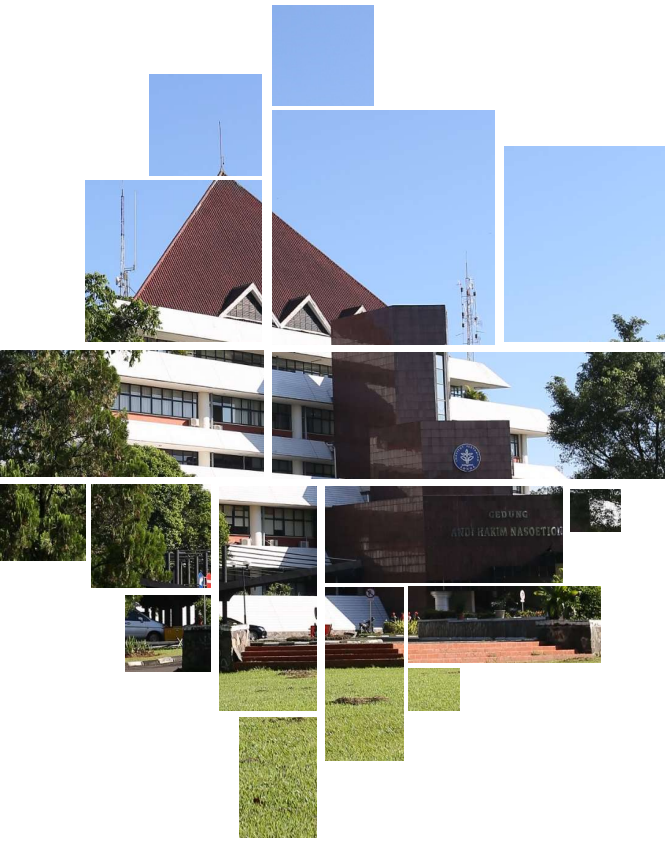
STA513 – Analisis Statistika untuk Bisnis, Ekonomi, dan Industri

Semester Ganjil 2020/2021



PERTEMUAN #1

Pendahuluan (sambungan)

disusun oleh:

Bagus Sartono

bagusco@gmail.com

0852-1523-1823

Prodi Statistika dan Sains Data

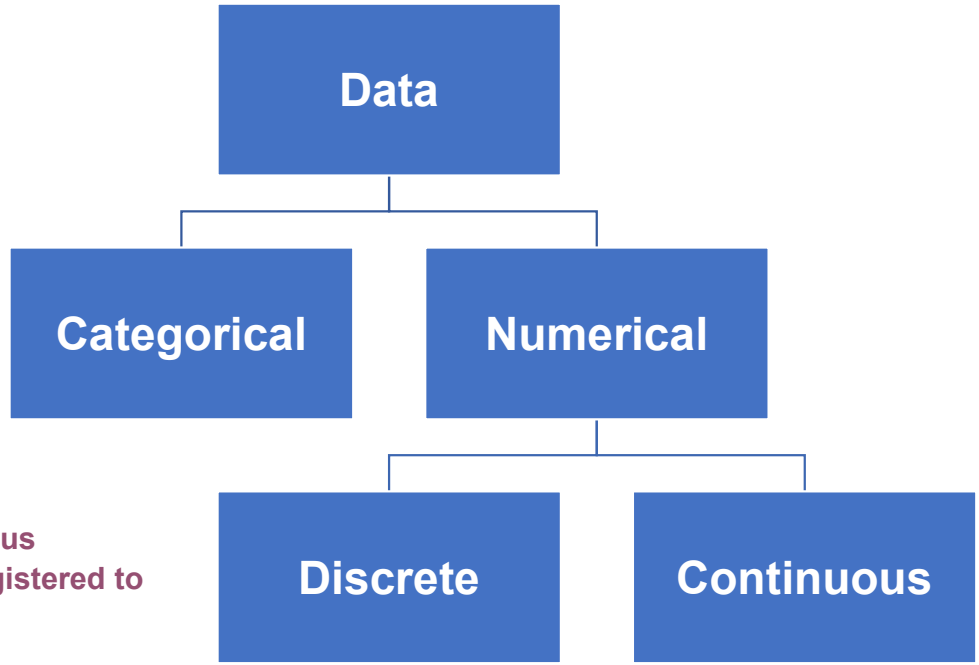
Fakultas Matematika dan Ilmu Pengetahuan Alam

