

PERTEMUAN I PENGANTAR STATISTIKA DAN ANALISIS DATA



DR.TUKIYAT,M.SI

081398222862

Program Studi Magister Teknik Informatika
Program Pascasarjana
Universitas Pamulang

PENGANTAR

- **Tujuan Pembelajaran**

- Mahasiswa mampu dan memahami tentang statistika dan ruang lingkupnya, konsep data dan analisis data untuk penyelesaian masalah serta penyajian data yang efektif.

- **Pokok Bahasan**

- 1) Data
- 2) Terminologi Statistics
- 3) Statistika Vs Statistik
- 4) Pengolahan Data
- 5) Penyajian Data

FAKTA

DATA adalah fakta – fakta yang mewakili suatu objek, keadaan, kondisi atau Peristiwa yang terjadi baik didalam atau diluar lingkungan organisasi yang dapat berupa angka, huruf, simbol atau kombinasi.



INFORMASI adalah **DATA** yang telah diolah menjadi suatu bentuk yang berarti dan bermanfaat dan dapat digunakan untuk pengambilan keputusan.



PENGAMBILAN KEPUTUSAN

Berdasar Pengumpulan:

1. Observasi
2. Studi Literatur
3. Kuesioner
4. Wawancara

Berdasar Banyaknya Data:

1. Sensus
2. Sampel

Penyajian Data

1. Tabel
2. Grafik (Piktogram, Batang, Garis, Lingkaran)
3. Kartogram

Berdasar Sumber Pengumpulan

- Data Primer : data yang diperoleh atau dikumpulkan oleh orang yang melakukan penelitian atau yang bersangkutan yang memerlukannya.
- Data Sekunder : data yang diperoleh atau dikumpulkan dari sumber-sumber yang telah ada.

Susunan

- Data acak atau data tunggal : data yang belum tersusun atau dikelompokkan ke dalam kelas-kelas interval
- Data berkelompok : data yang sudah tersusun atau dikelompokkan dalam kelas-kelas interval.

Berdasar Sifat

- Data kualitatif : data yang tidak berbentuk bilangan. Contoh : warna, jenis kelamin, status perkawinan
- Data kuantitatif : data yang berbentuk bilangan. Contoh : tinggi, umur, jumlah.

Skala Pengukuran

- Data Nominal : data yang diberikan pada objek atau kategori yang tidak menggambarkan kedudukan objek, tetapi hanya sekadar label atau kode saja.
- Data Ordinal : data yang penomoran objek atau kategorinya disusun menurut besarnya.
- Data Interval: data yang kategorinya dapat diurutkan berdasarkan suatu atribut yang memberikan informasi interval antara tiap obyek/kategori sama.
- Data Rasio : data yang memiliki sifat-sifat data nominal, data ordinal, dan data interval, dilengkapi dengan titik nol absolut dengan makna empiris. Angka yang ada pada data menunjukkan ukuran sebenarnya dari obyek/kategori yang diukur.

Berdasar Waktu

- Data berkala: data yang terkumpul dari waktu ke waktu untuk memberikan gambaran perkembangan suatu kegiatan
 - Contoh: Data perkembangan harga smartphone selama 10 bulan terakhir yang dikumpulkan tiap bulan.
- Data cross section: data yang terkumpul pada suatu waktu tertentu untuk memberikan gambaran perkembangan keadaan atau kegiatan pada waktu itu.
 - Contoh: Data harga Smartphone merk Samsung Galaxy A pada bulan Februari 2017

Data

Where Statistics Come From



Some Basic Terminology

- A **population** consists of all the items or individuals about which you want to draw a conclusion.
- A **sample** is the portion of a population selected for analysis.
- A **variable** is a characteristic of an item or individual.
- **Data** are the different values associated with a variable.
- A **parameter** is a numerical measure that describes a characteristic of a population.
- A **statistic** is a numerical measure that describes a characteristic of a sample.

Types of Statistics

Statistics

The branch of mathematics that transforms data into useful information for decision makers.



