



Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

DATA SCIENCE DARMAJAYA  
“YOUR BEST FUTURE IN DATA”

PERTEMUAN KE: 3

# Statistik Deskriptif

# KULIAH

OLEH: NURJOKO



# Learning Objectives

- Memahami Konsep Data
- Mampu menghitung dan menggunakan ukuran pusat
- Menggunakan Ukuran Sebaran
- Mampu membuat grafik dan visualisasi data yang tepat
- Mampu Mengidentifikasi Distribusi Data
- Mampu menginterpretasikan hasil statistik deskriptif
- Learning Objective 7



# DEFINISI

- **Statistika**

Ilmu mengumpulkan, menata, menyajikan, menganalisis, dan menginterpretasikan data menjadi informasi untuk membantu pengambilan keputusan yang efektif.

- **Statistik**

Suatu kumpulan angka yang tersusun lebih dari satu angka.

## Definisi Statistika Deskriptif

- Cabang statistik yang berkaitan dengan pengumpulan, penyajian, dan interpretasi data statistik untuk memberikan gambaran yang jelas dan ringkas tentang karakteristik dasar dari dataset.
- **Tujuan utama statistika deskriptif** adalah untuk menyajikan informasi data secara ringkas dan mudah dipahami agar pemahaman tentang data dapat diperoleh dengan cepat.



**Statistika deskriptif berkenaan** dengan bagaimana data dapat digambarkan atau disimpulkan secara:

- **Numerik** (misalnya menghitung rata-rata dan standar deviasi) atau
- **Grafis** (dalam bentuk tabel atau grafik), untuk mendapatkan gambaran sekilas mengenai data tersebut, sehingga lebih mudah dibaca dan bermakna.

**Data kasar (*raw data*)** diperoleh dari hasil pengukuran suatu variable pada sample yg diambil dari suatu populasi menggunakan teknik pengambilan sample tertentu

Langkah-langkah kegiatan statistika utk menangani data kasar :

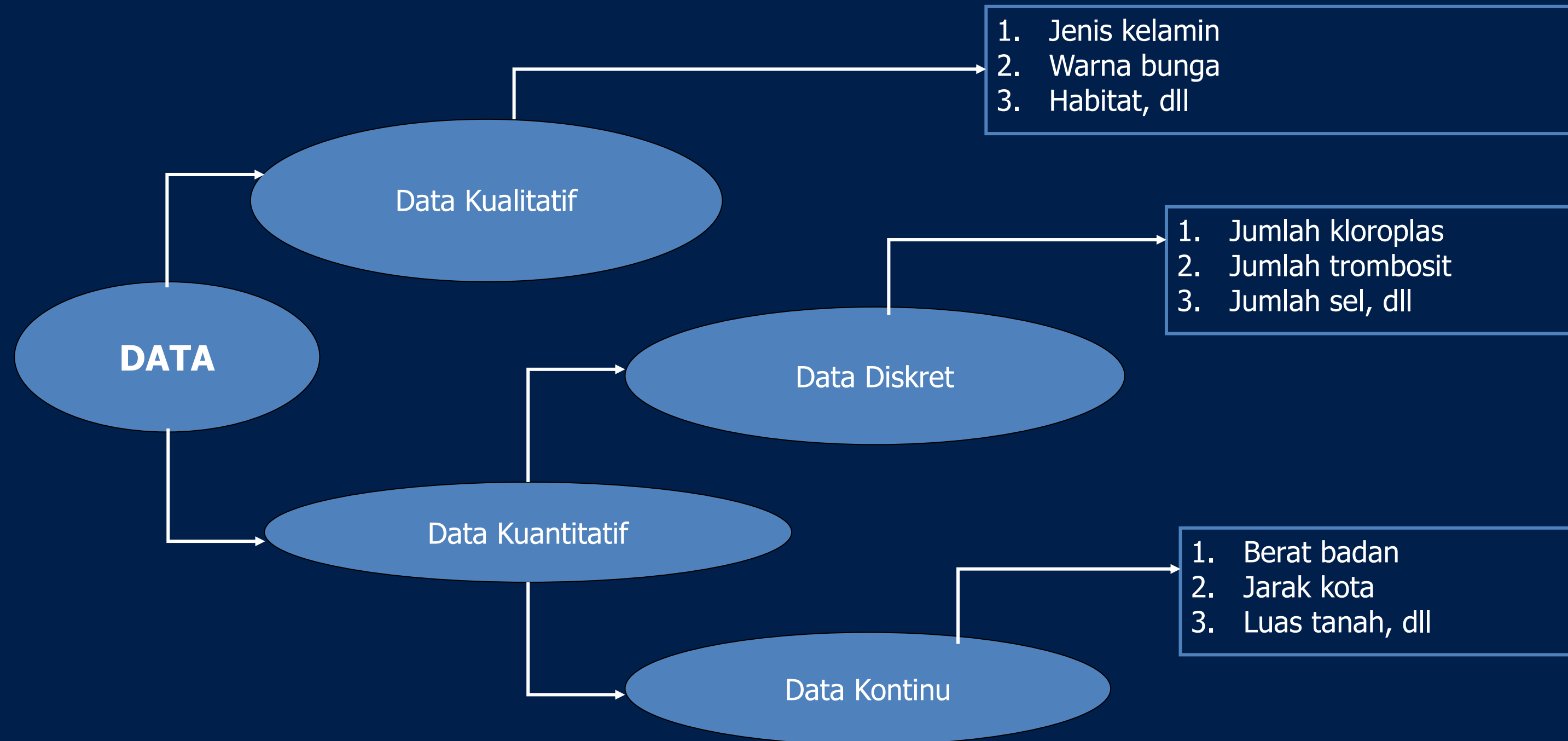
1. Pengumpulan data
2. Pengolahan data (diurutkan atau digolongkan)
3. Penyajian data dalam tabel atau grafik
4. Penafsiran sajian data
5. Analisa data
6. Penafsiran dan pengambilan kesimpulan
7. Pemanfaat penafsiran dan kesimpulan utk penentuan kegiatan penelitian Ibih lanjut.

