

# STA513 – Analisis Statistika untuk Bisnis, Ekonomi, dan Industri

Semester Ganjil 2020/2021

## PERTEMUAN #1 Pendahuluan

disusun oleh:  
**Bagus Sartono**  
[bagusco@gmail.com](mailto: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 —





IPB University  
— Bogor Indonesia —

# Tim Pengajar



Prof. Dr. Hari Wijayanto



Dr. Bagus Sartono



# Perkenalan: Bagus Sartono



1989 SDN Kencong 1, Jember – Jawa Timur  
1992 SMPN 1 Kencong, Jember – Jawa Timur  
1995 SMAN 2 Lumajang – Jawa Timur  
2000 Sarjana Sains, Statistika IPB  
2004 Magister Sains, Statistika IPB  
2012 PhD in Applied Economics, Universiteit Antwerpen

Google Scholar: <https://scholar.google.com/citations?user=Yc1l77YAAAAJ&hl=en>

Scopus: <https://www.scopus.com/authid/detail.uri?authorId=54382487700>

SINTA: <http://sinta.ristekbrin.go.id/authors/detail?id=5975411&view=overview>

Linked-In: <https://www.linkedin.com/in/bagus-sartono-17b85357/>



# Deskripsi Mata Kuliah

- **STA513 Statistika untuk Bisnis, Ekonomi dan Industri**
- Mata kuliah ini ditujukan untuk memberikan landasan tentang statistika yang berguna dalam penelitian di bidang bisnis, ekonomi dan industri. Materi meliputi metode pengumpulan data, pemodelan dan analisis data, prinsip pendugaan parameter dan pengujian hipotesis satu populasi serta dua populasi. Untuk pengumpulan data ditekankan pada percontohan (*sampling*), khususnya percontohan yang sering dilakukan dalam bisnis, ekonomi dan industri seperti contoh acak sederhana (*simple random sample*), contoh berstrata (*stratified sample*), contoh bertahap (*multistage sample*), serta contoh tak-berpeluang (*non-probaility sampling*), sedangkan untuk analisis ditekankan pada analisis regresi, analisis deret waktu dan peramalan, pengendalian mutu dan statistika non-parametrik.



# Silabus

NO	POKOK BAHASAN	SUB POKOK BAHASAN	PERTEMUAN
1	Pendahuluan	<ol style="list-style-type: none"><li>1. Pengertian Statistika</li><li>2. Apa itu Statistika?</li><li>3. Skala pengukuran peubah</li><li>4. Penyajian dan peringkasan data</li><li>5. Diagram batang, diagram titik, diagram dahan daun dan diagram kotak-garis</li><li>6. Mengapa contoh? BUKAN cacah lengkap?</li></ol>	1 x
2	Sebaran Penarikan Contoh	<ol style="list-style-type: none"><li>1. Ukuran pusat dan ukuran pencaran</li><li>2. Peubah acak dan sebaran normal</li><li>3. Sebaran percontohan (rataaan, proporsi contoh, ragam contoh)</li><li>4. Dalil limit pusat</li></ol>	1x
3	Inferensi Statistika	<ol style="list-style-type: none"><li>1. Pendugaan Parameter</li><li>2. Pengujian hipotesis dan selang kepercayaan (nilaitengah satu populasi, nilaitengah dua populasi)</li><li>3. Pembandingan dua populasi</li></ol>	2x



# Silabus

NO	POKOK BAHASAN	SUB POKOK BAHASAN	PERTEMUAN
4	One-Way ANOVA	<ol style="list-style-type: none"><li>1. Penggunaan one-way anova untuk pengujian kesamaan nilai tengah dua atau lebih populasi</li><li>2. Uji Beda Nyata Terkecil dan Uji Beda Nyata Jujur Tukey</li></ol>	1x
5	Statistika Non-Parametrik	<ol style="list-style-type: none"><li>1. Perlunya pendekatan uji non-parametrik</li><li>2. Uji Tanda</li><li>3. Uji Peringkat Bertanda Wilcoxon</li><li>4. Uji Mann-Whitney</li><li>5. Uji Kruskal-Wallis</li></ol>	1x
6	Analisis Data Kategorik	<ol style="list-style-type: none"><li>1. Tabel Kontingensi</li><li>2. Uji Kesamaan Proporsi Chisq-Test</li><li>3. Uji Kebebasan Chisq-Test</li></ol>	1x



# Silabus

NO	POKOK BAHASAN	SUB POKOK BAHASAN	PERTEMUAN
7	Teknik sampling	<ol style="list-style-type: none"><li>1. Mengapa bekerja dengan contoh</li><li>2. Teknik penarikan contoh berpeluang<ol style="list-style-type: none"><li>a. Simple random sampling</li><li>b. Systematic random sampling</li><li>c. Stratified random sampling</li><li>d. Multistage random sampling</li></ol></li><li>3. Teknik penarikan contoh tak-berpeluang<ol style="list-style-type: none"><li>a. Quota sampling</li><li>b. Snowball sampling</li><li>c. dll</li></ol></li></ol>	2x
8	Analisis Regresi Linear	<ol style="list-style-type: none"><li>1. Analisis Korelasi</li><li>2. Regresi Linear Sederhana</li><li>3. Pendugaan model regresi</li><li>4. Pengujian dan diagnostik model</li><li>5. Regresi linear berganda</li><li>6. Masalah multikolinearitas</li></ol>	2x



# Silabus

NO	POKOK BAHASAN	SUB POKOK BAHASAN	PERTEMUAN
9	Analisis Regresi Logistik Biner	1. Pengenalan model regresi logistik biner 2. Pendugaan koefisien model regresi logistik 3. Pengujian	1x
10	Pengantar Analisis Data Deret Waktu	1. Pengenalan data deret waktu 2. Pemulusan data deret waktu 3. Regresi trend	1x
11	Pengantar Statistika Pengendalian Mutu	1. Pengenalan statistika pengendalian mutu dan tools yang digunakan 2. Pengenalan bagan kendali 3. Pengenalan sampling penerimaan	1x