Institut Pertanian Bogor

2020

IPB University — Bogor Indonesia —

Tipe Data



Examples:

- Marital Status
- Are you registered to vote?
- Eye Color

(Defined categories or groups)

Examples:

- Number of Children
- Defects per hour

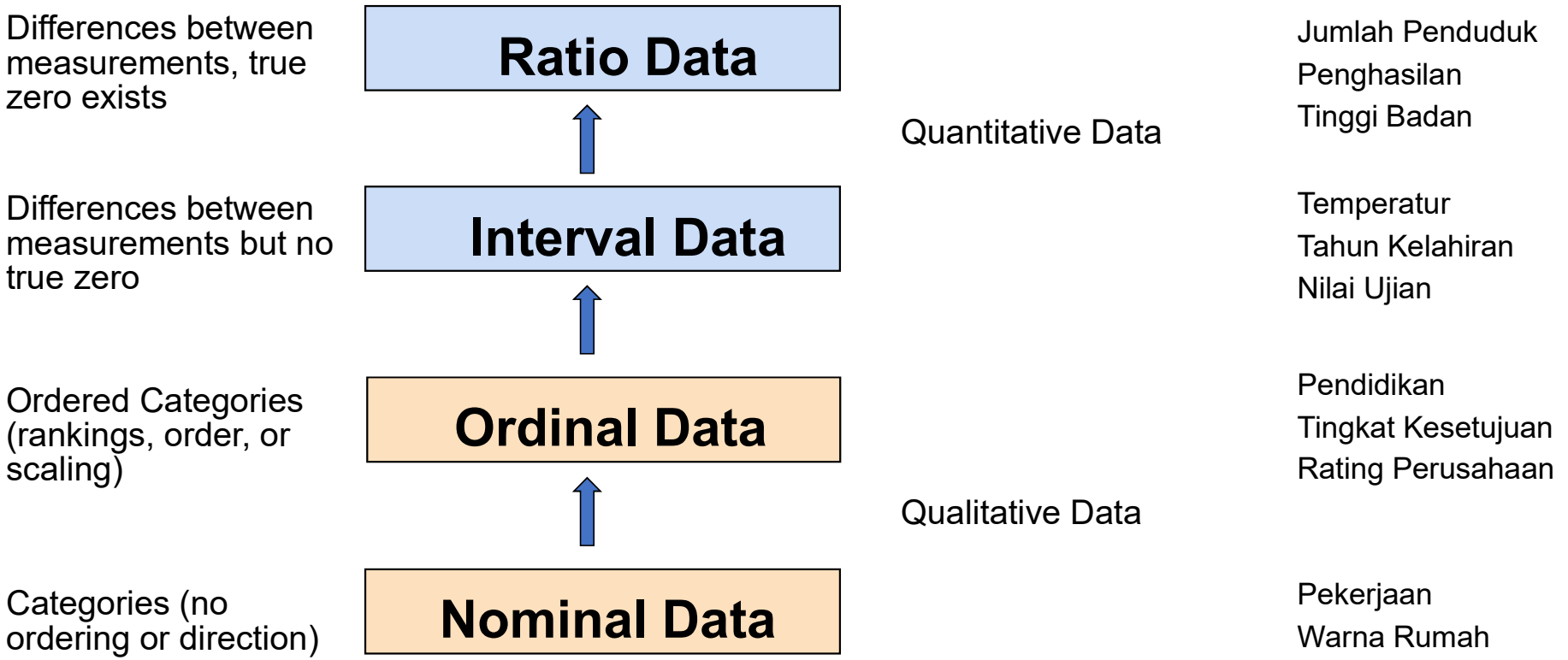
(Counted items)

