

Week 1 Introduction to Statistics

IMT01303404 Statistics and Probability

4 Credits

Even Semester 2023 / 2024

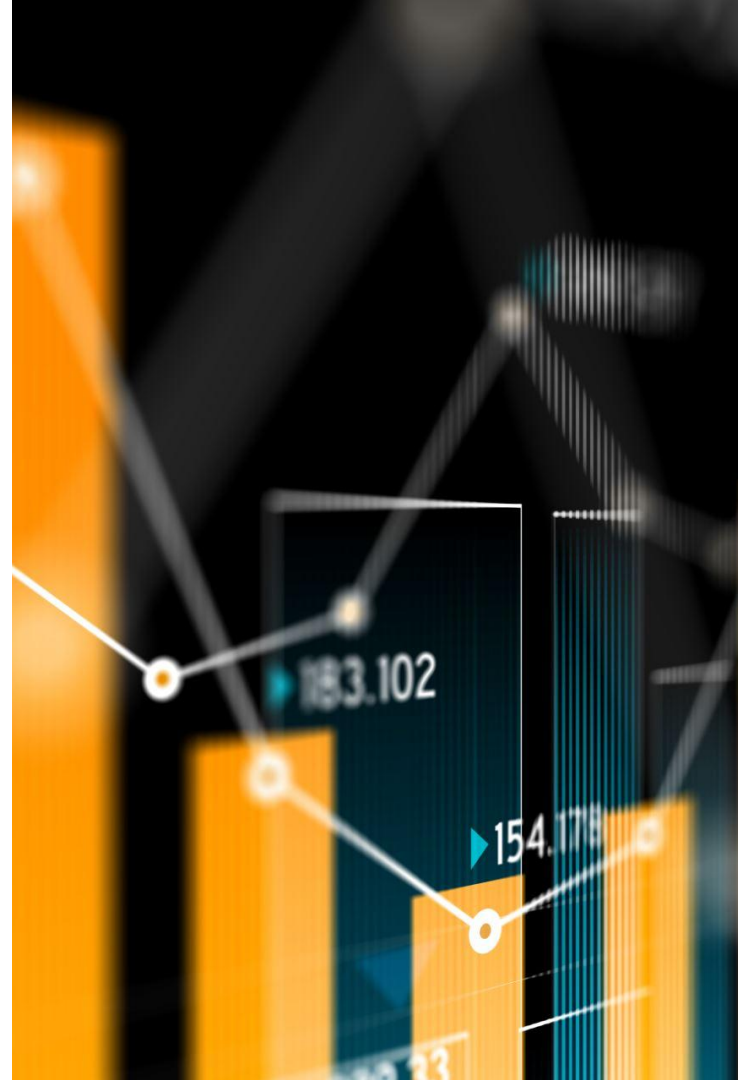
Dr. Trianggoro Wiradinata, M.Eng.Sc.
Richard Evan Sutanto, M.Sc., Ph.D.





Learning Objectives

- / LO1-1 Explain why knowledge of statistics is important
- / LO1-2 Define statistics and provide an example of how statistics is applied
- / LO1-3 Differentiate between descriptive and inferential statistics
- / LO1-4 Classify variables as qualitative or quantitative, and discrete or continuous
- / LO1-5 Distinguish between nominal, ordinal, interval, and ratio levels of measurement
- / LO1-6 List the values associated with the practice of statistics



Why Study Statistics

- Data are collected everywhere and require statistical knowledge to make the information useful
- Statistical techniques are used to make professional and personal decisions
- A knowledge of statistics is needed to understand the world and be conversant in your career
- In summary, statistics will help you make more effective personal and professional decisions





What is Meant by Statistics?

STATISTICS is The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

- ✓ It's more than presenting numerical facts

Example: The inflation rate for the calendar year was 0.7%. By applying statistics, we could compare this year's inflation rate to past observations of inflation. Is it higher, lower, or about the same. Is there a trend of increasing or decreasing inflation?

Types of Statistics



**DESCRIPTIVE
STATISTICS**



**INFERENTIAL
STATISTICS**



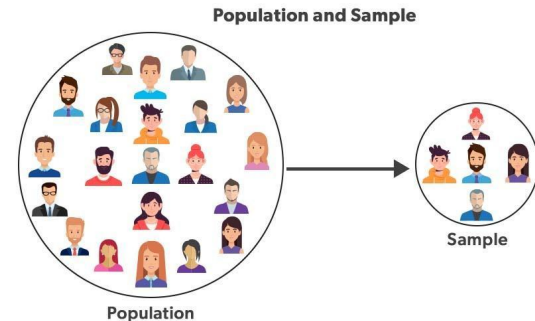
Descriptive Statistics

DESCRIPTIVE STATISTICS is method of organizing, summarizing, and presenting data in an informative way.