# Bahan Bacaan

- David R. Anderson; Dennis J. Sweeney; Thomas A. Williams; James J. Cochran. 2017. *Statistics for Business & Economics* (13rd ed). South Western Educational Publishing



# Outline

1. Pengertian Statistika
2. Beberapa istilah penting dalam statistika
3. Mengapa contoh? BUKAN cacah lengkap?
4. Skala pengukuran peubah
5. Penyajian dan peringkasan data
6. Diagram batang, diagram titik, diagram dahan daun dan diagram kotak-garis



## Survei Cak Lontong #1:

Saya men-survei 100 orang dan seratus-seratusnya akan memilih Pak Prabowo sebagai presiden

Survei-nya saya lakukan di kantor Gerindra



## Survei Cak Lontong #2:

Saya men-survei 100 lelaki mapan dan 100% belum pernah selingkuh.

Saat survei, responden didampingi istri masing-masing.

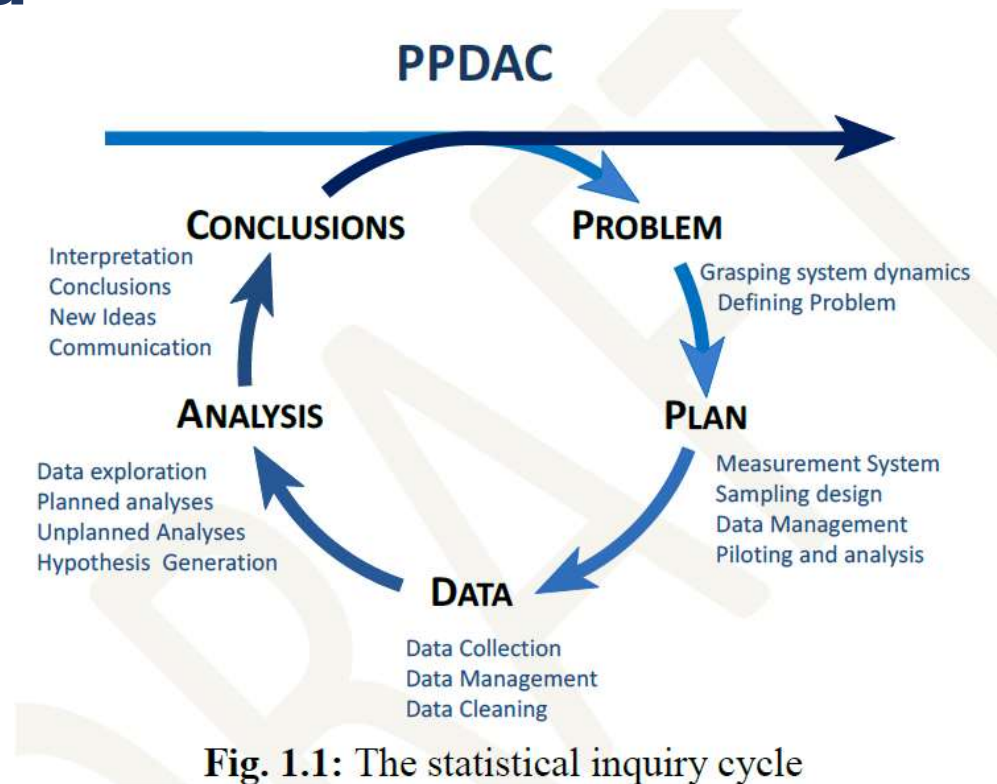




# Statistika

- ilmu yang mempelajari metode pengumpulan data, analisis data, dan penarikan kesimpulan yang didasarkan pada hasil analisis
- as defined by the American Statistical Association (ASA) “***is the science of learning from data, and of measuring, controlling and communicating uncertainty.***”

# Statistika



**Fig. 1.1:** The statistical inquiry cycle

C.J. Wild and M. Pfannkuch. 1999. Statistical Thinking in Empirical Enquiry .  
*International Statistical Review* 67: 223-265,



# Statistik di Sekitar Kita

-0,05

Inflasi, Agustus 2020

-5,32

Persen

Pertumbuhan Ekonomi, Triwulan II 2020

9,78

Persen

Persentase Penduduk Miskin, Maret 2020

4,99

Persen

Tingkat Pengangguran Terbuka, Februari  
2020

0,38

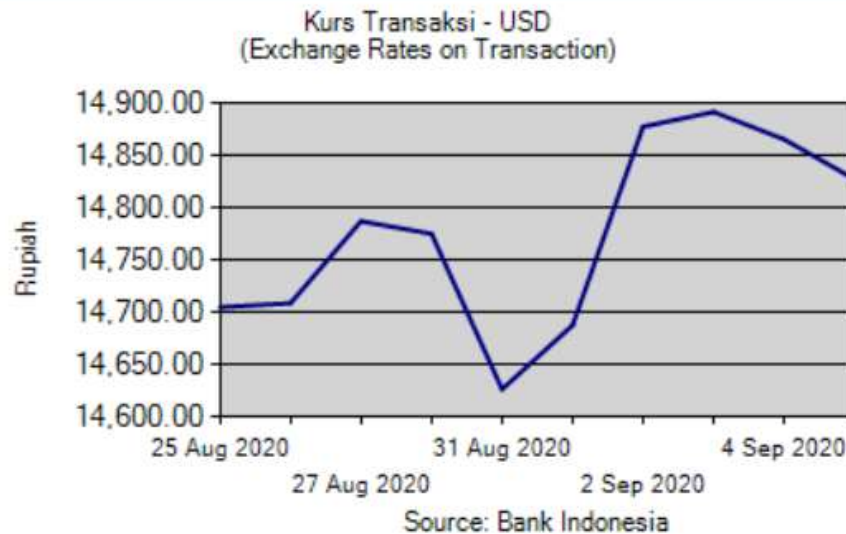
Gini Rasio, Maret 2020

71,92

Indeks Pembangunan Manusia, 2019

<https://www.bps.go.id/indikator/>  
[8 Sep 2020]