Examples:

- Weight
- Voltage

(Measured characteristics)

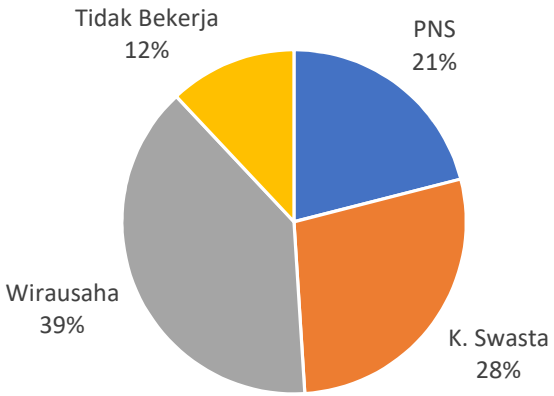
Skala Pengukuran - Scales of Measurement

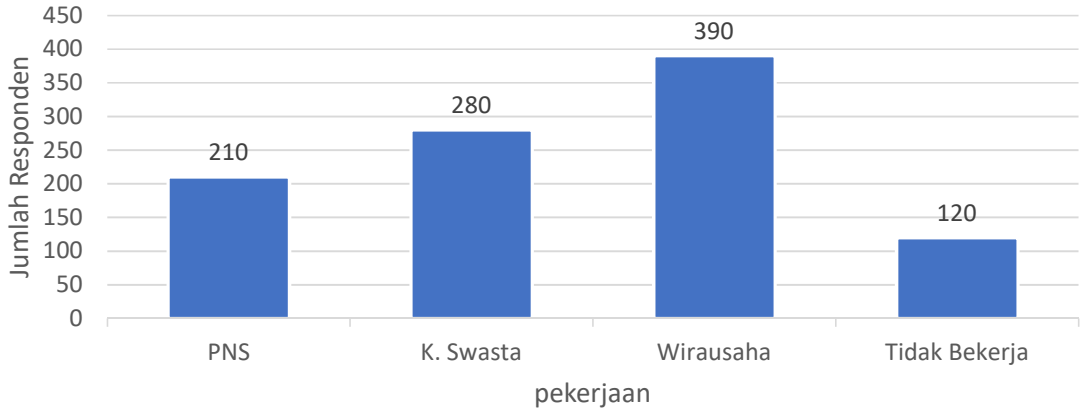


Peringkasan dan Penyajian Data Kategorik

- Tabel Sebaran Frekuensi
- Pie Chart
- Bar Chart
- Pareto Diagram

No	Pekerjaan	Frek	%
1	PNS	210	21%
2	K. Swasta	280	28%
3	Wirausaha	390	39%
4	Tidak Bekerja	120	12%
	Total	1000	





Peringkasan dan Penyajian Data Numerik

- Ukuran Pemusatan
  - Rata-Rata
  - Median
  - Modus
- Ukuran Penyebaran
  - Range: max – min
  - Ragam (variance) dan Simpangan Baku (standard deviation)
  - Jangkauan Antar Kuartil (inter quartile range): Q3 – Q1

Arithmetic Mean

- The arithmetic mean (mean) is the most common measure of central tendency
  - For a population of N values:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

