

Basic Statistics





Agenda hari ini

1. **Mengapa statistics itu penting**
2. **Populasi, sample, parameter, statistics**
3. **Tipe data dalam Statistics**
4. **Frekuensi dan histogram**
5. **Descriptive dan Inferential Statistics**
6. **Komponen dalam Descriptive Statistics**
7. **Usecase dan Praktek**



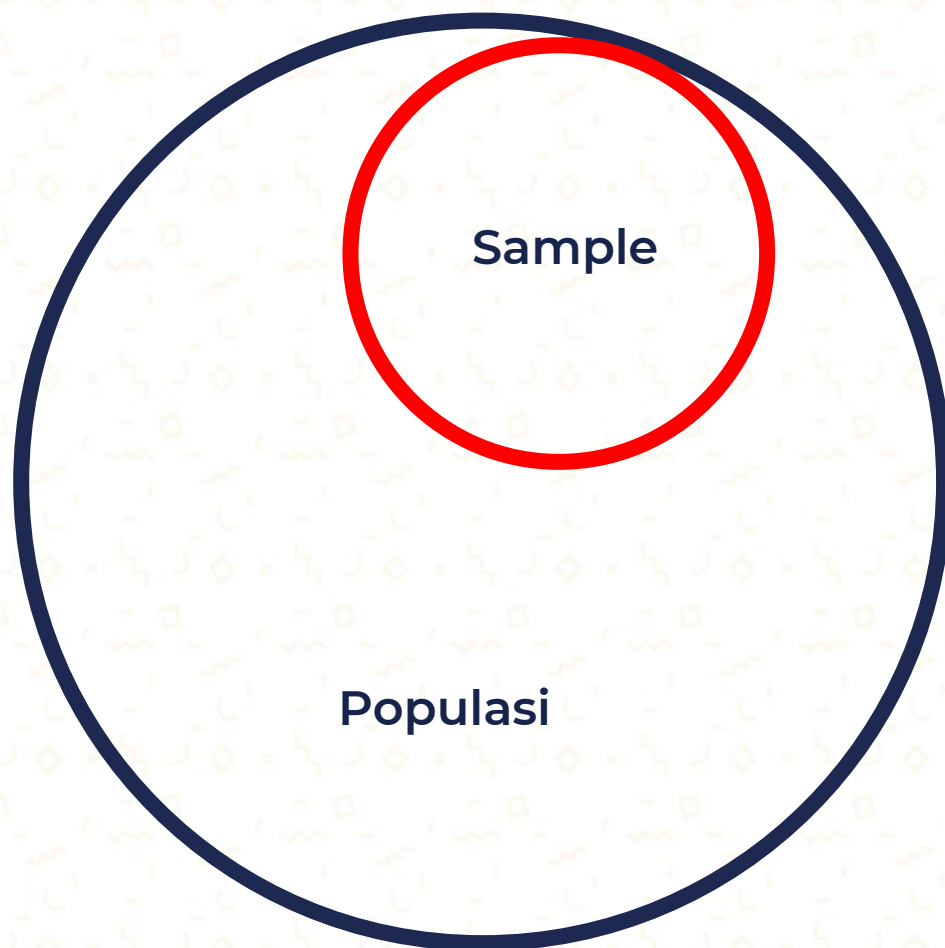
Mengapa statistics ini penting?



- *Data is everywhere*
- Bagaimana cara kita menginterpretasi data dalam jumlah besar dan membuat summary informasi yang bermakna?
- Di era informasi, data dibutuhkan dalam decision making, dan decision making terjadi dalam frekuensi yang sangat cepat dalam dunia bisnis
- Pemahaman terhadap statistics membantu dalam menilai suatu informasi apakah benar, salah, *make sense* atau tidak