# Statistik di Sekitar Kita



<https://www.bi.go.id/id/moneter/informasi-kurs/transaksi-bi/Default.aspx>







# Beberapa Istilah

- **Populasi**

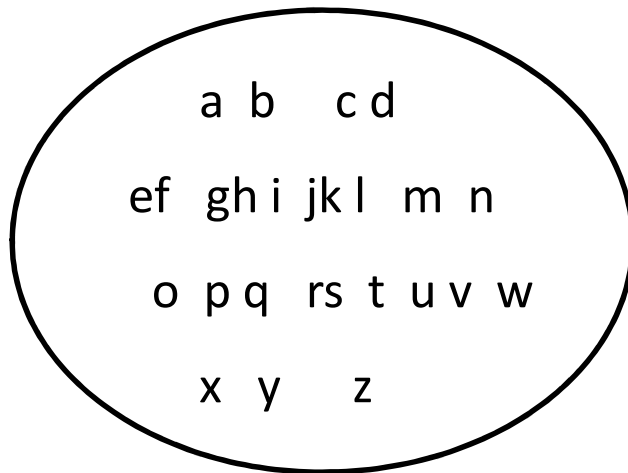
- is the collection of all items of interest or under investigation
  - $N$  represents the population size

- **Contoh**

- is an observed subset of the population
  - $n$  represents the sample size

# Beberapa Istilah

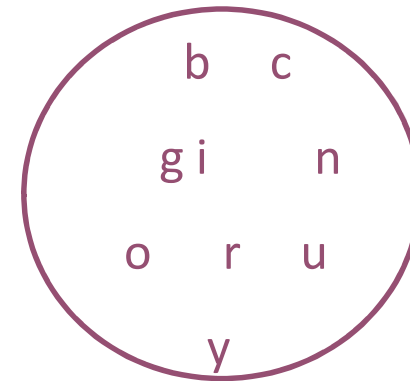
## Populasi



Nilai yang dihitung menggunakan data populasi disebut **parameter**

Dalam banyak hal... nilai parameter **tidak diketahui**

## Contoh



Nilai yang diperoleh dari perhitungan menggunakan data contoh disebut **statistik**

Statistik merupakan **penduga** bagi parameter



# Beberapa Istilah

- Sensus (cacah lengkap)
  - Proses mengumpulkan data dari seluruh anggota populasi
- Survei
  - Proses pengumpulan data dari contoh



Survei Sosial Ekonomi Nasional

# Mengapa Contoh?

- Bekerja dengan contoh menghemat sumberdaya (tenaga, waktu, biaya)
- Ada proses pengamatan yang bersifat merusak (misal pengamatan kualitas kandungan produk), tidak mungkin dilakukan pada seluruh objek.
- Ada situasi dimana populasi bertambah terus atau objeknya bergerak
- Mengamati dengan contoh sudah cukup untuk memperoleh gambaran mengenai populasi

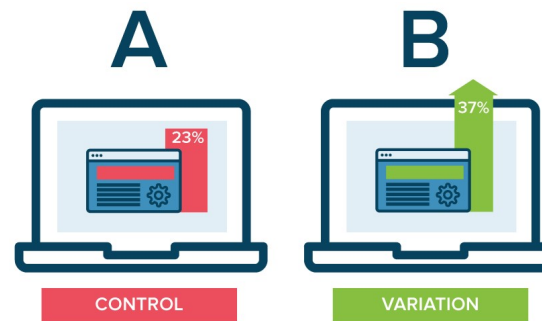


<https://dissolve.com/>

# Pengumpulan Data



survei



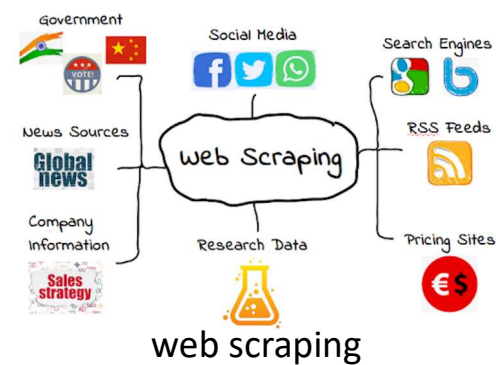
percobaan



observasi



database  
(administrasi, transaksi, tangkapan aktivitas)





# Statistika Deskriptif dan Inferensia

- **Statistika deskriptif**

- penyajian data (tabel, grafik atau gambar, ukuran deskriptif)
- merupakan upaya agar info dapat ditangkap dg jelas

- **Inferensi statistika**

- merupakan proses penarikan kesimpulan secara induktif

# Ilustrasi

- kajian efektivitas obat tekanan darah utk hipertensi moderat (90– 150)
- 389 pasien, dibagi dua
  - 196 sbg kelompok kontrol (*placebo*)
  - 193 sbg kelompok yang diberikan obat senyawa aktif

Uraian		Aktif	Kontrol
1.	Jumlah kasus dengan tekanan darah $\geq 130$	0	24
2.	Jumlah kasus kejadian hipertensif	37	89
3.	Jumlah kasus serangan jantung dan penyakit jantung koroner	35	38
4.	Jumlah kasus pembengkakan jantung	12	20
5.	Jumlah pasien	193	196

sumber: KA Notodiputro



Uraian		Aktif	Kontrol
1.	Jumlah kasus dengan tekanan darah $\geq 130$	0	24
2.	Jumlah kasus kejadian hipertensif	37	89
3.	Jumlah kasus serangan jantung dan penyakit jantung koroner	35	38
4.	Jumlah kasus pembengkakan jantung	12	20
5.	Jumlah pasien	193	196

- Dari 193 pasien kelompok aktif, 19% mengalami kejadian hipertensif
  - Ini merupakan statistika deskriptif (hanya fakta)
- Pasien yang mengonsumsi obat tekanan darah lebih kecil risiko utk terkena pembengkakan jantung drpd yang tidak mengonsumsi obat tersebut
  - Ini merupakan pernyataan inferensi (kita melampaui fakta!)





