Introduction to Statistics

Statistics is a branch of mathematics, which involves:

- Collecting data.
- Organizing and analyzing data.
- Drawing conclusions from data.

Applications of statistics occur nearly in all fields of study, including accounting, finance, management and management information systems.
For example, people in business use statistics to find out what shoppers like or dislike, what product to buy or manufacture, and how advertising affects sales. Managers use statistics to decide how many workers they should employ and what types of special education they should have. People use statistics in order to find out where best to invest money.

Example 1

Sales managers wanted to know what makes people want to attend trade shows. They conducted a statistical survey asking people to rank the factors in their decision to attend. These people were employed in banking, finance, insurance, food and beverage, and computers. Some of the things listed included general interest, speakers, previously attended trade shows, investigating new products, and meeting other professionals.

By developing the survey, conducting it, gathering the information, arranging the information, studying the results, and reaching conclusions, we are using statistics. That is, we:

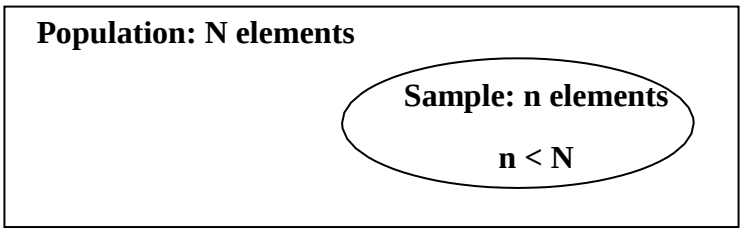
1. Collect the data.
2. Analyze the data.
3. Interpret the data and make a decision that will ensure the success of future trade shows.

Population & Sample

Population: The entire group of objects or individuals whose properties are to be analyzed. The population always contains all elements of the group that we would like to analyze. Depending on the application, a population could be the whole population of Canada, or the students of a college, or the buyers of a certain brand, etc.

Sample: A part or subset of the population used to gain information about the whole.

The following diagram shows the relationship between population (the big box) and the sample (the small circle inside the big box).



A population always yields accurate results. A measure calculated from the population is called a ***parameter***. However finding the exact population parameter is not always possible due to economic constraints like time and money. But as Samuel Johnson once said: "You don't have to eat the whole ox to know that the meat is tough." When the population cannot be examined, we can approximate the population parameter by calculating the same measure from a sample. A measure calculated from a sample is called a ***statistic***. A sample statistic is never accurate, but in many cases we can use it as a fair approximation of the real population parameter.

For example, in conducting quality control experiments to determine the average life of light bulbs a manufacturer might randomly sample only fifty light bulbs during daily production to evaluate the production process. Testing all of the light bulbs would not only be time consuming and expensive, but it also would be impossible, because after the tests the bulbs do not produce light any more, and we cannot sell them. In this case we must use a sample to evaluate the life time of the bulbs.

A **random sample** is a sample chosen from the whole population in such a way that every element in the population has the same chance of being chosen. For example, if we investigate what people think of driving standards in the UAE, it would be wrong to ask only taxi drivers. We must *randomly* ask people from all nationalities, male and female, young and old, drivers and non-drivers.
A carefully selected random sample can be a good representative of the population.

Example 2

Joe Free Zone has gained recognition in the past ten years of becoming the commercial center in USA. There are several hundreds of companies there. In its effort to improve services, the local government of New York randomly selected 62 companies operating in Joe free zone and requested them to fill out a detailed questionnaire about existing levels of services and proposed improvements.

(a) Which of the following is the population in this example?

i) All the factories in USA. ii) All the business centers in New York.

iii) All the factories in Boston. iv) All the factories in Joe free

(b) Describe the sample of the experiment.

Solution:

(a) iv) (b) The sample consists of the selected 62 companies operating in Joe free

Classification of Data

Data can be classified into two broad types: **qualitative** and **quantitative**.

When data are classified according to qualities or physical attributes, they are said to be *qualitative*. For example, gender, occupation, nationality, and color are qualitative. They are not measured with a scale; they represent categories only.

When data yield measurements that are numerical in nature, they are said to be *quantitative*. Ages, height, weight, speed, number of patients in a hospital are examples of quantitative data. Numerical data can be manipulated by arithmetic expressions (+, -, ×, ÷).

Some data can be described with numbers, still they are qualitative data. For example passport numbers, or telephone numbers are expressed with numerals, but they cannot be manipulated with arithmetic expressions, so they are *qualitative data*.

Example 3

Classify the following data as qualitative or quantitative

(a) The weather forecast for next week is: sunny, sunny, cloudy, foggy, cloudy, rainy, thunderstorms.

(b) The most popular color among NYU motorists in white.

