



LEARNING PROGRESS REVIEW

Week 8

Entropy Team

A large, light gray semi-circle graphic located in the bottom right corner of the slide.

OUR TEAM

Entropy Team



Adhang Muntaha Muhammad

<https://www.linkedin.com/in/adhangmuntaha/>



Aziz Fauzi

<https://www.linkedin.com/in/aziz-fauzi-a6904711b/>



Iwan Wahyu

<https://www.linkedin.com/in/iwan-wahyu-setyawan-506809183>



Marcellina Alvita F

<https://www.linkedin.com/in/marcellina-alvita-faustina-63a284226>



Ramadhan Luthfan

<https://www.linkedin.com/in/luthfan-mahathir-91369b18b>

DAFTAR ISI

1.

Basic Statistics

Materi statistik (dasar)

2.

Intermediate Statistics

Materi statistik (menengah)

3.

Advanced Statistics

Materi statistik (lanjutan)

01

BASIC STATISTICS

Materi statistik
(dasar)

Pentingnya Statistik



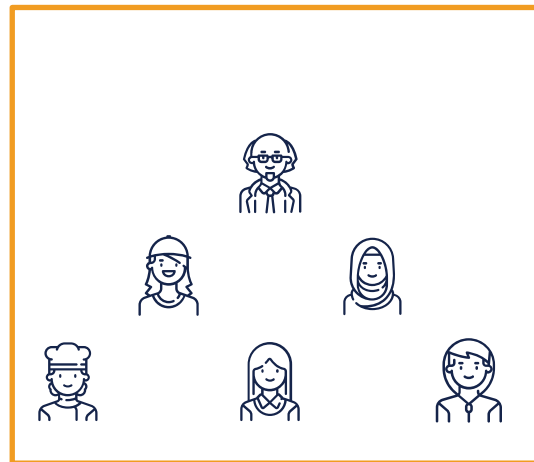
Statistik dapat digunakan untuk **membuat *summary*** dari data besar menjadi kecil, sehingga seolah-olah bisa **menggeneralisasi populasi**

Populasi dan Sampel

Populasi

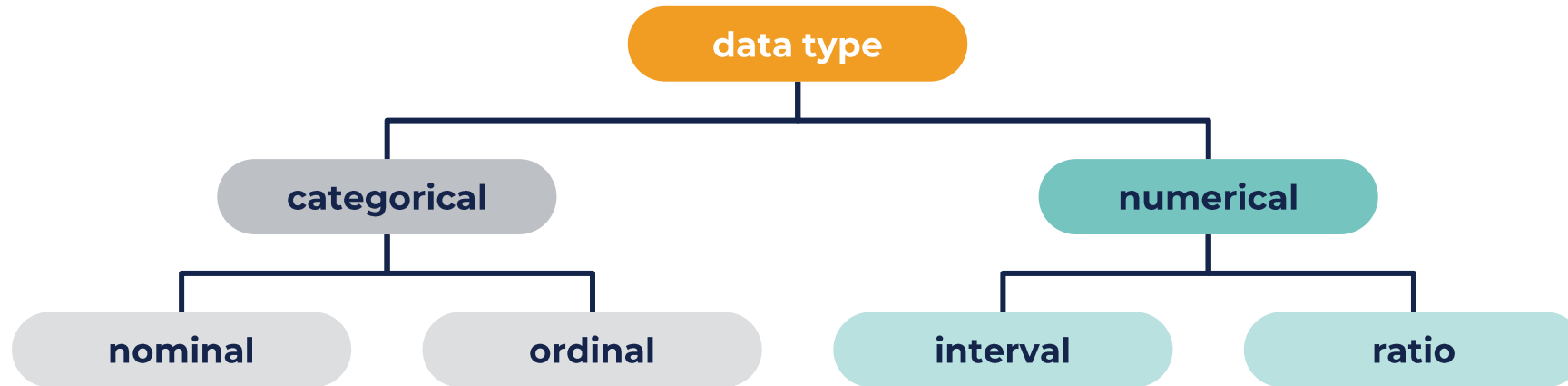


Sampel



- **Populasi** merupakan **keseluruhan** elemen observasi
- **Sampel** merupakan **bagian** (*subset*) dari populasi
- **Parameter** adalah **ukuran** yang menggambarkan keseluruhan **populasi**
- **Statistik** adalah **ukuran** yang menggambarkan **sampel**