# Data di sekitar kita



## JUMLAH TERPAPAR COVID-19 DI INDONESIA

Update 6 September 2020 Pukul 12.00 WIB

89.701

SUSPEK

27.979

SPESIMEN

34

PROVINSI

489

KAB/KOTA

⬆ 3.444

194.109

KONFIRMASI

⬆ 2.174

138.575

SEMBUH

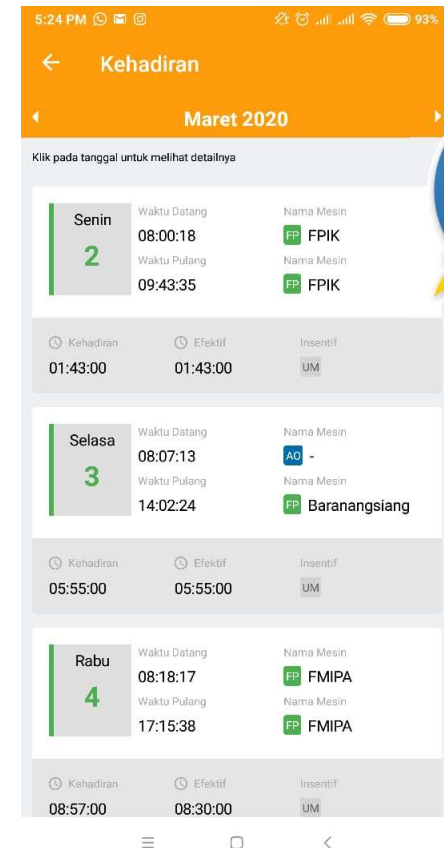
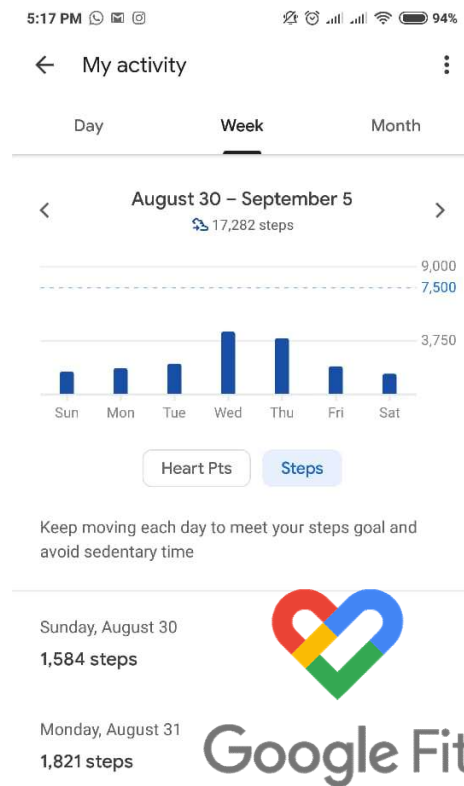
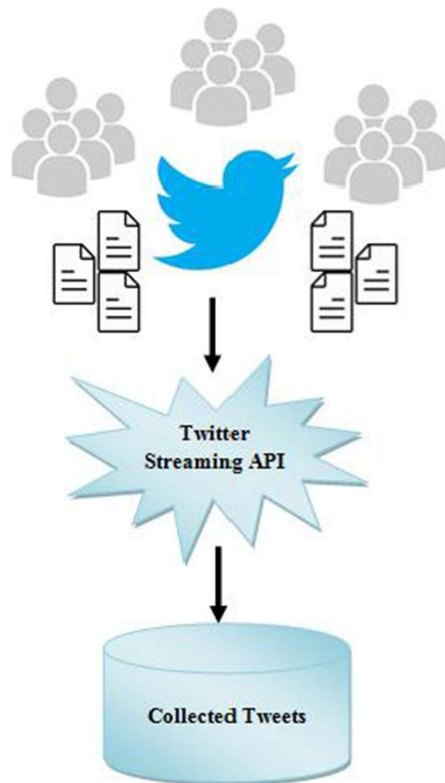
⬆ 85