- Poin 1,2,3,4,7 disebut **statistik deskriptif** (tanpa analisis, tanpa generalisasi, tanpa pengujian hipotesis, dan hanya melakukan perhitungan-perhitungan saja)
- Disajikan dalam bentuk tabel distribusi frekuensi (mean, modus, median), bar-diagram, histogram, polygon, dll
- Poin 1,2,3,4,5,6,7 disebut **statistik inferensial** (dg analisis, generalisasi, pengujian hipotesis)

▪

Uji  $t, z, F$





# Istilah dalam statistika

- **Obyek** = benda hidup atau mati yg diuji unsur-unsur, sifat dan kelakuannya melalui pengamatan, pengukuran dan penilaian guna mendapat info atau nilai-nilai yg berguna mengenai benda tsb
- **VARIABEL**  
Suatu sifat dari obyek atau unsur dari obyek yg dpt diamati atau diukur shg menghasilkan nilai, ukuran atau criteria lain yg dpt bervariasi
- **VARIATE**  
Angka/nilai ukuran/criteria lain yg dicapai suatu variabel pada suatu individu atau unit statistic

- **VARIASI**

Adanya perbedaan antar nilai/variante/ukuran dll dari suatu variabel pada populasi atau sampel

- **VARIABILITAS**

Kemungkinan utk bervariasi dr nilai suatu variable pd suatu populasi atau sample

- **PARAMETER**

suatu variabel terukur yg digunakan sbg criteria utk mengevaluasi suatu populasi atau sistem

# PENYAJIAN DATA

## a. Tabel

Bentuk penyajian ini merangkum data ke dalam bentuk baris atau kolom berupa kategori-kategori dan angka frekuensi. Terdapat 2 jenis, yaitu:

### 1. Tabel satu arah (one-way table)

Jumlah Responden Berdasarkan Umur

Kategori umur	Jumlah responden
< 20 tahun	10
21 – 30 tahun	15
31 – 40 tahun	25
> 40 tahun	10
<b>Total</b>	<b>60</b>

## 2. Tabulasi silang (lebih dari satu arah / two-way table)

Merupakan metode tabulasi untuk merangkum data dengan dua atau lebih variabel secara bersamaan, sisi (kolom) sebelah kiri dan baris atas menyatakan kelas untuk kedua variabel yang digunakan. Metode tabulasi silang digunakan jika :

- 1) salah satu variabel bersifat kualitatif dan lainnya kuantitatif
- 2) kedua variabel berupa variabel kualitatif
- 3) kedua variabel berupa variabel kuantitatif

Contoh tabulasi silang yang terdiri dari 2 variabel, yaitu umur dan jenis kelamin. Umur ada di baris dan jenis kelamin ada di kolom

Kategori umur	Jenis Kelamin	
	Laki-laki	Perempuan
< 20 tahun	5	5
21 – 30 tahun	10	5
31 – 40 tahun	12	13
> 40 tahun	2	8
Total	29	31

### 3. Tabel Distribusi Frekuensi

Pengelompokkan data menjadi tabulasi data dengan memakai kelas-kelas data dan dikaitkan dengan masing-masing frekuensinya

- **Kelebihan**

Dapat mengetahui gambaran secara menyeluruh

- **Kekurangan**

Rincian atau informasi awal menjadi hilang

# Contoh

## Distribusi Frekuensi Tinggi Badan 100 Mahasiswa

<u>Tinggi Badan</u>	<u>Frekuensi</u>
151-153	3
154-156	7
157-159	12
160-162	18
163-165	27
166-168	17
169-171	11
172-174	5

Sumber: Data buatan



## LIMIT, BATAS, NILAI TENGAH, DAN LEBAR KELAS

- Limit Kelas/Tepi Kelas

Nilai terkecil/terbesar pada setiap kelas

- Batas Kelas

Nilai yang besarnya satu desimal lebih sedikit dari data aslinya

- Nilai Tengah Kelas

Nilai tengah antara batas bawah kelas dengan batas atas kelas

- Lebar Kelas

Selisih antara batas bawah kelas dengan batas atas kelas

## CARA MEMBUAT TABEL DISTRIBUSI FREKUENSI

- 1) Tentukan Range atau jangkauan data (r)
- 2) Tentukan banyak kelas (k)

Rumus Sturgess :

$$k=1+3,3 \log n$$

- 3) Tentukan lebar kelas (c)

