# 1. Introduction to Types of Statistical Data

Welcome to the series of E-learning modules on Types of Statistical Data. In this module we are going to cover the different types of statistical data with the help of examples.

By the end of this session, you will be able to:

- Explain the meaning of Data
- Explain the following different types of data
    - Grouped and Ungrouped
    - Qualitative and Quantitative
    - Nominal and Ordinal
    - Interval and Rational
    - Discrete and Continuous
    - Time Series Data
    - Primary and Secondary

The word data is derived from the Latin word datum, which means 'something given' or a single piece of information. This could be a fact or a figure that is used specifically as a reference point or a single measurement or a derived result. For example in geography it is used as a specific location from which distances are measured.

Data is the plural of datum and means facts that are collected, summarized, analysed and interpreted. Today data is a more commonly used word than datum. In recent times we refer to singular pieces of information as data points rather than datum.

Thus, the information we collect about a topic or subject under investigation is called data. For example- Runs scored by Rahul Dravid in all the innings of his career may be referred to as data collected. Another example can be information collected by interviewing one thousand families in Bangalore about their weekly expenditure.

Data may refer to individual bits of information which is singular in nature or to the collection of information which is plural in nature. Data and information are frequently used interchangeably. We must remember that

data is the lowest level among the three. Data on its own carries no meaning. For data to become information it must be interpreted and take on a meaning. Data that is not subject to processing or any other manipulation is called raw data.

# 2. Grouped and Ungrouped Data

Grouped Data: The data which is in the tabular form and which has class intervals (or values) and frequencies is called grouped data.

Example: This example is in a tabular form and has class intervals and frequencies.

Here: The first column shows the Frequency Table where class intervals of age in years in shown.

The second column shows the frequency of a class is the number of numbers in that class.

**Figure 1**

| Age in Years<br>Class Intervals | Number of persons<br>Frequency |
|---|---|
| 0-10 | 20 |
| 10-20 | 39 |
| 20-30 | 42 |
| 30-40 | 34 |
| 40-50 | 12 |

Ungrouped Data: The raw data without class intervals and frequencies is called ungrouped data.

Example:

**Figure 2**

| Age in Years | 40 | 45 | 60 | 65 |
|---|---|---|---|---|

If ungrouped data is given to us we might wonder how we are going to group it.

Let’s take an example to understand this concept better-

We have the scores of Rahul Dravid for the last 10 innings he played. They are- 25, 1, 47, 9, 29, 5, 10, 68, 33 & 82 respectively. We can group this data with reference to the innings number (with respect to his entire career) that Rahul played-example

**Figure 3**

| Innings Number | Score |
|---|---|
| 605 | 25 |
| 604 | 1 |
| 603 | 47 |
| 602 | 9 |
| 601 | 29 |
| 600 | 5 |
| 599 | 10 |
| 598 | 68 |
| 597 | 33 |
| 596 | 82 |

We could group this data into various class intervals-

**Figure 4**

| Runs | Number of times scored |
|---|---|
| 0-9 | 3 |
| 10-49 | 5 |
| 50-99 | 2 |
| 100 and above | 0 |

When data is presented in this manner it gives the reader better information than when raw data is presented. Depending on the analysis to be done data can be grouped and presented in a form that will enable understanding.

Quantitative Data: Quantitative data is often referred to as the measurable data.

It measures length, height, area, volume, weight, speed, time, temperature, humidity, sound levels, cost, members, ages, etc. Each of these quantities will also have a measure. For example Weight of human beings is usually measured in Kilograms in India. Similarly world over age is measured in years. Certain measures of quantity are standard world over while certain others are country specific. For example in India we measure distances between two places in kilometres whereas in U.S.A it is measured in miles. We must be careful when quantifying data to ensure that it is converted into a common measure.

This type of data allows statisticians to perform various arithmetic operations, such as addition and multiplication, to find parameters of a population like mean or variance.

The observations represent counts or measurements, and thus all values are numerical. Each observation represents a characteristic of the individuals in a population or a sample.