Tipe Data



	Nominal	Ordinal	Interval	Ratio
Deskripsi	Kategori tidak memiliki bobot	Kategori memiliki bobot	Tidak memiliki <i>true zero</i> , sehingga bisa memiliki nilai negatif	Memiliki <i>true zero</i> , sehingga tidak memiliki nilai negatif
Contoh	Jenis kelamin	Tingkat kepuasan	Suhu	Usia

Domain Statistik

Descriptive Statistics

- Digunakan untuk **menggambarkan** atau merangkum **karakteristik** kumpulan **data**, sehingga pola dari data dapat dilihat
- Erat kaitannya dengan *central tendency* dan persebaran data

Inferential Statistics

- Digunakan untuk membuat **generalisasi** pada sebuah **populasi** menggunakan sampel
- Erat kaitannya dengan *sampling*, estimasi parameter, dan uji hipotesis

Statistik Deskriptif

Measure of central tendency

- Digunakan untuk melihat **seberapa memusat** suatu data
- Beberapa ukuran yang digunakan:
 - Rata-rata (*mean*)
 - Nilai tengah (*median*)
 - Modus (*mode*)

Measure of spread

- Digunakan untuk melihat **seberapa menyebar** suatu data
- Beberapa ukuran yang digunakan:
 - Minimum
 - Maksimum
 - Rentang
 - Kuantil
 - Varians
 - Standar deviasi

Rata-rata

```
in | print(data['age'].mean())
   | print(data['score'].mean())
out | 24.25
    | 67.5
```

- Digunakan untuk menghitung rata-rata
- Sangat **sensitif** terhadap **outlier**

	name	age	score
0	Entropy	25	80
1	Team	23	85
2	Digital	24	85
3	Skola	25	20



$$mean = \bar{x} = \frac{\sum_i^n x}{n}$$

$$\bar{x}_{age} = \frac{25 + 23 + 24 + 25}{4} = 24.25$$

$$\bar{x}_{score} = \frac{80 + 85 + 85 + 20}{4} = 67.5$$

Median

```
in | print(data['age'].median())  
   | print(data['score'].median())  
out | 24.5  
    | 82.5
```

- Digunakan untuk mencari nilai tengah
- **Tidak sensitif** terhadap **outlier**

	name	age	score
0	Entropy	25	80
1	Team	23	85
2	Digital	24	85
3	Skola	25	20



age	23	24	25	25
score	20	80	85	85

$$median_{age} = \frac{(24 + 25)}{2} = 24.5$$

$$median_{score} = \frac{(80 + 85)}{2} = 82.5$$

Modus

```
in | print(data['age'].median())  
   | print(data['score'].median())  
out | 0      25  
    | dtype: int64  
    | 0      85  
    | dtype: int64
```

