarga : 108, 112, 127, 118, 113. Maka Range-nya adalah 19.
2. Dari contoh pengemasan Jus Jeruk A dan B sebelumnya, diperoleh Range A = 0,12 dan Range B=0,26. Artinya, data B “lebih menyebar” dibanding data A.

Lanjutan....

- Range bukan ukuran keragaman yang baik terutama apabila ukuran populasinya besar.
- Range memperhatikan nilai-nilai ekstrem tetapi tidak mengatakan apapun untuk sebaran data diantara nilai-nilai ekstrem tersebut.

| | | | | | | | | | |
|---------|---|---|---|---|---|---|----|----|----|
| Gugus A | 3 | 4 | 5 | 6 | 8 | 9 | 10 | 12 | 15 |
| Gugus B | 3 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 15 |

- Gugus A & B mempunyai $\bar{X} = Me = 8$ tetapi data A sangat variatif dari 3-15, sedangkan data B datanya cenderung terletak di sekitar pusat datanya.

Sifat Range

1. Jika dari observasi x_i dibentuk observasi baru $y_i = x_i \pm a$ dan a bilangan positif sebarang, maka $Range(y_i) = Range(x_i)$.
2. Jika dari observasi x_i dibentuk observasi baru z_i dengan $z_i = bx_i$ dan b sebarang bilangan real maka $Range(z_i) = b \cdot Range(x_i)$

Sebaran Tengah

- Bagian tengah data dapat digunakan untuk mengetahui sebaran data.
- Sebaran tengah tidak terpengaruh oleh nilai ekstrim sehingga dapat dikatakan sebagai sebaran yang Robust (Tangguh).
- Rumus:

$$d_q(x) = q_A - q_B$$

Dengan q_A dan q_B berturut-turut menyatakan kuartil atas dan bawah.

Contoh

Diberikan data observasi sebagai berikut

217, 136, 295, 304, 428, 273, 261

Jika diurutkan diperoleh

136, 217, 261, 273, 295, 304, 428

Diperoleh

$$q_B = \text{data ke-} \frac{(n+1)}{4} = \text{data ke-2} = 217$$

$$q_A = \text{data ke-} \frac{3(7+1)}{4} = \text{data ke-6} = 304$$

Sehingga diperoleh $d_q(x) = 304 - 217 = 87$

- Sebaran tengah merupakan sebaran yang penting dan sering digunakan dalam eksplorasi data
- Hal ini disebabkan karena sebaran tengah mudah untuk diperoleh dan Tangguh terhadap outlier

Ragam / Variansi

- Diberikan data populasi sebanyak N data $x_1, x_2, x_3, \dots, x_N$
- Dengan rata-rata populasinya μ maka variansi dari data populasi tersebut didefinisikan dengan

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

Contoh...

Diberikan data sebagai berikut :

| | | | | | | | | | |
|---------|---|---|---|---|---|---|----|----|----|
| Gugus A | 3 | 4 | 5 | 6 | 8 | 9 | 10 | 12 | 15 |
| Gugus B | 3 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 15 |

Diperoleh $\sigma_A^2 = \frac{124}{9}$ dan $\sigma_B^2 = \frac{78}{9}$

Karena $\sigma_A^2 > \sigma_B^2$ artinya data A lebih beragam daripada data B.

Simpangan Baku/Standar Deviasi

$$SD = \sigma = \sqrt{\sigma^2}$$

- Contoh :

Dengan contoh sebelumnya, didapat simpangan baku untuk gugus

A dan gugus B berturut-turut adalah $\frac{2}{3}\sqrt{31}$ dan $\frac{2}{3}\sqrt{17}$

Ragam/Variansi Sampel

- Diberikan data $x_1, x_2, x_3, \dots, x_n$ maka variansi sampelnya adalah

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Dengan \bar{x} : *rata – rata sampel*

Contoh....

Harga 1 bungkus kopi instan dengan berat 200 gram di 4 toko kelontong yang dipilih secara acak di San Diego menunjukkan bahwa kenaikan harga dari bulan sebelumnya yaitu 12, 15, 17, 20 sen. Tentukan ragamnya!

Jawab :

$$\bar{x} = \frac{12 + 15 + 17 + 20}{4} = 16 \text{ sen}$$

$$s^2 = \frac{\sum_{i=1}^4 (x_i - 16)^2}{4 - 1} = \frac{(12 - 16)^2 + (15 - 16)^2 + (17 - 16)^2 + (20 - 16)^2}{3} = \frac{34}{3}$$

Sifat Variansi

Jika \bar{x} adalah hasil pembulatan suatu bilangan decimal maka akan diperoleh eror yang cukup besar sehingga perlu diberikan rumus yang baru yang tidak melibatkan penghitungan \bar{x} yaitu

$$s^2 = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

Sifat Variansi

1. Jika dari observasi x_i dibentuk observasi baru $y_i = x_i \pm a$, $a > 0$, dan \bar{x} menyatakan rata-rata observasi x_i dan $y_i = \bar{y} \pm a$ maka $Var(y) = Var(x)$
2. Jika dari observasi x_i dibentuk observasi baru $y_i = x_i \pm a$, $a > 0$, dan \bar{x} menyatakan rata-rata observasi x_i dan $y_i = \bar{y} \pm a$ maka

$$Var(y) = b \cdot Var(x)$$

Note Untuk Variansi

- Variansi memiliki kelebihan karena Seluruh data digunakan dalam penghitungan
- Variansi tidak menarik (secara eksploratif) karena terpengaruh oleh data outlier/ekstrim

Contoh

Tentukan ragam untuk data 3,4, 5, 6, 6, 7 yang merupakan banyaknya ikan trout yang tertangkap oleh 6 orang nelayan yang diambil secara acak pada 19 Juni 1981 di suatu Danau.

Jawab :

| Banyak ikan (x_i) | Nilai x_i^2 |
|-----------------------|---------------|
| 3 | 9 |
| 4 | 16 |
| 5 | 25 |
| 6 | 36 |
| 6 | 36 |
| 7 | 49 |
| Jumlah = 31 | Jumlah = 171 |

$$\begin{aligned} S^2 &= \frac{6 \cdot \sum_{i=1}^6 x_i^2 - (\sum_{i=1}^6 x_i)^2}{6(6-1)} \\ &= \frac{6 \cdot 171 - 31^2}{6.5} = \frac{65}{30} \\ &= 2,17 \end{aligned}$$

Simpangan Rata-rata / Deviasi Rata-rata

$$SR = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Contoh ;

Tentukan simpangan rata-rata dari data berikut : 7,6,8,7,6,10,5

Jawab :

$$\bar{x} = \frac{49}{7} = 7$$

$$SR = \frac{|7 - 7| + |6 - 7| + |8 - 7| + |7 - 7| + |6 - 7| + |10 - 7| + |5 - 7|}{7} = \frac{8}{7}$$