Chapter 1
Introduction to Statistics

MAS291 - STATISTICS & PROBABILITY

Ly Anh Duong

duongla3@fe.edu.vn

Table of Contents

What is Statistics?

1 Introduction

Statistics is the science of **collecting**, **organizing**, **analyzing**, and **interpreting DATA** in order to **make decisions**

Descriptive Statistics:

Involves **organizing**, **summarizing**, and **displaying** data.

e.g. Tables, charts, averages

Inferential Statistics:

Involves using **sample data** to draw conclusions about **a population**.

Big picture of Statistics

1 Introduction

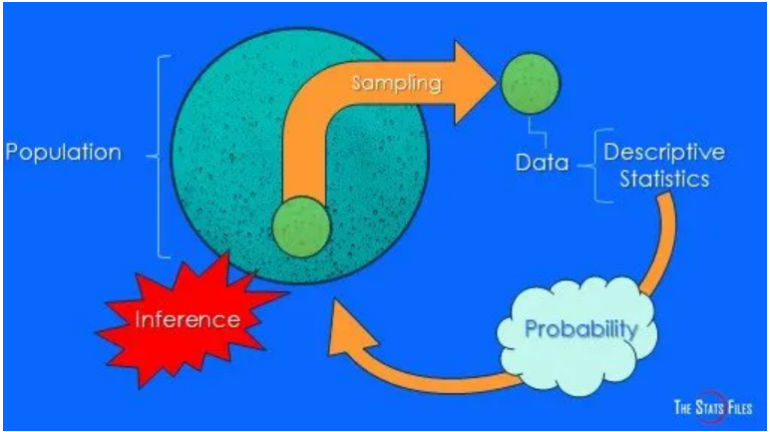


Table of Contents

2 Data

Statistical concepts

2 Data

- **Population (tổng thể):** the complete collection of all individuals to be studied.
- **Sample (mẫu):** Sub-collection of members selected from a population.
- **Data:** consist of information coming from observations, counts, measurements, or responses.
- **Parameter (tham số):** a numerical measurement describing some characteristic of a population.
- **Statistic:** a numerical measurement describing some characteristic of a sample.
- **Census (điều tra):** collection of data from every member of a population

Type of data

2 Data

Qualitative data

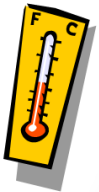
Major

Place of birth

Quantitative data

Age

Temperature



Discrete

Continuous

Table of Contents

3 Data Collection

The basis methods

3 Data Collection

- **Retrospective study (nghiên cứu hồi cố):** using historical data.
- **Observational study (nghiên cứu quan sát):** A researcher observes and measures characteristics of interest of part of a population.
- **Designed experiment (thực nghiệm):** A treatment is applied to part of a population and responses are observed.

Table of Contents

4 Errors

Errors

4 Errors

- **Sampling error:** the difference between a sample result and the true population result; such an error results from chance sample fluctuations.
- **Non-sampling error:** sample data incorrectly collected, recorded, or analyzed (such as by selecting a biased sample, using a defective instrument, or copying the data incorrectly).

Table of Contents

5 Problem

Problem

5 Problem

1. Explain the statistical terms as listed below:
a/ Population - Sample
b/ Parameter - Statistic
c/ Observational study - Experiment - Case study
d/ The type of observational study: Cross - sectional, Retrospective and Prospective
e/ Quantitative data - Qualitative data
f/ Discrete data - Continuous data
g/ Mechanistic model - Empirical model - Probability models
h/ Collecting data - Analysis data - Presentation data
i/ Random sample - Random variable

Problem

5 Problem

2. The US government wants in know how American citizens feel about the war in Iraq. They randomly select 500 citizens from each state and ask them about their feeling. What are the population and the sample?

3. Determine whether the given value is a statistic or a parameter.

a/ A sample of 120 employees of a company is selected, and the average is found to be 37 years.

b/ After inspecting all of 55,000 kg of meat stored at the Wurst Sausage Company, it was found that 45,000 kg of the meat was spoiled.

Problem

5 Problem

4. Is the study experimental or observational?
a/ A marketing firm does a survey to find out how many people use a product. Of the one hundred people contacted, fifteen said they use the product.
b/ A clinic gives a drug to a group of ten patients and a placebo to another group of ten patients to find out if the drug has an effect on the patients' illness.
5. Identify the type of observational study.
a/ A statistical analyst obtains data about ankle injuries by examining a hospital's records from the past 3 years.
b/ A researcher plans to obtain data by following those in cancer remission since January of 2015.
c/ A town obtains current employment data by polling 10,000 of its citizens this month.

Problem

5 Problem

6. Identify the number as either continuous or discrete.
a/ The total number of phone calls a sales representative makes in a month is 425.
b/ The average height of all freshmen entering college in a certain year is 68.4 inches.
c/ The number of stories in a Manhattan building is 22.
7. Classify each set of data as discrete or continuous.
a/ The number of suitcases lost by an airline.
b/ The height of corn plants.
c/ The number of ears of corn produced.
d/ The time it takes for a car battery to die.

8. Fill in the bank
a/ Observational Study is a basic method of ...
b/ Designed Study is a basic method of ...
c/ Retrospective Study, observational study and designed experiment are three basis methods of ...
d/ A designed experiment is a method of ...

Q&A

Thank you for listening!