# Statistics
# Tabular Representation of Statistical Data

**(1) The word 'data' means information. Its exact dictionary meaning is" given facts".**

**(2) Statistical data are of two types (i) primary (ii) secondary.**

**(i) Primary:**
**Example:** Population census reports are primary data because these are collected, compiled and published by the population census organization.

**(ii) Secondary:**
**Example:** Economics survey of England is secondary data because these are collected, by more than one organization like Bureau of statistics, Board of Revenue, the Bank etc…

**(3) The number of times an observation occurs in the given data is called the frequency of the observation.**
***For Example:*** If 4 students get 68 marks, then we say that the frequency of 68 is 4.
If this is done for all the observations and the data is rearranged from the lowest to the highest value; frequency distribution is obtained.

**(4) There are two types of frequency distribution:**
**(i) Discrete frequency distribution:**
***For Example:*** The marks scored by 40 students of class IX in mathematics are given below:
81,55,68,79,85,43,29,68,54,73,47,35,72,64,95,44,50,77,64,35,79,52,45,54,70,83,62,64,72,92,84,76,63,43,54,38,73,68,52,54.
Prepare a frequency distribution with class size of 10 marks.
**Solution:**
Here,
Minimum mark = 29
Maximum mark = 95
Now,
Range = 95-29= 66
It is given that the class of size is 10
Then, no. of classes = Range/ Class Size = 66/10 = 6.6
Frequency distribution of marks-

| Class Interval | Tally marks | Frequency |
|---|---|---|
| 29 - 39 | \|\|\|\| | 4 |
| 39 - 49 | 𝍸 | 5 |
| 49 - 59 | 𝍸 \|\|\| | 8 |
| 59 - 69 | 𝍸 \|\|\| | 8 |
| 69 - 79 | 𝍸 \|\| | 7 |
| 79 - 89 | 𝍸 \| | 6 |
| 89 - 99 | \|\| | 2 |
| Total | | 40 |

**(ii) Continuous frequency distribution:**

***For Example:*** Following are the ages of 360 patients getting medical treatment in a hospital on a day:

| Age (in years): | 10 - 20 | 20 - 30 | 30 - 40 | 40 - 50 | 50 - 60 | 60 - 70 |
|---|---|---|---|---|---|---|
| No. of Patients: | 90 | 50 | 60 | 80 | 50 | 30 |

**Solution:** Cumulative frequency distribution:

| Age | Number of Patients | Cumulative frequency |
|---|---|---|
| Less than 20 | 90 | 90 |
| Less than 30 | 50 | 140 |
| Less than 40 | 60 | 200 |
| Less than 50 | 80 | 280 |
| Less than 60 | 50 | 330 |
| Less than 70 | 30 | 360 |
| Total | 360 | |

**(5) In a discrete frequency distribution the cumulative frequency of a particular value of the variable is the total of all the frequencies of the values of the variable which are less than or equal to the particular value.**

**Solution:**

**Frequency Distribution:** Construction of this from the given raw data is done by the use of the method of tally marks. In the first column of the frequency table, we write all the possible values of the variable from the lowest to the highest.

**Cumulative Frequency**: It is corresponding to a class is the sum of all frequencies up to end including that class. A table which shows the commutative frequencies over various classes is called a cumulative frequency distribution table.

| Marks | Tally marks | freq | cumulative freq. |
|---|---|---|---|
| 0 - 5 | \|\|\|\| | 4 | 4 |
| 5 - 10 | 卌 | 5 | 9 |
| 10 - 15 | 卌 卌 \| | 11 | 20 |
| Total | | 20 | |

**(6) A table which displays the manner is which cumulative frequency are distributed over various classes is called a cumulative frequency distribution or cumulative frequency table.**

***For Example:*** The marks scored by 55 students in a test are given below:

| Marks: | 0 - 5 | 5 - 10 | 10 - 15 | 15 - 20 | 20 - 25 | 25 - 30 | 30 - 35 |
|---|---|---|---|---|---|---|---|
| No. of students | 2 | 6 | 13 | 17 | 11 | 4 | 2 |

Prepare a cumulative frequency table.