8.025

MENINGGAL



# Data di sekitar kita

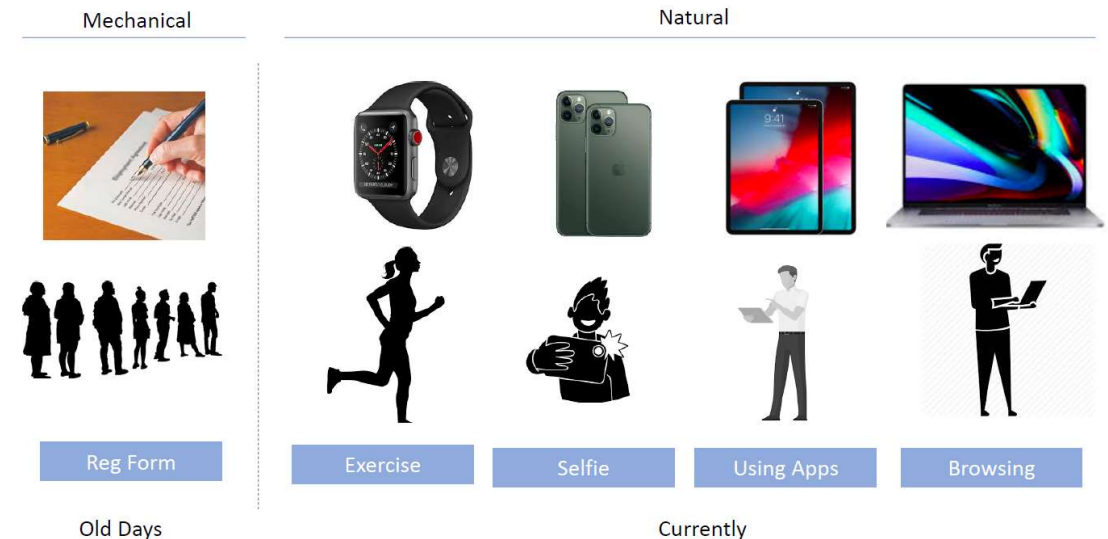


IPB Mobile

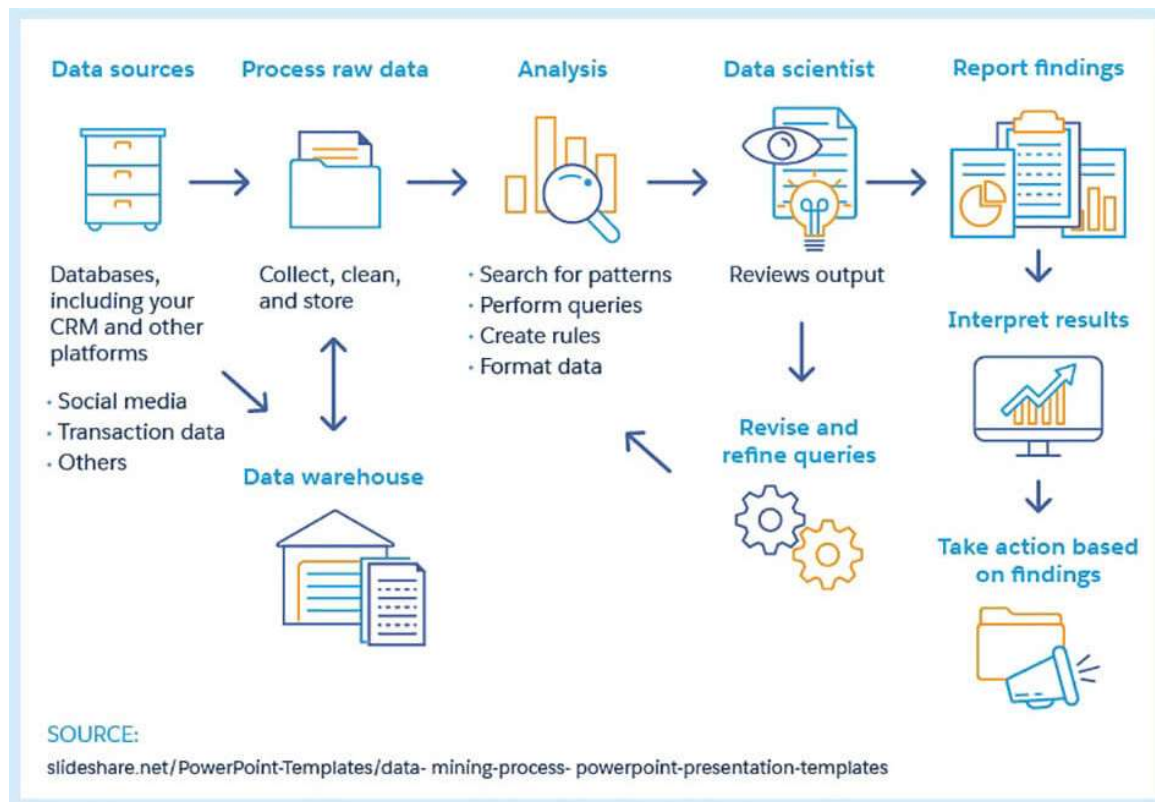
# Big Data

- Revolusi pengukuran:
  - tidak lagi selalu mekanistik, tapi juga bersifat natural
  - Yang mekanistik juga semakin cepat dilakukan, semakin banyak, dan variatif
- 3V's:
  - Volume: ukurannya semakin besar
  - Variety: tidak hanya angka, ada juga data teks, gambar, suara, video, dll
  - Velocity: bertambah dengan cepat

## Data Collections Activity



# Data Mining dan Analitika (Analytics)



- metode untuk pembuatan keputusan berdasarkan informasi dari database (besar)
- kombinasi pengetahuan statistika, pemodelan, dan ilmu komputer untuk mengubah data menjadi informasi
- tidak melupakan konteks (business flow)