- Digunakan untuk mencari nilai dengan **frekuensi terbanyak**

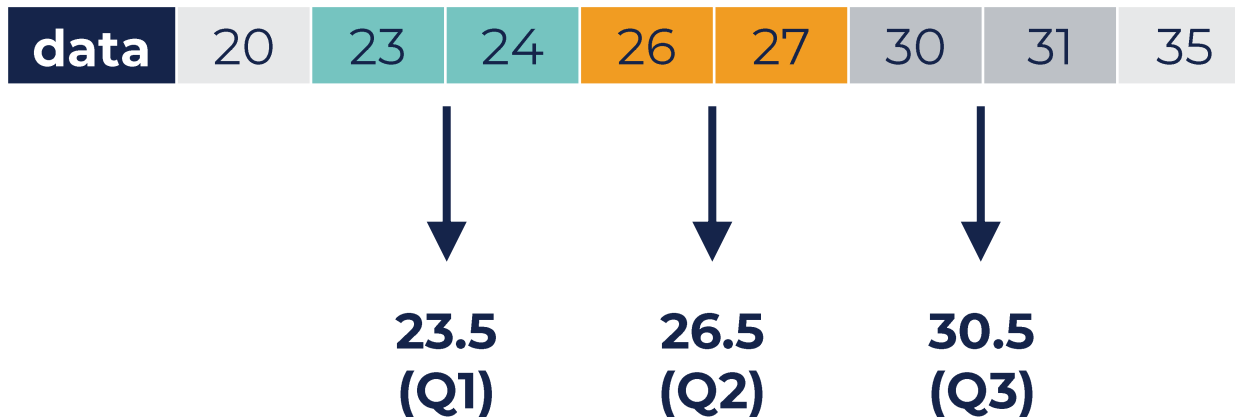
	name	age	score
0	Entropy	25	80
1	Team	23	85
2	Digital	24	85
3	Skola	25	20



age		score	
nilai	jumlah	nilai	jumlah
23	1	20	1
24	1	80	1
25	2	85	2

Kuantil

```
in | data.quantile(0.25, interpolation='midpoint')
   | data.quantile(0.50, interpolation='midpoint')
   | data.quantile(0.75, interpolation='midpoint')
out| 23.5
   | 26.5
   | 30.5
```



- Digunakan untuk mendefinisikan **bagian tertentu** dari kumpulan data
- **Kuartil** merupakan contoh **kuantil khusus** yang membagi data menjadi **4 bagian**
- **Interquartile range** merupakan selisih antara Q3 dengan Q1

Standar Deviasi

```
in | print(round(data['age'].std(),2))  
   | print(round(data['score'].std(),2))  
out | 0.96  
    | 31.75
```

- Digunakan untuk mengetahui **sebaran data**
- **Semakin besar** nilai standar deviasi, maka data **semakin tersebar**

	name	age	score
0	Entropy	25	80
1	Team	23	85
2	Digital	24	85
3	Skola	25	20

$$varians = \sigma^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n}$$

$$standar\ deviasi = \sigma = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n}}$$

02

INTERMEDIATE STATISTICS

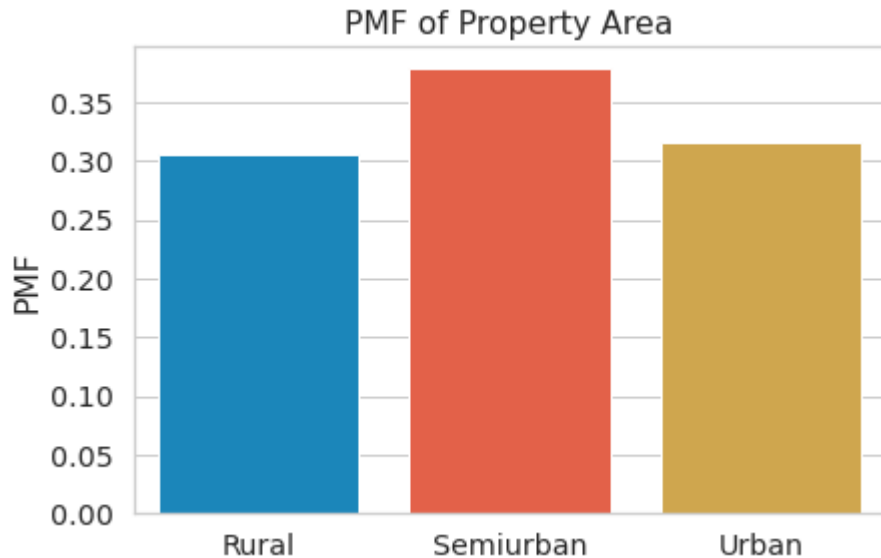
Materi statistik
(menengah)

