

# PENGUKURAN DATA

1

## 1. Terminology Populasi & Sampel

- **Populasi**: himpunan komplit dari individual, obyek atau nilai dari suatu pengamatan
  - Seringkali terlalu besar untuk dikaji secara keseluruhan
  - Mungkin nyata atau scr hipotesis saja (contoh hasil dari penelitian yg diulang sampai tak terhingga)
- **Sampel**: Sebuah himpunan bagian dari populasi
  - Sampel bisa dikatakan **random** (tiap anggota populasi memp. kesempatan yg sama untuk dipilih) atau **convenience** (apa yg tersedia).
  - Random selection attempts to ensure the sample is **representative** of the population.

2

## Peubah/Variabel

- **Peubah** adl besarnya pengukuran dalam sampel. Peubah bisa diklasifikasikan sbg:
  - **Kuantitatif** i.e. numerical
    - **Kontinu** (e.g. pH , level kolesterol)
    - **Diskret** (e.g. Banyaknya bakteri dlm koloni, banyaknya antrian)
  - **Kategori**
    - **Nominal** (e.g. gender, blood group)
    - **Ordinal** (**ranked** e.g. Ringan, sedang, atau berat). Peubah ordinal sering dicatat sbg kuantitatif.

3

## Peubah/Variabel

- Lebih jauh Peubah dapat diklasifikasikan sbg:
  - **Terikat/Dependent/Response**. **Peubah dari tujuan penelitian** (e.g. Tekanan darah dlm percobaan obat anti hipertensi). Tidak dikontrol oleh peneliti.
  - **Bebas/Independent/Predictor**
    - Disebut **Factor** jika dikontrol oleh peneliti.
    - **Covariate** jika tidak terkontrol.
- Jika nilai dari peubah tidak dapat diduga maka peubah dikatakan sbg **peubah acak (random variable)**

4

# Parameter & Statistik (Statistics)

- **Parameter**: Besaran yg menjelaskan karakteristik populasi. Pada umumnya tidak diketahui besarnya shg harus dilakukan statistika inferensia untuk parameter.
- **Statistik Deskriptif**: Besaran dan teknik yg digunakan untuk menjelaskan karakteristik sampel atau menggambarkan sampel e.g. mean, standard deviasi, box-plot

5

# 2. Mengukur Kecenderungan Nilai Tengah (Lokasi)

Pengukuran lokasi mengindikasikan dimana pada garis data nilai akan ditemukan. Pengukuran yg biasa dipakai adalah:

- (i) **Arithmetic Mean (Rataan)**,
- (ii) **Median**, dan
- (iii) **Mode (Modus)**

6

## Mean

- Diberikan  $x_1, x_2, x_3, \dots, x_n$  adl nilai sebenarnya dari peubah acak  $X$ , dari sampel berukuran  $n$ . **Arithmetic mean** dari sampel didefinisikan sbg:

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

7

## Contoh

Tekana darah systolic dari tujuh lelaki berumur menengah adl sbg:

151, 124, 132, 170, 146, 124 and 113.

$$\begin{aligned} \text{mean} \quad \bar{x} &= \frac{(151+124+132+170+146+124+113)}{7} \\ &= 137.14 \end{aligned}$$

8

# Median

- Jika data sampel diurutkan dari kecil ke besar, **median** adl
  - (i) Nilai tengah jika  $n$  ganjil, atau
  - (ii) Nilai antara antar dua nilai tengah jika  $n$  genap

9

# - $n$ ganjil

113, 124, 124, 132, 146, 151, and 170.

Median = 132.

Mode = 124.

10

# - $n$ genap

366, 327, 274, 292, 274 dan 230.

Urutkan data:

230, 274, 274, 292, 327 dan 366.

Median =  $(274+292) \div 2 = 283$ .

Mode = 274.

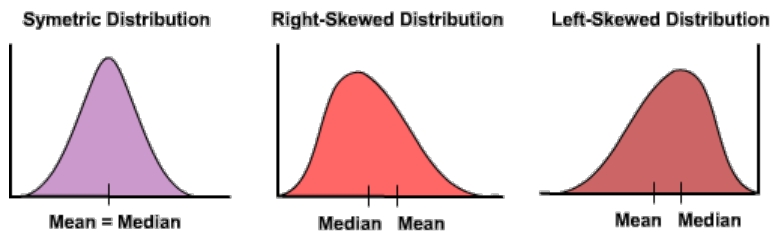
11

# Mean versus Median

- Jika histogram adl miring kekanan (right-skewed), jumlah sampel yg besar akan menaikkan nilai mean.
- Median tidak terpengaruh oleh besarnya sampel shg ia merupakan pengukuran yg lebih baik untuk titik pusat (sentralitas) apabila distribusi datanya miring.
- Jika mean=median=mode maka data dikatakan simetri
- Contoh: sampel mean = 98.28, median = 94.5, i.e. mean lebih dari median mengindikasikan bahwa distribusinya miring