Descriptive Statistics

Collecting, summarizing, and describing data

How to use only a few numbers or one graph to represent lots of data



Inferential Statistics

Methods for drawing conclusions (inferences) about a population from a sample

Descriptive Statistics

■ Collecting data

- e.g., Survey



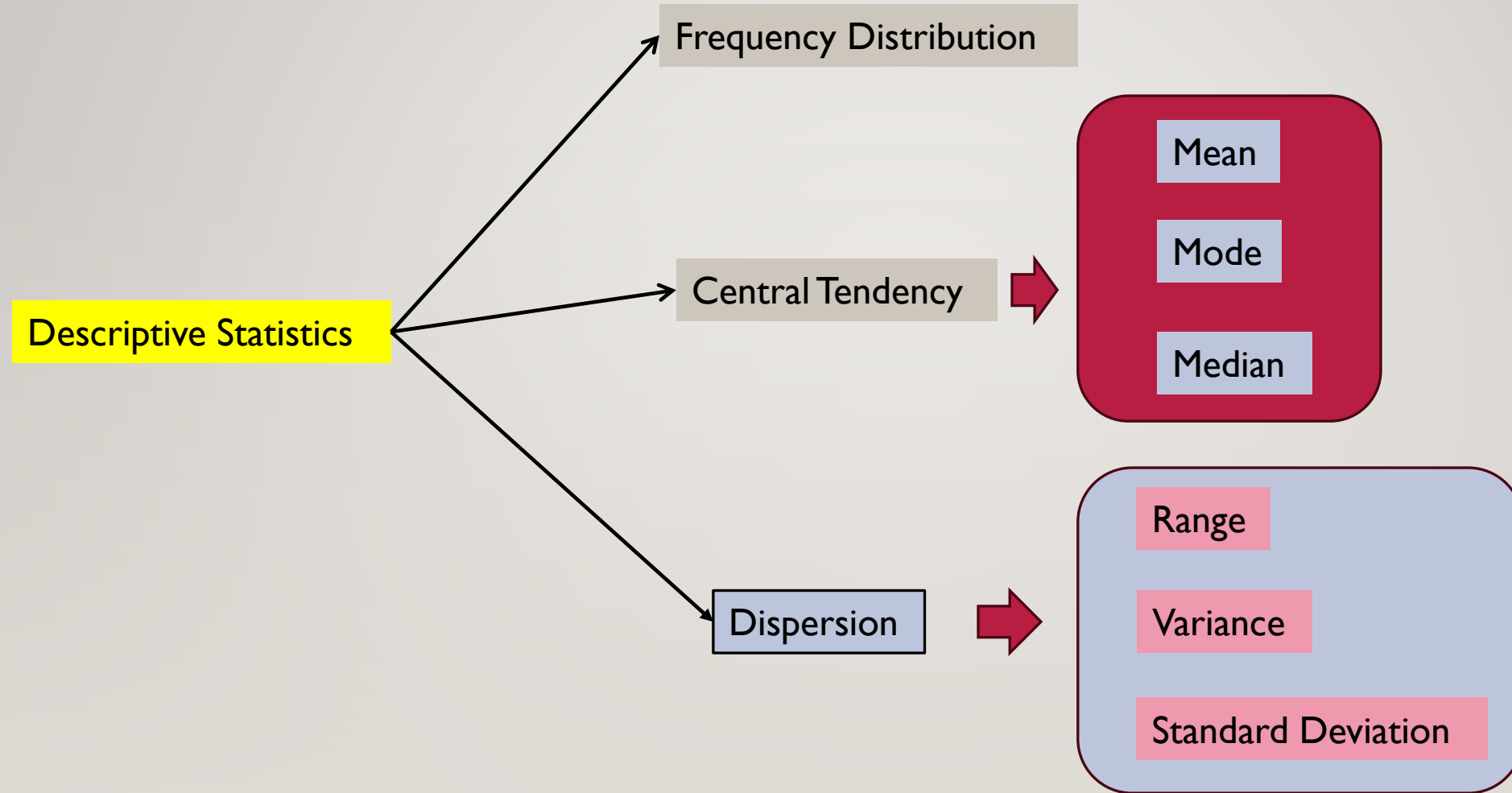
■ Cleaning / Organizing / Graphing data

- e.g., Tables and graphs

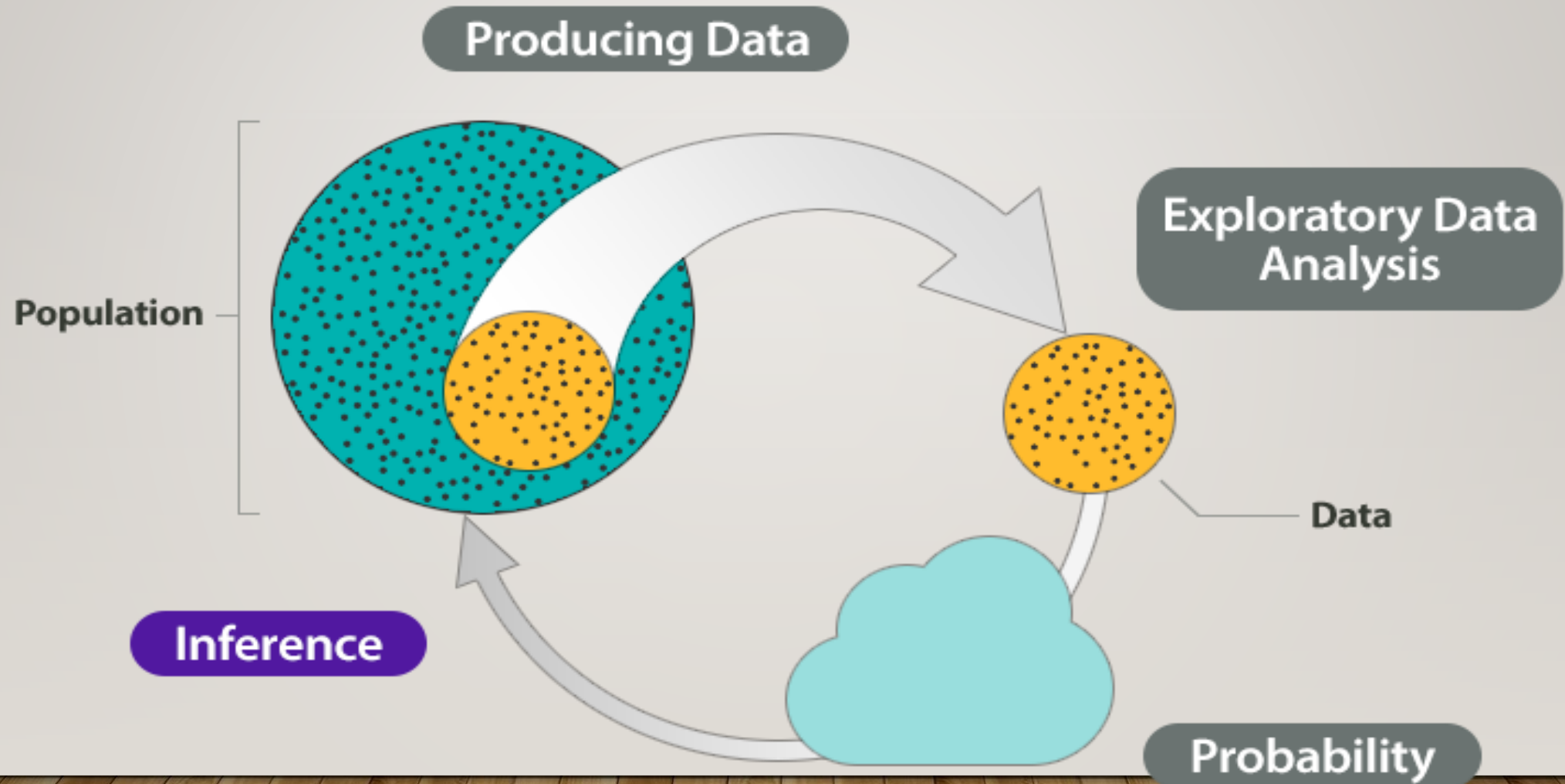


■ Summarizing data

- e.g., Sample mean = $\frac{\sum X_i}{n}$



Statistics Inference



Statistical Inference

Estimation

Point estimation

- Summarize the sample by a single value as an estimate of the population parameter.
- Ex. Average salary of junior data scientists is. 55,000 euros.

Interval estimation

- A range of values within which, we believe, the true population parameter lies with high probability.
- Ex. Average salary of junior data scientists is in the range of (52,000,55,000) With 95% confidence level.