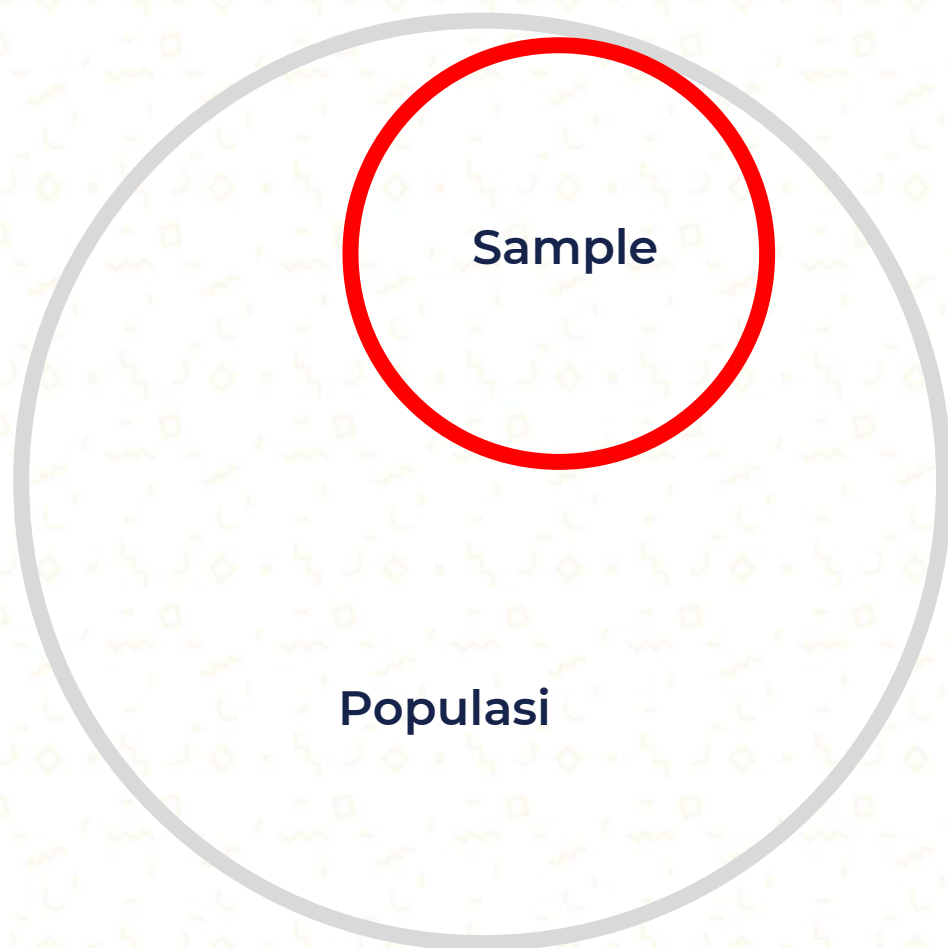
Populasi, sample, parameter, statistics



- Studi dalam Statistics erat kaitannya dengan studi terhadap sejumlah sample dari populasi
- Sample adalah bagian dari populasi
- Populasi menggambarkan keseluruhan anggota elemen yang menjadi observasi kita
- **Mengapa mempelajari sample?**