# Analitika

## DESCRIPTIVE

Explains what  
happened



## DIAGNOSTIC

Explains why it  
happened



## PREDICTIVE

Forecasts what might  
happen



## PRESCRIPTIVE

Recommends  
an action

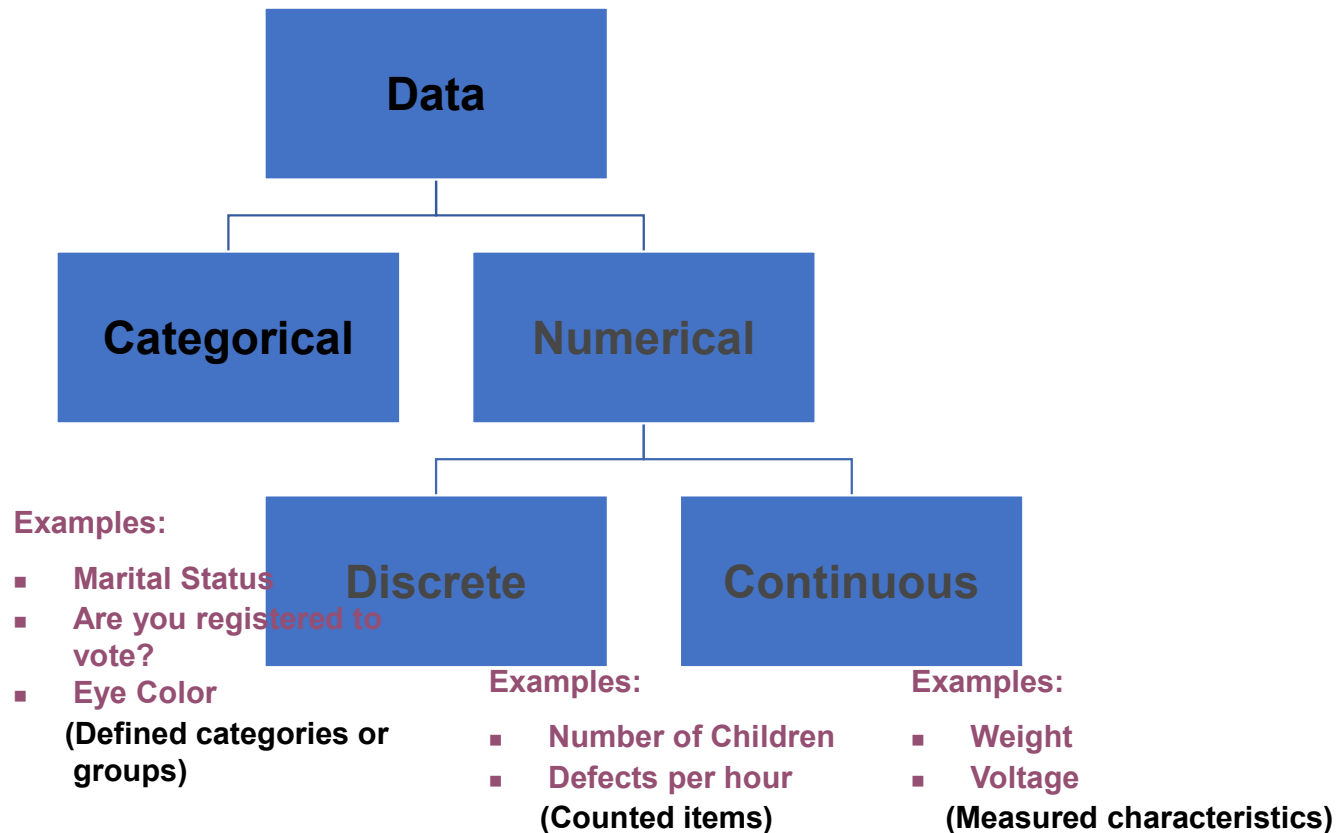




# Amatan (objek, individu) dan Peubah

- Amatan adalah individu sumber data yang dikumpulkan.
- Peubah (variable) adalah karakteristik yang menjadi minat studi dari amatan.
- Misal:
  - Amatan: mahasiswa
  - Peubah: Usia, Jenis Kelamin, Asal Daerah, Nilai Ujian
- Misal:
  - Amatan: desa
  - Peubah: luas wilayah, jumlah penduduk, keberadaan sungai, ketinggian lokasi
- Misal:
  - Amatan: tanaman padi
  - Peubah: tinggi tanaman, produktivitas, umur panen

# Tipe Data





# Skala Pengukuran - Scales of Measurement

Differences between measurements, true zero exists

**Ratio Data**



Differences between measurements but no true zero

**Interval Data**



Ordered Categories (rankings, order, or scaling)

**Ordinal Data**



Categories (no ordering or direction)

**Nominal Data**

Quantitative Data

Qualitative Data

Jumlah Penduduk  
Penghasilan  
Tinggi Badan

Temperatur  
Tahun Kelahiran  
Nilai Ujian

Pendidikan  
Tingkat Kesetujuan  
Rating Perusahaan

Pekerjaan  
Warna Rumah  
Spesies Tanaman





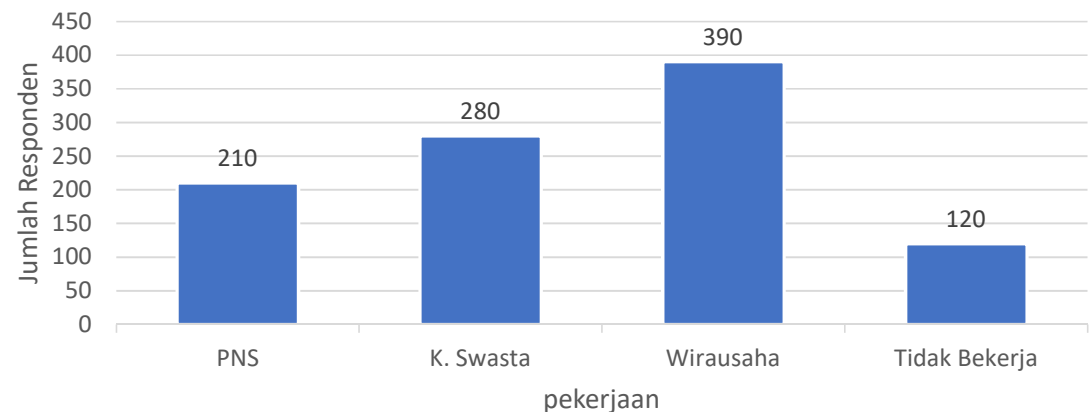
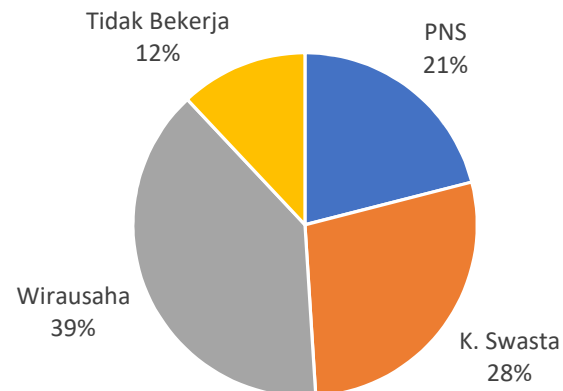
# Data Cross Sectional dan Data Deret Waktu

- **Cross-sectional Data:** diamati pada beberapa objek pada satu waktu yang sama atau hampir sama
- **Time Series Data:** diamati pada satu objek pada beberapa waktu (dengan interval yang teratur)
- Gabungan keduanya → **Panel Data**

# 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:  $\text{max} - \text{min}$
  - 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:

