

**KODE : D10J-1002**

# **ANALISIS DATA 1**

**PROGRAM STUDI ILMU AKTUARIA**

**SEMESTER : SATU**  
**SKS : TIGA (3)**  
**DOSEN : 1. Anna Chadidjah**  
**2. Hasna Afifah R**  
**TAHUN AKADEMIK : 2020-2021**



PROGRAM STUDI

# ILMU AKTUARIA

DEPARTEMEN STATISTIKA

FMIPA UNIVERSITAS PADJADJARAN

## Materi pertemuan ke 2

1. PENGUMPULAN DATA

2. PENYAJIAN DATA SECARA DESKRIPTIF

3. PENGOLAHAN DATA SECARA DESKRIPTIF

*Äk:t* | *U* *ariÄ*  
*npäd*

# POPULASI & SAMPEL

A **unit** is a single entity, usually a person or an object, whose characteristics are of interest.

The **population of units** is the complete collection of units about which information is sought.

A statistical **population** is the set of measurements (or record of some qualitative trait) corresponding to the entire collection of units about which information is sought.

A **sample** from a statistical population is the subset of measurements that are actually collected in the course of an investigation.

# METODE PENGUMPULAN DATA



## I. SENSUS

Proses mengumpulkan data dengan melibatkan semua anggota populasi (**kumpulan dari semua kemungkinan orang-orang, benda-benda dan ukuran lain dari objek yang menjadi perhatian**) pengamatan.

Catatan : data yang diperoleh disebut : [true value](#)

## II. SAMPLING

Proses mengumpulkan data dengan melibatkan sebagian anggota populasi pengamatan (**sampel**) yang dipilih secara acak.

Catatan : data yang diperoleh disebut : [estimate value](#)

*Data*

# STATISTIKA

Suatu cabang ilmu pengetahuan yang mempelajari proses sampling, pengolahan, analisa, dan pengambilan keputusan berdasarkan data

*Output :*

**Kebijakan, evaluasi, rekomendasi**

*Harapan : Output yang tepat dengan tingkat objektivitas informasi yang tinggi*

# STATISTIKA



- ❑ Pengumpulan data
  - ❑ Menyajikan data
  - ❑ Pengolahan data
  - ❑ Analisis data
  - ❑ Pengambilan keputusan terhadap data
- PENGELOLAAN  
DATA STATISTIK**