(c) The average height of NYU students is 155 cm.

(d) Most of Stanford students are smart.

(e) The average weight of a hamour fish is 1.6 kg.

(f) Crazy drivers drive at an average speed of 200 km/h.

Solution:

(a)	qualitative	(b)	qualitative	(c)	quantitative
(d)	qualitative	(e)	quantitative	(f)	quantitative

Variables

Quantitative data may be further classified into discrete and continuous data depending on the variable they describe.

A *variable* is something that can have different values when observed in different people, places, or things. For example, marital status (single, married, divorced), age, gender (male, female), occupation (teacher, manager, cashier), population density, quality of a manufactured product, prices and wages.

Variables may be **discrete** or **continuous**.

A discrete variable is a variable that can take only selected values. Some examples are the number of students in a class and the number of rooms in a house. There cannot be 15¼ students in the class, but there can be 15 or 16 students. *A discrete variable is mostly the result of a count.*

A continuous variable is one that can assume any value within a specified interval. Height of a student in the class, the time required to complete a task, weight of a shipment, the length of time before the first failure of a device, the profits or revenues of a company, are some examples. The height of a student cannot jump from 1.5 m to 1.6 m. It must assume every value between 1.5 m and 1.6 m at some time. *A continuous variable is mostly the result of a measurement.*

Example 4

Indicate which of the following are **continuous** and which are **discrete**:

(a) The number of children in a family.

(b) The weight of a person.

(c) The speed of a car.

(d) The number of plants in a garden.

(e) The capacity of an oil tank in liters.

(f) Number of units of an item held as inventory.

Solution:

(a)	discrete	(b)	continuous	(c)	continuous
(d)	discrete	(e)	continuous	(f)	discrete

Exercise 2.1

1. The Fishery Co. in San Francisco conducted a study on fishing practices among its 2366 registered fishermen in 1996. A random selection of 150 fishermen was asked to answer a group of questions. The answers gathered were organized, studied, and analyzed. A conclusion was reached to further guide these fishermen on safe fishing practices.

   (a) Describe the population.

   (b) Describe the sample.

2. A recent survey conducted by the NY Newspaper asked 1200 readers at random if they were satisfied with municipality services. 68% answered yes, 29% answered no, and 3% did not care.

   (a) Describe the population.

   (b) Describe the sample.

3. In order to open a new shopping center in Boston, a private company was hired to make a feasibility study. A survey was conducted among 1200 shoppers in different places in Boston. Based on the findings, the company allocated floor area to the different categories of food, men's fashion, women's fashion, playing grounds, etc…

   (a) Describe the population of this example. Does the population consist of all the residents in the USA?

   (b) Describe the sample of the example.

4. Indicate which of the following are *qualitative* and which are *quantitative variables*:

   (a) The religion of a person.

   (b) The color of a boat.

   (c) The speed of a plane.

   (d) The temperature of a room.

   (e) The amount of money deposited in a bank.

5. Shell wants to determine the number of its service stations that stay open 24 hours a day in the USA.

   (a) What is the population?

   (b) Are the data collected qualitative or quantitative?

6. Students enrolled at Harvard requested that the college begin a full 4-year program leading to a university degree. Reporters from the Boston Times and the USA News asked a group of students during lunchtime in the cafeteria whether they thought the college would agree.

   (a) What was the population in this survey?

   (b) Would the answers to the newspaper questions be from a sample or the population?

   (c) Would the results represent the opinions of all students in USA?

   (d) Would the results of this survey consist of quantitative or qualitative information?

7. Identify if the following observations are from *discrete* or *continuous* variables?

   (a) In a batch of 2000 radios, 14 were found to be defective.

   (b) Last week 6 students were absent from their Math 100 class.

   (c) A light truck weighs 2 tons.

   (d) On the last Math 100 test, there were 2 perfect scores.

   (e) A radar on I-95 Road spotted Joe driving at 180 km/h.

   (f) Mike spends an average of 3½ hours studying every day.

   (g) 254 students voted for Amy as the president of the student council.

   (h) Mary lost 4.8 kg on her latest diet.

Exercise 1 Answers

1 (a) 2366 registered fishermen (b) 150 fishermen

2 (a) USA (or New York) population (b) 1200 Readers

3 (a) all Boston residents (b) 1200 shoppers

4 (a) qualitative (b) qualitative (c) quantitative
(d) quantitative (e) quantitative

5 (a) all Shell service stations. (b) quantitative

6 (a) Harvard students (b) sample (c) no
(d) qualitative

7 (a) discrete (b) discrete (c) continuous
(d) discrete (e) continuous (f) continuous
(g) discrete (h) continuous