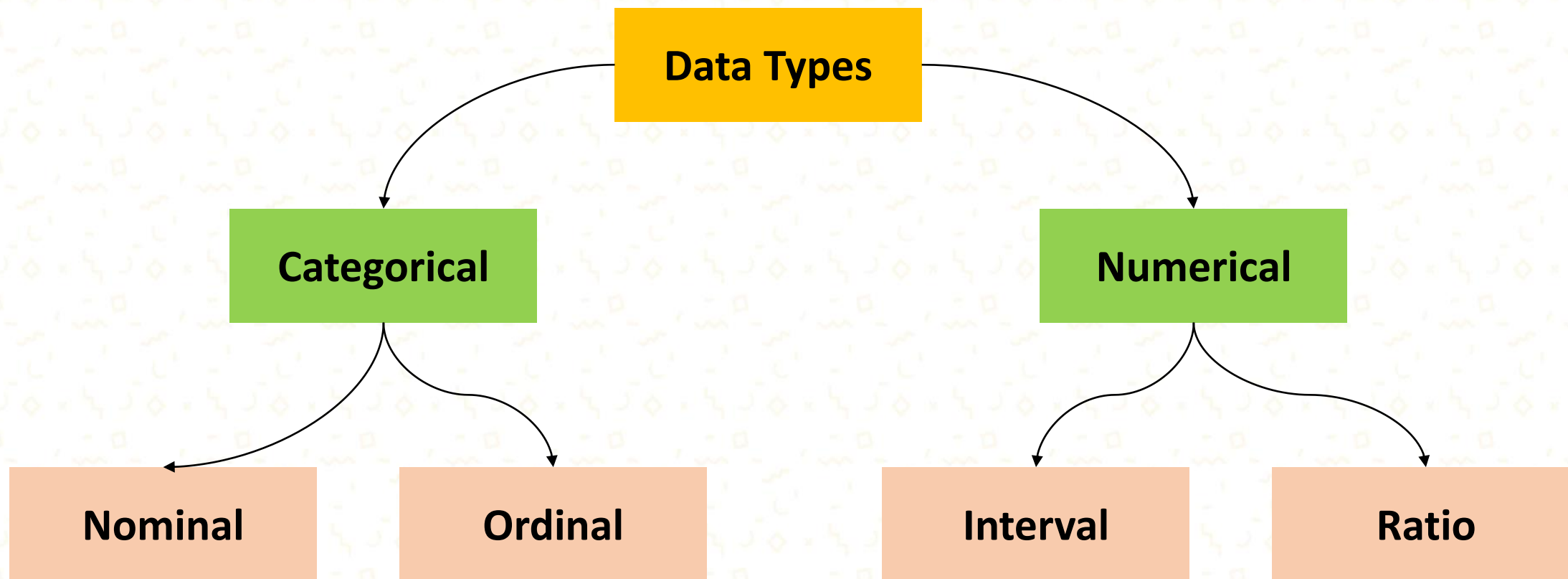
Populasi, sample, parameter, statistics



- Pada sample yang kita observasi, kita dapat mengukur *property* dari sample tersebut. Property ini disebut **Statistics** dari sample.
- Terkadang kita ingin mendapatkan statistics untuk keseluruhan Populasi. Hal ini disebut sebagai **Parameter**.
- Angka statistics digunakan sebagai **estimate** terhadap parameter, yang terkadang tidak diketahui nilainya



Tipe data dalam Statistics





Tipe data: Nominal dan Ordinal

- Tipe data **nominal** adalah data yang direpresentasikan sebagai kumpulan event atau obyek dalam kategori yang bersifat diskrit.
- Contoh data nominal :
 - Nama TV show di sebuah channel televisi
 - Jenis makanan di sebuah restoran
 - Kumpulan nama bank di Indonesia
- Tipe data **ordinal** mirip dengan data nominal, namun perbedaannya adalah data tersebut dapat diurutkan
- Contoh data ordinal :
 - Tingkat kepuasan terhadap pelayanan : Bad, Fair, Good, Excellent
 - Tingkat Pendidikan : SD, SMP, SMA, Sarjana



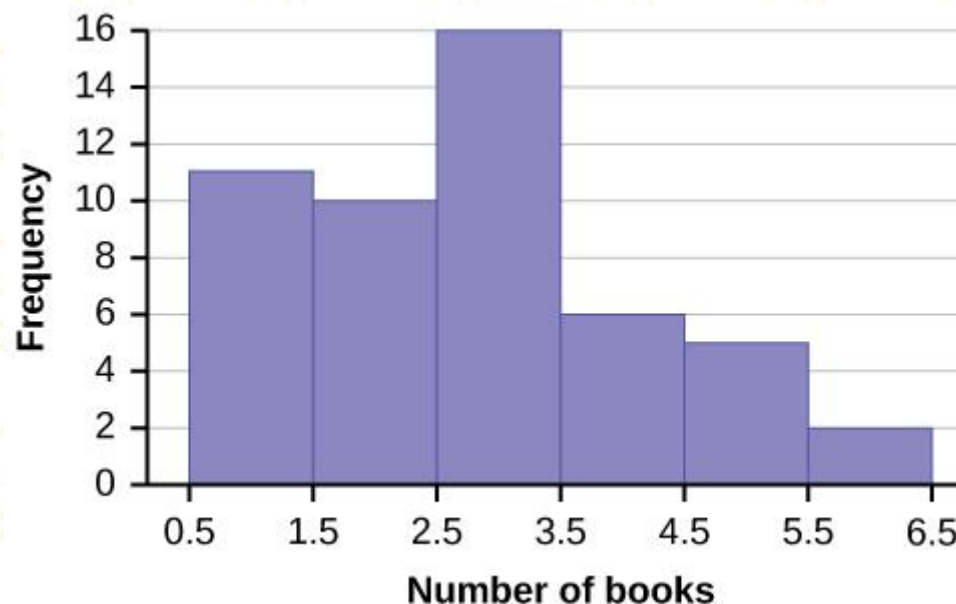
Tipe data: Interval dan Ratio

- Tipe data **interval** adalah data numerik yang tidak memiliki nilai true zero, artinya nilai bisa mencapai di bawah zero (negative).
- Contoh data interval :
 - Temperature
- Tipe data **ratio** mirip dengan data interval, namun data ini memiliki angka true zero.
- Contoh data ratio :
 - Usia
 - Income dalam rupiah
 - Tinggi badan