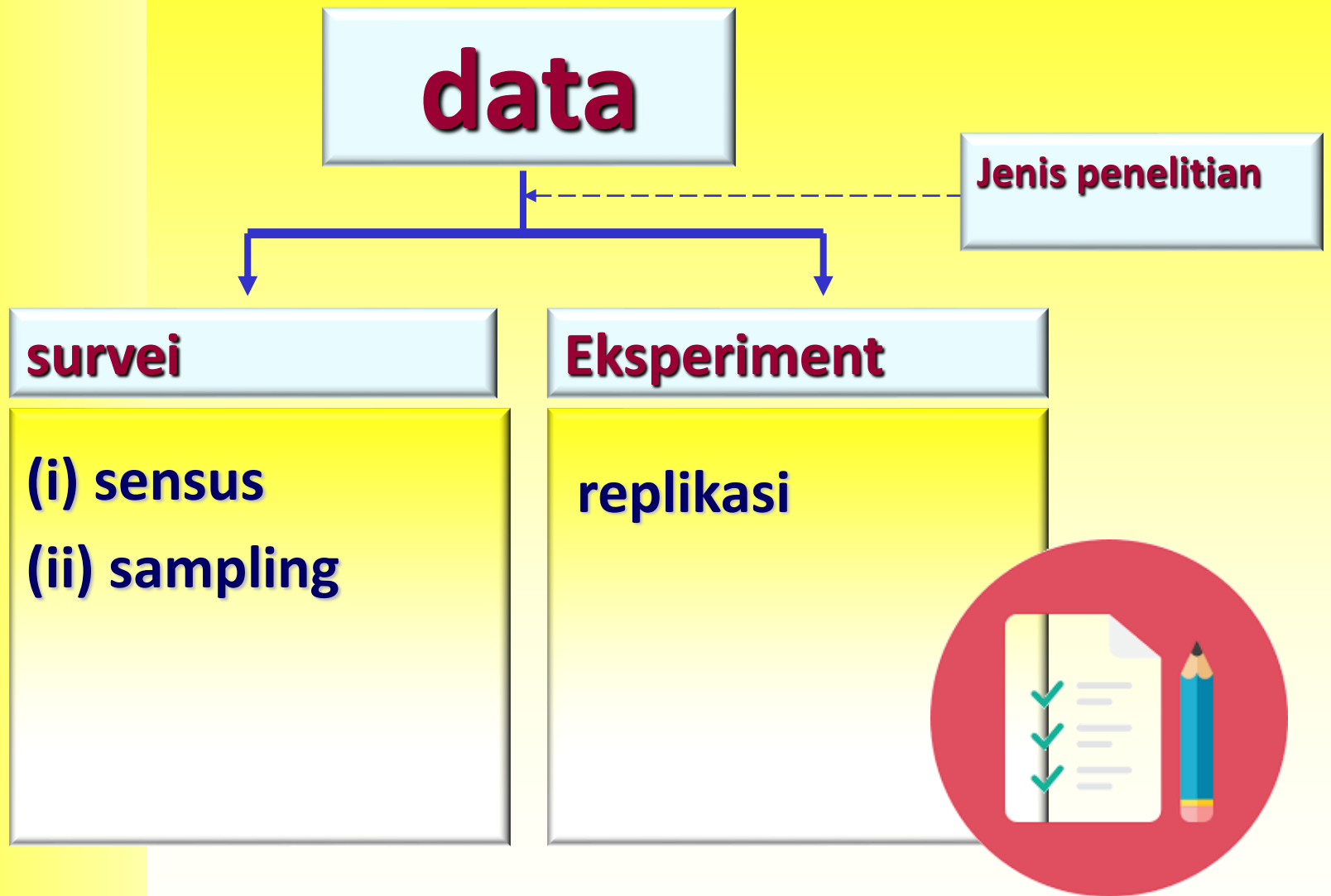
**Statistik** : Suatu kumpulan angka yang tersusun lebih dari satu angka

# PENGUMPULAN DATA / SAMPLING



- PROPORTIONAL  
RANDOM SAMPLING
- NON-PROPORTIONAL  
RANDOM SAMPLING

# CARA MEMPEROLEH DATA



# SAMPLING DESAIN SURVEI

- JUMLAH SAMPEL
- PEMILIHAN SAMPEL

PROBABILITY  
SAMPLING

NON  
PROBABILITY  
SAMPLING

# Probability Sampling



**Simple Random Sampling**



**Systematic Random Sampling**



**Stratified Random Sampling**



**Cluster Random Sampling**

## Simple random sampling



- (1) Undian
- (2) Tabel acak

**Memilih responden dari populasi yang homogen dimana setiap anggota populasi memiliki peluang yang sama untuk menjadi sampel**

**Kelebihan : (1) prosedurnya mudah untuk dilaksanakan, (2) tidak memerlukan proses pengolahan yang rumit**

**Kekurangan : (1) Mahal dan memerlukan waktu yang lama, (2) Memungkinkan munculnya sampling error yang tinggi, (3) kerangka sampel harus dibuat terlebih dahulu**

# **Systematic random sampling**

**Responden dipilih dengan sistematika tertentu.**

**Kelebihan : (1) mudah untuk dilaksanakan, (2) tidak diperlukan kerangka sampel, (3) simpel dan fleksibel**

**Kekurangan : Terjadinya data bias cukup tinggi**

# Stratified random sampling

**Responden dipilih dengan memperhatikan strata tertentu dari populasi**

**Kelebihan : (1) setiap responden dalam strata yang sama adalah homogen, (2) administrasi mudah**

**Kekurangan : (1) Strata dibuat tidak sesuai dengan tujuan krn tidak ada informasi awal, (2) Harus dibuat kerangka sampel yang terpisah dan berbeda untuk setiap strata.**

## Cluster random sampling

Jika populasi tersebar dalam beberapa wilayah (cluster) yang masing-masing memiliki karakteristik yang hampir sama, maka salah satu/beberapa wilayah tersebut dapat dipilih sebagai sampel

