

Descriptive Statistics With R Software

Introduction to Descriptive Statistics

::

Objectives, Steps and Basic Definitions

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Objectives of Statistical Analysis

What is Statistics?

It is

- **Science of turning data in information for decision making.**
- **Collection, analysis and drawing inferences from numerical facts, referred as data.**

Descriptive statistics

Starting point for knowledge discovery on the basis of data.

Data

Very important source of information.

Data can not speak. Data can not listen. Data can not understand our language.

Data has its own language like sign language.

Role of Statistics

Statistics is a language of data.

Statistics provides a scientific way to extract and retrieve the information hidden inside the data.

Statistics can not do miracles.

Statistics can not change the process or phenomenon.

Statistical tools provide forecasting but not like astrologer's parrot.

Role of Statistics

	Correct data	Wrong data
Correct statistical tool	Correct decision	Incorrect decision
Wrong statistical tool	Incorrect decision	Incorrect decision

Rule: Garbage in – Garbage out

Statistics : Cannot do everything in no time.

Tools can be developed but development needs time.

Descriptive statistics

Several components and tools- Graphical, analytical

Graphical tools- various type of plots

- 2D and 3D plots,**
- scatter diagram**
- Pie diagram**
- Histogram**
- Bar plot**
- Stem and leaf plot**
- Box plot etc.**

Use appropriate graphics.

Analytical tools

Central tendency of data	Dispersion in data
<ul style="list-style-type: none">• Mean• Median• Mode• Geometric mean• Harmoninc mean• Quantiles etc.	<ul style="list-style-type: none">• Variance• Standard deviation• Standard error• Mean deviation,• Absolute deviation• Range etc.

Analytical tools

Structure of data	Relationships in data
<ul style="list-style-type: none">• Symmetricity• Skewness• Kurtosis etc.	<ul style="list-style-type: none">• Correlation coefficient• Rank correlation• Multiple correlation coefficient• Partial correlation coefficient• Corralation ratio• Intraclass correlation• Linear Regression• Non linear regression etc.

Statistical thinking and Methods

Which of the tools to be used – Graphical or analytical?

Use both types of tools.

Graphical tools provide a visulization – First hand information.

Analytical tools – Quantitative information.

Both approaches work together and are inseperable.

Statistical thinking and Methods

Both – graphical and analytical tools – work together in a system of interconnected processes.

Variation exists in all processes.

Understanding the extent of variation and reducing it are the keys to success.

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.