Frekuensi dan histogram

| Nilai ujian | Jumlah siswa |
|-------------|--------------|
| 10 | 5 |
| 20 | 3 |
| 30 | 2 |
| 40 | 1 |
| 50 | 5 |
| 60 | 3 |
| 70 | 2 |



- Frekuensi dan histogram menggambarkan distribusi summary statistic yang melakukan grouping elemen dalam nilai yang sama
- Histogram membantu kita dalam melihat distribusi nilai secara summary dengan lebih mudah dibandingkan melihat dalam bentuk tabel



Pembagian domain statistics

Descriptive statistics

- Merupakan sebuah metode untuk mendeskripsikan dan menampilkan data secara summary dalam bentuk visual

Inferential statistics

- Dalam study statistic, seringkali cukup sulit untuk menentukan parameter dari sebuah populasi
- Inferential statistics adalah study untuk melakukan generalisasi terhadap suatu sample yang mewakili sebuah populasi



Descriptive statistics:

Komponen utama dalam Descriptive Statistics

- Pengukuran terhadap tendency nilai tengah (Measures of Central Tendency)
- Pengukuran variabilitas dalam data (Measures of Spread)

Measures of Central Tendency

- Mean
- Mode
- Median

Measures of Spread

- Range
- Min-max
- Quartiles/percentiles
- Standard deviation/variance



Descriptive statistics: Mean

- Metric yang paling populer digunakan dalam descriptive statistics
- Namun juga perlu kehati-hatian dalam interpretasi nilainya
- Secara matematis :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \bar{x} = \frac{\sum x}{n}$$

Bagaimana Mean bisa memberikan informasi yang *misleading*?

- **Mean sensitive terhadap outlier**

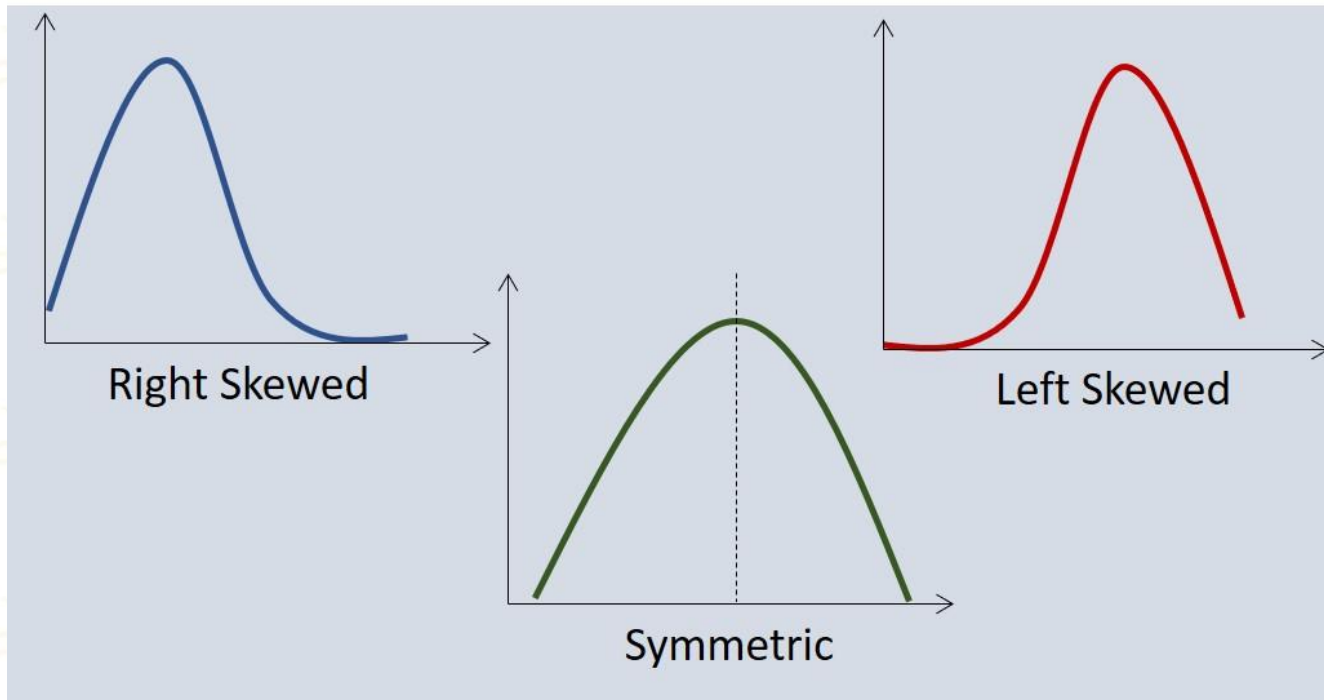
| Siswa | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------|----|----|----|----|----|----|----|----|----|
| Usia | 15 | 18 | 16 | 14 | 15 | 12 | 17 | 60 | 70 |