**Kelebihan** : (1) tidak memerlukan kerangka sampling, (2) sampel yang terpilih berada pada daerah yang berdekatan

**Kekurangan** : memungkinkan munculnya sampling error yang besar

# Non Probability Sampling

**Convenience / incidental Sampling**

**Judgment Sampling**

**Quota Sampling**

**Snowball Sampling**

Relatif tepat digunakan pada kondisi : (1) tahapan eksplorasi, (2) pengujian awal suatu angket, (3) berhadapan dengan populasi yang homogen, (4) Minimnya pengetahuan peneliti dalam bidang statistik, (5) adanya tuntutan akan kemudahan dari aspek operasional

## **Convenience / incidental Sampling**

**Responden adalah orang yang kebetulan ditemukan atau mudah untuk ditemui dan dijangkau**

**Kelebihan : (1) murah, (2) waktu cukup singkat, dan (3) responden mudah diakses**

**Kekurangan : (1) data bias, (2) tidak tepat digunakan jika banyaknya anggota populasi diketahui, dan (3) perlu kehatian-hatian yang tinggi dalam menerjemahkan hasil**

# Judgment Sampling



- (1) Expert sampling
- (2) Purposive sampling

Responden dipilih berdasarkan kriteria yang telah dirumuskan peneliti

**Kelebihan :** (1) baik digunakan jika sampel yang diambil kecil, (2) Memberikan wewenang kepada kemampuan dan pengalaman peneliti dalam menentukan sampel.

**Kekurangan :** menuntut kejelian dan kemampuan serta pengalaman yang tinggi dalam merumuskan sampel

# Quota Sampling

-  Merumuskan kategori kontrol/ quota dari populasi
-  Pemilihan sampel ( convenience atau judgment sampling)

## Kelebihan :

- (1) murah,
- (2) memberi keleluasaan kepada peneliti dalam menentukan elemen sampel

## Kekurangan :

- (1) Data bias,
- (2) Jumlah populasi harus besar

# Snowball Sampling

**Sampel diambil  
secara berantai  
dari populasi yang  
spesifik**

**Kelebihan :**  
**Bias relatif kecil**

**Kekurangan :**  
**(1) Mahal**  
**(2) waktu lama**

Menentukan jumlah sampel adalah “penting”, tetapi “jauh lebih penting” memilih sampel yang dapat menggambarkan populasinya

## Probability sampling

Jumlah populasi ( $N$ ) diketahui

$n$  minimum dihitung melalui rumus, diantaranya “SLOVIN”

$$n = \frac{N}{1 + N\varepsilon^2}$$

## Non-probability sampling

$n$  tidak dihitung melalui rumus, melainkan dikompromikan sesuai dengan metodenya dan kebutuhan.



# PENYAJIAN / DESCRIBING DATA



- Melalui Tabel dan Grafik
- Ukuran Pusat Data  
(*Measures of Center*)
- Ukuran Sebaran Data  
(*Measures of Variation*)