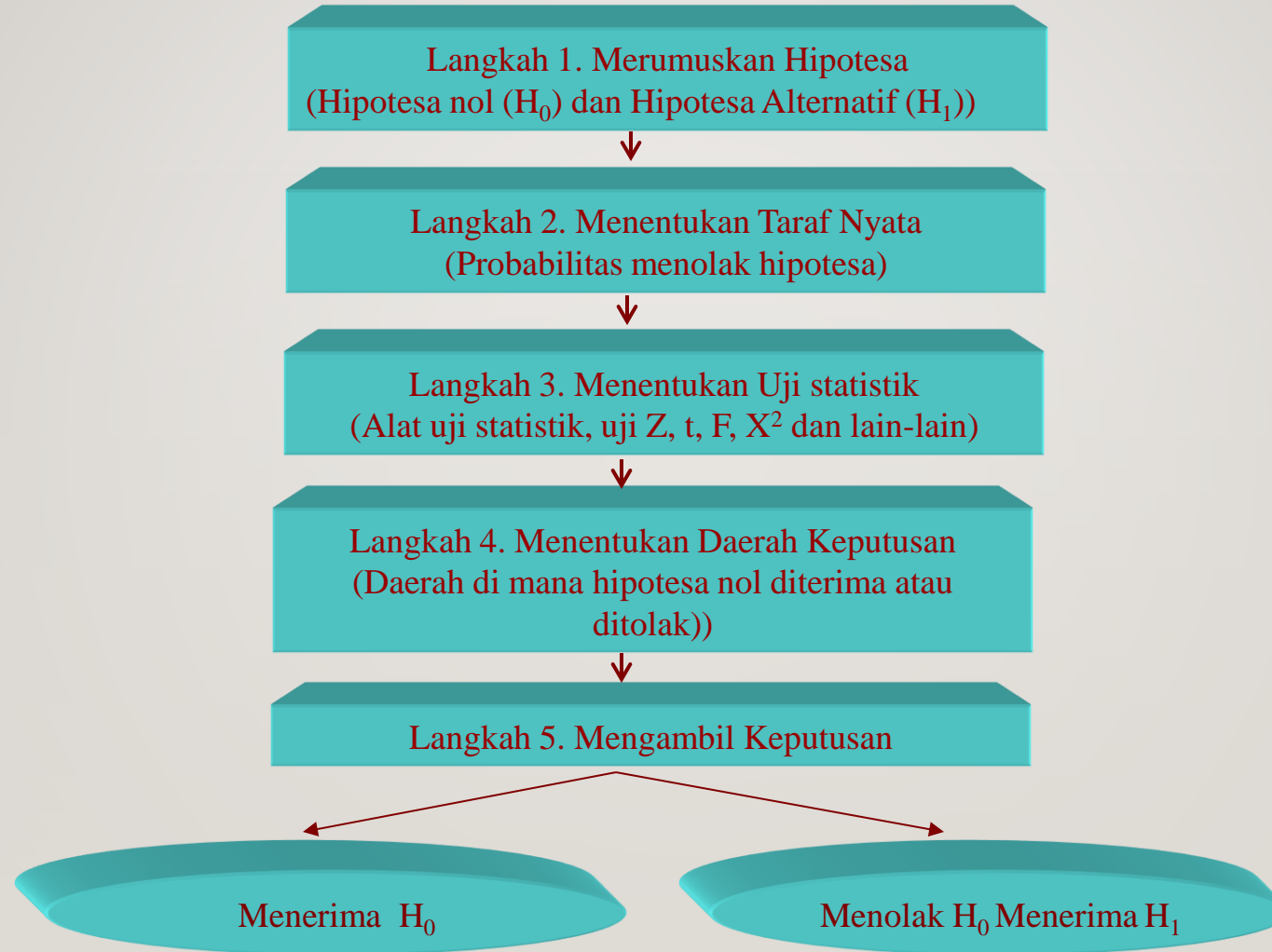
Testing of Hypothesis

- To decide whether a statement regarding population parameter is true or false, based on sample data.
- Ex. Claim: Average salary of junior data scientists is greater than.50,000 euros annually.

Why is statistical inference important?



Test Hypotesis Procedure



Type I and Type II Errors in Hypothesis Testing

Alpha (α) is the probability of making a Type I error →

We want α to be as low as possible!

Beta (β) is the probability of making a Type II error.

The power of a test is the probability ($1 - \beta$) of rejecting the null hypothesis when it is indeed false and hence should be rejected → **We want power to be as high as possible!**

Unfortunately, α and β are interrelated. So, it's necessary to balance the two types of errors.

The level of α along with the sample size will determine the level of β for a particular research design.

In practice, we usually set α at 1%, 5%, or 10%.

Null Hypothesis	Researcher	
	Does Not Reject H_0	Rejects H_0
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