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

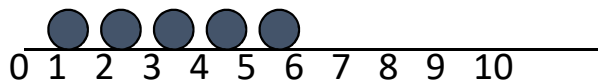
Observed values

Sample size

# Arithmetic Mean

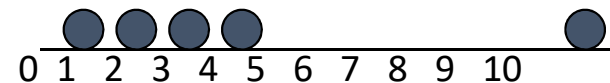
(continued)

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



Mean = 3

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

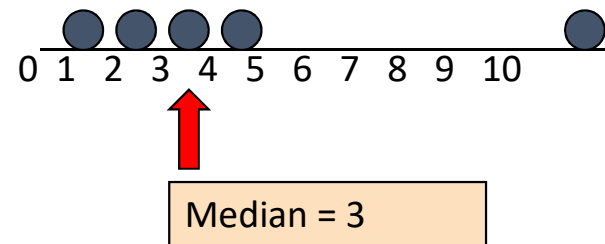
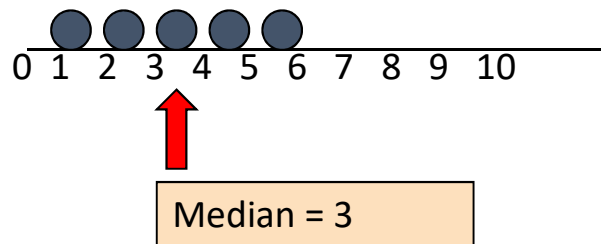


Mean = 4

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

# Median

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



- Not affected by extreme values



# Finding the Median

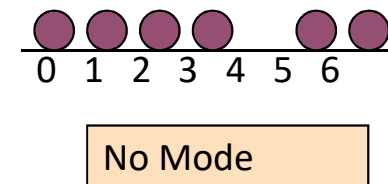
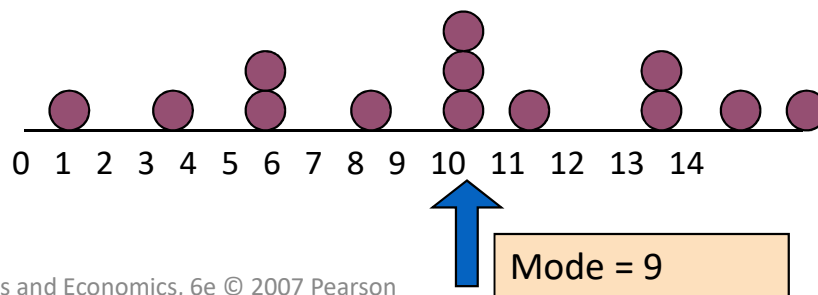
- The location of the median:

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

- 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

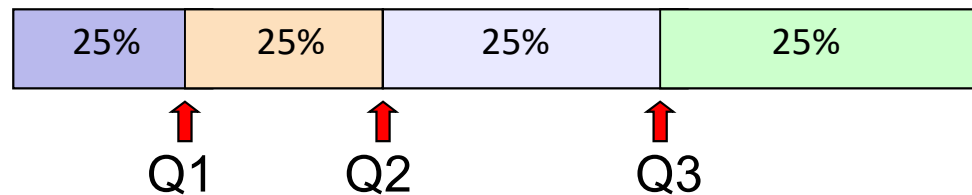
- A measure of central tendency
- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical data
- There may may be no mode
- There may be several modes





# 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 2<sup>nd</sup> and 3<sup>rd</sup> values,

so  $Q_1 = 12.5$

# Variance

- Average of squared deviations of values from the mean

- Population variance:

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

Where  $\mu$  = population mean

$N$  = population size  $\mu$

$x_i$  =  $i^{\text{th}}$  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

$x_i$  =  $i^{\text{th}}$  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:

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

- Sample standard deviation:

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

# Sebaran Data Numerik: Histogram

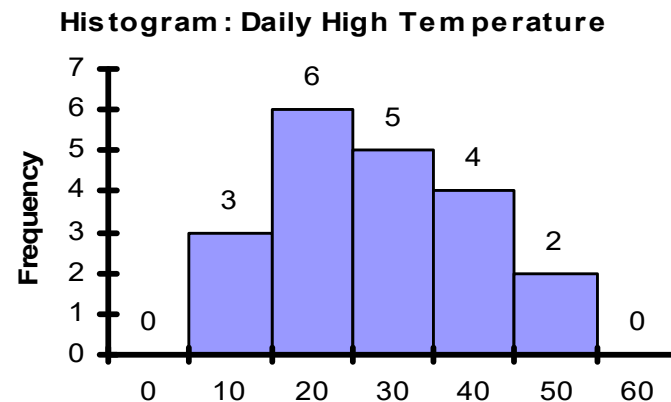
**Example:** A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature

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

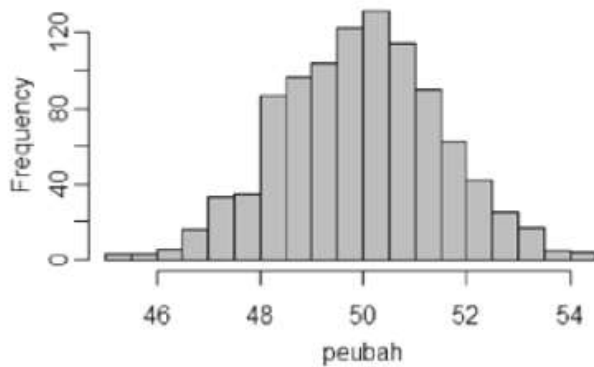
**Data in ordered array:**

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

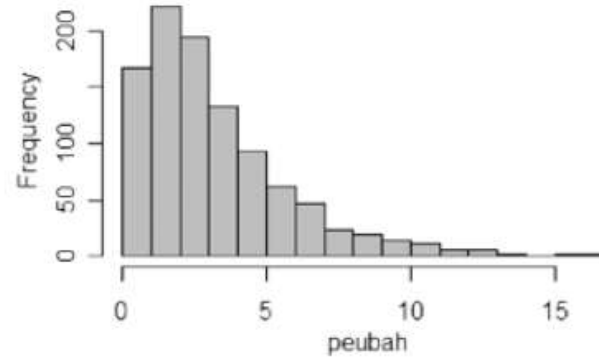
Interval	Frequency
10 but less than 20	3
20 but less than 30	6
30 but less than 40	5
40 but less than 50	4
50 but less than 60	2



