Statistics & Probability

Introduction

Original goal of statistics was to collect data about population based on population samples.

By population we mean a group of all existing components available for observation during statistical research. For example:

If a statistical research is performed about physical hight of 15-year old girls, the population will be all girls currently aged 15.

Considering the fact that the number of population members is usually high, the research will be based on the so-called **sample examination** where only part of the population is used. The examined part of the population is called a **sample**. What's really important is to make a definite selection that is as representative of the whole group as possible.

There are several ways to achieve it. To avoid of omitting some elements of the population the so-called **random sample** is used in which each element of population has the same chance of being selected.

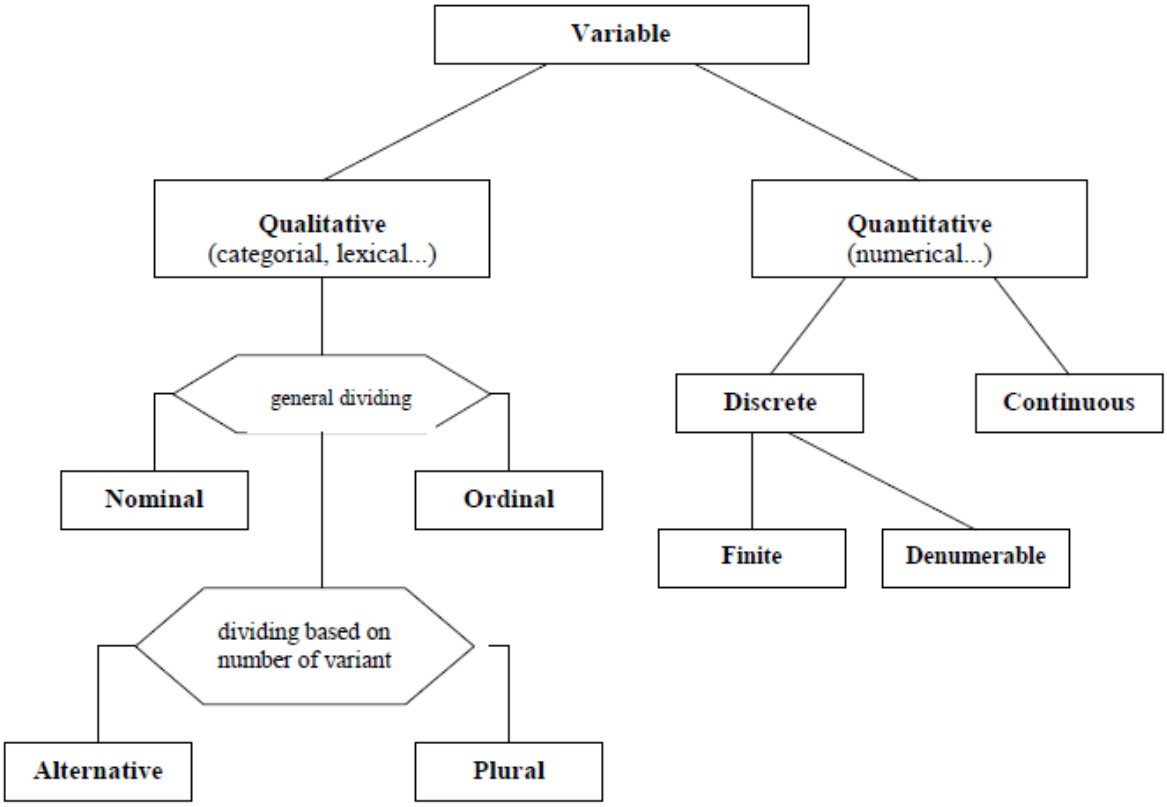
It goes without saying that sample examination can never be as accurate as examining the whole population. Why do we do prefer it then?

1. To save time and minimize costs (especially for large populations).
2. To avoid damaging samples in destructive testing (some tests like examining cholesterol in blood etc., lead to the permanent damage of examined elements).
3. Because the whole population is not available.

Now that you know that statistics can describe the whole population based on information gathered from a population sample we will move on to Exploratory Data Analysis (EDA).

Data we observe will be called **the variables** and their values **variable variants**. **EDA** is often the first step in revealing information hidden in a large amount of variables and their variants.

Because the way of processing variables depends most on their type, we will now explore how variables are devided into different categories. The variables division is shown in the following diagram.


Variable
Qualitative
(categorial, lexical...)
Quantitative
(numerical...)
general dividing
Discrete
Continuous
Nominal
Ordinal
Finite
Denumerable
dividing based on
number of variant
Alternative
Plural