# Penyajian data melalui tabel dan grafik

1. Frekuensi Relatif

2. Pie Chart

3. Pareto diagram

4. line diagram

5. Histogram

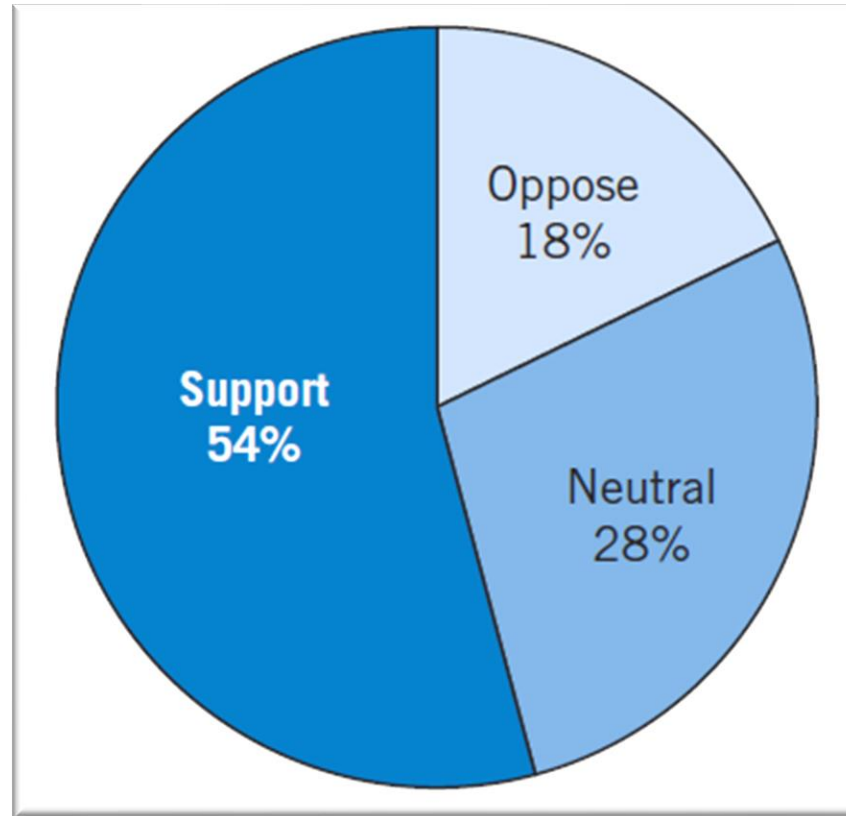
6. Stem-and-Leaf Display

# 1. FREKUENSI RELATIF

$$\text{Relative frequency of a category} = \frac{\text{Frequency in the category}}{\text{Total number of observations}}$$

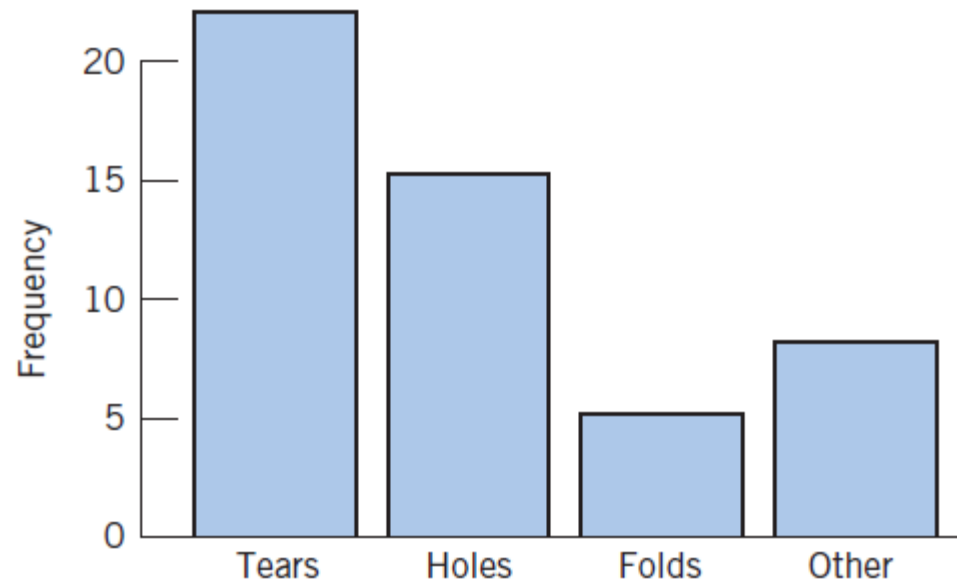
Responses	Frequency	Relative Frequency
Support	152	$\frac{152}{280} = .543$
Neutral	77	$\frac{77}{280} = .275$
Oppose	51	$\frac{51}{280} = .182$
Total	280	1.000