Probabilitas



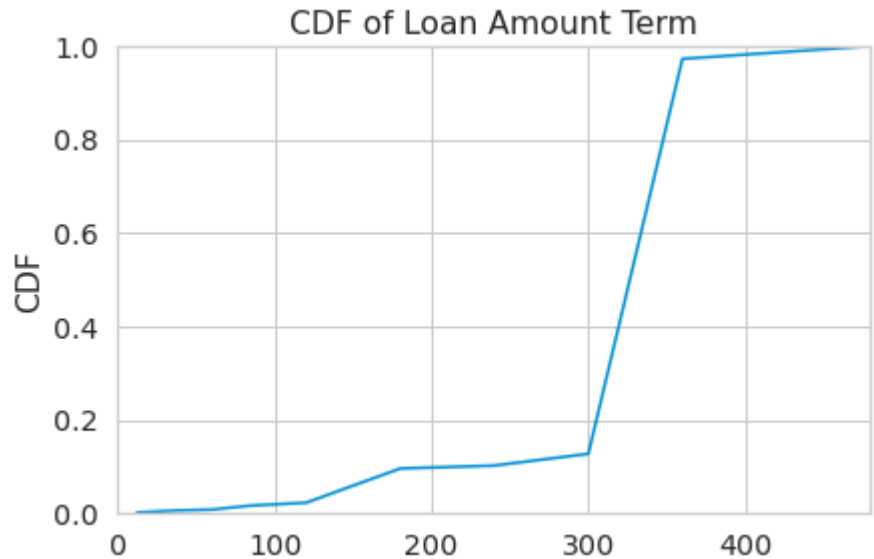
- Probabilitas digunakan untuk **mengukur** seberapa besar **kemungkinan** suatu peristiwa akan terjadi
- Probabilitas memiliki nilai dari **0 sampai 1**
 - Probabilitas 0 : tidak mungkin terjadi
 - Probabilitas 1 : pasti terjadi

Probability Mass Function



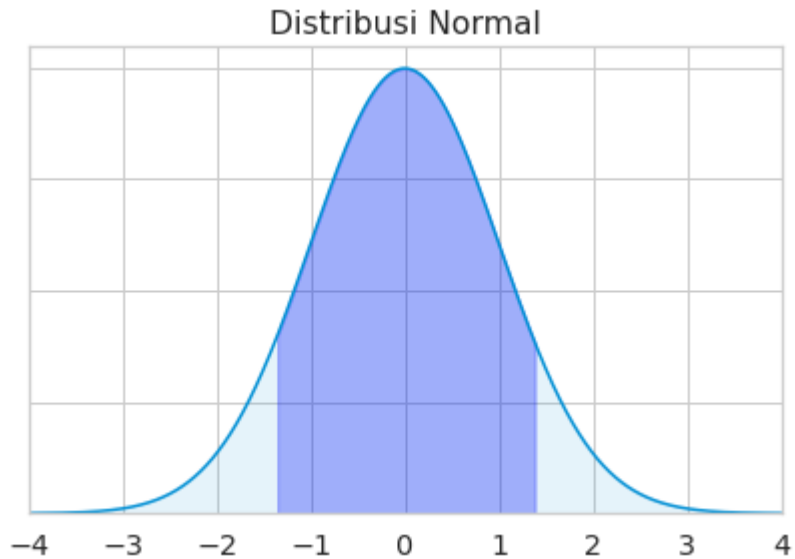
- Digunakan untuk melihat **frekuensi** dari **data diskrit**
- Frekuensi direpresentasikan dalam **bentuk persentase**
- Cocok digunakan untuk data yang **variasinya sedikit**

Cumulative Distribution Function



- Digunakan untuk melihat **perubahan** dari data diskrit, kontinu, atau campuran
- Perubahan direpresentasikan dalam **bentuk persentase**
- Cocok digunakan untuk data yang **variasinya banyak**