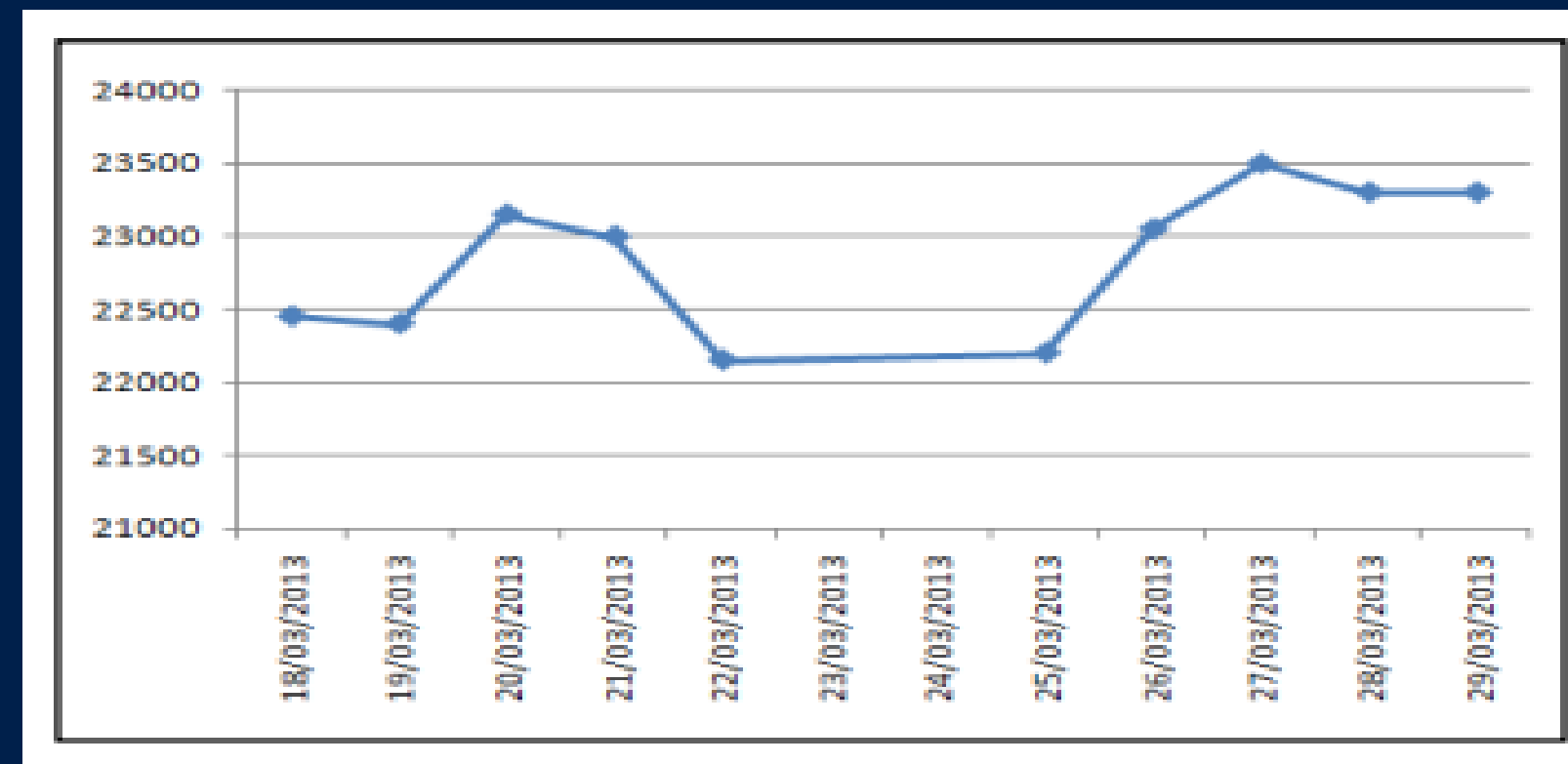
$$c=r/k$$

- 4) Tentukan limit bawah kelas pertama dan kemudian batas bawah kelasnya
- 5) Tambah batas bawah kelas pertama dengan lebar kelas untuk memperoleh batas atas kelas
- 6) Tentukan limit atas kelas
- 7) Tentukan nilai tengah kelas
- 8) Tentukan frekuensi

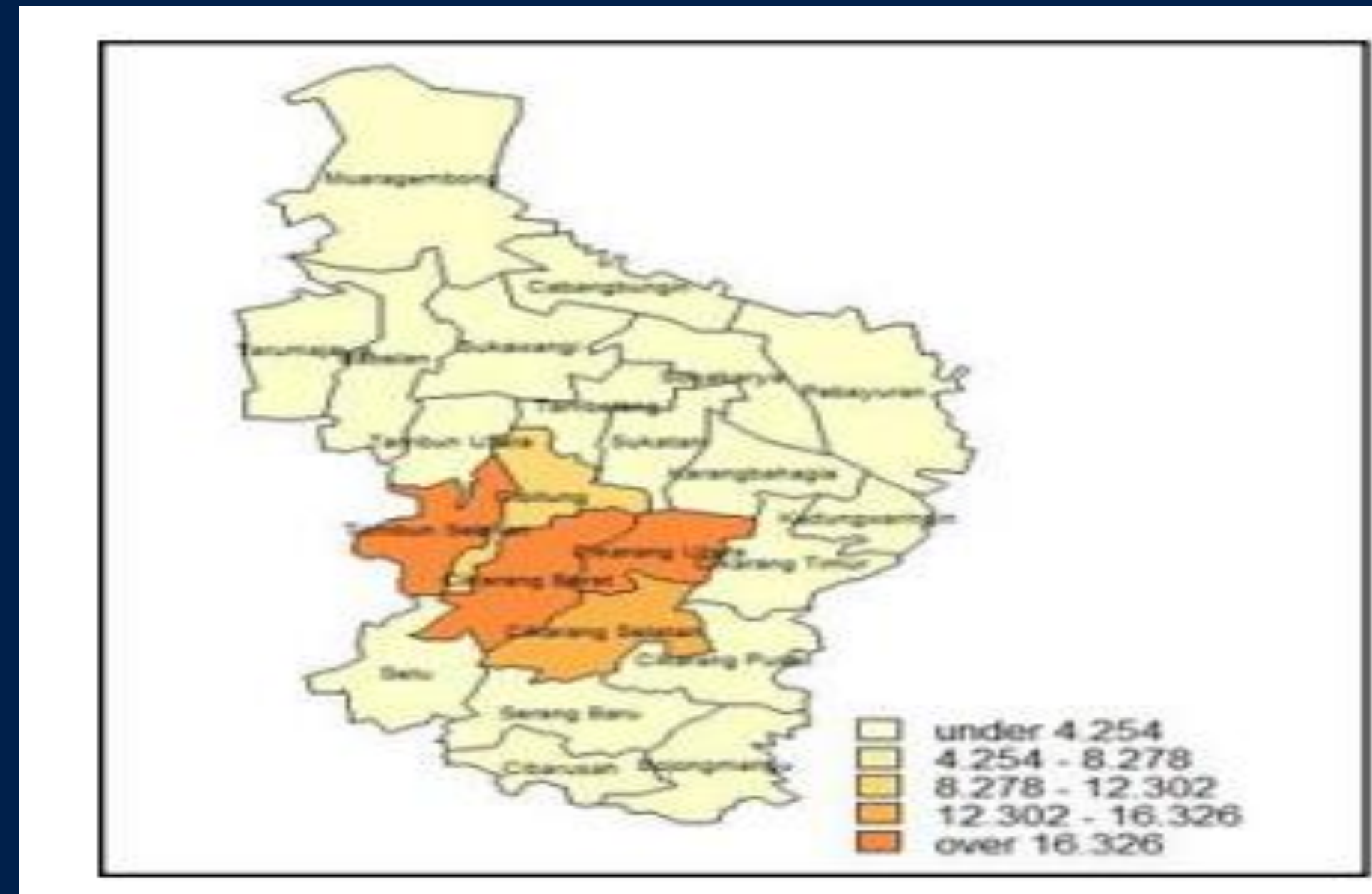
## b. Grafik

Penyajian data dari tabel menjadi grafik-grafik yang lebih menarik dan informatif