Mean = 26.33 (reasonable kah?)



Descriptive statistics: Mean

- Mean juga sensitive ketika data yang diberikan skewed, atau cenderung tidak memiliki distribusi normal
- Semakin skewed, mean bisa jadi kehilangan kemampuan untuk memberikan gambaran nilai tengah dari suatu data





Descriptive statistics: Median

- Median dapat membantu menyelesaikan isu representasi data nilai tengah dengan Mean ketika terdapat outlier
- Median diukur dengan cara :
 - Urutkan elemen numeric dari urutan terkecil s/d terbesar
 - Tentukan banyaknya elemen dalam data tersebut, misal ukuran datanya sebesar **n**
 - Apabila N ganjil, maka median adalah :

$$\text{Median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ observation}$$

- Apabila N genap, maka median adalah :

$$\text{Median} = \frac{\frac{n}{2}^{\text{th}} \text{ obs.} + \left(\frac{n}{2} + 1 \right)^{\text{th}} \text{ obs.}}{2}$$



Descriptive statistics: Median

- Ilustrasi untuk $N = \text{ganjil}$

| | | | | | | |
|----|----|----|----|----|----|----|
| 32 | 35 | 40 | 42 | 50 | 58 | 60 |
|----|----|----|----|----|----|----|

- Median = elemen ke $(7+1)/2 = 4$. Elemen ke-empat adalah **42**

- Ilustrasi untuk $N = \text{genap}$

| | | | | | |
|----|----|----|----|----|----|
| 32 | 35 | 40 | 42 | 50 | 58 |
|----|----|----|----|----|----|

- Median = elemen ke $(6)/2$ dan ke $(6)/2+1$ dijumlah, dan dibagi 2.

$$(40+42)/2 = \mathbf{41}$$



Descriptive statistics: Median

- Median dapat membantu menyelesaikan isu representasi data nilai tengah dengan Mean ketika terdapat outlier
- Median diukur dengan cara :
 - Urutkan elemen numeric dari urutan terkecil s/d terbesar
 - Tentukan banyaknya elemen dalam data tersebut, misal ukuran datanya sebesar **n**
 - Apabila N ganjil, maka median adalah :

$$\text{Median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ observation}$$

- Apabila N genap, maka median adalah :

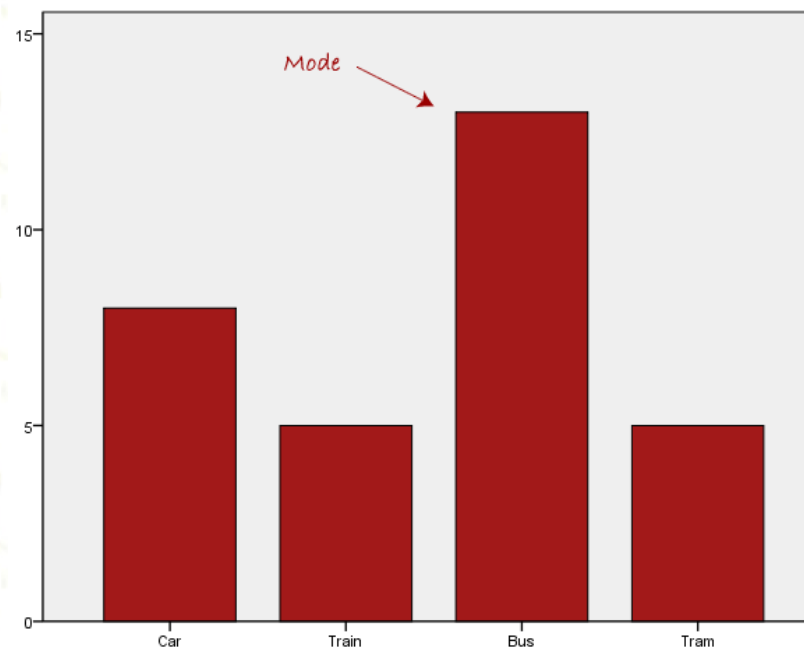
$$\text{Median} = \frac{\frac{n}{2}^{\text{th}} \text{ obs.} + \left(\frac{n}{2} + 1 \right)^{\text{th}} \text{ obs.}}{2}$$