**Solution:** Cumulative frequency table

| Marks | No of Students | Cumulative frequency |
|---|---|---|
| Less than 5 | 2 | 2 |
| Less than 10 | 6 | 8 |
| Less than 15 | 13 | 21 |
| Less than 20 | 17 | 38 |
| Less than 25 | 11 | 49 |
| Less than 30 | 4 | 53 |
| Less than 35 | 2 | 55 |
| Total | 55 | |

# Statistics
# Graphical Representation of Statistical Data

**(1) Following are the methods of graphical representation of data:**

**(i) Bar graphs:**

***For Example:*** The population of Delhi state in different census years is as given below:

| Census year | 1961 | 1971 | 1981 | 1991 | 2001 |
|---|---|---|---|---|---|
| Population in Lakhs | 30 | 55 | 70 | 110 | 150 |

Represent the above information with the help of a bar graph.

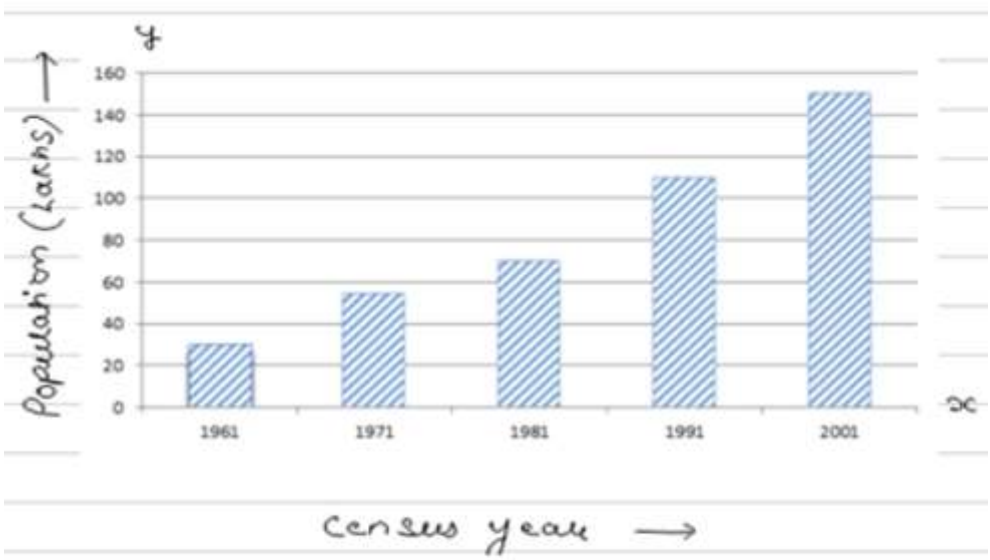
**Solution:** To represent the given data by a pictograph follow the below steps:

Step 1: Draw a horizontal and vertical line.

Step 2: Mark 5 point on horizontal line at equal distance for census year.

Step 3: Erect rectangles of the same width at marked points.

Step 4: The height of the rectangles are proportional to the numerical values of the given population.




**(ii) Histogram:**

***For Example:*** Construct a histogram for the following data:

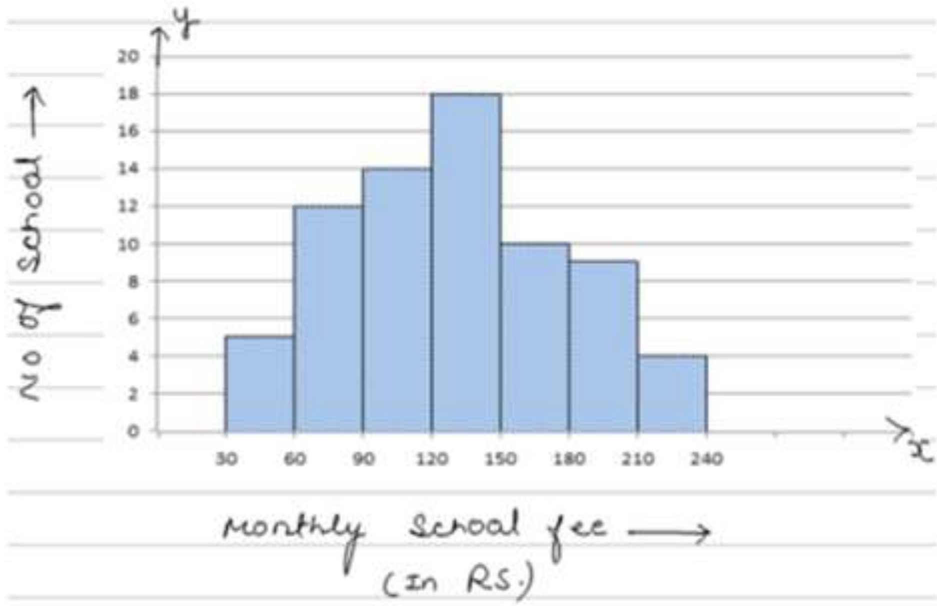
| Monthly School fee (in Rs): | 30-60 | 60-90 | 90-120 | 120-150 | 150-180 | 180-210 | 210-240 |
|---|---|---|---|---|---|---|---|
| No of Schools | 5 | 12 | 14 | 18 | 10 | 9 | 4 |