1. Grafik garis (line chart), untuk melihat pertumbuhan atau perkembangan suatu kejadian.

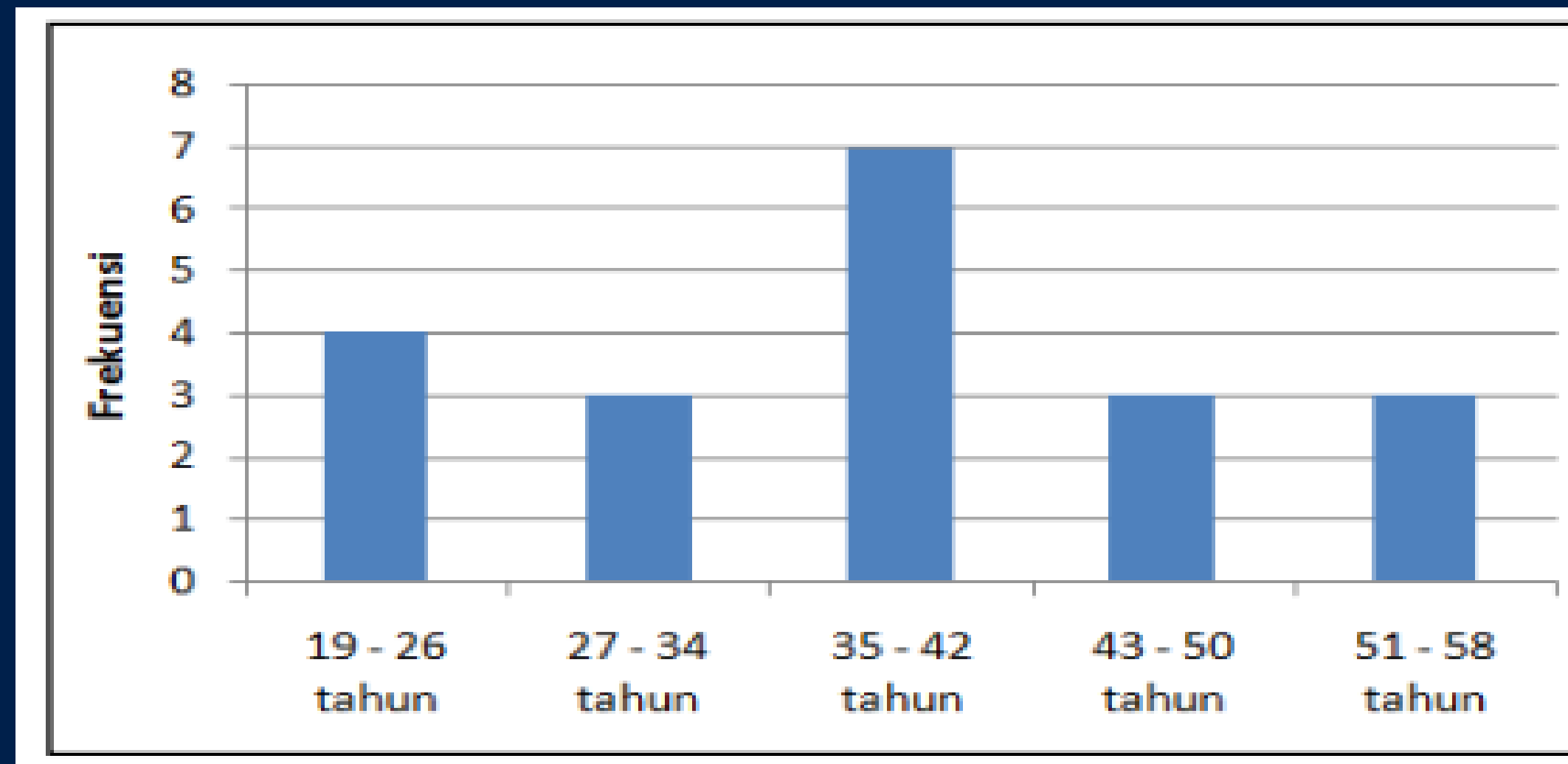


2. Grafik peta, untuk melihat atau menunjukkan lokasi suatu wilayah beserta atribut atau karakteristiknya.



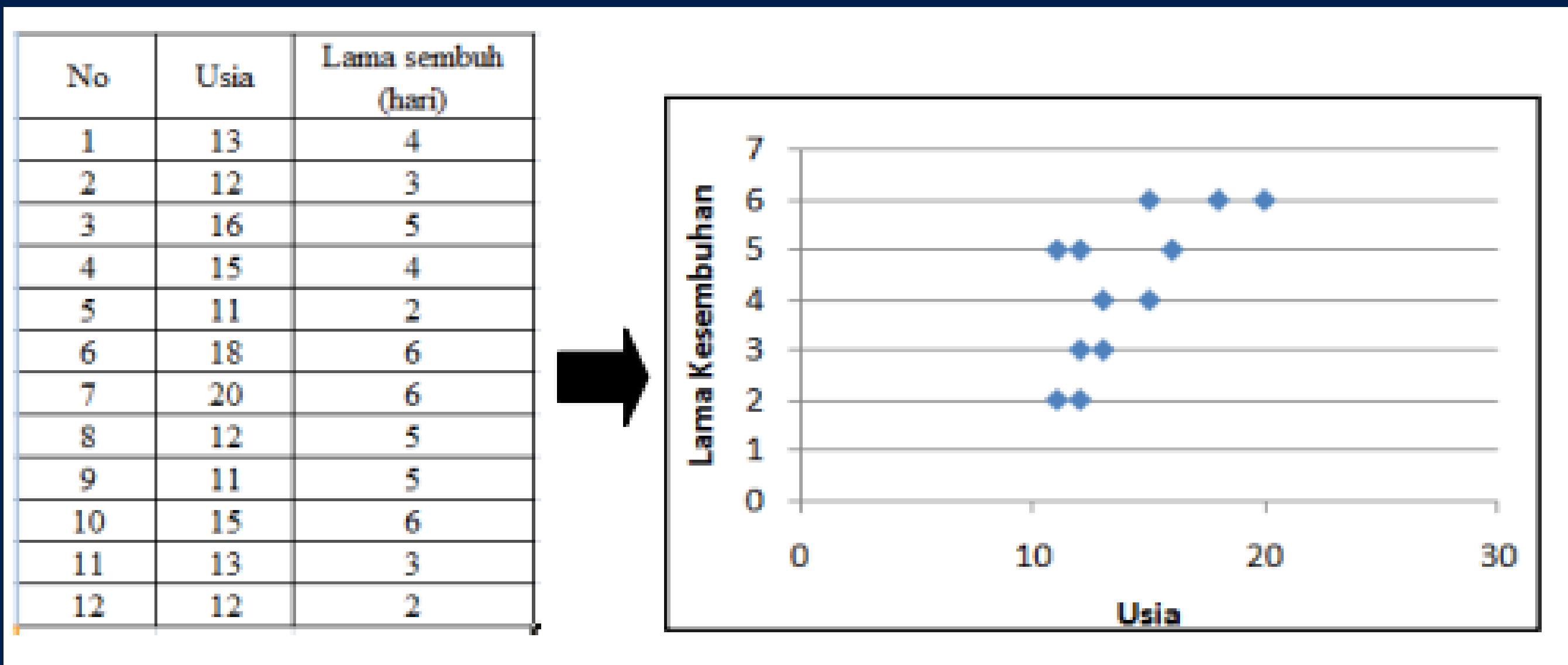
Grafik Peta Persentase Kontribusi Produk Domestik Regional Bruto di Kab. Bekasi

