Chapter 1

Introduction

Outline of Chapter 1

Reference: Chapter 1 of Elementary Statistics by ACS

STATISTICS

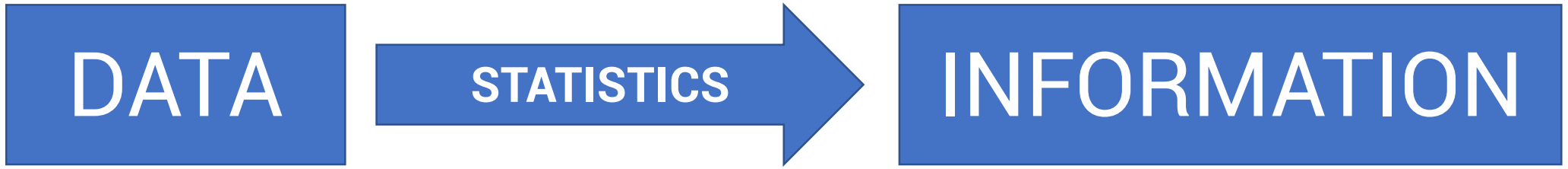
Introduction

Definition

Statistics is the branch of science that deals with the **collection**, **organization**, **analysis**, **interpretation** and **presentation** of data.

Why Study Statistics?

- Information empowers!



- Better decisions and intelligent choices supported by real life data

Statistical Inquiry

Definition

Statistical Inquiry is a designed research that provides information needed to solve a research problem.

- "Information" is processed data
- What distinguishes a statistical inquiry from other types of research is the use of data that have been collected and analyzed in answering the research problem.
  - ➢ We collect data about a particular collection of elements, process the data, and answer the research problem.

Statistical Inquiry: Example

Problem 1: What is the distribution of Stat 101 students in this class in terms of age (in years)?

Problem 2: Who will win in the 2022 Presidential Elections?

Problem 3: What is the most profitable product offered by the company in the previous year?

Population and Sample

Definition

Population is the collection of all elements under consideration in a statistical inquiry.

Sample is a subset of the population.