Population values

Population size

  - For a sample of size n:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

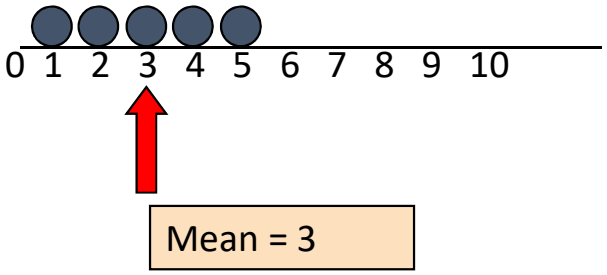
Observed values

Sample size

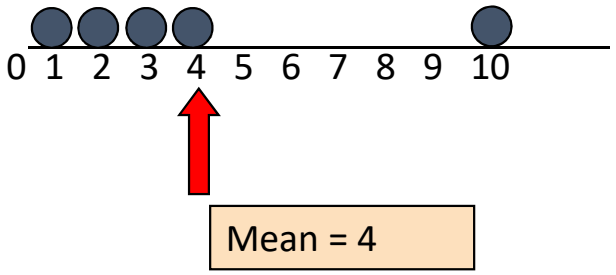
Arithmetic Mean

(continued)

- The most common measure of central tendency
- Mean = sum of values divided by the number of values
- Affected by extreme values (outliers)



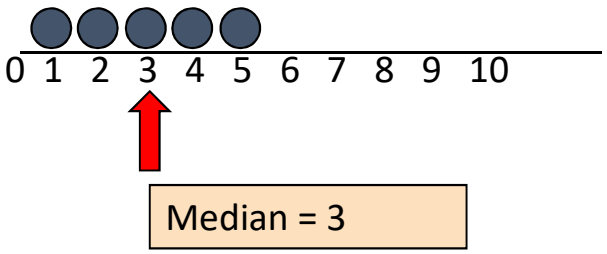
$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

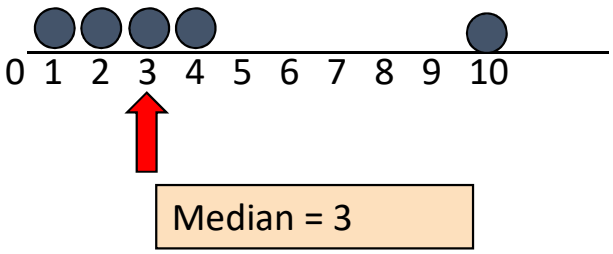


$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

Median

- In an ordered list, the median is the “middle” number (50% above, 50% below)





- Not affected by extreme values

Finding the Median

- The location of the median:

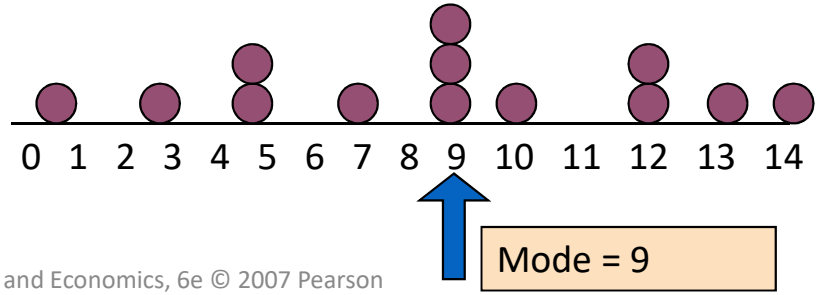
$$\text{Median position} = \frac{n+1}{2} \text{ position in the ordered data}$$

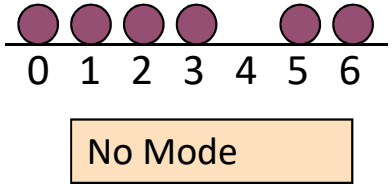
  - If the number of values is odd, the median is the middle number
  - If the number of values is even, the median is the average of the two middle numbers

- Note that $\frac{n+1}{2}$ is not the *value* of the median, only the *position* of the median in the ranked data

Mode

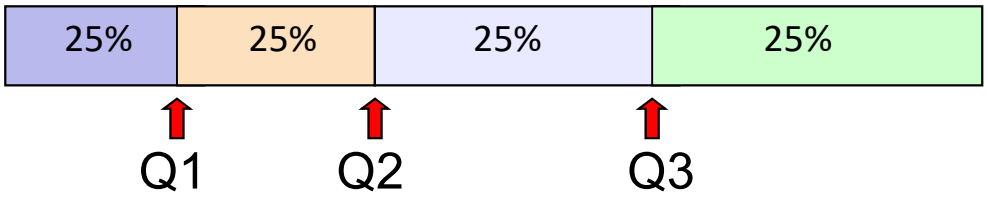
- A measure of central tendency
- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical data
- There may may be no mode
- There may be several modes