Descriptive statistics: Mode

- Mode adalah elemen yang memiliki **frequency terbanyak** dalam suatu data numerik
- Apabila mean dan median tidak bisa digunakan dalam data yang berbentuk categorical, Mode bisa digunakan untuk data categorical

| Jenis mobil | Jumlah pemilik |
|---------------|----------------|
| BMW | 20 |
| Mercedes Benz | 10 |
| Honda | 40 mode |
| Toyota | 5 |



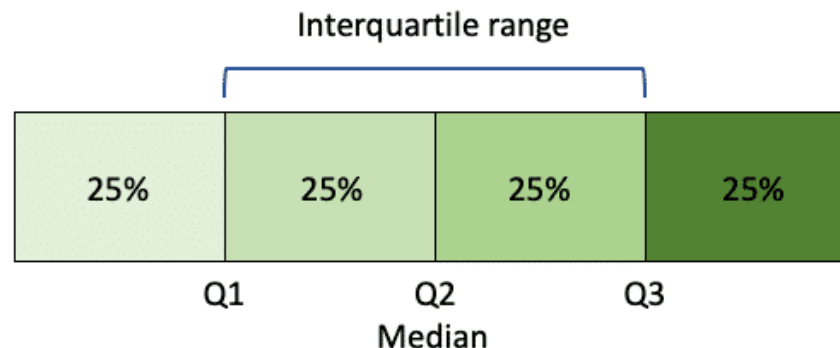


Descriptive statistics: Quartiles, Percentiles, Deciles

- Mean dan median memberikan nilai tengah, atau nilai paling banyak dalam sebuah data.
- Seringkali, kita membutuhkan informasi yang lebih banyak, seperti, nilai 10% tertinggi dari data, nilai 70% tertinggi dari data, dsb
- Untuk pertanyaan-pertanyaan tersebut, kita bisa jawab dengan : **quartiles, percentiles, atau deciles.**



Descriptive statistics: Interquartile Range



- Interquartile range memberikan gambaran mengenai kumpulan nilai tengah (middle fifty) dari sebuah data.
- Penggunaan range Min-Max terkadang bisa terpengaruh oleh outlier
- IQR cenderung lebih tidak sensitive terhadap outlier dibandingkan metric Range lain seperti Min-Max



Descriptive statistics: Standard deviation

- Mean, median, modes, memberikan informasi tentang nilai tengah dari suatu data
- Bagaimana apabila kita ingin mengetahui sebaran dari data?
- Kita bisa menggunakan variance / standard deviation
- Secara matematis :
- Contoh ilustrasi

$$\text{variance} = \sigma^2 = \frac{\sum (x_r - \mu)^2}{n}$$

$$\text{standard deviation } \sigma = \sqrt{\frac{\sum (x_r - \mu)^2}{n}}$$

| Nilai | (x-mean) | (x-mean)^2 |
|----------|----------|-------------|
| 10 | -17.8 | 316.84 |
| 15 | -12.8 | 163.84 |
| 30 | 2.2 | 4.84 |
| 43 | 15.2 | 231.04 |
| 32 | 4.2 | 17.64 |
| 10 | -17.8 | 316.84 |
| 14 | -13.8 | 190.44 |
| 24 | -3.8 | 14.44 |
| 56 | 28.2 | 795.24 |
| 44 | 16.2 | 262.44 |
| Jumlah | | 2313.6 |
| N | | 10 |
| Variance | | 231.36 |
| Stdev | | 15.21052267 |

Standard Deviation 15.21052267
Mean 27.8

QnA?



**Thank
YOU**