- Descriptive statistics can be used to organize data into a meaningful form
- You can summarize data and provide information that is easy to understand
- Example: There are a total of 46,837 miles of interstate highways in the U.S. The interstate system represents 1% of the nations roads but carries more than 20% of the traffic. Texas has the most interstate highways and Alaska doesn't have any.

POPULATION The entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest.

SAMPLE A portion or part of the population of interest.



POPULATION	SAMPLE
<ul style="list-style-type: none"> The measurable quality is called a parameter. 	<ul style="list-style-type: none"> The measurable quality is called a statistic.
<ul style="list-style-type: none"> The population is a complete set. 	<ul style="list-style-type: none"> The sample is a subset of the population.
<ul style="list-style-type: none"> Reports are a true representation of opinion. 	<ul style="list-style-type: none"> Reports have a margin of error and confidence interval.
<ul style="list-style-type: none"> It contains all members of a specified group. 	<ul style="list-style-type: none"> It is a subset that represents the entire population.

Before we discuss Inferential Statistics

Sample vs. Population





Sample is preferable when ...

01

Total cost to collect all data is way out of budget

02

Possibility of object destruction

03

Not possible to collect all data from population

Using a sample to learn something about a population is done extensively in business, agriculture, politics, and government.

EXAMPLE: Television networks constantly monitor the popularity of their programs by hiring consulting company to sample the preferences of TV viewers.



Inferential Statistics

INFERENTIAL STATISTICS is method used to estimate a property of a population based on a sample.

- Inferential statistics can be used to estimate properties of a population by analyzing its sample
- You can make decisions based on a limited set of data
- Example: Among 5,000 university students we take 400 students to interviewed as a sample, as a result 300 students stated healthy food is important. We infer that 75% of all students think that healthy food is important, hence 75% of food vendor must meet healthy food criteria.



Types of Variables

QUALITATIVE VARIABLE An object or individual is observed and recorded as a non-numeric characteristic or attribute.

Examples: gender, city of birth, eye color, etc.

QUANTITATIVE VARIABLE A variable that is reported numerically.

Examples: balance in your bank account, the life of a car battery, the number of people employed by a company, etc.



Types of Variables

- Quantitative variables can be *discrete* or *continuous*
- Discrete variables are typically the result of **counting**
 - Values have “gaps” between the values
 - Examples: the number of bedrooms in a house, the number of students in a statistics course
- Continuous variables are usually the result of **measuring** something
 - Can assume any value within a specific range
 - Examples: the air pressure in a tire, duration of flights from Surabaya to Jakarta

