**STATISTICS**

**Basic concepts:**

* Statistics is a branch of Mathematics which deals with collection of data, classification of the data and interruption.
* The word ‘**Statistic**’ is derived from the Latin word “Status”, the Italian word “Stata” and the Greek word “Statisic” which means ‘A political state’.
* Father of Statistic is ‘**Sir Ronald A Fisher’**.
* The information which is collected is called ‘*Data*’.
* The objects in the data are called ‘*Observations*’.
* No. of time the observation occur is called “*Frequency*” of the observation.
* Generally, in Statistics the data may be numerical or graphical or pictorial.
* When the data is collected by investigator himself with a definite objective, then the data obtained is called “*Primary data*”.
* When the data is gathered from a source which already had the information stored, then that data obtained is called “*Secondary data*”.
* There are different types to present a data. Some of them are a) Raw data

b) Ungrouped frequency distribution

c) Grouped frequency distribution

d) Pictograms

e) Histograms

f) Pie charts

g) Frequency polygons

h) Frequency curve

i) Cumulative frequency curves

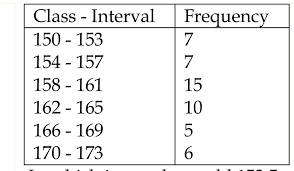
* If we present the data without any changes, then it is called ‘*Raw data’*.

ex: 12, 24, 15, 45, 25, 34, 42, 19.

* If we present the data with classification, then it is called ‘Classified data’. The classified data will be present in two type of tables. They are ungrouped frequency distribution and grouped frequency distribution.
* The table having observations and frequency is called ‘*Ungrouped frequency distribution’*.



* The table having class intervals and frequency is called ‘Grouped frequency distribution’.



* We have two types of class intervals. They are Inclusive classes and Exclusive classes.
* The classes like 0 – 10, 10 – 20, 20 – 30,……. are called ‘*Exclusive classes’*.
* The classes like 0 - 9, 10 – 19, 20 – 29,….. are called ‘*Inclusive classes’*.
* Any class is represented by two numbers. The least number is called ‘*Lower Limit*’ of the class.
* The highest number is called ‘*Upper Limit*’ of the class.

ex: For the class 10 – 19

Lower limit = 10

Upper limit = 19

* In exclusive classes, the lower limit belongs to the class and the upper limit does not belong to the class.
* In inclusive classes, both lower and upper limits belong to the class.
* The average of lower limit of a class and upper limit of its previous class is called ‘*Lower Boundary*’ of the class.
* The average of upper limit of a class and lower limit of its next class is called ‘*Upper Boundary*’ of the class.

ex: given classes are 0-9, 10-19, 20-29,30-39,……

for the class 20-29

Lower boundary = = = 19.5

Upper boundary = = = 29.5

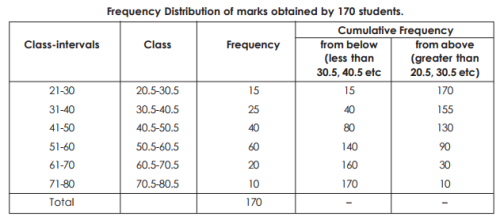
* The lower boundary of a class is equal to the upper boundary of its previous class.
* The upper boundary of a class is equal to the lower boundary of its next class.
* For exclusive classes, the limits and boundaries are same
* The average of boundaries of a class is called ‘*Class mark*’ or ‘*Mid value*’ of the class. It is denoted by ‘x’.

Class mark (x) =

* The difference of boundaries of a class is called ‘*Class size*’ or ‘*Length of the class’*. It is denoted by ‘h’.

Class size (h) = upper boundary – lower boundary

* Class size is also equal to difference of limits.
* Sum of the frequencies from the beginning to the upper boundary of a class is called ‘Less than cumulative frequency’ of the class.
* Sum of the frequencies from the ending to the lower limit of a class is called ‘*Greater than cumulative frequency*’ of the class.



* In raw data, the difference of the highest value and lowest value is called ‘*Range of the data’*.

Range = Highest value – lowest value’

* Range of first ‘n’ natural numbers is n – 1.

**Central tendency values:**

* For interruption of a data, We use a numerical value which is called ‘Central tendency value’.
* We have three types of central tendency values. They are

1. Arithmetic Mean

2. Median

3. Mode

**Arithmetic Mean:**

* This is the most common and widely used measure of central tendency.
* It depends on all observations.
* It is affected by extreme values.

**Arithmetic Mean of Raw data:**

* The ratio of sum of observations and number of observations is called ‘Arithmetic Mean’ of the data.
* We can find A.M of a raw data in two methods. They are

1. Direct method

2. Deviation method

* Direct method

A.M = =

* Deviation method

A.M = A

here A – Assumed mean

x – observation

n – number of observations

**Arithmetic Mean of ungrouped frequency distribution:**

* We can find A.M of ungrouped frequency distribution in two methods. They are

1. Direct method

2. Deviation method

* Direct method

A.M =

here f – frequency of the observation

x – observation

* Deviation method

A.M = A +

here A – Assumed mean

f – frequency of the observation

d – deviation

d = x – A

x – observation

**Arithmetic Mean of grouped frequency distribution:**

* We can find A.M of grouped frequency distribution in three methods. They are

1. Direct method

2. Deviation method (Assumed mean method)

3. Step deviation method

* Direct method

A.M =

here f – frequency of the class

x – class mark of the class

* Deviation method

A.M = A +

here A – Assumed mean

f – frequency of the class

d – deviation of the class

d = x – A

x – class mark of the class.