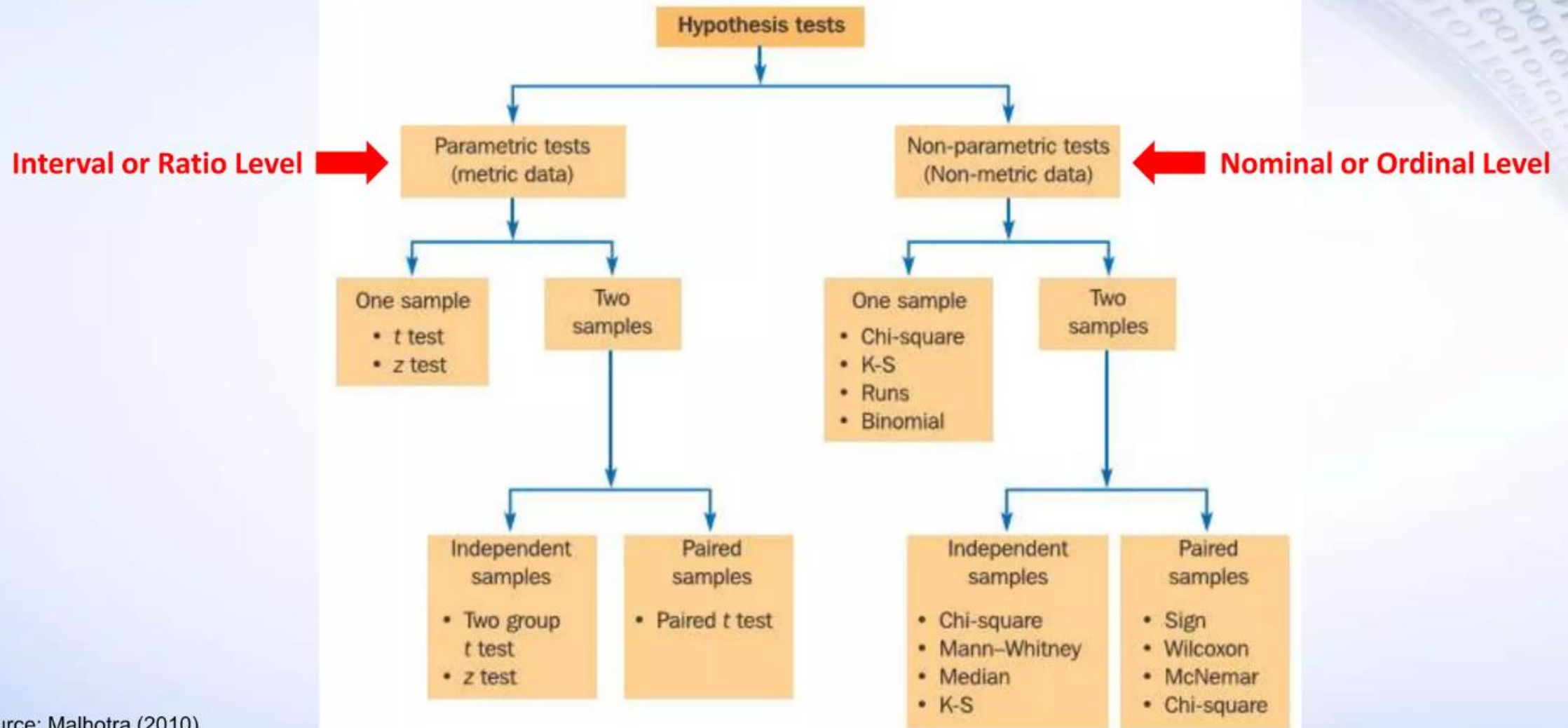
The risk of both α and β can be controlled by increasing the sample size.

For a given level of α , increasing the sample size will decrease β , and hence increasing the power of the test ($1 - \beta$).

Think of sample size as a magnifying glass.



Hypothesis Tests Related to Differences



STATISTIKA VS STATISTIK

- Statistika

ilmu pengetahuan yang mempelajari tentang cara-cara pengumpulan fakta, pengolahan serta analisis pemberatan kesimpulan dan penafsiran kesimpulan yang cukup beralasan berdasarkan fakta dan analisa yang dilakukan



- Statistik :

Kumpulan angka untuk menerangkan sesuatu baik angka yang belum tersusun maupun angka yang sudah tersusun dalam suatu daftar atau grafik.

Fakta : berbentuk angka/ keterangan lain yang disusun/disajikan dalam daftar/diagram untuk menggambarkan persoalan sesuai dengan fakta yang ada, agar khalayak dapat memahami dan mengerti tentang gambaran situasi yang sebenarnya

PERAN STATISTIK

- **Model matematika** lebih mudah dalam merumuskan masalah (lebih singkat dan logis) struktur masalah akan lebih terungkap dan masalah dalam mengidentifikasikan.
- **Pemecahan masalah** menurut metode matematika dan pengujiannya yang menggunakan cuplikan (sampel) dapat diteruskan dengan metode statistika populasi dibantu dengan komputer.
- Model matematika dapat melihat apakah asumsi-asumsi yang dipergunakan dalam penelitian dapat dipenuhi atau tidak.