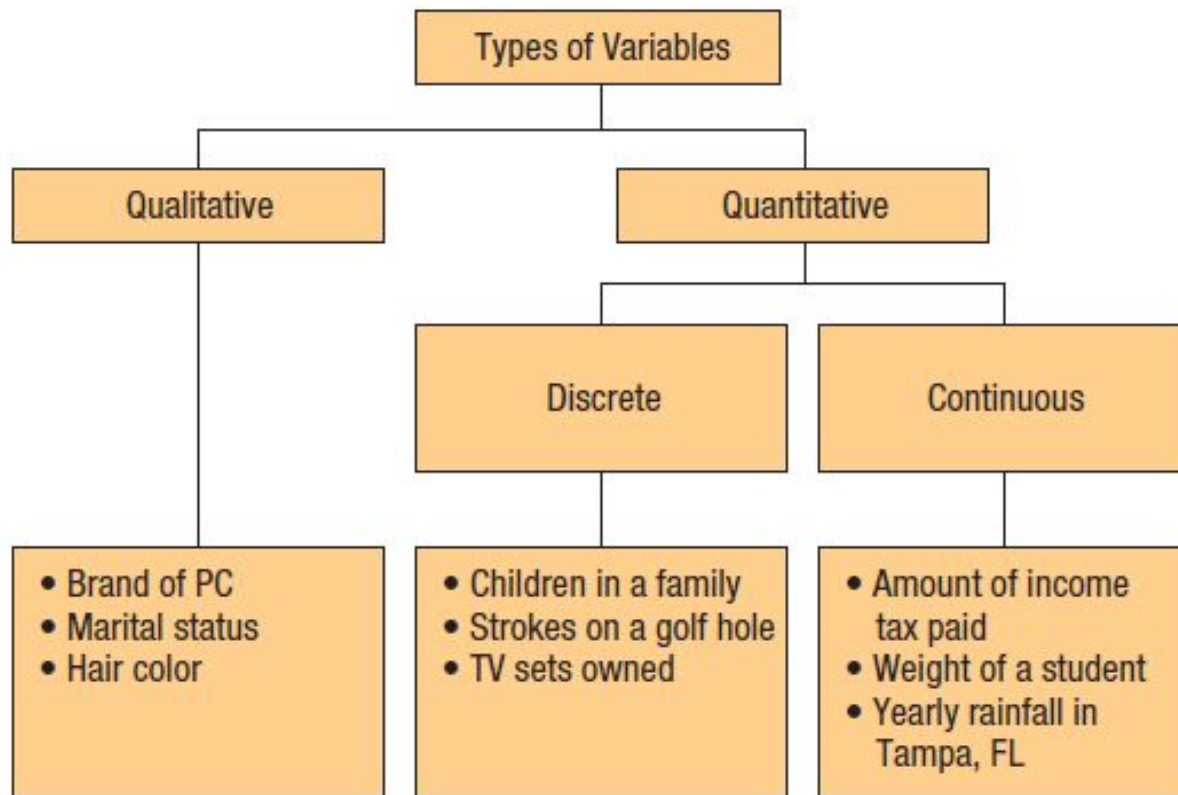


CHART 1-2 Summary of the Types of Variables



Levels of Measurement

01

NOMINAL

Named variables

ORDINAL

Named + ordered variables

02

03

INTERVAL

Named + ordered + proportionate interval between variables

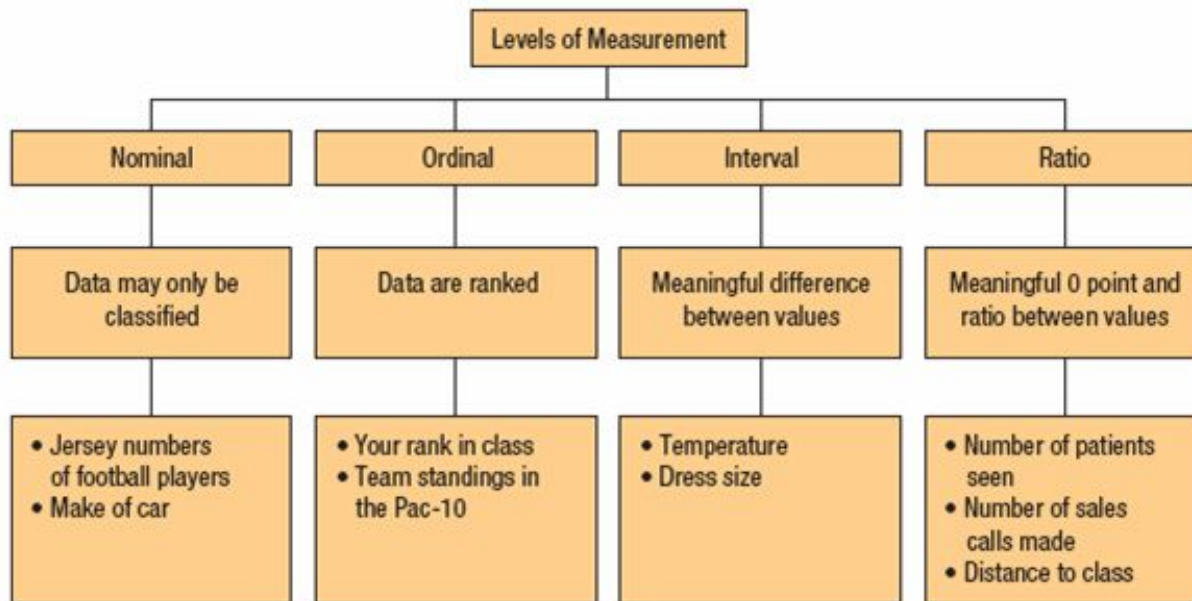
RATIO

Named + ordered + proportionate interval between variables
+ Can accommodate absolute zero

04

The level of measurement determines the type of statistical analysis that can be performed

Summary for Levels of Measurement



Why do we need to know Levels of Measurement?

- The levels of measurement dictates the calculations that can be done to summarize and visualize the data.
- To determine the statistical tests that should be performed on the data.



Nominal Level

- Nominal is the lowest level of measurement. Nominal scales classify data into distinct categories without any inherent order, so it doesn't matter which one comes first.
- For example, a nominal scale could classify students by gender (male or female) or department (Informatics, Design, or Business)

Ordinal Level

- Ordinal scales are used to classify data into distinct categories that have a natural order (can be sorted).
- For example, an ordinal scale could be used to classify people by their level of satisfaction with a product (very satisfied, satisfied, neutral, dissatisfied, very dissatisfied), another example is level of education (primary, junior high, senior high, undergraduate, postgraduate)

Interval Level

- The interval scale has an order (just like ordinal data) and there is exact difference between the two values. Interval scale doesn't have absolute zero (true zero)
- The classic example of an interval scale is the temperature in Celsius. We can clearly say that 50°C degrees is greater than 40°C. That means there is an order. We can also say that the difference between 50°C and 40°C degrees is the same as between 70°C and 60°C (same interval). Also **0°C degrees does not mean "no temperature"**.

