# Chapter I

# Introduction

**D. 1. 1. (Statistics)**
Statistics as a branch of applied mathematics is the science of collecting, organising, presenting, analysing, and interpreting data.
It comprises

1. *Descriptive Statistics*
2. *Inductive* (or *Inferential*) *Statistics.*

*Descriptive Statistics* is used to *describe* the basic features of the data in a study. It provides simple summaries about a *population* or a *sample*.

Departing from a *sample* of observations (a sub-collection of all observations), *Inductive* (or *Inferential) Statistics* attempts to make general conclusions about the so-called (*target*) *population* (consisting of all possible observations). The inductive statistics makes intensive use of probability theory and methods.
Whereas the measured characteristics of a sample are called *sample statistics*, the measured characteristics of a statistical population are called *population parameters*.

**Ex. 1. 1.**
1) Statements where descriptive statistics has been used:

1. Increase in the annual sales of cars in the country was 10% of the sales of the previous year.
2. There are 9589 regular users of Internet in a university of 10000 students.

2) Statements where inferential statistics has been used:

1. The chance of getting a defective TV from a batch of 600 TV's is 0.03.
2. There is no evidence that the home price in a small German town is higher than 350000.00 €.

**D. 1. 2. (*(Target) Population*)**
A *population* consists of all elements –individuals, items, or objects – whose characteristics are being studied.

**D. 1. 3. (*Sample*)**
A portion of the population selected for study is referred to as a *sample*.

**Ex. 1. 2.**
1) Examples of population data:

1. The result of *every* person who took the SAT in a given year.
2. The ballots of *every* person who voted in a given election.
3. Prices of petrol at *every* petrol station in the country.

2) Examples of sample data:

1. The results of 2000 students who took the SAT in a given year.
2. The choices made by 5% of all people who voted in a given election.
3. Prices of petrol at 1 out of every 10 petrol stations in the country.

**D. 1. 4. (*Representative Sample*)**
A sample that represents the characteristics of the population as closely as possible is called a *representative sample*.

**D. 1. 5. (*Random Sample, Simple Random Sample*)**
A sample drawn in such a way that each element of the population has a chance of being selected is called a *random sample*.
If all samples of the same size selected from a population have the same chance of being selected, we call it a *simple random sample*.

**D. 1. 6. (Statistical Units, Characteristics and their Attributes)**

(Statistical) Unit: = "Subject of statistical investigation".

(Statistical) Characteristic: = "Object of statistical investigation"

(Statistical) Attributes: = "Properties of statistical characteristics".

**Ex. 1. 3.**

| Characteristic | Unit | Attributes |
|---|---|---|
| family status | HTW students | single, married, divorced, widow(er) |
| gender | HTW students | man, woman |
| weight | FHTW students | 50 – 120 kg |
| marks in mathematics | BIB students | very good, good, sufficient, ... |

**R. 1. 2. (Classification of Statistical Characteristics)**
There are two ways of classifying statistical characteristics: either into qualitative and quantitative or according to the type of measurement scale.

**D. 1. 7.**

$$\text{characteristic} := \begin{cases} 1.\ \text{qualitative} \begin{cases} 1.1.\ \text{nominal} \\ 1.2.\ \text{ordinal} \end{cases} \\ 2.\ \text{quantitative} \begin{cases} 2.1.\ \text{continuous} \\ 2.2.\ \text{discrete} \end{cases} \end{cases}$$

1.
The attributes of a *qualitative (categorical) characteristic* are characterised by their *kind*.

1. 1.
A *nominal characteristic* is measured on a so-called *nominal scale*. There is *no* hierarchy of attributes; the distances between the attributes can *not* be interpreted.

(*Examples*: eye colour, occupation, social status, religious affiliation, gender).

1. 2.
An ordinal *characteristic* is measured on a so-called *ordinal scale*. There *is* a hierarchy of attributes; the distances between the attributes can *not* be interpreted.

(*Examples*: drinking habits, educational institutions, marks).

2.
The attributes of a *quantitative characteristic* (or a *variate* or a *variable*) are characterised by their *magnitude*. There *exists* not only a hierarchy of their attributes (values), but also the distances between them *can* be interpreted.

2. 1.
*Continuous variates* may possess (at least theoretically) *any* value within an *interval*.

(*Examples*: temperature, length, age)

2. 2.
*Discrete variates* may possess only certain *distinct* attributes (values). The number of the values can be finite or countably infinite.

(*Examples*: number of bad items in a sample, number of men at a university, monthly income in €)

**D. 1. 8. (*Types of Measurement Scale*)**

1. *Nominal characteristics* allow for only qualitative classification. They can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories.

   (*Examples*: gender, colour, “race”)

2. *Ordinal characteristics* allow us to rank the items we measure in terms of which has less and which has more of the quality represented by the characteristic, but still they do not allow us to say "how much more."

   (*Examples*: socio-economic status of families. For example, we know that upper-middle is higher than middle but we cannot say that it is, for example, 18% higher.)

3. *Interval characteristics* allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them.

   (*Examples*: temperature measured in degrees Fahrenheit or Celsius .We can say that a temperature of 40 degrees is higher than a temperature of 30 degrees, and that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees.)

4. *Ratio characteristics* have all properties of interval characteristics. In addition to those properties, they feature an identifiable absolute zero point. Thus they allow for statements such as one characteristic is twice more as another.

   (*Examples*: measures of time or space.)

*(Last revised: 26.03.09)*