**Solution:** To represent the given data by a histogram follow the below steps:

Step 1: Draw a horizontal and vertical line.

Step 2: Taking class interval on horizontal line and corresponding frequencies on vertical line.

Step 3: Construct rectangles to obtain the histogram of given frequency distribution as shown in fig




### (iii) Frequency polygon:

***For Example:*** The following table gives the distribution of IQ's (intelligence quotients) of 60 pupils of class V in a school:

| IQ's: | 125.5 to 13. 25 | 118.5 to 125.5 | 111.5 to 118.5 | 104.5 to 111.5 | 97.5 to 104.5 | 90.5 to 97.5 | 83.5 to 90.5 | 76.5 to 83.5 | 69.5 to 76.5 | 62.5 to 69.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of pupils | 1 | 3 | 4 | 6 | 10 | 12 | 15 | 5 | 3 | 1 |

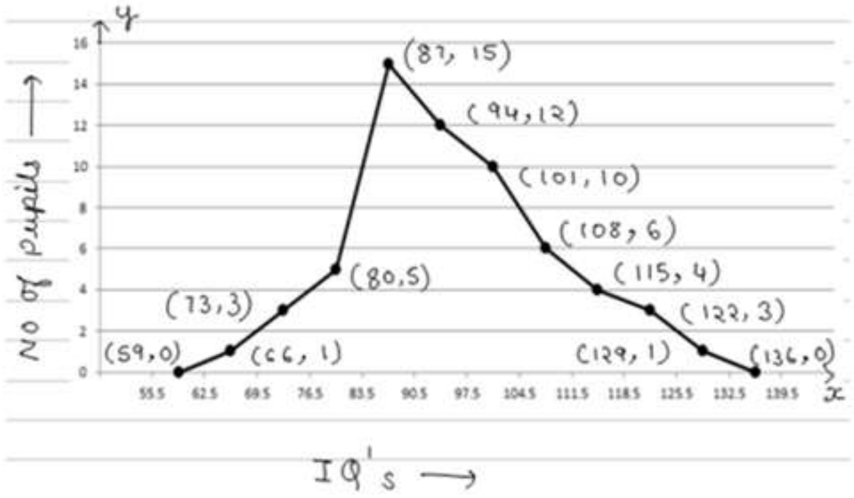
Draw a frequency polygon for the above data.

**Solution:** To represent the given data by frequency polygon follow the below steps:
Step 1: We need to find class mark.
We know, Class Mark = (Upper limit + Lower limit)/2.

| Expenditure | Class Mark | No. of workers |
|---|---|---|
| 62.5 – 69.5 | 66 | 1 |
| 69.5 – 76.5 | 73 | 3 |
| 76.5 – 83.5 | 80 | 5 |
| 83.5 – 90.5 | 87 | 15 |
| 90.5 – 97.5 | 94 | 12 |
| 97.5 – 104.5 | 101 | 10 |
| 104.5 – 111.5 | 108 | 6 |
| 111.5 – 118.5 | 115 | 4 |
| 118.5 – 125.5 | 122 | 3 |
| 125.5 – 132.5 | 129 | 1 |

Step 2: Draw a horizontal and vertical line.
Step 3: Taking class interval on horizontal line and corresponding frequencies on vertical line.

Step 4: Plot (66, 1), (73, 3), (80, 5), (87, 15), (94, 12), (101, 10), (108, 6), (115, 4), (122, 3), (129, 1)
Step 5: Join the plotted points.
Step 6: The end point (66, 1) and (129, 1) are joined to the midpoint (59, 0) and (136, 0) respectively of imagined class intervals to obtain the frequency polygon.