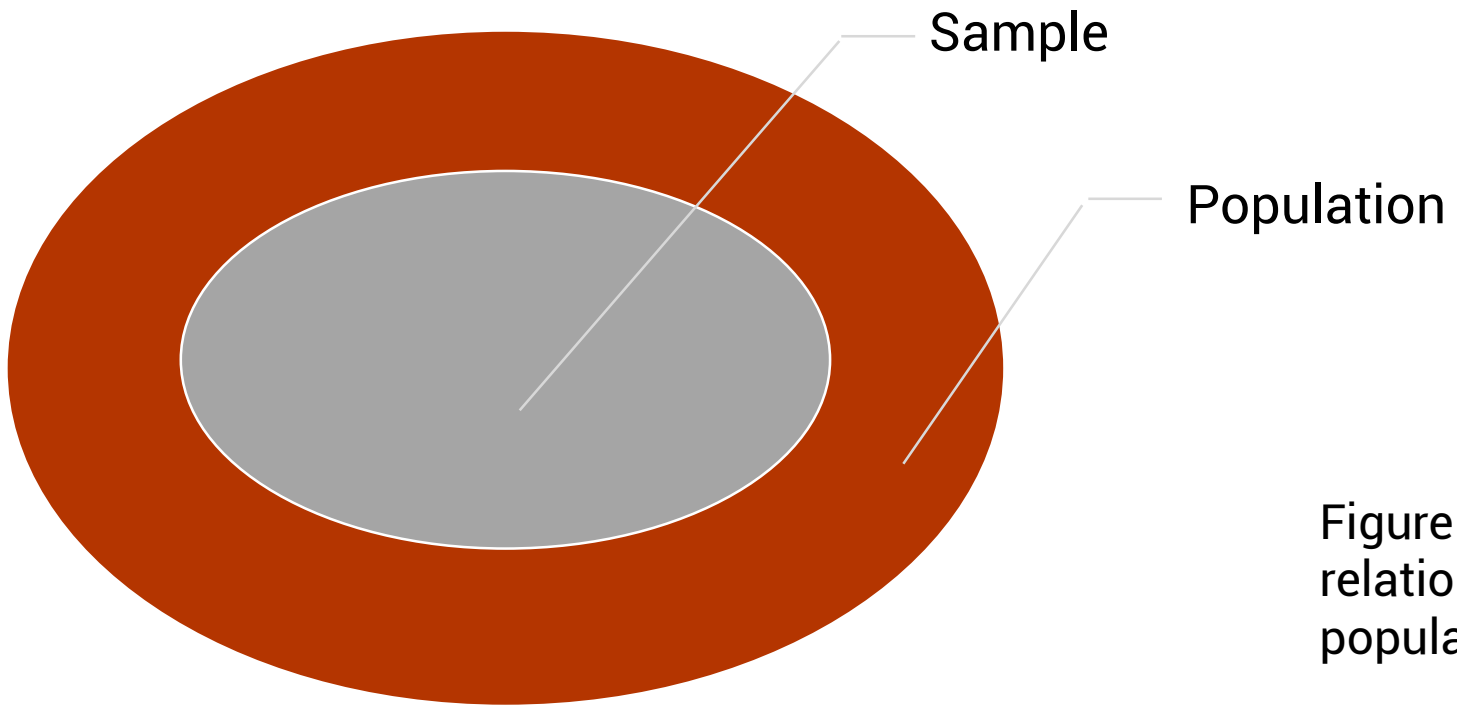


Figure 1. Diagram depicting the relationship between the population and the sample.

Population and Sample: Example

Problem 1: What is the distribution of Stat 101 students in terms of age (in years)?

- **Population**: Collection of Stat 101 students
- **Possible Sample**: Stat 101 students in the 1st row

Problem 2: Who will win in the 2022 Presidential Elections?

- **Population**: Set of all registered voters
- **Possible Sample**: 500 registered voters each from Luzon, Visayas, and Mindanao

Population and Sample

- The scope of the study is a major consideration in choosing the population of interest
- Due to (budgetary) constraints, the researchers may opt to delimit the scope of the study
- Possibly, consider a smaller or more specific population of interest
- The elements of the population can be individuals, objects, animals, geographic areas, among others.

Population and Sample: Example

- **Problem**: What is the average expenditure of households in Metro Manila?
- **Population**: set of all households in Metro Manila

- **Problem**: What is the average expenditure of households in Quezon City?
- **Population**: set of all households in Quezon City

Example of population with people as elements:
- set of farmers in Central Luzon

Examples of population with animals/objects as elements:
- collection of milkfish cultured in Luzon
- set of fluorescent bulbs manufactured for a month

Examples of population with areas as elements:
- set of barangays in Metro Manila
- collection of fishponds in Pampanga

Variable, Observation and Data

Definition

The **variable** is a characteristic or attribute of the elements in a collection that can assume different values for the different elements.

An **observation** is a realized value of a variable.

Data is the collection of observations.

- In a Statistical Inquiry, we will measure at least one characteristic of the elements of the population under study.

Variable, Observation and Data

Variable	Possible Observations
• S = Sex of a student	Male, Female
• E = Employment status	Temporary, Permanent, Contractual
• I = Monthly income (Php)	$i \geq 0$
• N = Number of Children	$n = 0, 1, 2, \ldots$
• H = Height of a player (cm)	$h > 0$

Example

Elements: students in an Elementary Statistics course

Variable: X = age of a student in an Elementary Statistics class

The ages of all the 35 students in this class were determined and recorded as follows:

16, 16, 16, 16, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 19, 19, 19, 19, 19, 19, 20, 20, 20, 21, 21, 22

This collection of observations, or realized values of X, is our **data**.

Example

A doctor claims that three tablespoons of pure virgin coconut oil daily can reduce weights of women. To test the doctor's claim, a researcher studied two groups of 25 women aged 35 to 40 years with weights between 130 to 140 pounds. He administered the three tablespoons of pure virgin coconut oil daily for a period of three months to one group of women only. After three months, he weighed the two groups of women.

a) Identify the population of interest.

Collection of all women aged 35 to 40 years with weights between 130 to 140 pounds

Example

A doctor claims that three tablespoons of pure virgin coconut oil daily can reduce weights of women. To test the doctor's claim, a researcher studied two groups of 25 women aged 35 to 40 years with weights between 130 to 140 pounds. He administered the three tablespoons of pure virgin coconut oil daily for a period of three months to one group of women only. After three months, he weighed the two groups of women.

b) What are the two samples?

A group of 25 women who will take 3 tablespoons of PVCO daily and a group of 25 women who will not

Example

A doctor claims that three tablespoons of pure virgin coconut oil daily can reduce weights of women. To test the doctor's claim, a researcher studied two groups of 25 women aged 35 to 40 years with weights between 130 to 140 pounds. He administered the three tablespoons of pure virgin coconut oil daily for a period of three months to one group of women only. After three months, he weighed the two groups of women.

c) What is the variable of interest?

Differences in weight (before and after) of the group who will take PVCO and those who will not

Summary Measure

A **summary measure** is a single numeric figure that describes a particular feature of the whole collection.

Definition

The **parameter** is a summary measure describing a specific characteristic of the population.

The **statistic** is a summary measure describing a specific characteristic of the sample.

Summary Measure: Example

Consider the case where our population consists of 35 students in an Elementary Statistics class. The parameter of interest is P = proportion of students in this class with Facebook account.

$$P = \frac{no.\ of\ elements\ possessing\ a\ certain\ characteristic}{no.\ of\ elements\ in\ the\ collection}$$

$$P = \frac{no.\ of\ students\ in\ this\ class\ with\ Facebook\ Account}{no.\ of\ students\ in\ this\ class}$$

The variable under study is X, whether or not the student owns a Facebook Account. This variable has only two possible values: With or Without.