Quartiles

- Quartiles split the ranked data into 4 segments with an equal number of values per segment



- The first quartile, Q_1, is the value for which 25% of the observations are smaller and 75% are larger
- Q_2 is the same as the median (50% are smaller, 50% are larger)
- Only 25% of the observations are greater than the third quartile

Quartile Formulas

Find a quartile by determining the value in the appropriate position in the ranked data, where

First quartile position: $Q_1 = 0.25(n+1)$

Second quartile position: $Q_2 = 0.50(n+1)$
(the median position)

Third quartile position: $Q_3 = 0.75(n+1)$

where n is the number of observed values

Quartiles

- Example: Find the first quartile

Sample Ranked Data: 11 12 13 16 16 17 18 21 22

(n = 9)

Q_1 = is in the 0.25(9+1) = 2.5 position of the ranked data

so use the value half way between the 2nd and 3rd values,

so $Q_1 = 12.5$

Variance

- Average of squared deviations of values from the mean
  - Population variance:

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

Where μ = population mean

N = population size

x_i = i^{th} value of the variable x

Sample variance:

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

Where $\bar{X}$ = arithmetic mean

n = sample size

X_i = i^{th} value of the variable X

Standard Deviation

- Most commonly used measure of variation
- Shows variation about the mean
- Has the same units as the original data

- Population standard deviation:

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

Sample standard deviation:

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

Sebaran Data Numerik: Histogram

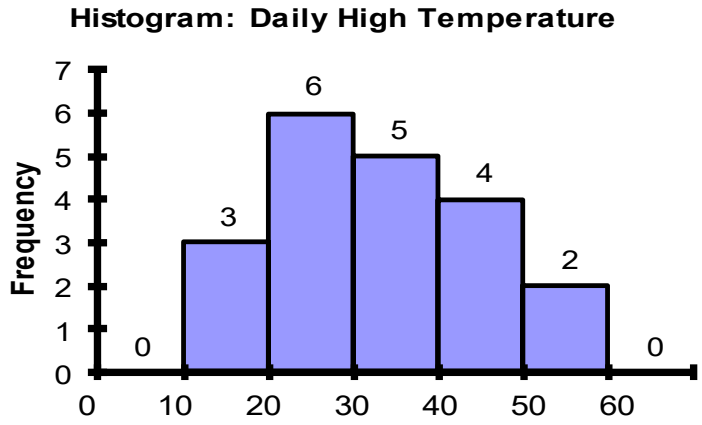
Example: A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27

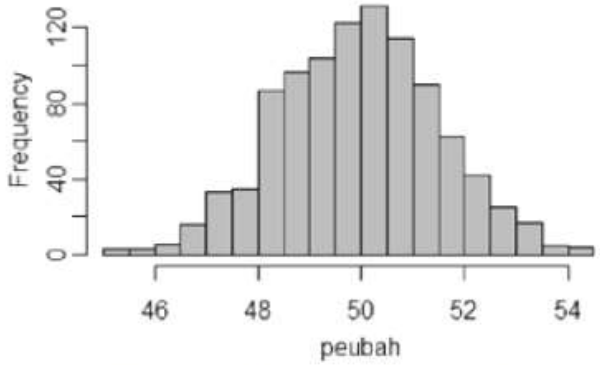
Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