Ratio Level

- Ratio scales are used to measure data on a continuous scale, with a meaningful zero point and the ability to compare ratios of different measurements. Ratio scales have a meaningful zero point and support the comparison of ratios, so it is possible to perform all mathematical operations on this type of data.
- Examples of ratio scales include weight, height, etc. A 10 Kg is heavier than 5 Kg. The difference between 10 Kg and 5 Kg is the same as the difference between 100 Kg and 95 Kg. Also, we have an absolute zero here. **A weight of 0 Kg means that there is "no weight"**.



Exercise 1

Discrete or Continuous

- a. [D / C] Jumlah mahasiswa yang belajar di kelas
- b. [D / C] Tinggi rata-rata mahasiswa Informatika
- c. [D / C] Jumlah Bahasa yang dikuasai
- d. [D / C] Berat smartphone
- e. [D / C] Waktu yang dibutuhkan untuk lari 100m
- f. [D / C] Kecepatan kereta
- g. [D / C] Jumlah anjing di Citraland
- h. [D / C] Jumlah alumni entrepreneur dari UC
- i. [D / C] Jam bangun pagi
- j. [D / C] Jumlah bintang dari kepuasan pelanggan di app store



Exercise 2 Sample or Population

- a. [S / P] A Fast-Growing FMCG company in Surabaya wants to know consumer's response on their newest product
- b. [S / P] An owner of a startup company with 20 employees desires to know the average body mass index of the employees
- c. [S / P] A biscuit company wishes to know the freshness of its products on the market

Exercise 3 Levels of Measurement

Letakkan variabel dibawah ini ke Tabel A dan Tabel B

- Konsumsi BBM Mobil
- Jenis Kelamin
- Luas Halaman
- Merk Minuman
- Kecepatan Lari
- Hasil akhir TOEFL
- Peringkat Kelas
- Jumlah Laptop di kelas
- Penghasilan per Bulan

Tabel A

	Discrete	Continuous
Qualitative		
Quantitative		

Tabel B

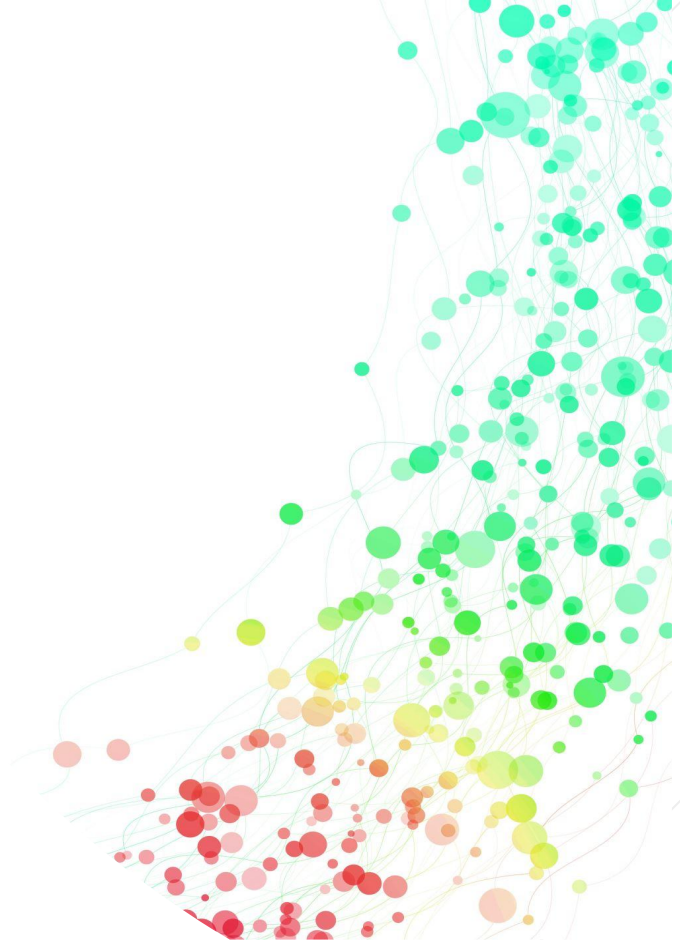
	Discrete	Continuous
Nominal		
Ordinal		
Interval		
Ratio		

Python IDE

Hands on Tools



Please choose one of tools above to complete hands-on activities.





Please obtain Basic Python Certificate

<https://www.kaggle.com/learn/intro-to-programming>

in Two Weeks



Post it in your LinkedIn Account and don't forget to tag your lecturers and asdos. Please add [#IMT_Statistics](#) in your post.

Next Meeting > Numerical Measures

INFORMATICS

