# Descriptive Statistics With R Software

## Introduction to Descriptive Statistics

## ::

## Variables and Types of Data

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

# 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 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.**

# Discrete Variables

**Variables can take a finite number of values.**

**Informally speaking, variables are "counted“.**

**Example:**

- **Number of children in a family – 1, 2, 3, etc.**
- **Number of branches of a school in a city – 4, 6, 7 etc.**

# Continuous Variables

**Variables can take an infinite number of values.**

**Informally speaking, variables are “measured“ and not “counted“ .**

**Example:**

- **Length of a road is 1.5 kms.**
- **Length of a road is 1.52 kms.**
- **Length of a road is 1.521 kms.**
- **…**

# Grouped Data

**Sometimes the original values of data are grouped or the data is available in the form of groups.**

**Original values in a group may not be known.**

**Only the category to which the values belong to is known.**

## Primary Data

**Data originally collected by an investigator for the first time for any statistical investigation.**

## Secondary Data

**Data which has already been collected by some person or agency for any statistical investigation.**

**Some data which is primary for one may be secondary for other.**

# Source of Primary Data

1. **Direct personal investigation**

2. **Indirect oral investigation**

3. **Questionaire received through postal mail, email, e-forms (google forms), online surveys etc.**

4. **Questionaire sent through surveyors.**

# Sources of Secondary Data

1. **Published sources**

2. **Data collected from survey agencies**

3. **Places where public reports the data, e.g.; muncipelities.**