PERLUNYA STATISTIKA

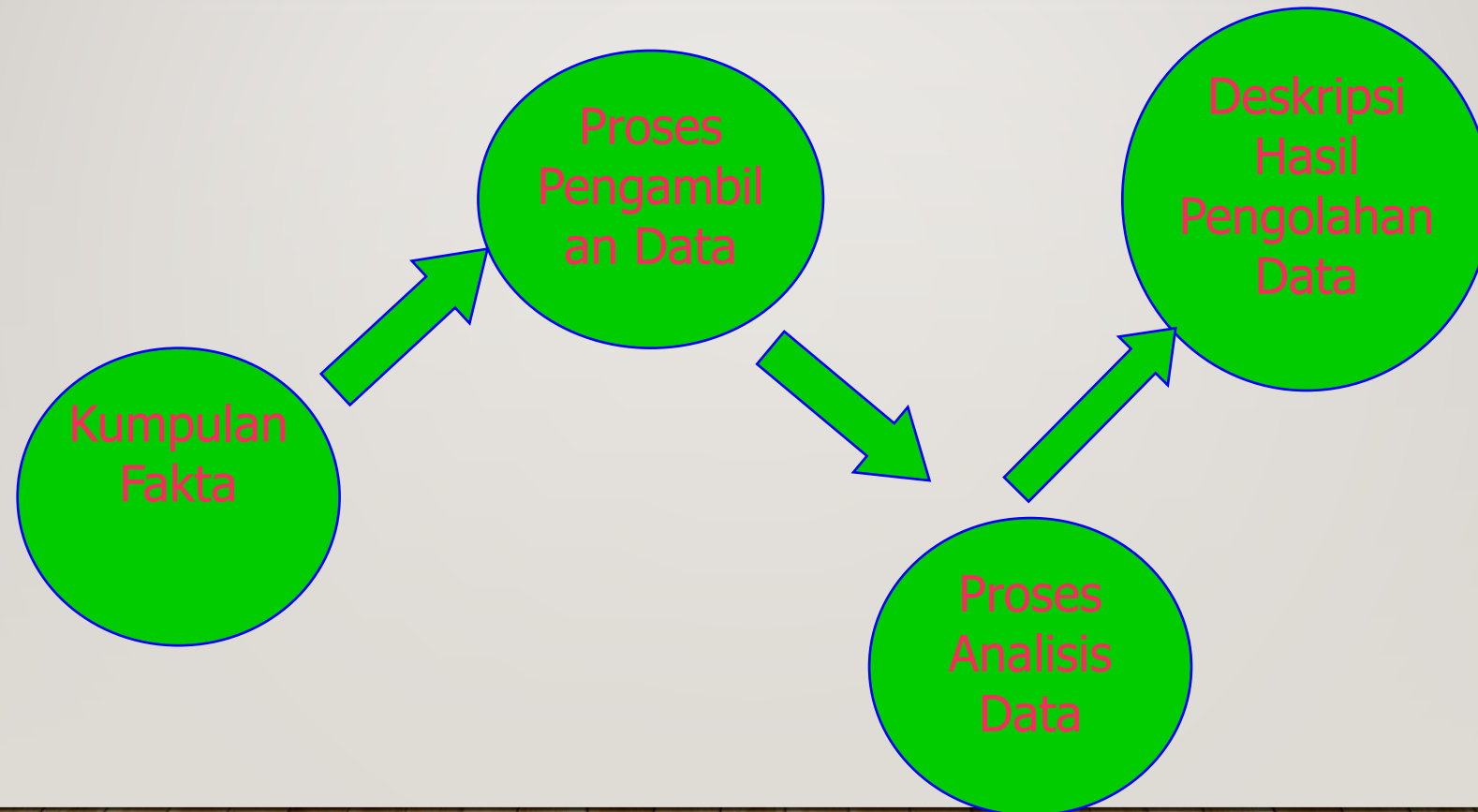
1. Menjelaskan hubungan antar variable
2. Membuat rencana dan ramalan
3. **Mengatasi berbagai perubahan**; ---? Perluay mengantisipasi perubahan kebijakan karena faktoe eksternal; antisiapasi harga saahm karena pandemic, dampak kenaikan BBM antispasi kenaikan upah buruh di perusahaan
4. Membuat keputusan yang lebih baik

WHAT IS DATA ANALYSIS ?