- Batang (bar graph), berfungsi untuk melihat distribusi atau perbandingan nilai, frekuensi, atau persentase di setiap kelas (kategori).



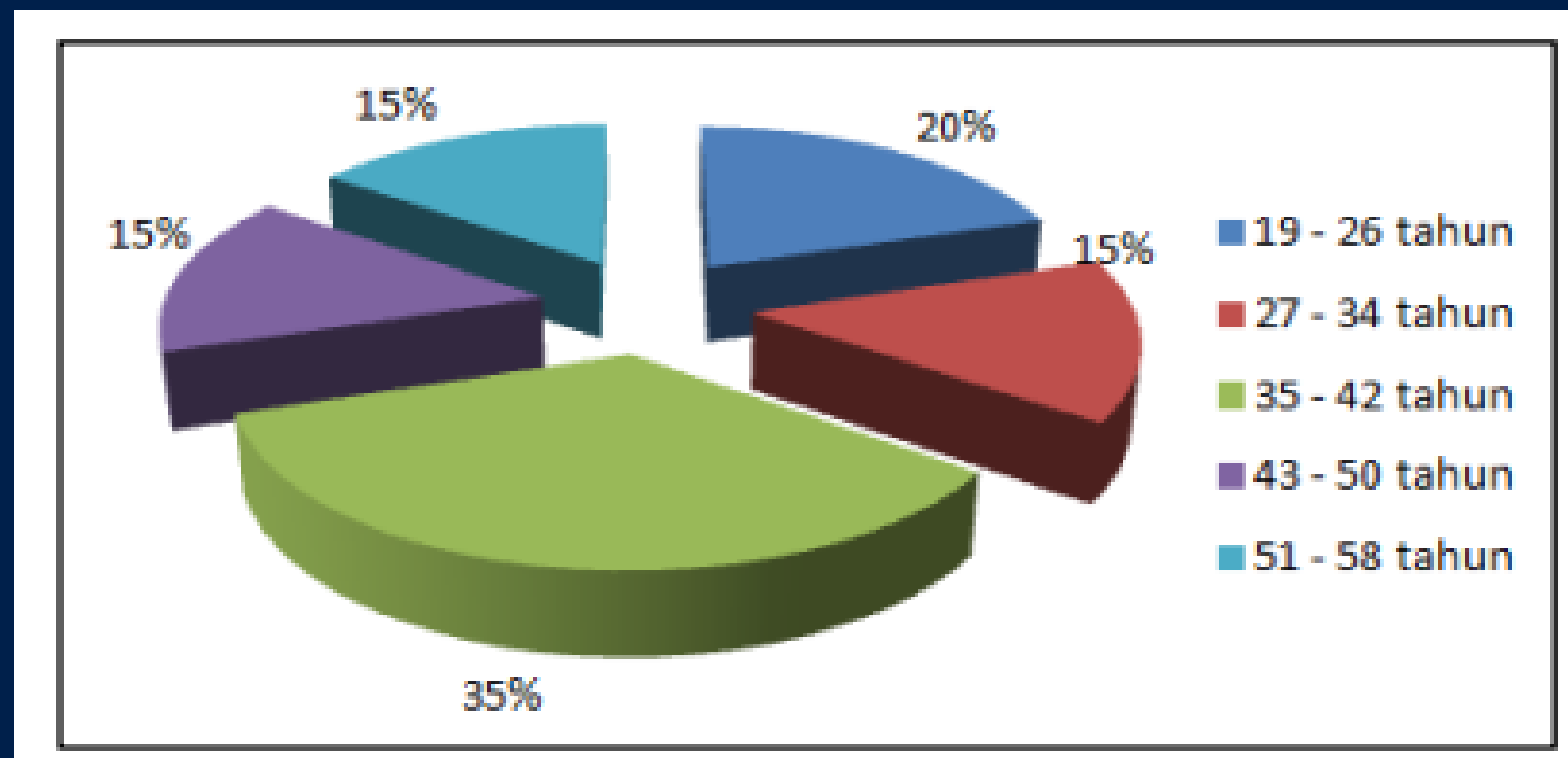
Grafik Batang Frekuensi Responden Berdasarkan Umur

4. Grafik titik Merupakan metode persentasi secara grafis untuk menggambarkan hubungan antara dua variabel kuantitatif (Santoso, 2004).