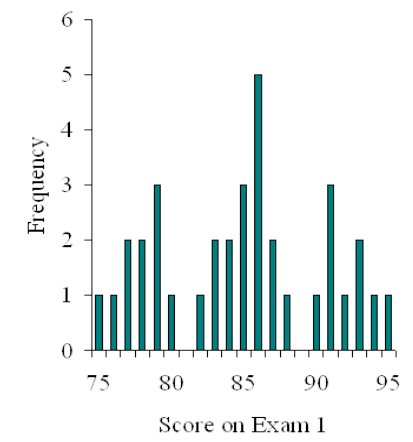
12

# Mean vs Median

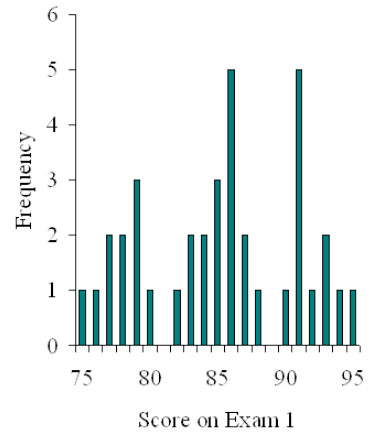


# Mode

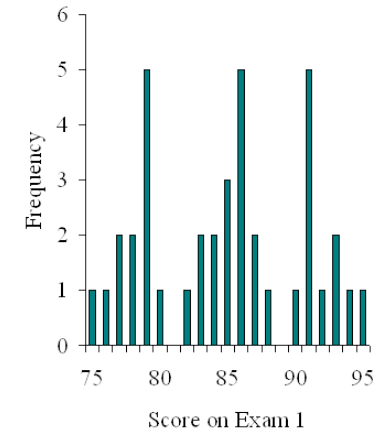
**mode** adl nilai yang paling sering terjadi



# Distribusi Bimodal



# Distribusi Multimodal



## Kapan Mode digunakan?

Mode digunakan ketika data berskala nominal

Hanya satu-satunya pengukuran pemusatan yang sesuai untuk data berskala nominal

## 3. Mengukur Penyebaran

- Pengukuran penyebaran merupakan karakteristik seberapa menyebar distribusi data, i.e. Bagaimana data menyebar.
- Pengukuran yg biasa digunakan:
  1. Range
  2. Variance & Standard deviation
  3. Coefficient of Variation (atau relative standard deviation)
  4. Inter-quartile range

18

## Range

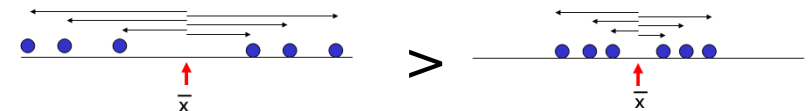
- **Range** sampel adl beda antara data terbesar dan terkecil dalam observasi
- Mudah dihitung;
  - Contoh kasus tekanan darah: min=113 dan max=170, range=57 mmHg
- Berguna untuk kasus tertentu (tergantung data) □
- Sensitif terhadap nilai ekstrim □

19

## Sampel Varians (Variance)

- **Sample variance**,  $s^2$ , adl arithmetic mean dari beda (deviasi) terhadap sampel mean yang dikuadratkan:

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



20

## Standard Deviasi (Deviation)

- **Sample standard deviation,  $s$** , adl akar dari varians

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

$s$  mempunyai pengukuran yg sama dengan peubah X.

