**Introduction to Statistics**

**Meaning of Statistics**

Everyday we come across a lot of information in the form of facts, numerical figures,

tables, graphs, etc. These are provided by newspapers, televisions, magazines and

other means of communication.

Examples

* cricket batting or bowling averages,
* profits of a company,
* temperatures of cities,
* polling results, and so on.

These facts or figures, which are numerical or otherwise, collected with a definite purpose are called *data*. Data is the plural form of the Latin word *datum*.

Plural: Data

Singular: Datum

Our world is becoming more and more information oriented. Every part of our

lives utilises data in one form or the other. So, it becomes essential for us to know how

to extract meaningful information from such data. This extraction of meaningful information is studied in a branch of mathematics called *Statistics*.

**Origin of Statistics**

The word ‘statistics’ appears to have been derived from the Latin word ‘status’

meaning ‘a (political) state’. In its origin, statistics was simply the collection of data on

different aspects of the life of people, useful to the State.

Over the period of time, however, its scope broadened and statistics began to concern itself not only with the collection and presentation of data but also with the interpretation and drawing of inferences from the data. Statistics deals with collection, organisation, analysis and interpretation of data.

**Meaning of “Statistics”**

The word ‘statistics’ has different meanings in different contexts.

Let us observe the following sentences:

1. May I have the latest copy of ‘Population Statistics of India’.

2. I like to study ‘Statistics’ because it is used in day-to-day life.

In the first sentence, statistics is used in a plural sense, meaning numerical data. These

may include the people of India, age, gender by states etc.

In the second sentence, the word ‘statistics’ is used as a singular noun, meaning the subject which deals with the collection, presentation, analysis of data as well as drawing of meaningful conclusions from the data.

**Main Topics**

We will look at three main topics here;

* **Collection of Data**
* **Presentation of Data**
* **Graphical Representation of Data**

**Collection of Data**

How is the data collected? What mechanisms exist?

* Websites
* Application forms
* Surveys

Legitimate or illegitimate (privacy/ethical concerns)

Sample heights of students

|  |  |  |
| --- | --- | --- |
| SI | Name | Height |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**What are the sources of Data? Classification of data sources**

**Primary Data**

When the information was collected by the investigator herself or himself with a definite objective in her or his mind, the data obtained is called *primary data*.

How is it collected? – Physical Observation, testing, questionnaires, surveys, personal interviews

**Secondary Data**

When the information was gathered from a source which already had the information stored, the data obtained is called *secondary data*.

Such data, which has been collected by someone else in another context, needs to be used with great care ensuring that the source is reliable.

Who uses secondary data?

**Presentation of Data**

*What are the main objective or principles while presenting the data?*

* Present them in a form which is meaningful
* Easily understood and
* Gives its main features (average, trend) at a glance.

Consider the marks obtained by 10 students in a mathematics test as

given below:

55 36 95 73 60 42 25 78 75 62

This is an instance of *raw data.* No real presentation here.

(eye-ball the data and answer the following questions)

* Highest
* Lowest
* Range

**Frequency table**

**1.**

**Construct a simple frequency table**

Simple form first.

Necessary when there is more data – eye-balling is not practical

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 10 | 20 | 36 | 92 | 95 | 40 | 50 | 56 | 60 | 70 |
| 92 | 88 | 80 | 70 | 72 | 70 | 36 | 40 | 36 | 40 |
| 92 | 40 | 50 | 50 | 56 | 60 | 70 | 60 | 60 | 88 |

To make the data readable, construct a frequency table.

**2.**

**Construct a Frequency table with class intervals**

Let us now consider the following frequency distribution table which gives the weights of 38 students of a class:

Frequency Table with Class Intervals

|  |  |
| --- | --- |
| **Weights (in kgs)** | **Nos** |
| 31-35 | 9 |
| 36-40 | 5 |
| 41-45 | 14 |
| 46-50 | 3 |
| 51-55 | 1 |
| 56-60 | 2 |
| 61-65 | 2 |
| 66-70 | 1 |
| 71-75 | 1 |
| Total | 38 |

New data comes in

* 35.5 kg
* 40.5 kg

Convert open-ended class intervals to close-ended class intervals

|  |  |
| --- | --- |
| **Weights (in kgs)** | **Nos** |
| 30.5-35.5 | 9 |
| 35.5-40.5 | 5+1 |
| 40.5-45.5 | 14+1 |
| 45.5-50.5 | 3 |
| 50.5-55.5 | 1 |
| 55.5-60.5 | 2 |
| 60.5-65.5 | 2 |
| 65.5-70.5 | 1 |
| 70.5-75.7 | 1 |
| Total | 38 |

By convention, we consider 35.5 in the class 35.5 - 40.5 and not in 30.5 - 35.5. Similarly, 40.5 is considered in 40.5 - 45.5 and not in 35.5 - 40.5.

**Graphical Representation of Data**

* **Bar graph**
* **Histogram**
* **Frequency polygon**

**Bar Graph**

Background

Have you seen graphs in the newspapers, television, magazines, books etc.?

Discussion

What is the purpose?

Why graphs?

The purpose of the graph is to show numerical facts in visual form so that they can be understood quickly, easily and clearly. Thus graphs are visual representations of data collected.

Data can also be presented in the form of a table; however a graphical presentation is easier to understand. This is true in particular when there is a trend or comparison to be shown.

Bar Graph

A bar graph is a pictorial representation of data in which usually bars of uniform width are drawn with equal spacing between them on one axis (say, the *x*-axis), depicting the variable.

The values of the variable are shown on the other axis (say, the *y*-axis) and the heights of

the bars depend on the values of the variable.

**Sample data for bar graph**

In a particular section of Class IX, 40 students were asked about the months of their birth and the following graph was prepared for the data so obtained:

|  |  |
| --- | --- |
| **Months** | **Number of Students** |
| Jan | 3 |
| Feb | 4 |
| Mar | 2 |
| Apr | 2 |
| May | 5 |
| June | 1 |
| July | 2 |
| Aug | 6 |
| Sep | 3 |
| Oct | 4 |
| Nov | 4 |
| Dec | 4 |

*Draw the bar graph*

**Questions**

* How many students were born in the month of November?
* In which month were the maximum numbers of students born?

**Draw a bar graph for the following**

A family with a monthly income of Rs. 20,000 had planned the following expenditures per month under various heads:

|  |  |
| --- | --- |
| **Heads** | **Expenditure**  **(Rupees Thousands)** |
| Grocery | 4 |
| Rent | 5 |
| Education of children | 5 |
| Medicine | 2 |
| Fuel | 2 |
| Entertainment | 1 |
| Miscellaneous | 1 |

*Draw the bar graph*