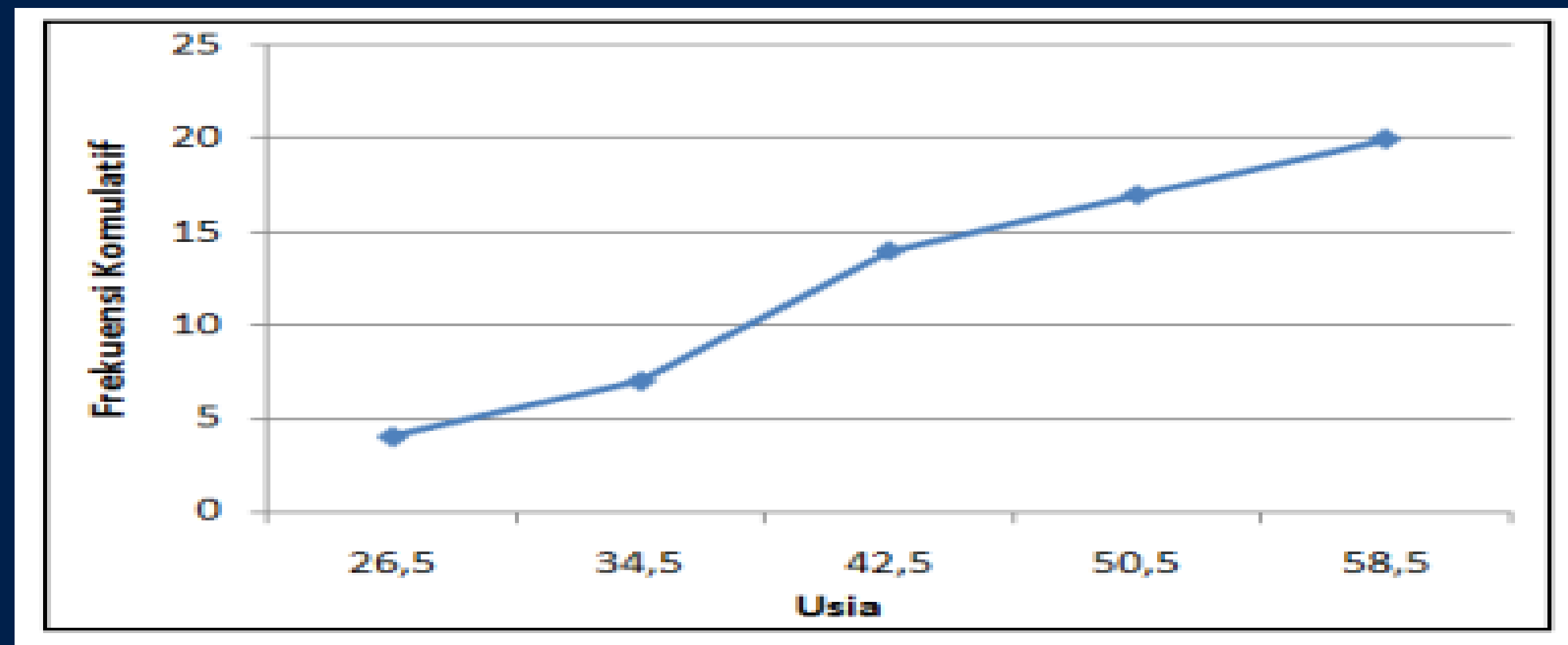
Grafik Titik Usia dan Lama Kesembuhan

5. Grafik Lingkaran (pie chart) Grafik ini dapat berfungsi untuk melihat perbandingan (dalam persentase atau proporsi).



Pie Chart Umur Responden

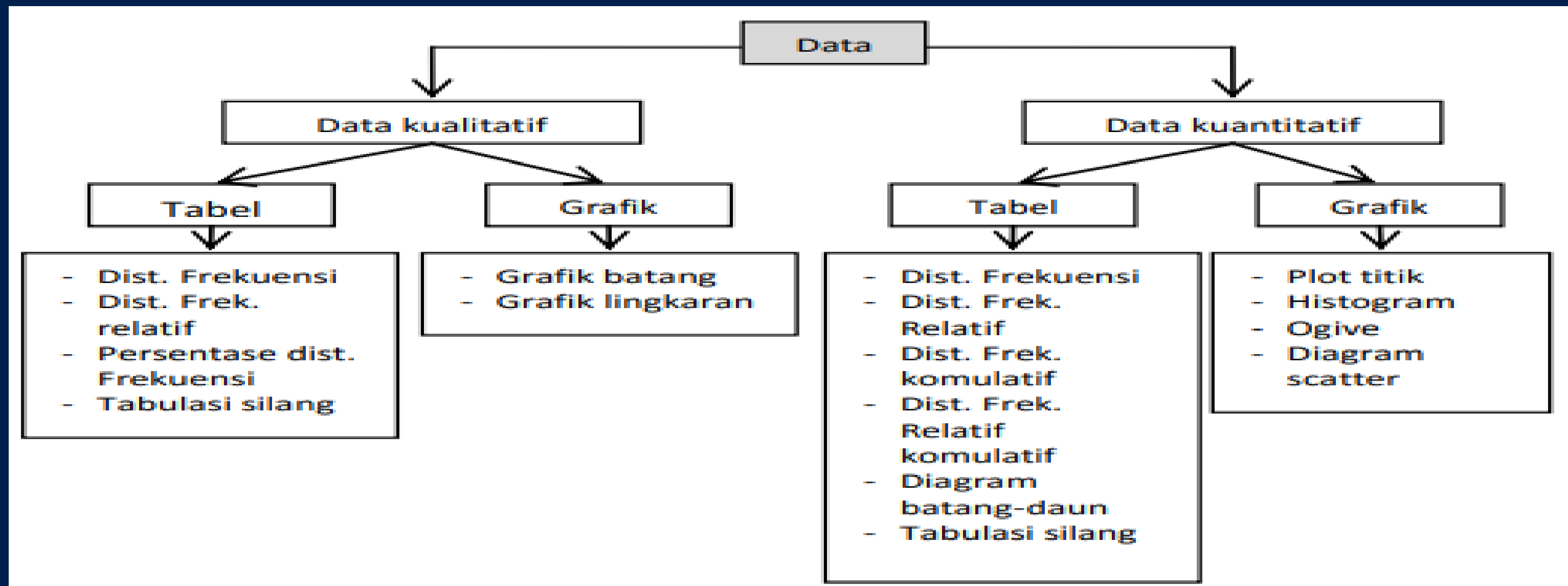
6. Ogive Merupakan grafik yang menyajikan nilai kumulatif. Sumbu horizontal merupakan nilai data, sedangkan sumbu vertikal adalah dapat berupa frekuensi kumulatif, frekuensi relatif kumulatif, atau persen frekuensi kumulatif.



