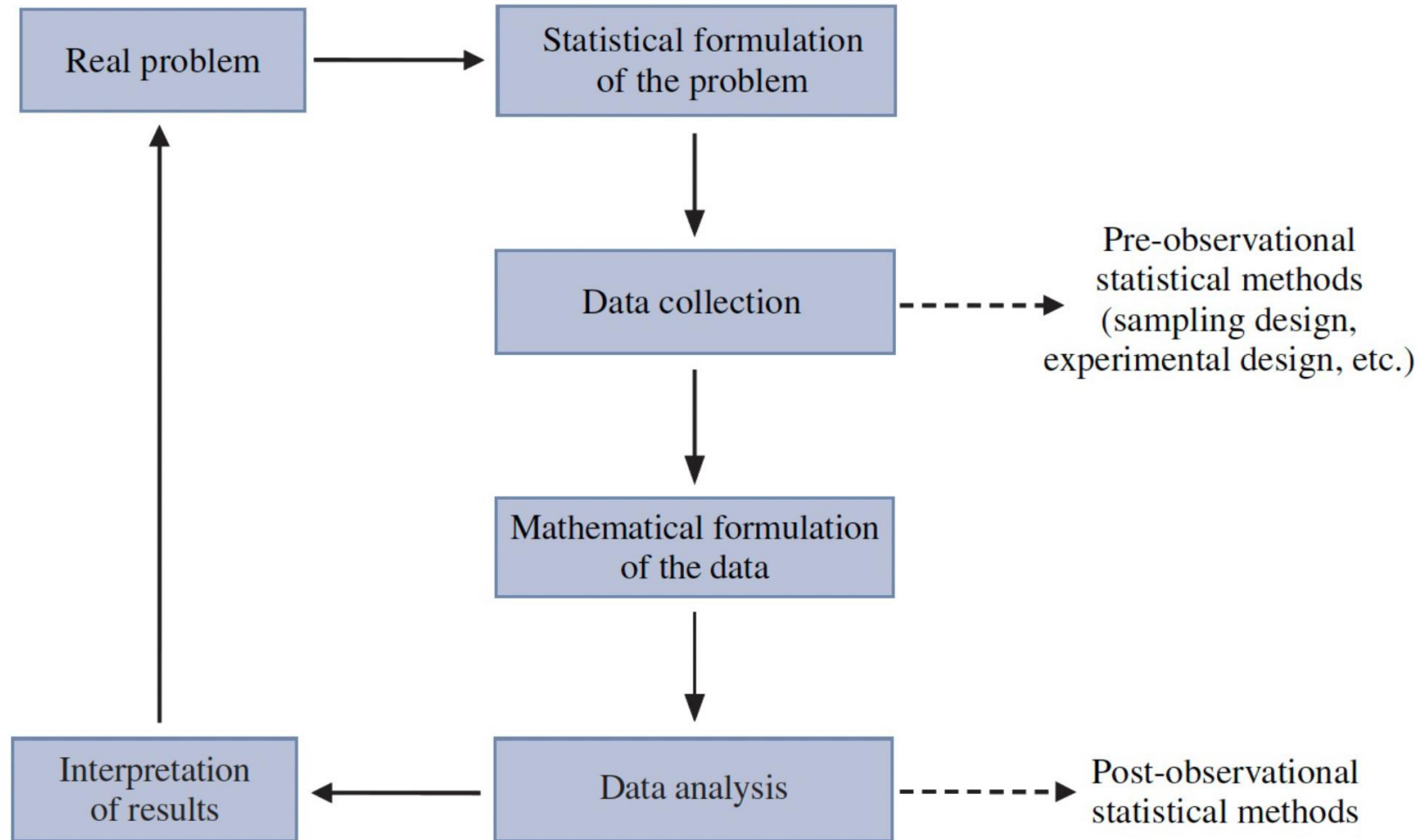# Introduction to statistics

What's statistics?

Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting and presenting empirical data.

Statistics is a highly interdisciplinary field; research in statistics finds applicability in virtually all scientific fields and research questions in the various scientific fields motivate the development of new statistical methods and theory.

Statistics and knowledge discovery problem

Terminology

population is the set of all entities we are interested in studying. It may be a collection of objects, transactions, events, people, etc..

sample is a subset of the units of a population obtained through a process of selection for the purpose of investigating the characteristics of the population.