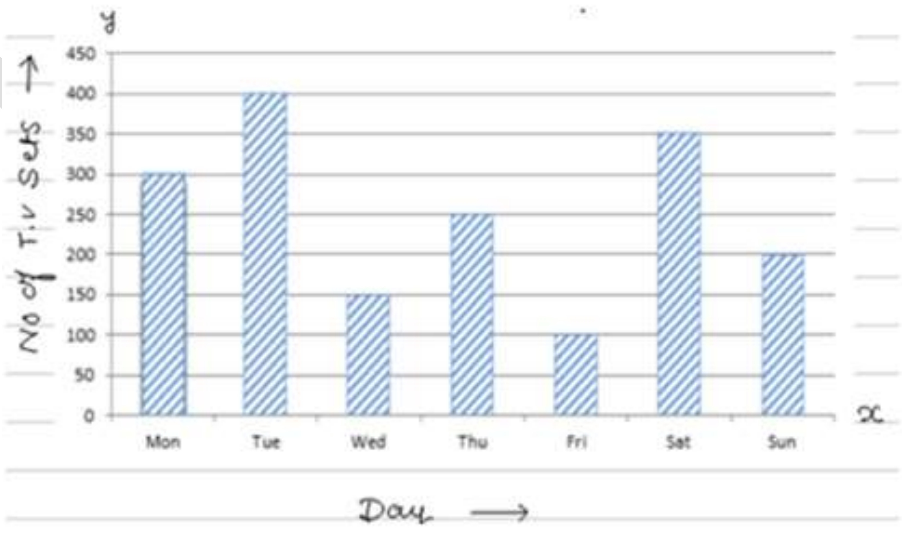

**(2) A bar graph is a pictorial representation of the numerical data by a number of bars (rectangles) of uniform width erected horizontally or vertically with equal spacing between them. Each rectangle or bar represents only one value of the numerical data and so there are as many bars as the number of values in the numerical data. The height or length of a bar indicates on a suitable scale the corresponding value of the numerical data.**

***For Example:*** The following table shows the daily production of T.V. sets in an industry for 7 days of a week:

| Day | Mon | Tue | Wed | Thurs | Fri | Sat | Sun |
|---|---|---|---|---|---|---|---|
| No. of T.V. sets | 300 | 400 | 150 | 250 | 100 | 350 | 200 |

Represent the above information by a pictograph.

**Solution:** To represent the given data by a pictograph follow the below steps:
Step 1: Draw a horizontal and vertical line.
Step 2: Mark 7 days on horizontal line at equal distance.
Step 3: Erect rectangles of the same width at marked points.
Step 4: The height of the rectangles are proportional to the numerical values of the given data of no of TV sets.

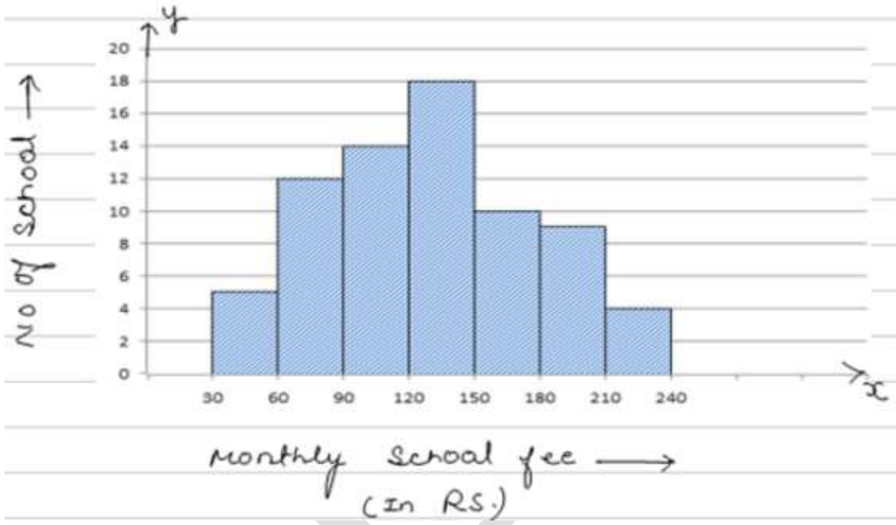


**(3) A histogram or frequency histogram is a graphical representation of a frequency distribution in the form of rectangles with class intervals as bases and heights proportional to corresponding frequencies such that there is no gap between any two successive rectangles.**