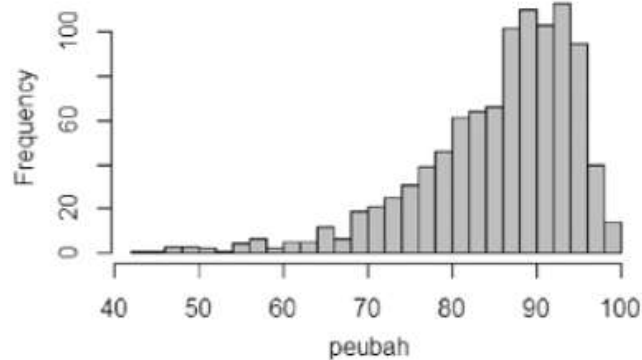
# Bentuk Tipikal Sebaran Data



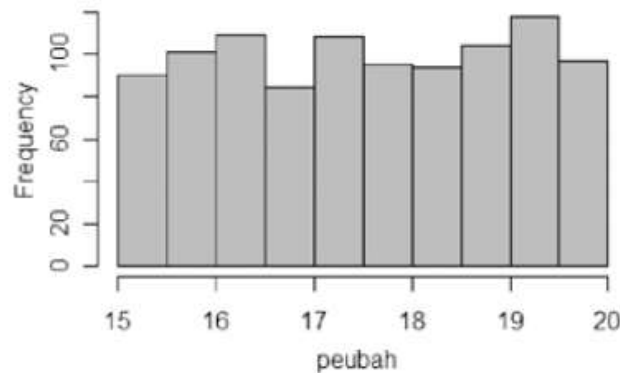
(a) histogram dengan pola simetrik



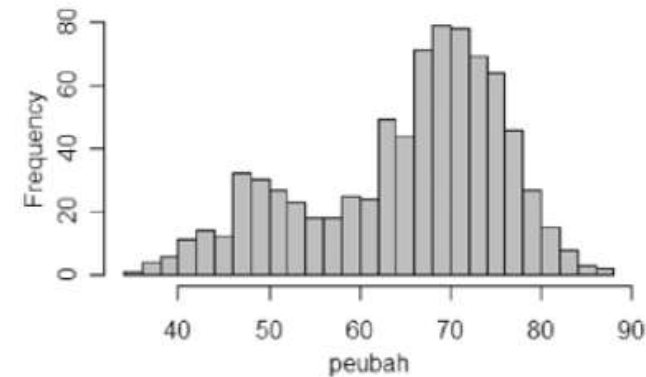
(b) histogram dengan pola menjulur ke kanan



(c) histogram dengan pola menjulur ke kiri

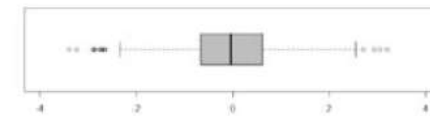
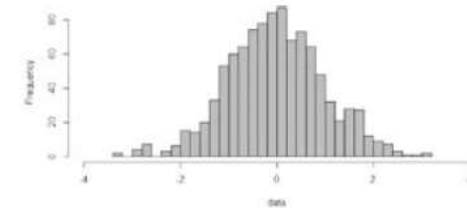
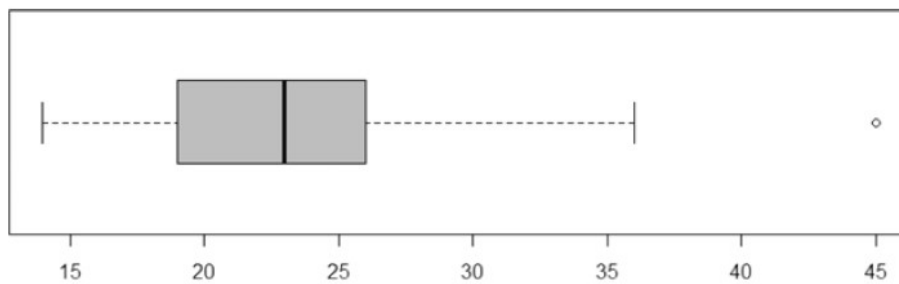


(d) histogram dengan sebaran seragam

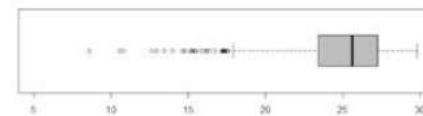
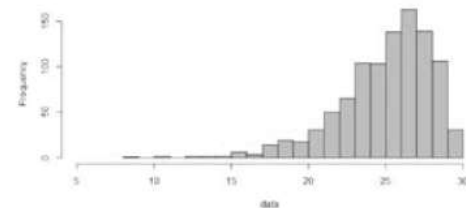


(e) histogram yang berpola bimodal

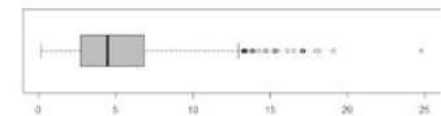
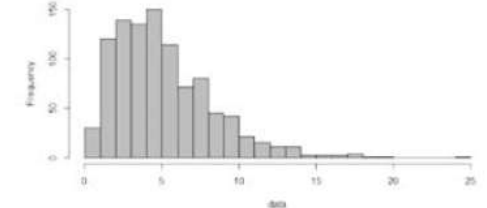
# Sebaran Data Numerik: Boxplot



(a)



(b)



(c)

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  $BA = Q3 + \frac{3}{2} (Q3 - Q1)$
  - batas bawah  $BB = Q1 - \frac{3}{2} (Q3 - Q1)$
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