# 3. Quantitative Data & Classification of Quantitative Data

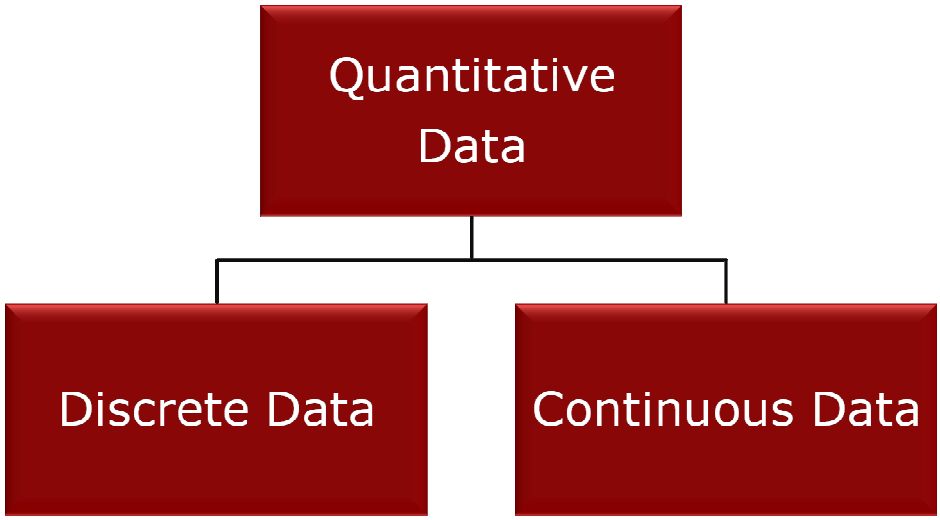
Quantitative data can be represented visually in graphs, histograms, tables and charts.

Example:

1. Height of children is easily measurable and countable.
2. Pulse rate of a human being is easily countable as the values are always numerical.
3. Similarly, we can also count the marks scored by the students in the aptitude test.
4. And also measure the time taken to finish the test.

Quantitative Data can be classified as

- Discrete Data
- Continuous Data




Discrete Data: The data which assumes only some specified values in a given range is called discrete data.

Discrete data has following characteristics:

- Discrete data applies only to numerical data
- It is finite in nature
- It is countable
- It has space between given values
- It is exact

Example:

1. The number of eggs a chicken lays in 24 hours
2. Number of children in a family
3. Number of bicycles sold in a day
4. Number of students passing a stats exam

Continuous Data: The data which assumes all the values in the range is called continuous data.

Characteristics of continuous data are:

- Continuous data applies only to numeric data
- It has an infinite number of choice
- This data cannot be counted
- It has no space over a range of values

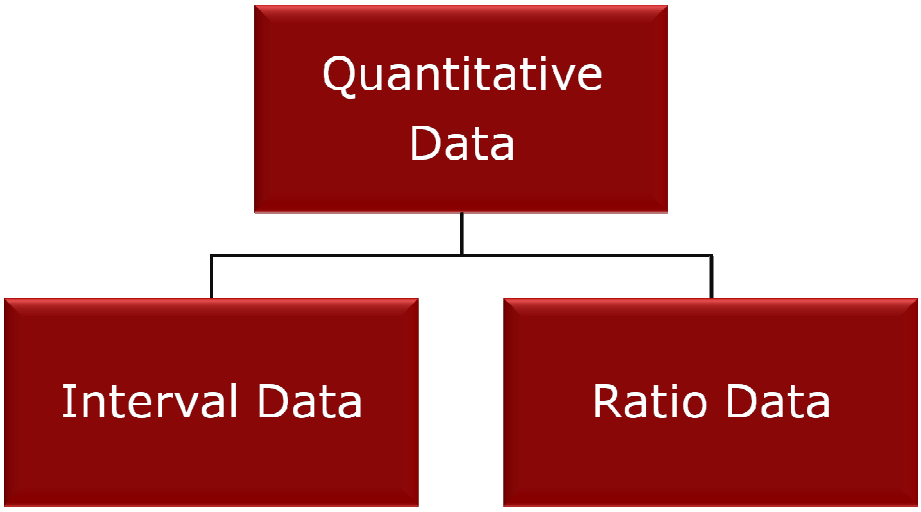
Example:

1. Weight of eight new born babies obtained from the records of hospital.
2. Quantity of milk that a cow produces in 24 hours.
3. The height of a basket ball player.
4. The current temperature outside.