Statistical units are the individual members of the population. They are the entities for which data are collected.

Variable is any characteristic or property of the statistical units that we want to analyze. It can "vary," taking different values from one unit to another (i.e. the colour of the unit).
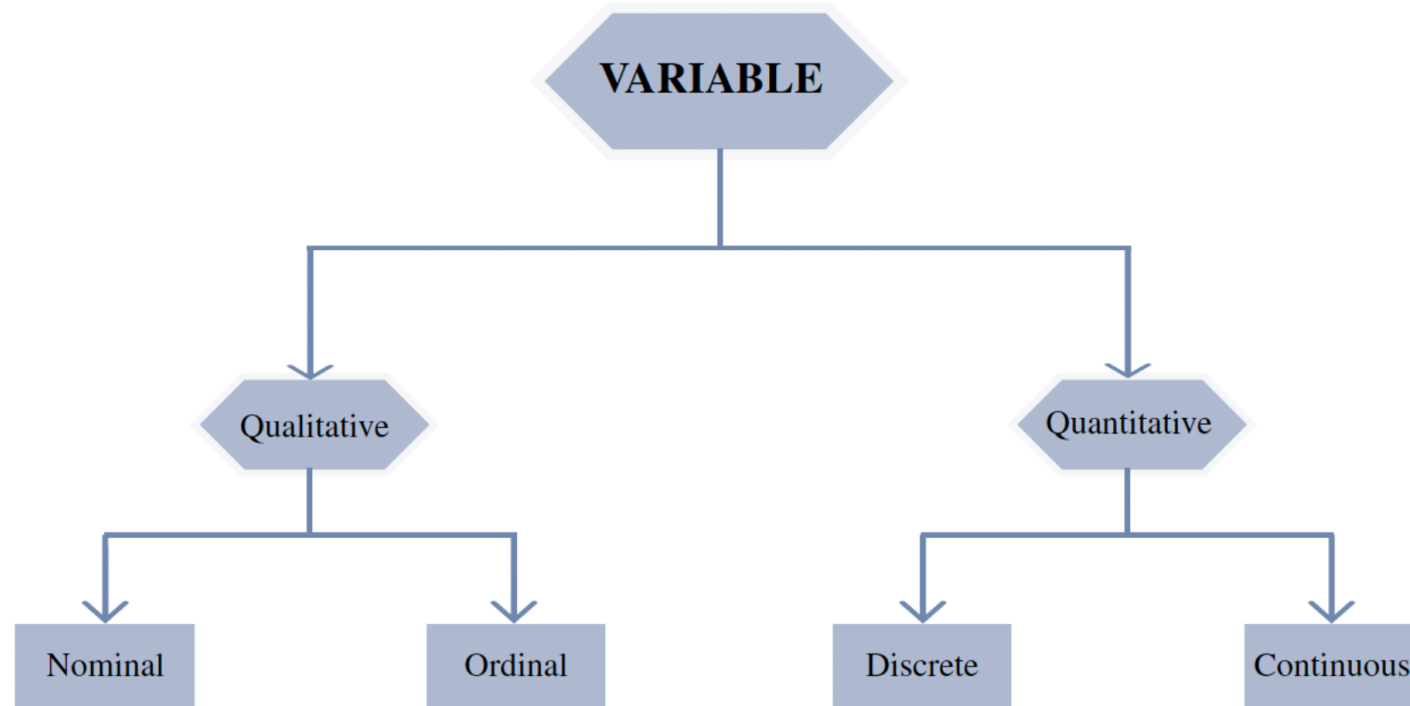
Notation

X indicates the variable under consideration

x indicates the value that X takes on a generic unit

n indicates the size of the sample i.e. the number of units under consideration

Type of variables

Descriptive vs inferential stats

Statistical data analysis requires that the phenomena of interest can be represented as mathematical variables and thus measured across groups of units

Descriptive statistics is the process of **quantitatively** describing or summarizing data using a variety of techniques known as statistics.

Inferential statistics is the process of using data to infer properties of an underlying population that cannot be fully observed, utilizing a sample taken from it.

Inferential statistics strictly relies on probabilistic (mathematical) models that describe the entire population, whereas descriptive statistics does not rely on such assumptions