

1 Data Exploration

Informative analyses to formulate and refine hypotheses for

- feature subset selection
- selection of tools and techniques
- further data collection

1.1 Distribution of a single variable

In statistics, a quantity that we measure or classify.

1.1.1 Populations and samples

We study properties of one or more samples to characterize a wider, much larger, population.

1.1.2 Data sets

Collection of values of raw variables.

1.1.3 Variables and observations

An observation is a list of values of all variables for a single member of a population.

1.1.4 Types of data

Numeric variables are measurable, quantitative, may be continuous (fractional) or discrete (integer).

Categorical variables are classified, qualitative, may be ordinal (ranked) or nominal (without ordering).

Scale can be interval (time) or ratio (length, weight).

Cross-sectional data is measured at certain point of time while time series data is measured at various points in time.

Longitudinal data combines cross-sectional with time series.