# 4. Further Classification of Quantitative Data

Quantitative data may be further refined into two distinct groups:

- Interval Data
- Ratio Data




Quantitative Interval Data has following characteristics:

- It is ranked
- It has a precise difference between units of measure
- It defines values in such a way that the interval between two data values is meaningful
- It does not have a meaningful zero

Example:

1. Temperature comparison in degree Fahrenheit.
2. Level of happiness rated from 1 to 10.

Quantitative Ratio Data has following characteristics:

- This data possesses the characteristics of the interval measure, except that a meaningful zero exists
- True ratios exist when the same information is collected from two or more subjects
- The ratio of the data values is meaningful

Example:

- Physical body measurements
    - Height
    - Weight
- Money or monetary value

# 5. Qualitative Data

Qualitative Data (Categorical): Data that can be placed into distinct partitions or categories according to some defining characteristic or attribute is called qualitative data.

Example: In this table the class interval shows the different categories like rich, middle and poor to some attributes like boys and girls.

**Figure 4**

| Class | Boys | Girls | Total |
|---|---|---|---|
| Rich | 46 | 89 | 135 |
| Middle | 47 | 35 | 82 |
| Poor | 20 | 10 | 30 |
| **Total** | **113** | **134** | **247** |

Another example we can study is of Top Television viewing choices among 100 males and 100 females surveyed. Here the quality refers to the preference of a type of program over another type of program. Each male and female had only 1 choice to make out of the 3 types of programs given to them .Hence in qualitative terms we can see that males prefer sports and females prefer serials and as a whole among the 200 people surveyed games shows are the most preferred viewing choice among the 3 categories given.

**Figure 5**

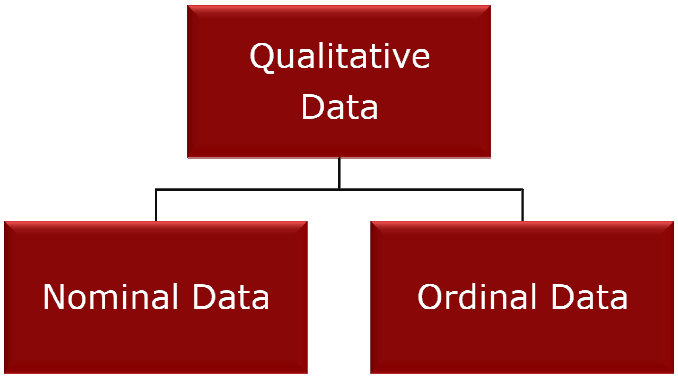
| Gender | Sports | Family Serials | Game Shows | Total |
|---|---|---|---|---|
| Men | 57 | 6 | 37 | 100 |
| Women | 10 | 58 | 32 | 100 |
| Total | 67 | 64 | 69 | 200 |

Colours like: White, Red, and Blue

Vegetables like: Carrots, Cucumbers and Tomatoes.

Qualitative Data may be further divided into two groups:

- Nominal Data
- Ordinal Data



Qualitative Nominal Data: Classifies data into mutually exclusive (overlapping), exhaustive categories in which no order or ranking (better or worse) can be imposed on the data.

Qualitative nominal data may be either numeric or non numeric.

In qualitative nominal numeric data:

- The number value indicates the category of a data element
- And the meaningful arithmetic cannot be performed upon the data

Example:

1. Area codes used with telephones:
   Area code 608 plus area code 715 is meaningless.
2. Course Numbers:

   Course number 84-240 minus 804-201 makes no numerical sense.

In Qualitative nominal non numeric data, the description of the data indicates the category of an element.

Example:

1. Sex: Male, Female.
2. Marital Status: Married, Single, Separated, Divorced, Widowed.
3. College Courses: Mathematics, English, History, Science.