***For Example:*** Construct a histogram for the following data:

| Monthly School fee (in Rs): | 30-60 | 60-90 | 90-120 | 120-150 | 150-180 | 180-210 | 210-240 |
|---|---|---|---|---|---|---|---|
| No of Schools | 5 | 12 | 14 | 18 | 10 | 9 | 4 |

**Solution:** To represent the given data by a histogram follow the below steps:
Step 1: Draw a horizontal and vertical line.
Step 2: Taking class interval on horizontal line and corresponding frequencies on vertical line.
Step 3: Construct rectangles to obtain the histogram of given frequency distribution as shown in fig.

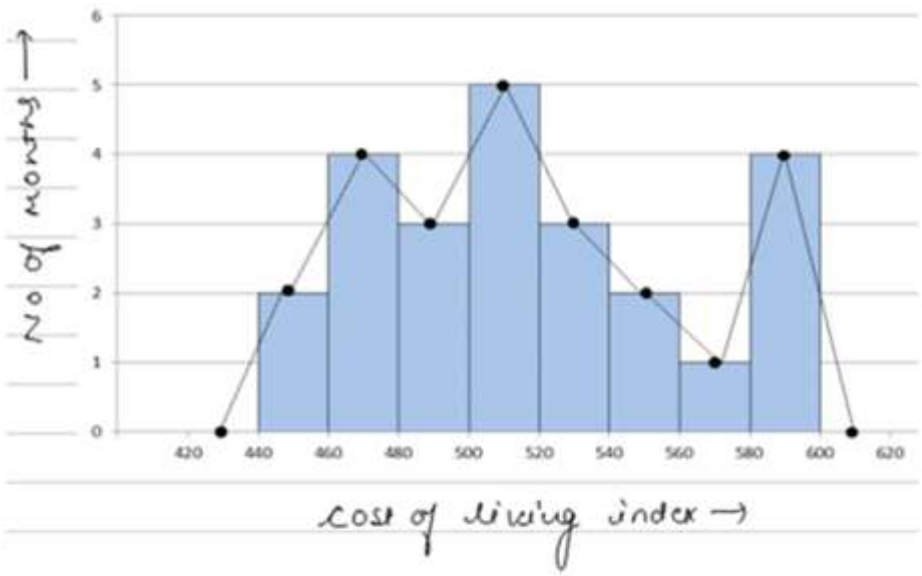



**(4) A frequency polygon of a given frequency distribution is another method of representing frequency distribution graphically.**

***For Example:*** Draw, in the same diagram, a histogram and a frequency polygon to represent the following data which shows the monthly cost of living index of a city in a period of 2 years:

| Cost of living Index: | 440-460 | 460-480 | 480-500 | 500-520 | 520-540 | 540-560 | 560-580 | 580-600 |
|---|---|---|---|---|---|---|---|---|
| No. of months | 2 | 4 | 3 | 5 | 3 | 2 | 1 | 4 |

**Solution:** To represent the given data by a histogram follow the below steps:
Step 1: Draw a horizontal and vertical line.
Step 2: Taking class interval on horizontal line and corresponding frequencies on vertical line.
Step 3: Construct rectangles to obtain the histogram of given frequency distribution as shown in fig.
Step 4: Obtain the midpoints of the upper horizontal side of each rectangle.
Step 5: Join these midpoints of the adjacent rectangles of the histogram by line segment.
Step 6: Obtain the midpoints of two class intervals of zero frequency i.e. on X axis, one adjacent to the first, on its left and one adjacent to the last, on its right.
These class intervals are known as imagined class intervals.
Step 7: Complete the polygon by joining the midpoints of first and last intervals to the midpoints of imagined class interval adjacent to them.