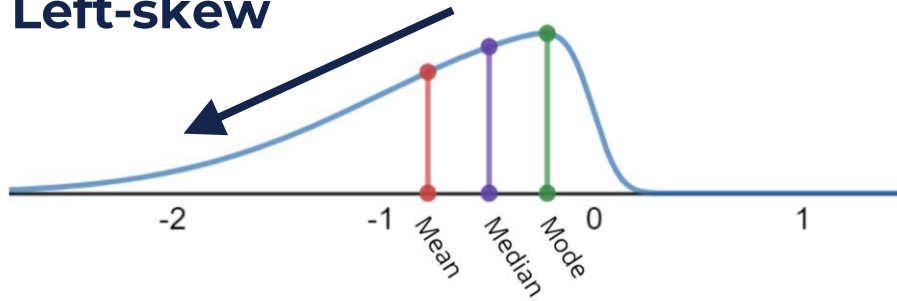
Distribusi Normal



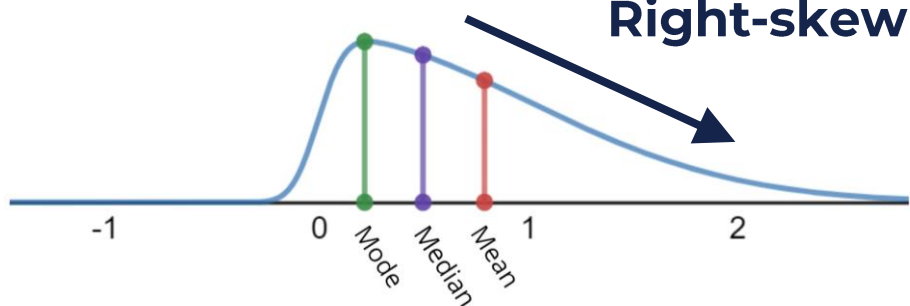
- Distribusi normal memiliki bentuk kurva seperti lonceng (*bell-shaped*) yang **simetris**
- Distribusi normal memiliki nilai **rata-rata**, **median**, dan **modus** yang **hampir sama**

Skewness

Left-skew

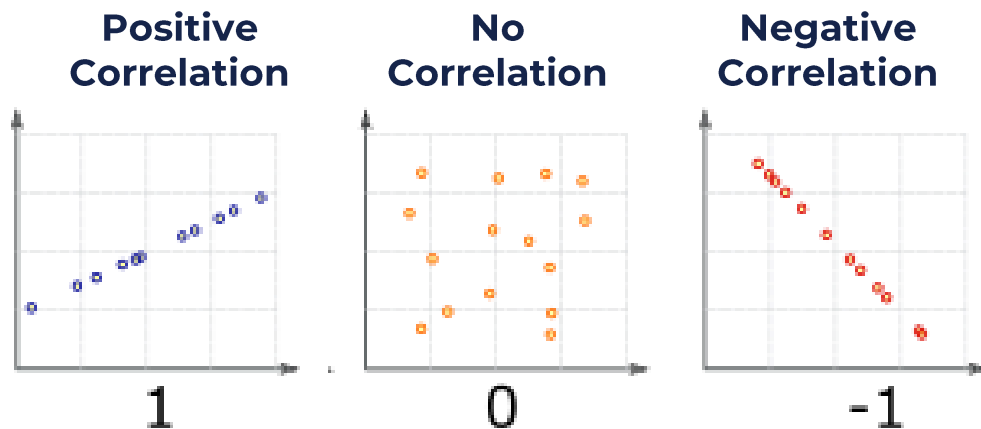


Right-skew



- *Skewness* merupakan **kecondongan** dari data
- Kecondongan tersebut membuat nilai **rata-rata, median, dan modusnya tidak mirip**
- ***Left-skew*** : nilai melandai ke arah kiri
- ***Right-skew*** : nilai melandai ke arah kanan