Qualitative Ordinal Data: It classifies data into categories that can be ranked. This data also provides information about relative comparisons and has no precise differences between the ranks.

Qualitative Ordinal data may be either numeric or non numeric.

In qualitative ordinal numeric data, the value permits ranking or ordering of data by means of numbers.

# 6. Time Series Data

Time Series Data: Time Series data is a series of observations of a phenomenon recorded at successive points of time is called time series data. It is chronological arrangement of statistical data regarding the phenomenon.

Example: This table shows the recorded price of gold for the last few years.

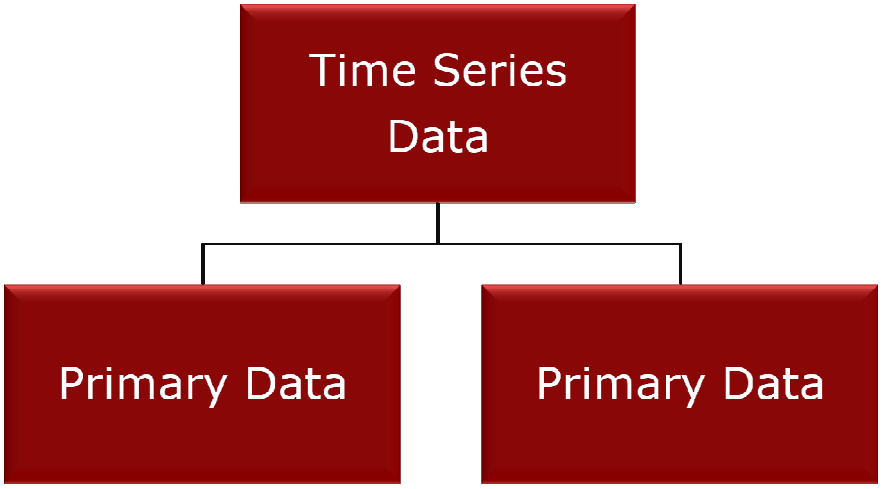
**Figure 7**

| Year | 1970 | 1980 | 1990 | 2000 | 2010 |
|---|---|---|---|---|---|
| Price per gram (Rs.) | 200 | 500 | 700 | 1000 | 2000 |

Time series data was famously used by former US presidential candidate and climate change activist Al Gore to show us how the Earth is getting affected by global warming. He made a presentation where he showed the trends in rising temperatures in various parts of the world in recent years.

Time series data is divided into two parts:

- Primary Data
- Secondary Data




Primary Data: Data that has been collected from first-hand-experience is known as primary data. Primary data has not been changed or altered by human beings; therefore its validity is greater than secondary data.

Example: The data collected from the Experiments and Surveys are the primary data. Like:

1. Mean Temperature of the earth in the year 2011.
2. Population of India in 2000.

Sources of Primary Data: Sources for primary data are limited and at times it becomes difficult to obtain data from primary source because of either scarcity of population or lack of cooperation. Regardless of any difficulty one can face in collecting primary data; it is the most authentic and reliable data source.

Secondary Data: Data collected from a source that has already been published in any form is called as secondary data. The review of literature in any research is based on secondary data. It is mostly collected from books, journals and periodicals.

Sources of Secondary Data:

Example:

1. Grade earned in school exam: A+, B+, C+, D+
2. Awards in a judged contest: First Place, Second Place, Third Place

In qualitative ordinal non numeric data, the data description permits ranking or ordering of data.

Example:

1. Household income: low, middle, high.
2. Size of an automobile: Sub-compact, Compact, Mid-size, Full-size, Limousine, Stretch limousine

Secondary data is often readily available. After the expense of electronic media and internet the availability of secondary data has become much easier.

Some of the Published Printed Sources are:

1. Books
2. Journals/periodicals:
3. Magazines/Newspapers:
4. Published Electronic Sources:
5. e-journals
6. General websites
7. Weblogs

Some of the Unpublished Personal Records are:

1. Diaries
2. Letters
3. Government Records

4. Census Data/population statistics
5. Health records
6. Educational institutes records
7. Public Sector Records:
8. NGO's survey data
9. Other private companies records