

STATISTICS

TERMS RELATED TO STATISTICS

Meaning

Statistics is the branch of science which deal's with the collection or arrangement of data, classification or tabulation of data, analysis of data and drawing a conclusion from the analysis of the required objectives.

Statistical Data

Examples of Types of Data

-
- **Primary Quantitative Data:**
 - Questionnaires
 - Structured Interviews
 - **Primary Qualitative Data:**
 - Participant Observation
 - Unstructured Interviews
 - **Secondary Quantitative Data:**
 - Official statistics
 - **Secondary Qualitative Data:**
 - Letters, articles, newspapers

Statistical Data refers to the set of numerical facts collected for the purpose of an investigation is called statistical data. There are two types of data:

1. Primary data

The data which are naturally obtained by the investigator himself for the first time for his own use is called primary data. Primary data are also called First-handed data.

2. Secondary data

The data which was collected by someone from the help of primary data is called secondary data, is artificial in nature. They are also called as Second-handed data in nature. The secondary data can be expressed in two types. They are:

a) Raw or Ungrouped data:

Raw or Ungrouped data are those types of data which are obtained in original form. For example, look the following means of 10 students in class test in Optional Maths.

20	65	92	86	35
57	78	83	40	50

b) Array:

Simply, array refers to the arrangement of data in ascending or descending of data order. Generally, Arrays are the data which are put in the form of a table which is also called the presentation of data. If the number of times an observation occurs, then it is called frequency of distribution.

For example,

20, 15, 30, 10, 25, 30, 35

Now,

Arranging this data in ascending order,

10, 15, 20, 25, 30, 30, 35

Considering it in tabular form,

Variable	10	15	20	25	30	35
Frequency	2	5	8	11	3	7

Frequency Distribution

A tabular arrangement of data showing the frequency of each observation is called a frequency distribution. If the table shows the data with their corresponding frequency is called Frequency Distribution.

The frequency distribution may be divided into following types:

a) Individual Series:

If the items are listed individually then it is called individual series. In individual series, there is no frequency.

For example,

Students	2	4	6	8	10
Marks	3	5	7	9	11

b) Discrete Series:

If the series are formed by a discrete variable then it is called as a discrete series. In this series, the variable is taken the only exact measurement.

For example,

Marks	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
No. of students	4	8	12	16	20

c) Continuous Series:

If the variable is represented in a continuous way then it is called continuous series. In this series, variable lies in two groups which are called class interval.

For example,

Marks	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
No. of students	4	8	10	14	16

Ac
Go

Cumulative Frequency

The sum of all the frequencies of all the previous class and given class is known as cumulative frequency. If the class interval is shown together with cumulative frequencies in the table then it is called cumulative frequency table.

Measures of the Central Tendency

Mean, Median, Quartiles and Mode are the measures of central tendency.

MEAN

Arithmetic Means

$$\text{arithmetic mean} = \frac{\sum_{n=1}^k x_n}{k}$$

Arithmetic Mean

If the total sum observation is divided by a total number of observations, then it is called arithmetic mean. It is denoted by \bar{X} (Read as X-bar)

$$\therefore \text{Arithmetic Mean} = \frac{\text{Total sum of observation}}{\text{Total no. of observation}}$$

For example,

Arithmetic mean of 1, 3, 7, 11, & 13

$$\begin{aligned} &= \frac{1+3+7+11+13}{5} \\ &= \frac{35}{5} \\ &= 7 \end{aligned}$$

1. Calculation of Mean for individual series

The mean of individual series is calculated by adding all the observation and dividing the sum by the total number of observation.

If $x_1, x_2, x_3, \dots, x_n$ are be n variants value of variable a. Then arithmetic mean is denoted by \bar{X}

$$\text{So, } \bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum X}{n}$$

Where $\sum X$ = sum of n observation or items

n = no. of observations or items

X = variable

2. Calculation of Mean for discrete series

Mean for discrete series can be calculated by

$$\begin{aligned} \bar{X} &= \frac{\text{sum of the product of } f \text{ and } x}{\text{sum of } f} \\ &= \frac{\sum fx}{N} \end{aligned}$$

3. Calculation of Mean for continuous series

For calculating mean in continuous series the following formulae is used:

$$\bar{X} = A + \frac{\sum fx}{N} \times i$$

where, F = Frequency and A = mid-value

MEDIAN

The median is a variate value which divides the given data into two equal parts. If the data are arranged either in ascending or descending order, the middlemost number is median.