-
- Data analysis is the process of collecting, modeling, and analyzing data using various statistical and logical methods and techniques (**Analisis data adalah proses pengumpulan, pemodelan, dan analisis data menggunakan berbagai metode dan teknik statistik dan logis**)
 - Data analysis is the practice of working with data to glean useful information, which can then be used to make informed decisions (**Analisis data adalah praktik bekerja dengan data untuk mengumpulkan informasi yang berguna, yang kemudian dapat digunakan untuk membuat keputusan berdasarkan informasi**)



ALUR PENGOLAHAN DATA





5 Type of Analytics

1. Descriptive: What is happening?

- Correct Data
- Effective Exploratory data analysis

2. Diagnostic: Why is it happening?

- Finding the causes
- Separating all the patterns

3. Predictive: What is likely to happen?

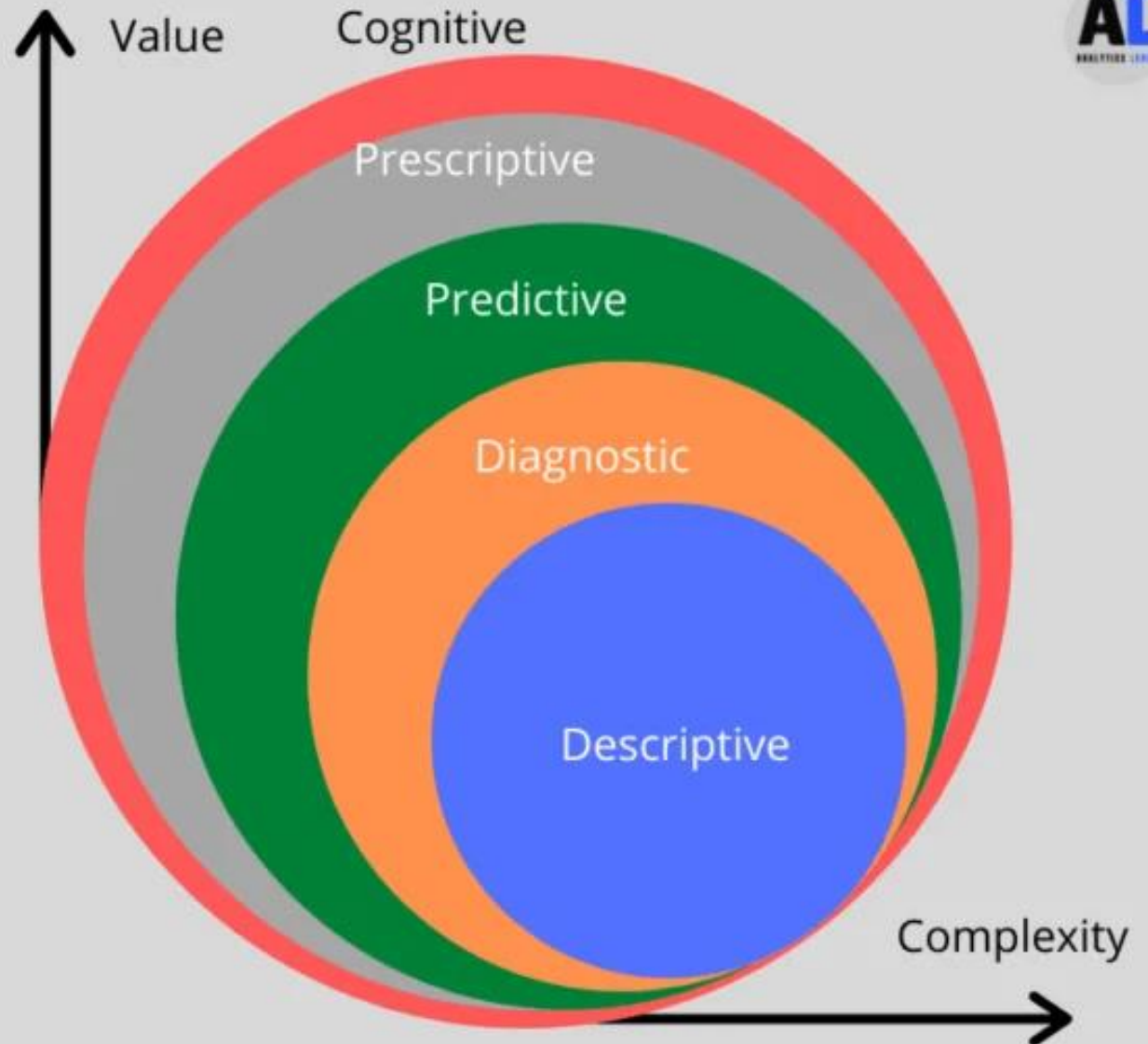
- Choosing the right algorithm
- Bulding the right business strategies

4. Prescriptive: What do I need to do?

- Using the advance analytics
- Recommended actions

5. Cognitive Analytics

- Neurological and Behavioral analysis



Descriptive

What happened?