Korelasi



- Korelasi merupakan **hubungan antarvariabel**
- **Jenis korelasi:**
 - Korelasi positif
 - Korelasi negatif
 - Korelasi netral (tidak ada korelasi)
- **Korelasi tinggi** yaitu ketika nilai korelasi mendekati 1 atau -1
- Korelasi **bukan penanda** adanya **sebab-akibat** (*correlation is not causation*)

03

ADVANCED STATISTICS

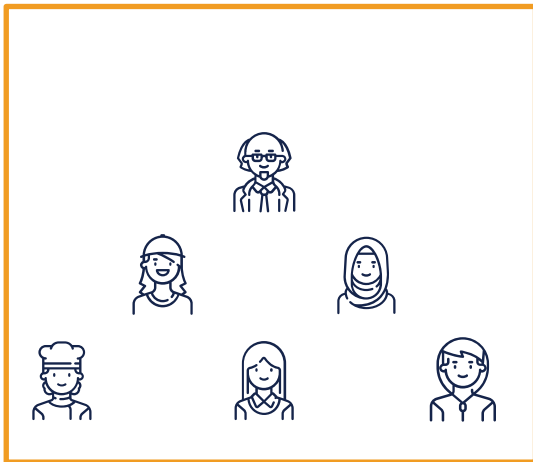
Materi statistik
(lanjutan)