## 2. PIE CHART



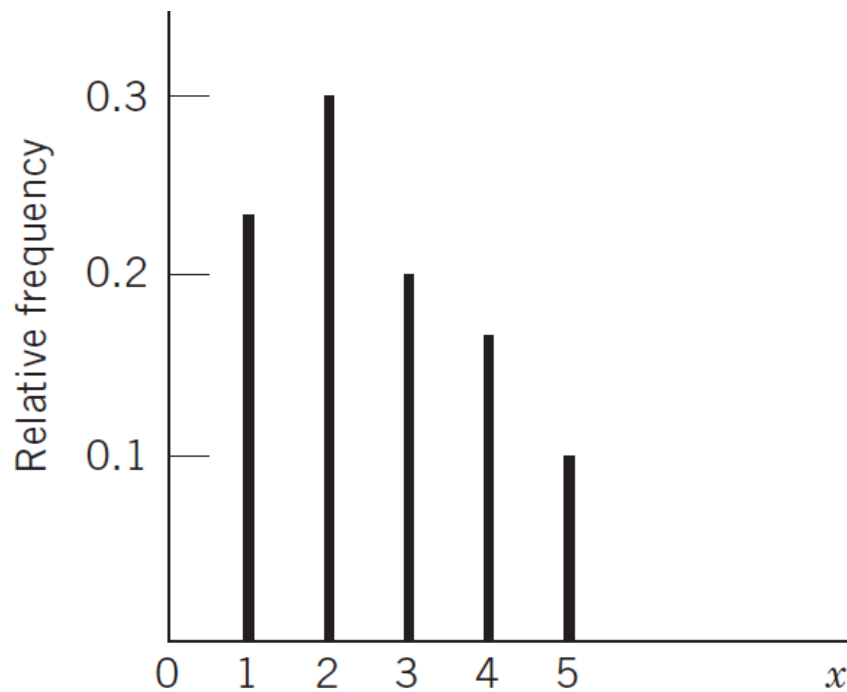
### 3. PARETO DIAGRAM

A **Pareto diagram** is a powerful graphical technique for displaying events according to their frequency. According to Pareto's empirical law, any collection of events consists of only a few that are major in that they are the ones that occur most of the time.



## 4. LINE DIAGRAM

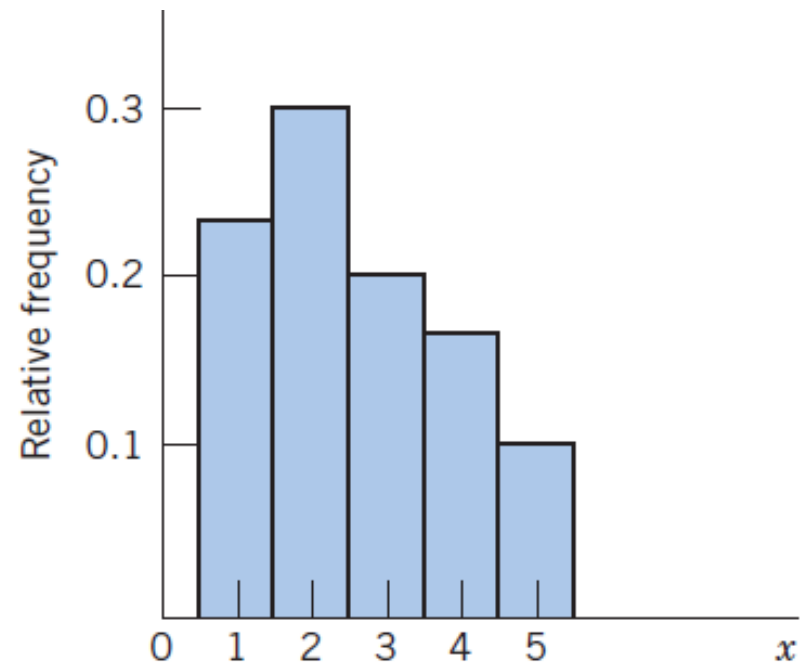
**line diagram**, we draw a vertical line at each value and make the height of the line equal to the relative frequency.



(a) Line diagram

## 5. HISTOGRAM

A **histogram** employs vertical rectangles instead of lines. These rectangles are centered at the values and their areas represent relative frequencies. Typically, the values proceed in equal steps so the rectangles are all of the same width and their heights are proportional to the relative frequencies as well as frequencies.



## 6. STEM-AND-LEAF DISPLAY

A **stem-and-leaf display** provides a more efficient variant of the histogram for displaying data, especially when the observations are two-digit numbers. This plot is obtained by sorting the observations into rows according to their leading digit.