No of months →
0
1
2
3
4
5
6
420
440
460
480
500
520
540
560
580
600
620
cost of living index →

# Statistics – Measures of Central Tendency

**(1) Arithmetic mean (AM), Geometric mean (GM), Harmonic mean(HM), Median and Mode are various measures of central tendency.**

**(i) Arithmetic mean:**

***For Example:*** Find the mean of 994, 996, 998, 1002 and 1000.

**Given,** No. of values n = 5

We know, mean $(\bar{x}) = \dfrac{x_1 + x_2 + \text{..........} + x_n}{n}$

$$\text{So, mean} = \frac{994 + 996 + 998 + 1000 + 1002}{5}$$

$$= \frac{4990}{5}$$

$$= 998$$

Hence, mean of the given numbers is 998

**(ii) Geometric mean:**

If we have a series of n positive values such as {x_1}, {x_2}, {x_3}, .....,{x_k} $x_1, x_2, x_3, \ldots\ldots\ldots, x_k$ are repeated $f_1, f_2, f_3, \ldots, f_k$ times respectively then geometric mean will become:

G.M of X $= \overline{X} = \sqrt[n]{x_1^{f_1}.x_2^{f_2}.x_3^{f_3}\ldots\ldots.x_k^{f_k}}$ (For Grouped Data)

Where $n = f_1 + f_2 + f_3 + \ldots f_k$

**Example:** Find the Geometric mean of the values 10, 5, 15, 8, 12

**Solution:** Here x1=10, x2=5, x3=15, x4=8, x5=12 and n=5

G.M of X= $\overline{X} = \sqrt[5]{10 \times 5 \times 15 \times 8 \times 12}$

$$\overline{X} = \sqrt[5]{72000 = 9.36}$$

**(iii) Harmonic mean:** Harmonic mean is quotient of "number of the given values" and "sum of the reciprocals of the given values".

Harmonic mean in mathematical terms is defined as follows:

| For Ungrouped Data | For Grouped Data |
|---|---|
| H.M of X = $\bar{X} = \dfrac{n}{\sum\left(\frac{1}{x}\right)}$ | H.M of X = $\bar{X} = \dfrac{\sum f}{\sum\left(\frac{f}{x}\right)}$ |

**Example:** Calculate the harmonic mean of the numbers: 13.5, 14.5, 14.8, 15.2 and 16.1

**Solution:** The harmonic mean is calculated as below:

| X | 1/X |
|---|---|
| 13.2 | 0.0758 |
| 14.2 | 0.0704 |
| 14.8 | 0.0676 |
| 15.2 | 0.0658 |
| 16.1 | 0.0621 |
| total | $\sum\left(\frac{1}{x}\right) = 0.3417$ |

$H.M \text{ of } X = \bar{X} = n\sum (1x)$
$H.M \text{ of } X = \bar{X} = 50.3417 = 14.63$

### (iv) Median:

***For Example:*** Find the median of this data: 83,37,70,29,45,63,41,70,34,54
**Solution:** Arrange the data in ascending order, we get-
29, 34, 37, 41, 45, 54, 63, 70, 70, 83
Here, the number of observation n = 10 (even)
Now, medians is-

$$\Rightarrow \frac{valueofthe(\frac{n}{2})^{th}observation + valueof(\frac{n}{2}+1)^{th}observation)}{2}$$

$$\Rightarrow \frac{valueofthe(\frac{10}{2})^{th}observation + valueof(\frac{10}{2}+1)^{th}observation)}{2}$$

$$\Rightarrow \frac{valueofthe(5)^{th}observation + valueof(6)^{th}observation)}{2}$$

$$\Rightarrow \frac{45+54}{2}$$

$$\Rightarrow \frac{99}{2}$$

= 49.5
Hence the value of median is 49.5.

### (v) Mode:

***For Example:*** Find out the mode of the following marks obtained by 15 students in class:
Marks: 4,6,5,7,9,8,10,4,7,6,5,9,8,7,7.
**Solution:** Arrange the data in the form of a frequency table-

| Value | Tally marks | frequency |
|---|---|---|
| 4 | II | 2 |
| 5 | II | 2 |
| 6 | II | 2 |
| 7 | IIII | 4 |
| 8 | II | 2 |
| 9 | II | 2 |
| 10 | I | 1 |

Since. the value of 7 occurs maximum number of times i.e 4.
Hence, the mode value is 7

**(2) (i) If** $x_1, x_2, x_3, \ldots, x_n$ are n values of a variable X, then the arithmetic mean of these values is given by $\overline{X} = \dfrac{x_1 + x_2 + x_3 + \ldots\ldots x_n}{n}$ $or,$ $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