Sampling

Populasi

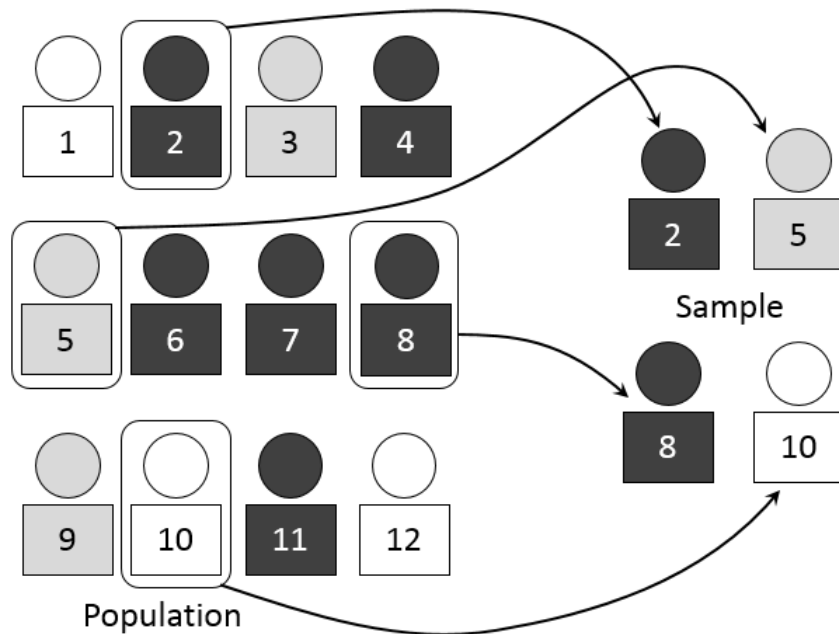


Sampel



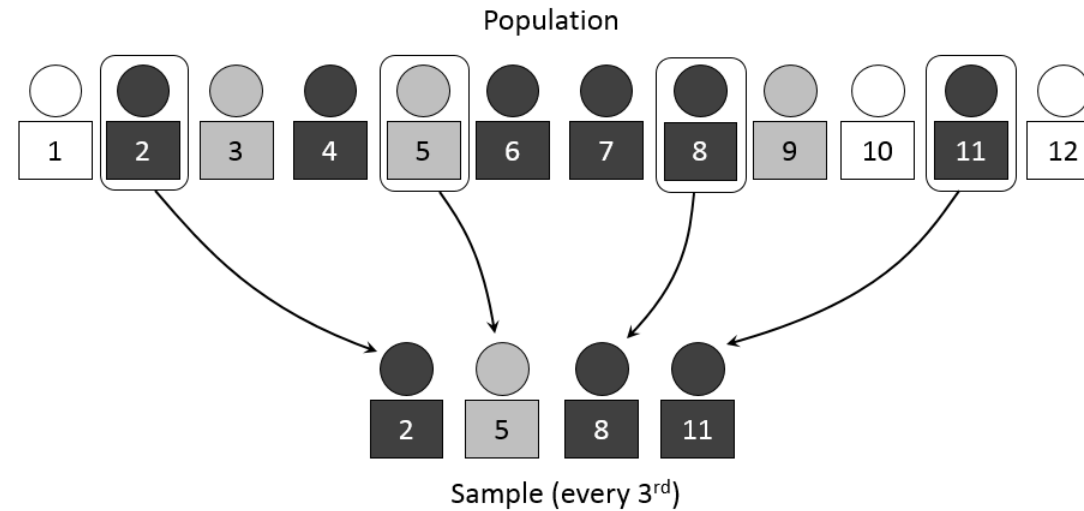
- *Sampling* merupakan metode **pemilihan sampel** (*subset*) untuk **memperkirakan** karakteristik (*parameter*) **populasi**
- *Sampling* dilakukan karena dapat **menghemat waktu, biaya**, dan lebih **praktis** dibanding mengumpulkan keseluruhan data

Simple Random Sampling



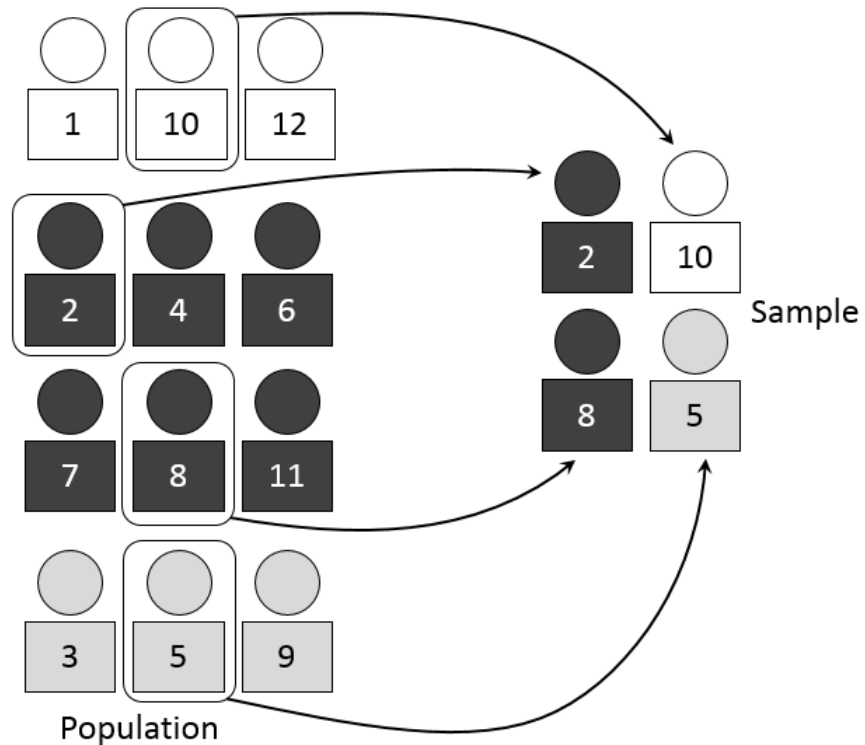
- Setiap elemen dari populasi memiliki probabilitas yang sama untuk dipilih

Systematic Sampling



- *Sampling* dilakukan dengan menggunakan pola tertentu, misal memilih setiap elemen dengan interval 3