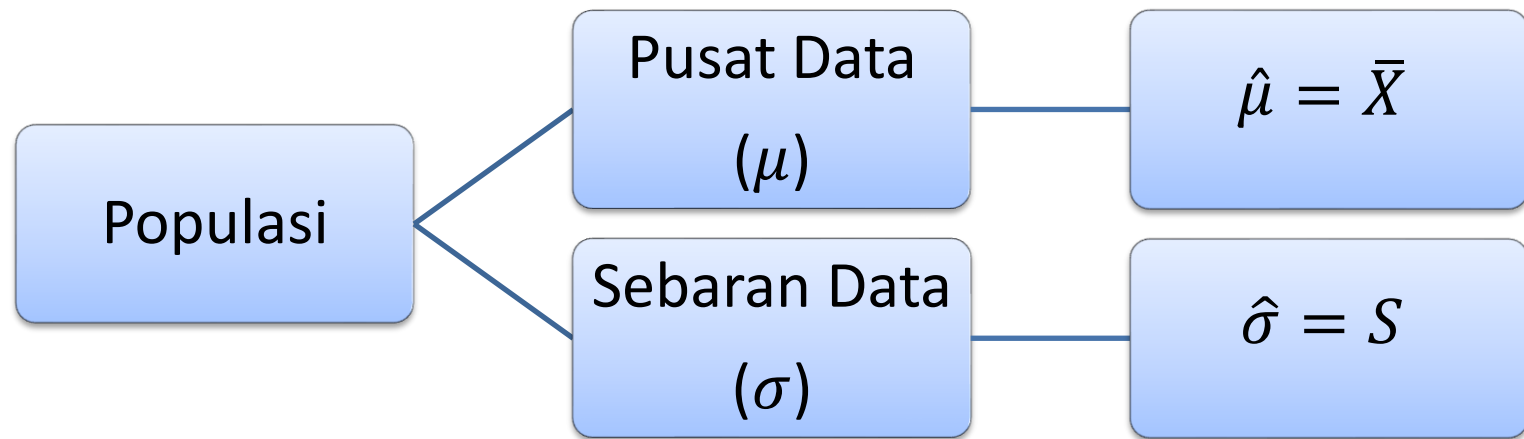
**TABLE** Stem-and-Leaf Display for

0	
1	
2	
3	7
4	289
5	35789
6	022345689
7	01234556778899
8	00134456789
9	0023589

# Data Klaim Asuransi di beberapa RS kota Bandung

No	Rumah Sakit	Banyak Klaim dalam satu bulan	Total Besar Klaim dalam satu bulan
1	RS. HASAN SADIKIN	2.208	Rp 30.000.000.000
2	RS. ADVENT	151	Rp 1.000.000.000
3	RS AL ISLAM	486	Rp 4.000.000.000
4	RS. IMMANUEL	425	Rp 2.700.000.000
5	RS SANTO YUSUP	261	Rp 1.300.000.000
6	RS. SANTOSA	189	Rp 1.500.000.000
7	RS RAJAWALI	138	Rp 600.000.000
8	RS. MUHAMMADYAH	169	Rp 700.000.000
9	RS. KEBONJATI	121	Rp 500.000.000
10	RS. MITRA KASIH	202	Rp 800.000.000
11	RS SARININGSIH	194	Rp 600.000.000
12	RS. PINDAD	160	Rp 600.000.000

# Karakteristik Populasi



# Statistik

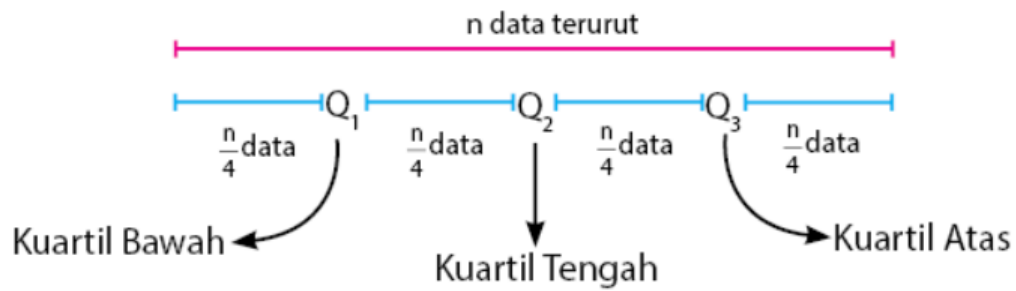
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

