

Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

Lecture 9

Basic Terminologies of Exploratory Statistical Data Analysis

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from
<http://home.iitk.ac.in/~shalab/sp>



Statistical thinking and Methods

Using the information gained by the tools of descriptive statistics and combining them together to reach to a meaningful conclusion to depict the information hidden inside the data is the objective of any statistical analysis.

Proper interpretation of inferences drawn is important.

Inferences are drawn from the data.

Data generating process - Non deterministic, Random

Why collect data?

To verify theoretical findings.

Draw inferences just on the basis of collected data.

Developing statistical models, which can be further used for policy decisions, classification, forecasting etc.

Steps Involved

- **Identify objective(s) of Statistical Analysis**
- **How to get data?**
 - **Laboratory experiment,**
 - **survey,**
 - **primary data,**
 - **secondary data**
- **Use appropriate statistical tool**
- **Correct, valid and meaningful interpretation of the result.**

Observations

The units on which we measure the data are called observations.

For example, number of persons, cars, monthly expenditure on food etc.

Population

Collection of all the units is called population.

Example:

Objective: To find average age of all the female students in class 10 in a school on the basis of a sample.

Population: All the female students in class 10 in the school.

Population

Example:

Objective: To find how many female employees have salaries more than the respective male employees in a company on the basis of a sample.

Population: All the male and female employees in the company.

Sample

Sample is a subset of the population.

Selection of observations in the sample from the population is made in such a way that the sample is representative.

Sample is representative.

Representative sample

All characteristics present in the population are also present in the sample.

Sample

Various sampling schemes like

- simple random sampling,
- stratified sampling,
- cluster sampling,
- systematic sampling,
- multiphase sampling,
- multistage sampling etc.

are used to obtain a good sample.

Variables

Once a research question and the population of interest are identified, the observations are collected on a statistical variable.

Any information of interest is captured in such a variable.

Variables

Let the variable be represented by X .

Number of variables can be one or more than one.

- **Statistical analysis with one variable – univariate analysis.**
- **Statistical analysis with more than one variables – multivariate analysis.**

Values on Variables

Observations are collected on variables.

For example:

- If X is gender, then it takes 3 values – male, female, transgender
- If X is country in Asia, then it takes values – India, Bangladesh, China, Thailand etc.
- If X is any odd number, then it takes values – 1, 3, 5,...

Values on Variables

The values of a variable X are denoted by x

For example, let variable X is height of students.

Suppose height of two students is measured as 150 cms. and 160 cms.

Values on Variables

Then

height = 150 cms. and

height = 160 cms.

are the two values of height.

The two values of X are represented as

$x_1 = 150$ cms. and

$x_2 = 160$ cms.

Variables

Two types of variables:

Quantitative variables – Discrete and continuous

Qualitative variables

Quantitative Variables

Represent measurable quantities.

Values of X can be obtained.

Values of these variables can be ordered in a logical and natural way.

Quantitative Variables

Examples:

- **Sizes of shirt – 39, 40, 42, etc.**
- **Per kilo prices of vegetables – Rs. 30, Rs. 35, Rs. 45 etc.**
- **Number of colleges in a city – 8, 12, 15 etc.**
- **Heights of children – 1.2 m, 1.23 m, 1.32 m etc.**

Qualitative Variables

Represent non-measurable quantities.

Values x of variables cannot be ordered in a logical and natural ways.

Qualitative Variables

Examples:

- **Names of cities – Kanpur, Mumbai, Kolkata etc**
- **Colours of hair – Black, white, brown etc.**
- **Tastes of food – Sweet, salty, neutral etc.**
- **Performance – Good, excellent, bad etc.**

Qualitative Variables

Usually, numbers are assigned to qualitative variables.

Examples:

Variable : taste – sweet, salty, neutral.

Assign 1 to sweet.

Assign 2 to salty.

Assign 3 to neutral.