Stratified Sampling



- Dapat dilakukan jika terdapat beberapa kategori berbeda
- Setiap elemen dikelompokkan berdasarkan kategori tertentu, kemudian dilakukan *random sampling* dari masing-masing kategori tersebut

Uji Hipotesis



Uji hipotesis (*hypothesis testing*) merupakan metode dalam *inferential statistics* untuk **menentukan** apakah suatu **hipotesis** akan **diterima atau ditolak**

Istilah dalam Uji Hipotesis

- **Null hypothesis (H_0)**
Hipotesis awal yang dapat diterima atau ditolak berdasarkan hasil uji hipotesis
- **Alternative hypothesis (H_1)**
Hipotesis yang diterima jika H_0 ditolak
- **Confidence interval**
Rentang di mana H_0 akan diterima, dihitung berdasarkan persentase *confidence level*
- **Significance level (alpha, α)**
Probabilitas membuat keputusan yang salah ketika H_0 benar
- **Critical value**
Batas wilayah penerimaan hipotesis yang ditentukan oleh alpha
- **P-value**
Nilai uji hipotesis. Jika p-value < alpha, maka H_0 ditolak dan H_1 diterima

Metode Uji Hipotesis

Z-test

Digunakan jika:

Varians dari populasi
diketahui

ATAU

Ukuran sampel
lebih dari 30

T-test

Digunakan jika:

Varians dari populasi
tidak diketahui

DAN

Ukuran sampel
kurang dari 30

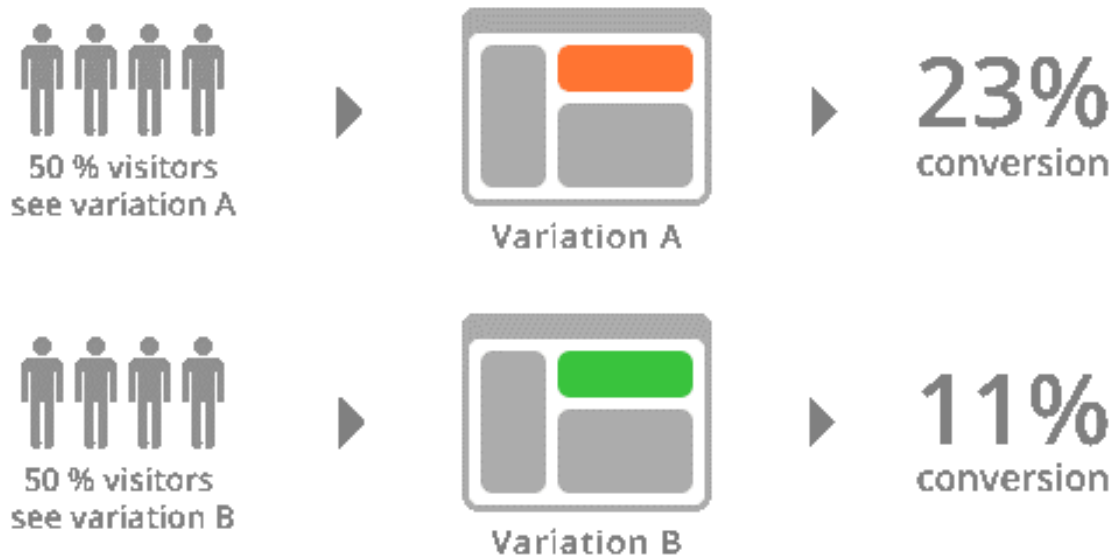
Chi-square

Digunakan untuk melihat
apakah ada ***dependency***
(ketergantungan)
antarvariabel

Sumber:

<https://www.analyticsvidhya.com/blog/2020/06/statistics-analytics-hypothesis-testing-z-test-t-test/>

AB Testing



- **Pengujian** dilakukan terhadap **2 variasi** (variasi A dan B)
- Pengujian dilakukan untuk **membandingkan performa** dari kedua variasi tersebut

THANKS

Entropy Team

CREDITS: This presentation template was originally created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**