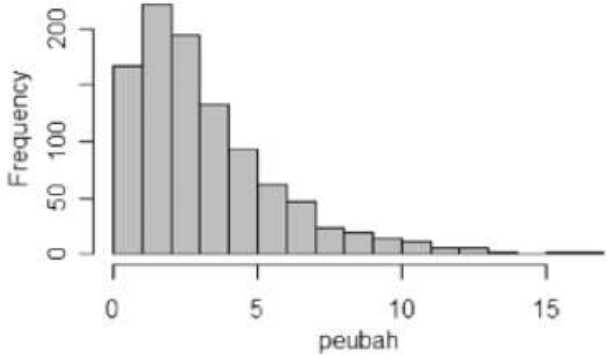
Interval	Frequency
10 but less than 20	3
20 but less than 30	6
30 but less than 40	5
40 but less than 50	4
50 but less than 60	2



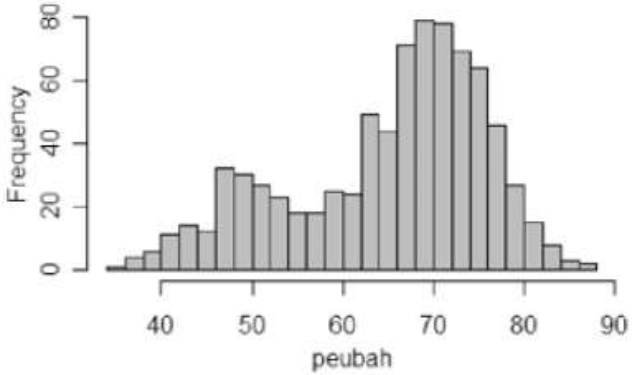
Bentuk Tipikal Sebaran Data



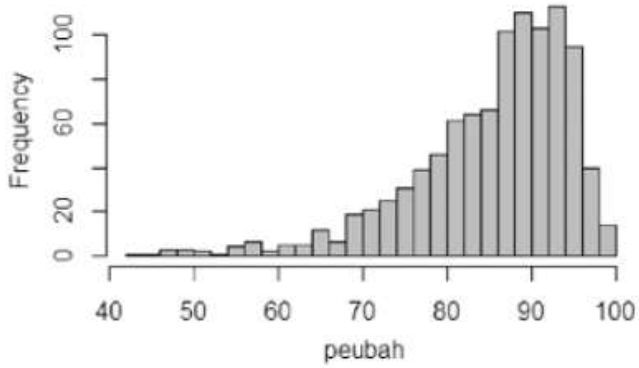
(a) histogram dengan pola simetrik



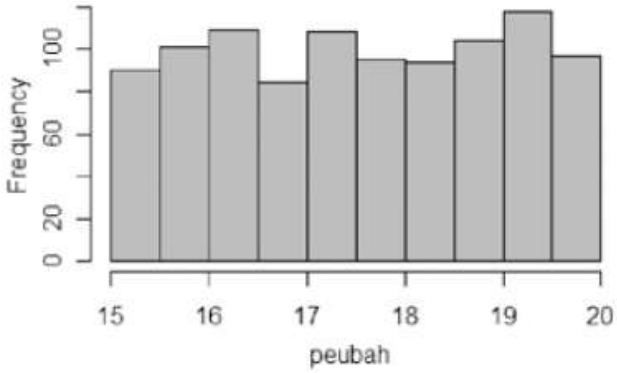
(b) histogram dengan pola menjulur ke kanan