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

# Dispersi

- $R = X_{max} - X_{min}$
- $d_q = Q_3 - Q_1$
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- $S = \sqrt{\text{Variansi}}$

# Kuartil



# Kuartil untuk $n$ ganjil

Jenis Kuartil	Rumus Kuartil Data Tunggal
Kuartil Bawah	$Q_1 = x_{\frac{1}{4}(n+1)}$
Kuartil Tengah	$Q_2 = x_{\frac{1}{2}(n+1)}$
Kuartil Atas	$Q_3 = x_{\frac{3}{4}(n+1)}$

# Kuartil untuk $n$ genap

1. Carilah nilai yang menjadi nilai tengahnya (median).
2. Membagi data di sebelah kiri median menjadi dua bagian yang sama dan menghasilkan kuartil bawah.
3. Membagi data di sebelah kanan median menjadi dua bagian yang sama dan menghasilkan kuartil atas

Contoh data dana yang dibutuhkan untuk menata kawasan kumuh di 12 kota

Kota	Dana untuk menata kawasan kumuh
A	113.75
B	99.59
C	111.79
D	300.78
E	101.66
F	124.13
G	90
H	20.19
J	99.92
J	112.49
K	135.72
L	102.09

Untuk mengetahui karakteristik dari data tersebut carilah taraf dan dispersinya

- Mean
- Median
- Modus
- Range
- Jarak antar Kuartil
- Variansi
- Standar deviasi

## Taraf

- $\bar{X} =$
- $Md =$
- *Modus*

## Dispersi

- $R =$
- $dq =$
- $S^2 =$
- $S =$

# Box Plots and Outliers

# Box Plot

We need 5 numbers, called the 5 number summary:

1. minimum value
2. Q1
3. median
4. Q3
5. maximum value

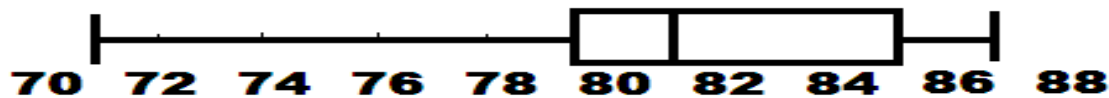
# Harry Potter Data

The 5 number summary is:

Minimum	70.9
Q1	78.9
Median	81.35
Q3	84.45
Maximum	86.2

# Drawing the plot free hand

- Use a scaled line to draw the box plot
- Approximate locations of the values.



- We say this is **SKEWED LEFT** since the longer tail is on the left side.

# Inter-quartile Range

The inter-quartile range (IQR) is the value of  $Q3 - Q1$

# Outliers

- An outlier is an unusual score, relative to the data set. It will influence the mean and the standard deviation.
- A data set might have no outliers, one outlier, or several outliers.
- Two types: Mild and Extreme

# Mild Outliers

Look for data values (x) that are either:

1.

2.  $Q3 + 1.5 \text{ IQR} < x \leq Q3 + 3 \text{ IQR}$

# Extreme Outliers

Look for values that are either:

1.

2.

# Harry Potter data - checking

# Example: Ages of actresses at the time they first won the Oscar

50	44	35	80	26	28	41
21	61	38	49	33	74	30
33	41	31	35	41	42	37
26	34	34	35	26	61	60
34	24	30	37	31	27	39
34	26	25	33			

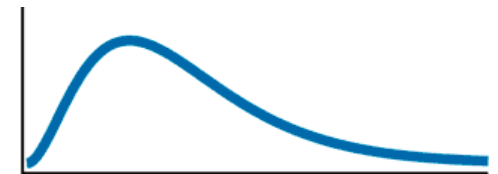
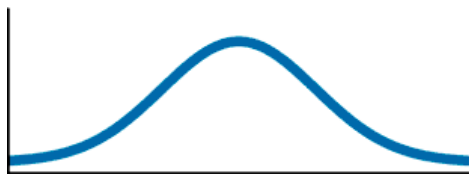
# 5 number summary for actresses

- Min =
- Q1 =
- Med =
- Q3 =
- Max =

# Actresses

- Check for outliers. Identify type.

# Comparison of Boxplots and histograms



*Bell-shaped*

*Uniform*

*Skewed*