Suppose that among the 35 students, 28 have a Facebook account. Thus,

$$P = \frac{no.\ of\ students\ in\ this\ class\ with\ Facebook\ Account}{no.\ of\ students\ in\ this\ class} = \frac{28}{35} = 0.8$$

Summary Measure: Example

Suppose we were not able to collect data from all the 35 students. Instead, we only took a sample of 10 students from this class. Among the 10 students in the sample, 7 have Facebook accounts. Can you compute for the parameter, P?

We cannot compute for the parameter, P=proportion of students in the population with Facebook accounts but we can compute for $\hat{P}$ (read as "P hat"), where $\hat{P}$ = proportion of students in the sample with Facebook accounts as follows:

$$\hat{P} = \frac{no.\ of\ students\ with\ Facebook\ Account\ in\ the\ sample}{no.\ of\ students\ in\ the\ sample} = \frac{7}{10} = 0.7$$

The proportion of students in our sample with Facebook account is an example of a statistic because it is a summary measure describing a characteristic of the sample.

Summary Measure: Example

Suppose we redefine the population as the collection of all students enrolled in all sections of Elementary Statistics so that the class consisting of 35 students earlier is now just a sample taken from this new population.

Is the earlier computed proportion of 0.8 a parameter or a statistic?

Example

Mr. Donaldo Chan, a candidate for vice-mayor in Orion, Bataan, wants to find out if there is a need to intensify his campaign efforts against his opponents. He requested the services of a group of students to interview 1,000 of the 3,000 registered voters of Orion, Bataan. The survey results showed that 75% of the 1,000 voters in the sample will vote for him as vice-mayor.

a) Identify the population and the sample.

b) Identify the variable of interest.

c) Identify the parameter and/or the statistic.

Example

a) Identify the population and the sample.

Population: Collection of all registered voters in Orion, Bataan.

Sample: A group of 1000 registered voters in Orion, Bataan.

b) Identify the variable of interest.

- Who will the respondent vote for; or
- Whether the respondent will vote for Donaldo Chan or not

c) Identify the parameter and/or the statistic.

Statistic: 75%

Two Major Fields of Statistics

Applied Statistics

- concerned with **procedures** and **techniques** used in the collection, presentation, organization, analysis, and interpretation of data.
- allows us to **select and properly implement** the **most appropriate statistical methods** that will provide solutions to the research problem

Theoretical or Mathematical Statistics

- concerned with the development of the **mathematical foundations** of the methods used in applied statistics.
- permits us to **understand the rationale** behind the methods we use in analysis and to **establish new theories** that will validate the use of new statistical methods or modifications of existing statistical methods in solving research problems that are more complex.

Two Major Areas of Applied Statistics

Descriptive Statistics

- includes all the techniques used in organizing, summarizing and presenting the **data on hand**.
- cannot be used to make generalizations about the population if the data on hand is simply sample data.

Inferential Statistics

- includes all the techniques used in analyzing the **sample data** that will lead to **generalizations about a population** from which the sample came from.
- We arrive at our conclusions under conditions of uncertainty because we use partial information only. Conclusions will be subject to some error. Probability theory will help us understand the possible errors that can be committed.

Example: Descriptive Statistics

- Given the **daily sales performance** for a product for the previous year, we can draw a **line chart** or a **column chart** to emphasize the upward or downward movement of the series.
- Likewise, we can use descriptive statistics to calculate a **quantity index per quarter** to **compare** the sales per quarter for the previous year.

Example: Inferential Statistics

Election polls make use of inferential statistics to predict the winners for the coming election based on data collected from a sample of registered voters.

Example: Krystal Surveys on the 1998 Presidential Elections

Sample size: 500 Metro Manila

2387 Other provinces

2887

	Actual (based on quick count)	Estimate (based on survey)
Estrada	40%	34%
De Venecia	16%	14%

Exercise

For each of the following problems below, decide whether the method that will be used belongs in the field of **descriptive statistics** or **inferential statistics**.

#2. Janine wants to determine the variability of her six exam scores in Algebra.

#6. A car manufacturer wishes to estimate the average lifetime of batteries by testing a sample of 50 batteries.

#8. a) A marketing research group wishes to determine the number of families not eating three times a day in the sample used for their survey.

#8. b) A marketing research group wishes to determine the number of families in the Philippines not eating three times a day based on the sample used for their survey.

#9. A politician wants to determine the total number of votes his rival obtained in the past election based on his copies of the tally sheet of electoral returns.

Reading Assignment

Statistical Inquiry

- Definition of statistical inquiry
- Possible objectives in a statistical inquiry
- Steps in conducting a statistical inquiry