For example,

1, 3, 8, 10, 13, 16, 20

The middle number 10 is median.



CALCULATION OF MEDIAN

Individual Series

M = Size of the $\left(\frac{N+1}{2}\right)^{\text{th}}$ item

Discrete Series

M = Size of the $\left(\frac{N}{2}\right)^{\text{th}}$ item

Continuous Series

M = Size of the $\left(\frac{N}{2}\right)^{\text{th}}$ item

$$M = L_1 + \frac{\frac{N}{2} - cf}{f} \times C$$

Where,

- o M = Median
- o N = Number of Observations
- o L = Lower limit of median class
- o cf = Cumulative frequency of the class preceding median class
- o f = frequency of median class observation
- o i = Class interval

Calculation of median for individual series

The formula for calculating median for individual series is given by;

Median (Md) = Position of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item

where, N = No. of observation or items.

The steps for calculation of median for individual series:

1. Arrange the data in ascending or descending order.
2. Count the number of terms (n).

Calculation of median for a discrete data

Let, $x_1, x_2, x_3, \dots, x_n$ be the variable values and $f_1, f_2, f_3, \dots, f_n$ be their corresponding frequency respectively.

Then, Median (Md) = Position of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item

where, N = total sum of a frequency

Calculation of median for a continuous data

To calculate the median for a continuous data, the following steps are applied:

1. Prepare cumulative frequency table
2. Find the value of $\left(\frac{N}{2}\right)^{\text{th}}$ items which gives the median class.
3. Use the formula, Median = $L + \frac{\frac{N}{2} - c.f}{f} \times i$

where,

L = lower limit of median class

c.f = cumulative frequency

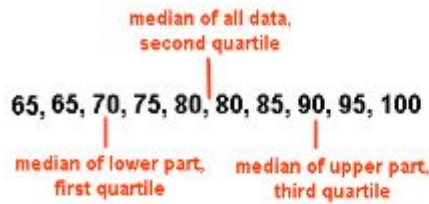
f = frequency

i = class interval

QUARTILE

A quartile is the variate value which divides the value of given frequency into four equal parts. There are three quartiles, they are: -

1. Lower Quartile (Q_1)
2. Middle Quartile (Q_2)
3. Upper Quartile (Q_3)



Quartiles can be calculated as following ways: -

Individual series

After arranging the data in ascending order, the quartiles are calculated by applying the following formula:

Lower quartile (Q_1) = value of $(\frac{N+1}{4})^{\text{th}}$ item.

Upper Quartile (Q_3) = value of $3(\frac{N+1}{4})^{\text{th}}$ item.

Where, N = Total sum of frequency.

Discrete Series

When the data are arranged in ascending order, the quartile is calculated by the following formula:

Lower quartile (Q_1) = value of $(\frac{N+1}{4})^{\text{th}}$ item

Upper quartile (Q_3) = value of $3(\frac{N+1}{4})^{\text{th}}$ item

Where N = Total sum of frequency

Continuous Series

For continuous series, the following formulae are used.

Lower quartile class = value of $3(\frac{N}{4})^{\text{th}}$ item

$$\text{Exact Lower Quartile } (Q_1) = L + \frac{\frac{N}{4} - C.F}{f} \times i$$

where,

L = Lower limit of corresponding class

f = frequency

c.f = cumulative frequency

i = class size

N = Total sum of frequency

Upper quartile class = value of $3(\frac{N}{4})^{\text{th}}$ item

$$\text{Exact upper quartile } (Q_3) = L + \frac{\frac{3N}{4} - C.F}{f} \times i$$

MODE

Mode is a number which occur or appears most often or frequently in a set of numbers. It is that value which has the highest frequency. It is denoted by M_0 .

Calculation of Mode in:

Individual series

In an individual series, mode can be found in by inspection. If the value occurring maximum number of times, then that is a modal value.

For example, 1,2,14,3,3,5,2,3

Variable	No. of occurence
1	2
2	2
3	3
4	1
5	1

In the above table, a variable value 3 repeats maximum number of times i.e. 3
∴ Mode (M_0) = 3

Discrete Series

In case of discrete series also, mode can be obtained by inspection only. Mode occurs with maximum frequency . For example,

No. of student	4	5	6	7	8
Marks	10	20	30	15	25

Here, maximum frequency os 30. So, corresponding modal size is 6.