(e) histogram yang berpola bimodal

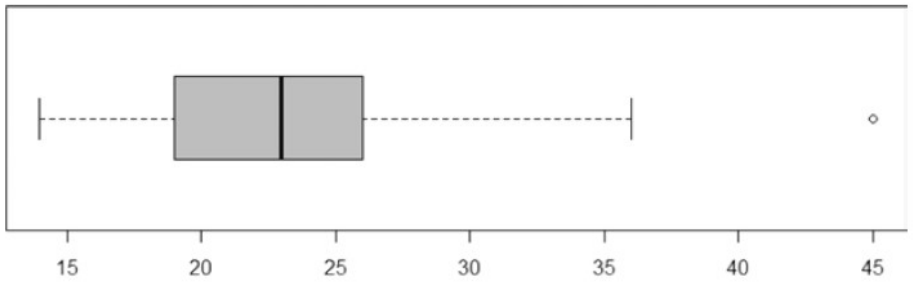


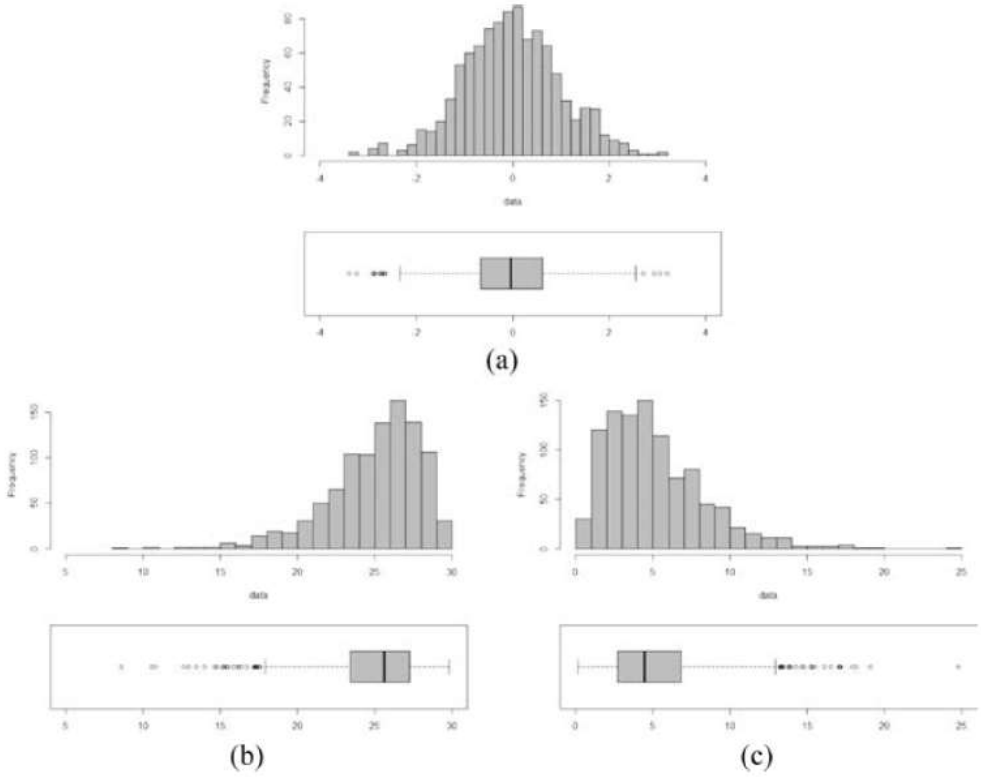
(c) histogram dengan pola menjulur ke kiri



(d) histogram dengan sebaran seragam

Sebaran Data Numerik: Boxplot





Gambar 1.9 Boxplot yang berpadanan dengan histogram

Prosedur Pembuatan Boxplot

1. Hitung beberapa statistik, meliputi:
   - statistik lima serangkai (Min, Q1, Q2, Q3, Max)
   - batas atas BA = Q3 + 3/2 (Q3-Q1)
   - batas bawah BB = Q1 - 3/2 (Q3-Q1)
2. Deteksi keberadaan pencilan (*outlier*), yaitu data yang nilainya kurang dari BB atau data yang lebih besar dari BA.
3. Gambar kotak horizontal, dengan batas kiri Q1 sampai batas kanan Q3, dan letakkan tanda garis di tengah kotak pada posisi Q2.
4. Tarik garis ke kanan, mulai dari Q3 sampai data terbesar di dalam batas atas.
5. Tarik garis ke kiri, mulai dari Q1 sampai data terkecil di dalam batas bawah.
6. Tandai pencilan dengan lingkaran kecil.

Terima Kasih

IPB University
Bogor Indonesia

IPB University

Bogor Indonesia

Inspiring Innovation with Integrity

in Agriculture, Ocean and Biosciences for a Sustainable World