***For Example:*** Find the mean of 994, 996, 998,1002 and 1000.

**Solution:** No. of values n = 5

We know, mean $(\bar{x}) = \dfrac{x_1 + x_2 + x_3 + \ldots\ldots x_n}{n}$

So, mean $= \dfrac{994 + 996 + 998 + 1000 + 1002}{5}$

$= \dfrac{4990}{5}$

$= 998$

Hence, mean of the given numbers is 998

**(ii)** If a variate X take values $x_1, x_2, x_3, \ldots, x_n$ with corresponding frequencies **f1, f2, f3, ....., fn** respectively, then the arithmetic mean of these values is given by

$$\overline{X} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots\ldots f_nx_n}{f_1 + f_2 + \ldots\ldots f_n} \; or, \; X = \frac{\sum_{i=1}^{n} f_i x_i}{N}, \text{ where } N = \sum_{i=1}^{n} f_i$$

***For Example:*** Calculate the mean for the following distribution

| x | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|
| f | 4 | 8 | 14 | 11 | 3 |

**Solution:** Calculation of the Arithmetic mean-

| $x_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|
| 5 | 4 | 20 |
| 6 | 8 | 48 |
| 7 | 14 | 98 |
| 8 | 11 | 88 |
| 9 | 3 | 27 |
| | $N = \Sigma f_i = 40$ | $\Sigma f_i x_i = 281$ |

Now, mean $= \sum \dfrac{f_i x_i}{f_1} = \dfrac{281}{40} = 7.025$

Hence, value of mean is 7.025

**(3) If $\bar{X}$ is the mean of n observations $x_1, x_2, \ldots x_n$, then**

**(i) The algebraic sum of the deviations about $\overline{X}$ is 0, i.e.** $\sum_{i=1}^{n}\left(x_i - \overline{X}\right) = 0$

***For Example:*** Duration of sunshine (in hours) in Amritsar for first 10 days of August 1997 as reported by the Meteorological Department are given below:
9.6, 5.2, 3.5, 1.5, 1.6, 2.4, 2.6, 8.4, 10.3, 10.9

Verify that $\sum_{i=1}^{10}\left(x_i - \overline{X}\right) = 0$

**Solution:** We have to verify that $\sum_{i=1}^{10}\left(x_i - \bar{x}\right) = 0$

Taking LHS,

$\Rightarrow \quad \sum_{i=1}^{10}\left(x_i - \bar{x}\right)$

$\Rightarrow \quad \sum_{i=1}^{10} x_i - \sum_{i=1}^{10} \bar{x}$

$\Rightarrow$ (9.6 + 5.2 + 3.5 + 1.5 + 1.6 + 2.4 + 2.6 + 8.4 + 10.3 + 10.9) – 10 × $\bar{x}$

$\Rightarrow$ 56 – 10 x 5.6

$\Rightarrow$ 56 – 56

$\Rightarrow$ 0 = RHS   Hence Proved

**(ii) Prove that the mean of the observations $x_1 \pm a, x_2 \pm a, \ldots\ldots, x_n \pm a$ is $\overline{X} \pm a$**

Given, $\bar{x} = \dfrac{x_1 + x_2 + \ldots\ldots + x_n}{n}$ .....(i)

The observation are $(x_1 + a), (x_2 + a), \ldots\ldots, (x_n + a)$

Mean = $\dfrac{(x_1 + a) + (x_2 + a) + \ldots\ldots + (x_n + a)}{n}$

$= \dfrac{x_1 + x_2 + \ldots\ldots + x_n + n \times a}{n}$

$= \dfrac{x_1 + x_2 + \ldots\ldots + x_n}{n} + \dfrac{n \times a}{n}$

From Equation (1); we get

Mean = $\left(\bar{x} + a\right)$

So that given statement is true.

**(iii) Prove that the mean of the observations $ax_1, ax_2, \ldots\ldots, ax_n$ is $a\overline{X}$**

***For Example:*** Mean of $x_1, x_2, \ldots\ldots, x_n = \dfrac{x_1, x_2, \ldots\ldots, x_n}{n}$

But mean = $\bar{x}$ (given)

$\bar{x} = \dfrac{x_1, x_2, \ldots\ldots, x_n}{n}$ ........(1)

The observation are $ax_1, ax_2, \ldots\ldots, ax_n$

Mean = $\dfrac{ax_1 + ax_2 + \ldots\ldots + ax_n}{n}$

$= \dfrac{a(x_1 + x_2 + \ldots\ldots + x_n)}{n} = a\bar{x}$

Thus, the given statement is true.