Analisis Deskriptif memberikan deskripsi yang terjadi pada masa lalu atau mendeskripsikan obyek dengan narasi sesuai dengan data yang tersedia dan membantu bisnis memahami kinerjanya dengan menyediakan konteks untuk membantu pemangku kepentingan menafsirkan informasi.



Descriptive Analytics

A descriptive statistic (in the count noun sense) is a summary statistic that quantitatively describes or summarizes features from a collection of information, while descriptive statistics (in the mass noun sense) is the process of using and analyzing those statistics

- Wikipedia



Predictive

What is likely to
happen in the future?

Analisis Prediktif memprediksi apa yang paling mungkin terjadi di masa depan dan memberi pengetahuan/wawasan yang dapat ditindaklanjuti berdasarkan informasi tersebut.



Predictive Analytics

Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modeling, and machine learning that analyze current and historical facts to make predictions about future or otherwise unknown events.



- Wikipedia



Diagnostic

Why did it happen?

Analitik Diagnostik mengambil data deskriptif selangkah lebih maju dan membantu peneliti dalam memahami mengapa sesuatu terjadi di masa lalu.



Diagnostic Analytics

Diagnosis is the identification of the nature and cause of a certain phenomenon. Diagnosis is used in many different disciplines, with variations in the use of logic, analytics, and experience, to determine "cause and effect". In systems engineering and computer science, it is typically used to determine the causes of symptoms, mitigations, and solutions.

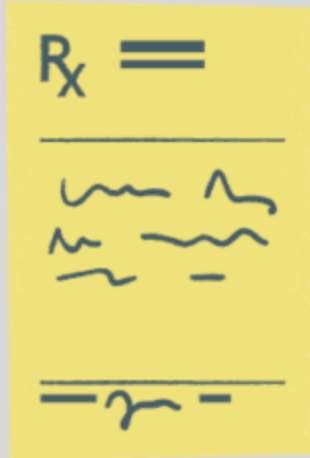
- Wikipedia



Prescriptive

What is the best
course of action?

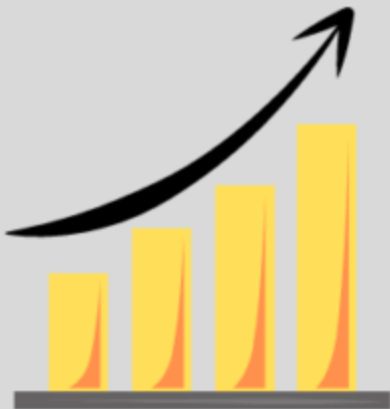
Prescriptive Analytics memberikan rekomendasi mengenai tindakan yang akan memanfaatkan prediksi dan memandu tindakan yang mungkin untuk memberikan solusi.



Prescriptive Analytics

Prescriptive analytics is the fourth phase of business analytics, which also includes descriptive and predictive analytics. Referred to as the "final frontier of analytic capabilities," prescriptive analytics entails the application of mathematical and computational sciences and suggests decision options to take advantage of the results of descriptive and predictive analytics.

- Wikipedia



PROSEDUR PENGOLAHAN DATA :

A. **PARAMETER** : Berdasarkan parameter yang ada statistik dibagi menjadi

- Statistik **PARAMETRIK** : berhubungan dengan inferensi statistik yang membahas parameter-parameter populasi; jenis data interval atau rasio; distribusi data normal atau mendekati normal.
- Statistik **NONPARAMETRIK** : inferensi statistik membahas parameter-parameter populasi; jenis data nominal atau ordinal; distribusi data tidak diketahui atau tidak normal

B. **JUMLAH VARIABEL** : berdasarkan jumlah variabel dibagi menjadi

- Analisis **UNIVARIAT** : hanya ada 1 pengukuran (variabel) untuk n sampel atau beberapa variabel tetapi masing-masing variabel dianalisis sendiri-sendiri. Contoh : korelasi motivasi dengan pencapaian akademik.
- Analisis **MULTIVARIAT** : dua atau lebih pengukuran (variabel) untuk n sampel di mana analisis antar variabel dilakukan bersamaan. Contoh : pengaruh motivasi terhadap pencapaian akademik yang dipengaruhi oleh faktor latar belakang pendidikan orang tua, faktor sosial ekonomi, faktor

PENGOLAHAN DATA

- Dengan cara manual
- Dengan bantuan software :
 - ✓ SPSS
 - ✓ R Program
 - ✓ Eviews
 - ✓ Weka
 - ✓ Rapidminer
 - ✓ SmartPLS
 - ✓ Dan Lain-lain

TERIMA KASIH

Selanjutnya belajar tentang Distribusi Frekwensi