21

## Contoh

Data	Deviation	Deviation <sup>2</sup>
151	13.86	192.02
124	-13.14	172.73
132	-5.14	26.45
170	32.86	1079.59
146	8.86	78.45
124	-13.14	172.73
113	-24.14	582.88
Sum = 960.0	Sum = 0.00	Sum = 2304.86
$\bar{x} = 137.14$		

22

## Contoh (contd.)

$$\sum_{i=1}^7 (x_i - \bar{x})^2 = 2304.86$$

$$\begin{aligned} \text{maka, } s &= \sqrt{\frac{2304.86}{7-1}} \\ &= 19.6 \end{aligned}$$

23

## Koefisien Variasi Coefficient of Variation

- **Coefficient of variation (CV)** atau **relative standard deviation (RSD)** adl standar deviasi sampel yang diekspresikan sbg persentase dari mean i.e.

$$CV = \left( \frac{s}{\bar{x}} \right) \times 100$$

- CV tdk terpengaruh oleh perubahan perkalian skala
- Akibatnya, ia merupakan cara yg berguna untuk membandingkan penyebaran data pada peubah-peubah yg berbeda skala pengukurannya

24



## Contoh

Tingkat denyut nadi dari 12 individu yg dirutkan naik adl:

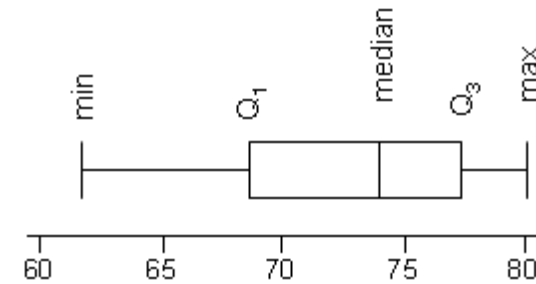
62, 64, 68, 70, 70, 74, 74, 76, 76, 78, 78, 80

$Q_1 = (68 + 70) : 2 = 69$ ,  $Q_3 = (76 + 78) : 2 = 77$

$IQR = (77 - 69) = 8$

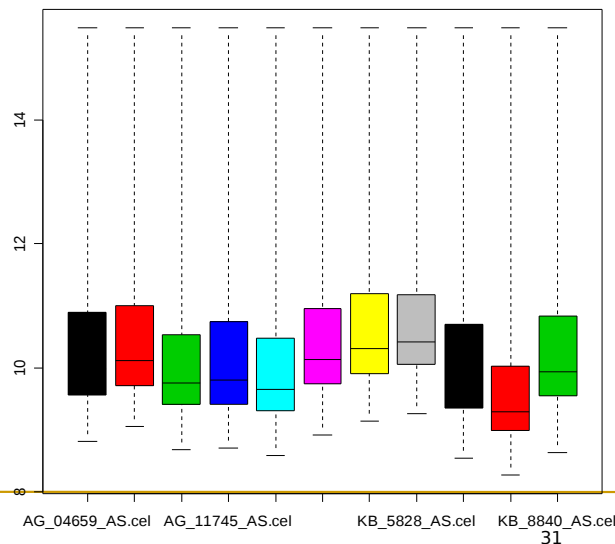
29

## Contoh: Box-plot



30

## Contoh: Box-plots dari intensitas from 11 gene expression arrays



31

## Outlier

- **Outlier** adl kejadian dimana sebuah observasi tidak terlihat tergolong bersama data yg lain
- Outlier bisa terjadi karena pencatatan yg salah, alat pengukuran yg salah, atau hal lainnya
- Memerlukan penanganan khusus

32

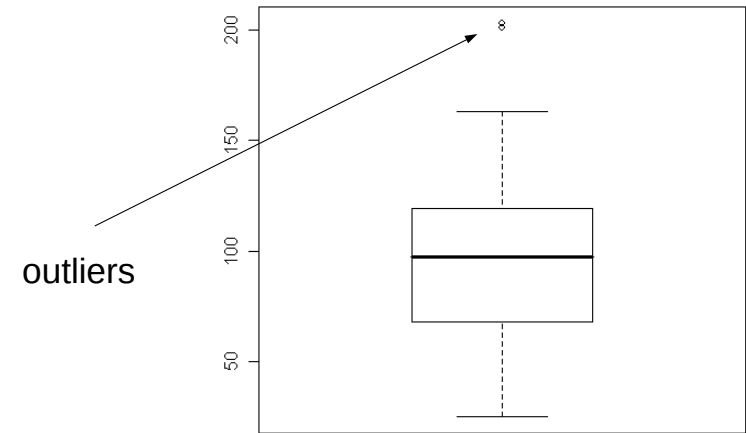


# Outlier Boxplot

- Definisi ulang dari batas atas dan bawah dari boxplot (garis kumis (whisker)) sbg:  
Batas Bawah =  $Q_1 - 1.5 \cdot IQR$ , dan  
Batas Atas =  $Q_3 + 1.5 \cdot IQR$
- Perhatikan bahwa garis mungkin tidak mencapai batas-batas ini
- Jika data point  $<$  batas bawah atau  $>$  batas atas, data point dianggap sbg outlier.

33

# Contoh



34