**(iv) Prove that the mean of the observations $\frac{x_1}{a}, \frac{x_2}{a}, \ldots\ldots \frac{x_n}{a}$ is $\frac{\overline{x}}{a}$**

**Proof:** We have, $\overline{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)$

Let $\overline{X}'$ be the mean of $\frac{x_1}{a}, \frac{x_2}{a}, \ldots\ldots \frac{x_n}{a}$. Then,

$$\overline{X}' = \frac{1}{n}\left(\frac{x_1}{a} + \frac{x_2}{a} + \ldots\ldots + \frac{x_n}{a}\right)$$

$$\overline{X}' = \frac{1}{n}\left(\frac{x_1 + x_2 + \ldots.. + x_n}{a}\right)$$

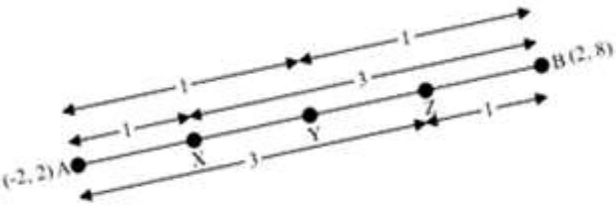
$$\overline{X}' = \frac{1}{a}\left(\frac{x_1 + x_2 + \ldots.. + x_n}{n}\right)$$

$$\overline{X}' = \frac{1}{a}\left[\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\right] = \frac{1}{a}\left(\overline{X}\right)$$

$$\overline{X}' = \frac{\overline{X}}{a}$$

**(4) Media of a distribution is the value of the variable which divides the distribution into two equal parts.**

***For Example:*** Find the coordinates of the points which divide the line segment A(-2, 2) & B(2, 8) into four equal parts.




**Solution:** From the figure, it can be observed that points P, Q, R are dividing the line segment in a ratio 1:3, 1:1, 3:1 respectively.

Coordinates of P = $\left(\frac{1\times2+3+(-2)}{1+3}, \frac{1\times8+3\times2}{1+3}\right)$

$= \left(-1, \frac{7}{2}\right)$

Coordinates of Q = $\left(\frac{2+(-2)}{2}, \frac{2+8}{2}\right)$

$= (0, 5)$

Coordinates of R = $\left(\frac{3\times2+1\times(-2)}{3+1}, \frac{3\times8+1\times2}{3+1}\right) = \left(1, \frac{13}{2}\right)$

(5) If $x_1, x_2, ......x_n$ are n values of a variable arranged in ascending or descending order, then

Median = value of $\left(\frac{n+1}{2}\right)^{th}$ observation, if n is odd

Median = (value of $\left(\frac{n}{2}\right)^{th}$ observation + value of $\left(\frac{n}{2}+1\right)^{th}$ observation)/2, if n is even

***For Example:***

(i) Find the median of this data: 83, 37, 70, 29, 45, 63, 41, 70, 34, 54

**Solution:** Arrange the data in ascending order, we get-

**29, 34, 37, 41, 45, 54, 63, 70, 70, 83**

**Here, the number of observation n = 10 (even)**

Now, medians is-

$$\Rightarrow \frac{value of the (\frac{n}{2})^{th} observation + value of (\frac{n}{2}+1)^{th} observation)}{2}$$

$$\Rightarrow \frac{value of the (\frac{10}{2})^{th} observation + value of (\frac{10}{2}+1)^{th} observation)}{2}$$

$$\Rightarrow \frac{value of the (5)^{th} observation + value of (6)^{th} observation)}{2}$$

$$\Rightarrow \frac{45+54}{2}$$

$$\Rightarrow \frac{99}{2}$$

= 49.5

Hence the value of median is 49.5.

**(ii) Find the median of this data: 15, 6, 16, 8, 22, 21, 9, 18, 25**

**Solution:** Arrange the data in the ascending order, we get-

**6, 8, 9, 15, 16, 18, 21, 22, 25**

Here, the number of observation n=9 (odd)

Now, median = $Value\ of \left(\frac{n+1}{2}\right)^{th} observation$

$= Value\ of \left(\frac{9+1}{2}\right)^{th} observation$

$= Value\ of \left(\frac{10}{2}\right)^{th} observation$

$= Value\ of\ (5)^{th} observation$

= 16

Hence, value of median is 16.