Ogive Data Umur Responden

## Prosedur Penggunaan Tabel dan Grafik

Dengan jenis data yang berbeda maka bentuk penyajian tabel dan grafik pun berbeda atau langkah-langkahnya pun berbeda





# UKURAN PEMUSATAN DAN LETAK DATA

Merupakan nilai tunggal yang mewakili semua data atau kumpulan pengamatan dimana nilai tersebut menunjukkan pusat data.

Yang termasuk ukuran pemusatan :

1. Rata-rata hitung
2. Median
3. Modus
4. Rata-rata ukur
5. Rata-rata harmonis

# Rata-rata hitung

Rumus umumnya :

$$\text{Rata - rata hitung} = \frac{\text{Jumlah semua nilai data}}{\text{Banyaknya nilai data}}$$

1. Untuk data yang tidak mengulang

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\Sigma X}{n}$$

2. Untuk data yang mengulang dengan frekuensi tertentu

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_n X_n}{f_1 + f_2 + \dots + f_n} = \frac{\Sigma fX}{\Sigma f}$$

## Dalam Tabel Distribusi Frekuensi

Interval Kelas	Nilai Tengah (X)	Frekuensi	fX
9-21	15	3	45
22-34	28	4	112
35-47	41	4	164
48-60	54	8	432
61-73	67	12	804
74-86	80	23	1840
87-99	93	6	558
		$\Sigma f = 60$	$\Sigma fX = 3955$

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{3955}{60} = 65,92$$

## 2. MEDIAN

Untuk data berkelompok

$$\text{Med} = L_0 + c \left( \frac{\frac{n}{2} - F}{f} \right)$$

$L_0$  = batas bawah kelas median

$F$  = jumlah frekuensi semua kelas sebelum kelas yang mengandung median

$f$  = frekuensi kelas median

## Contoh :

<u>Interval Kelas</u>	<u>Frekuensi</u>
9-21	3
22-34	4
35-47	4
48-60	8
61-73	12
74-86	23
87-99	6
	<u><math>\Sigma f = 60</math></u>

Letak median ada pada data ke 30, yaitu pada interval 61-73, sehingga :

$$L_0 = 60,5$$

$$F = 19$$

$$f = 12$$

$$\text{Med} = 60,5 + 13 \left( \frac{\frac{60}{2} - 19}{12} \right) = 72,42$$

## 3. MODUS

Untuk data berkelompok

$$\text{Mod} = L_0 + c \left( \frac{b_1}{b_1 + b_2} \right)$$

$L_0$  = batas bawah kelas modus

$b_1$  = selisih antara frekuensi kelas modus dengan frekuensi tepat satu kelas sebelum kelas modus

$b_2$  = selisih antara frekuensi kelas modus dengan frekuensi tepat satu kelas sesudah kelas modus

## Contoh :

<u>Interval Kelas</u>	<u>Frekuensi</u>
9-21	3
22-34	4
35-47	4
48-60	8
61-73	12
74-86	23
87-99	6
	$\Sigma f = 60$

Data yang paling sering muncul adalah pada interval 74-86, sehingga :

$$L_0 = 73,5$$

$$b_1 = 23 - 12 = 11$$

$$b_2 = 23 - 6 = 17$$

$$\text{Mod} = 73,5 + 13 \left( \frac{11}{11 + 17} \right) = 78,61$$

## HUBUNGAN EMPIRIS ANTARA NILAI RATA-RATA HITUNG, MEDIAN, DAN MODUS

Ada 3 kemungkinan kesimetrian kurva distribusi data :

- 1) Jika nilai ketiganya hampir sama maka kurva mendekati simetri.
- 2) Jika  $\text{Mod} < \text{Med} < \text{rata-rata hitung}$ , maka kurva miring ke kanan.
- 3) Jika  $\text{rata-rata hitung} < \text{Med} < \text{Mod}$ , maka kurva miring ke kiri.

Jika distribusi data tidak simetri, maka terdapat hubungan :

**Rata-rata hitung-Modus = 3 (Rata-rata hitung-Median)**

$$\bar{X} - \text{Mod} = 3(\bar{X} - \text{Med})$$



# KUARTIL, DESIL, PERSENTIL

## 1. Kuartil

Kelompok data yang sudah diurutkan (membesar atau mengecil) dibagi empat bagian yang sama besar.

Ada 3 jenis yaitu

kuartil pertama ( $Q_1$ ) atau kuartil bawah,  
kuartil kedua ( $Q_2$ ) atau kuartil tengah, dan  
kuartil ketiga ( $Q_3$ ) atau kuartil atas.

## Untuk data tidak berkelompok

$$Q_i = \text{nilai ke } - \frac{i(n+1)}{4}, i = 1, 2, 3$$

## Untuk data berkelompok

$$Q_i = L_0 + c \left( \frac{\frac{in}{4} - F}{f} \right), i = 1, 2, 3$$

$L_0$  = batas bawah kelas kuartil

$F$  = jumlah frekuensi semua

kelas sebelum kelas kuartil  $Q_i$

$f$  = frekuensi kelas kuartil  $Q_i$

## Contoh :

Interval Kelas	Nilai Tengah (X)	Frekuensi
9-21	15	3
22-34	28	4
35-47	41	4
48-60	54	8
61-73	67	12
74-86	80	23
87-99	93	6
		$\Sigma f = 60$

$Q_1$  membagi data menjadi 25 %

$Q_2$  membagi data menjadi 50 %

$Q_3$  membagi data menjadi 75 %

Sehingga :

$Q_1$  terletak pada 48-60

$Q_2$  terletak pada 61-73

$Q_3$  terletak pada 74-86

## 2. Desil

Kelompok data yang sudah diurutkan (membesar atau mengecil) dibagi sepuluh bagian yang sama besar.