* Step deviation method

A.M = A + × h

here A – Assumed mean

f – frequency of the class

u = =

d – deviation of the class

x – class mark of the class

h – class size

**Median:**

* Median is most stable measure of central tendency values.
* Median is not affected by extreme values.
* Median is not depended on all observations.
* If we change the least or highest value of a data, then there is no change in the median value.

**Median of raw data:**

* The middle most observation of a data is called ‘Median’ of the data.
* To find the median of the data, first we have to write the observations in ascending order or descending order.
* If the number of observations (n) is odd number, then th observation is median of the data.
* If the number of observations (n) is even number, then the average of th observation and th observation is median of the data.

**Median of ungrouped frequency distribution:**

* To find the median of ungrouped frequency distribution, we have to follow these steps:

Step 1: Find the L.C.F of the observations.

Step 2: Calculate sum of the frequencies and call it as ‘n’.

Step 3: Find the value of .

Step 4: Take the observation as median of the data whose

L.C.F is greater and near to the value.

**Median of grouped frequency distribution:**

* To find median of grouped data, first select a class whose L.C.F is greater and near to value as ‘*median class’*.
* We can find the median of grouped frequency distribution by using the following formula.

Median = L + × h

here L – Lower boundary of median class

n – sum of frequencies

cf – Less than cumulative frequency of the

class preceding the median class

f – frequency of the median class.

h – class size.

**Mode:**

* The most frequently occurring observation is called ‘Mode’ of the data.
* It depends neither on number of observations nor values of all observations.
* It is used to analyze both numerical and verbal data.
* Some data have no mode.

Ex: Mode of first 10 natural numbers.

* There may be 2 or more modes for some data.
* The data having one mode is called ‘Unimodal data’.
* The data having two modes is called ‘Bimodal data’.

**Mode of ungrouped frequency distribution:**

* The highest frequency observation of ungrouped data is taken as mode of the data.

|  |  |
| --- | --- |
| Observation | Frequency |
| 10 | 3 |
| 12 | 7 |
| 14 | 12 |
| 16 | 19 |
| 18 | 8 |
| 20 | 4 |

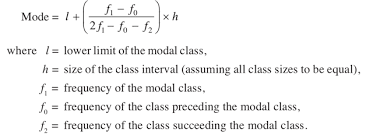
Ex:

Highest frequency = 19

Mode of the data = 16

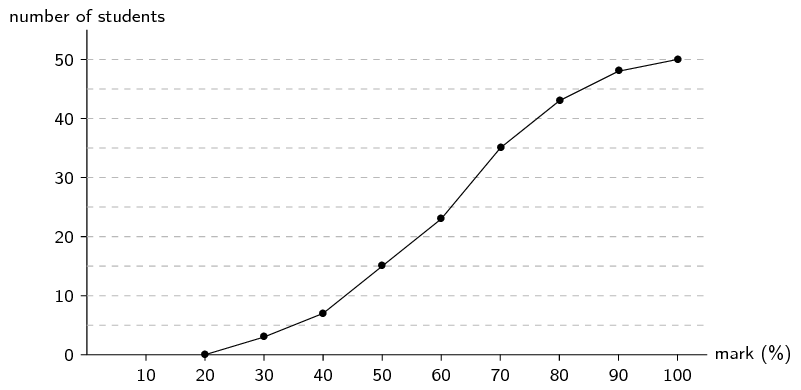
**Mode of grouped frequency distribution:**

* To find mode of grouped frequency distribution, we have to select the class having highest frequency as ‘*Modal class’*.
* Now by using the following formula, we can find mode of the data.

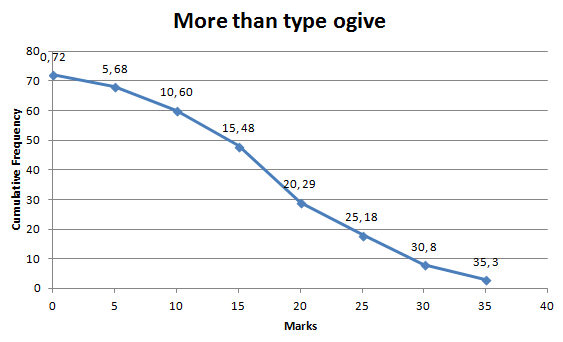


**Graphical representation of Cumulative Frequency Distribution:**

* The curves drawn for less than or greater than cumulative frequencies and the boundaries of frequency distribution are called ‘*Cumulative frequency curves*’ or ‘*Ogive*’ curves.
* The curve drawn with upper boundaries (on X – axis) and less than cumulative frequency (on Y – axis) is called ‘*Less than cumulative curve’ or ‘Less than Ogive’*.

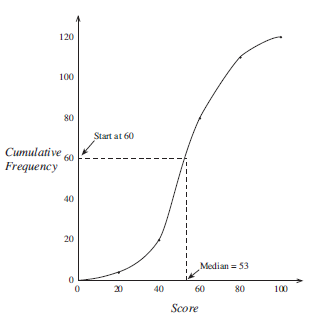


* The curve drawn with lower boundaries (on X-axis) and greater than cumulative frequency (on Y – axis) is called ‘*Greater than cumulative curve’ or ‘Greater than Ogive*’.



**Obtaining Median from the Ogive curves:**

* In less than Ogive, locate the value of on Y – axis. From this point, draw a line parallel to X – axis cutting the curve at a point. From this point, draw a perpendicular line to the X – axis. The foot of the perpendicular determines the MEDIAN of the data.



* Draw both ogives on the same axes. ‘The two Ogives will intersect each other at a point. From the point, if we draw a perpendicular on the X – axis, the point at which it cuts the X – axis gives us the MEDIAN.