Untuk data tidak berkelompok

$$D_i = \text{nilai ke } - \frac{i(n+1)}{10}, i = 1, 2, 3, \dots, 9$$

Untuk data berkelompok

$$D_i = L_0 + c \left( \frac{\frac{in}{10} - F}{f} \right), i = 1, 2, 3, \dots, 9$$

$L_0$  = batas bawah kelas desil

$D_i$

$F$  = jumlah frekuensi semua kelas sebelum kelas

desil  $D_i$

$f$  = frekuensi kelas desil  $D_i$

### 3. Persentil

Untuk data tidak berkelompok

$$P_i = \text{nilai ke } - \frac{i(n+1)}{100}, i = 1, 2, 3, \dots, 99$$

Untuk data berkelompok

$$P_i = L_0 + c \left( \frac{\frac{in}{100} - F}{f} \right), i = 1, 2, 3, \dots, 99$$



# Penyebaran Data

- 1. Range**
- 2. Simpangan Rata-Rata**
- 3. Simpangan Baku**
- 4. Jangkauan Semi Inter Kuartil**
- 5. Jangkauan Persentil**
- 6. Angka Baku**
- 7. Koefisien Variasi**

## Jangkauan (range)

→ selisih antara nilai maksimum dan nilai minimum yang terdapat dalam data.

Jangkauan dapat dihitung dengan rumus:

$$R = X_{\text{maks}} - X_{\text{min}}$$

Contoh :

Tentukan range dari data : 10,6,8,2,4

Jawab :

$$\rightarrow R = X_{\text{maks}} - X_{\text{min}} = 10 - 2 = 8$$

## SIMPANGAN RATA-RATA (SR)

nilai rata-rata hitung harga mutlak simpangan-simpangannya

### Jenis Simpangan Rata-rata

#### a. Data Tunggal

Rumus simpangan rata-rata untuk data tunggal

dengan

$S_R$  = simpangan rata-rata  
 $\bar{x}$  = rata-rata  
 $x$  = nilai data ke- $i$   
 $n$  = banyaknya data

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

#### b. Data kelompok

Rumus simpangan rata-rata untuk data kelompok

dengan  $S_R$  = simpangan rata-rata  
 $\bar{x}$  = rata-rata  
 $x$  = nilai data ke- $i$   
 $f_i$  = frekuensi data ke- $i$   
 $n$  = banyaknya data

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

## SIMPANGAN BAKU STANDAR DEVIASI (S)

akar dari jumlah deviasi kuadrat dari bilangan-bilangan tersebut dibagi dengan banyaknya bilangan atau akar dari rata-rata deviasi kuadrat

a. Data tunggal

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

atau

$$S = \sqrt{\frac{\sum x^2}{n} - \left[ \frac{\sum x}{n} \right]^2}$$

Tentukan simpangan baku dari data :

2,3,5,8,7.

Jawab :

$$\bar{x} = \frac{2 + 3 + 5 + 8 + 7}{5} = 5$$

$x$	$(x - \bar{x})$	$(x - \bar{x})^2$
2	-3	9
3	-2	4
5	0	0
8	3	9
7	2	4
		<b>26</b>

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$\sqrt{\frac{26}{5}}$$

$$\sqrt{5,2}$$

## Ragam (variansi)

→ jumlah kuadrat dari selisih nilai data observasi dari nilai rata-ratanya, kemudian dibagi dengan jumlah observasinya

a. Data Tunggal

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

b. Data Kelompok

Rumus ragam dan simpangan baku data kelompok.

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

# Angka Baku

Dalam statistika, untuk membandingkan dua keadaan atau lebih diperlukan nilai simpangan baku. Sedangkan, dasar yang digunakan untuk membandingkan dua keadaan atau lebih tersebut biasa disebut dengan angka baku. Angka baku dirumuskan dengan:

## Keterangan:

- $Z$  = angka baku
- $x$  = variabel atau data
- $\bar{x}$  = mean atau rata-rata
- $SD$  = simpangan baku

$$Z = \frac{x - \bar{x}}{SD}$$

## Contoh

Diketahui nilai  $x = 70$ ,  $\bar{x} = 50$ , dan  $SD = 15$ . Tentukan angka bakunya!

## Penyelesaian:

$$x = 70, \bar{x} = 50,$$

$$\text{dan } SD = 15.$$

maka

$$\begin{aligned} Z &= \frac{x - \bar{x}}{SD} = \frac{70 - 50}{15} \\ &= \frac{20}{15} = 1,333 \end{aligned}$$

# Koefisien Variasi

*Koefisien variasi* adalah variasi dalam bentuk relatif. Koefisien variasi merupakan perbandingan dua nilai antara standar deviasi dengan mean dan dikalikan dengan 100%.

Koefisien variasi dirumuskan sebagai berikut:

$$KV = \frac{SD}{\bar{x}} \times 100\%$$

**Keterangan:**

- $KV$  = koefisien variasi
- $SD$  = standar deviasi
- $\bar{x}$  = mean

$$\begin{aligned} &= \frac{12,042}{73,5} \times 100\% \\ &= 0,1638 \times 100\% \\ &= 16,38\% \end{aligned}$$

Jadi, koefisien variasinya adalah 16,38%.

Nilai	$f$
50 - 59	8
60 - 69	10
70 - 79	16
80 - 89	11
90 - 99	5
Jumlah	50

Tentukan koefisien variasinya!

**Penyelesaian:**

Dari tabel di atas, diperoleh mean ( $\bar{x}$ ) = 73,5 dan standar deviasinya ( $SD$ ) = 12,042

$$KV = \frac{SD}{\bar{x}} \times 100\%$$



# CONCLUSION

Fill in .....



# REFERENCES

Fill in IEEE Style



Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

# THANK YOU!!

DATA SCIENCE DARMAJAYA "YOUR BEST FUTURE IN DATA"