Statistics

STATISTICS

MEASURES OF CENTRAL TENDENCY

It is that single value which may be taken as the most suitable representative of the data. This single value is known as the average. Averages are generally the central part of the distribution and therefore they are also called the measures of central tendency.

It can be divided into two groups :

(a) MATHEMATICAL AVERAGE :

I. Arithmetic mean or mean

II. Geometric mean

III. Harmonic mean

(b) POSITIONAL AVERAGE :

I. Median

II. Mode

ARITHMETIC MEAN

Arithmetic mean of ungrouped data :

If x_1, x_2 , ..., x_n be n observations, then their arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} \quad \text{or} \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Arithmetic mean of discrete grouped data :

Let x_1 , x_2 , ..., x_n be n observation and let f_1, f_2 , ..., f_n be their corresponding frequencies, then their

$$\text{mean } \bar{x} = \frac{f_1x_1 + f_2x_2 + ... + f_nx_n}{f_1 + f_2 + ... + f_n} \quad \text{or} \quad \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Arithmetic mean of continuous grouped data :

Take mid points of given classes as x_i and use formula as given for discrete grouped data.

Short cut method :

Assumed Mean Method :

If the values of x or (and) f are large, the calculation of arithmetic mean by the previous formula used, is quite tedious and time consuming. In such case we take the deviation from assumed mean A which is in the middle or just close to it in the data.

Then $\bar{x} = A + \frac{\sum f_i \ d_i}{\sum f_i}$

where $d_i = x_i - A$ = deviation for each observation

This method is nothing but shiffting of origin from zero to the assumed mean on the number line.

Solved Examples

Ex. 1 Find the mean of the following frequency distribution

x:	5	15	25	35	45	55
f:	12	18	27	20	17	6

Sol. Here $N = \sum f_i = 100$. On taking assumed mean a = 35.

$\Sigma f_i d_i = \Sigma f_i (x_i - 35) = 12(-30) + 18(-20) + 27(-10) + 20(0) + 17(10) + 6(20) = -700$

$$\therefore \quad \bar{x} = a + \frac{\sum f_i d_i}{N} = 35 + \frac{-700}{100} = 28$$

Step Deviation Method or Change of Scale

Sometimes during the application of shortcut method of finding the mean, the deviation d_i are divisible by a common number h (say). In such case the arithmetic is reduced to a great extent taken by-

$$d_i = \frac{x_i - a}{h}, \; i = 1, 2, ... n$$

$$\therefore \quad \bar{x} = a + h\left(\frac{\sum f_i d_i}{\sum f_i}\right)$$

Note :

* a and h can be any numbers but if the lengths of class intervals are equal then h may be taken as width of the class interval.
* In particular if each observation is multipled or divided by a constant, the mean is also multiplied or divided by the same constant.

Weighted Arithmetic Mean

If $w_1, w_2, w_3 ......, w_n$ are the weight assigned to the values $x_1, x_2, x_3 ....., x_n$ respectively, then the weighted average is defined as-

$$\text{weighted A.M.} = \frac{w_1x_1 + w_2x_2 + ... + w_nx_n}{w_1 + w_2 + ... + w_n}$$

$$= \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Solved Examples

Ex.2 Find the weighted mean of first n natural numbers when their weights are equal to their squares.

Sol. Weighted mean $= \dfrac{1.1^2 + 2.2^2 + ... + n.n^2}{1^2 + 2^2 + ... + n^2}$

$$= \frac{1^3 + 2^3 + ... + n^3}{1^2 + 2^2 + ... + n^2} = \frac{n^2(n+1)^2}{4} \cdot \frac{6}{n(n+1)(2n+1)}$$

$$= \frac{3n(n+1)}{2(2n+1)}$$

Combined Mean

If $\bar{x}_1$ and $\bar{x}_2$ be the means of two related groups having n_1 and n_2 items respectively then the combined mean $\bar{x}$ of both the group is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}.$$

If there are more than two groups then

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + ...}{n_1 + n_2 + n_3 + ...}$$

Ex.3 The mean income of a group of person is Rs.400. Another group of persons has mean income Rs.480. If the mean income of all the persons in the two groups together is Rs. 430, then find the ratio of the number of persons in the groups.

Sol. $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$\because \; \bar{x}_1 = 400, \; \bar{x}_2 = 480, \; \bar{x} = 430$

$$\therefore \quad 430 = \frac{n_1(400) + n_2(480)}{n_1 + n_2}$$

$$\Rightarrow 30n_1 = 50n_2 \qquad \Rightarrow \qquad \frac{n_1}{n_2} = \frac{5}{3}$$

PROPERTIES OF ARITHMETIC MEAN

(i) If $\bar{x}$ is the mean of $x_1, x_2, ..., x_n$, then mean of $ax_1 + b, \; ax_2 + b, ... ax_n + b$ is $a\bar{x} + b$.

(ii) Arithmetic mean is dependent on change of origin & scale.

MERITS OF ARITHMETIC MEAN

(i) It is rigidly defined.
(ii) It is based on all the observation taken.
(iii) It is calculated with reasonable ease.
(iv) It is least affected by fluctuations in sampling.
(v) It is based on each observation and so it is a better representative of the data.
(vi) It is relatively reliable
(vii) Mathematical analysis of mean is possible.

DEMERITS OF ARITHMETIC MEAN

(i) It is severely affected by the extreme values.
(ii) It cannot be represented in the actual data since the mean does not coincide with any of the observed value.
(iii) It cannot be computed unless all the items are known.

Solved Examples

Ex.4 Find mean of data 2, 4, 5, 6, 8, 17.

Sol. Mean = $\frac{2+4+5+6+8+17}{6} = 7$

Ex.5 Find the mean of the following distribution :

x :	4	6	9	10	15
f:	5	10	10	7	8

Sol. Calculation of Arithmetic Mean

x_i	f_i	f_ix_i
4	5	20
6	10	60
9	10	90
10	7	70
15	8	120
	$N = \Sigma f_i = 40$	$\Sigma f_ix_i = 360$

$\therefore$ Mean = $\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{360}{40} = 9$

Ex.6 Find the mean wage from the following data :

Wage (in Rs) :	800	820	860	900	920	980	1000
No. of workers :	7	14	19	25	20	10	5

Sol. Let the assumed mean be A = 900 and h = 20.

Calculation of Mean

Wage (in Rs) x_i	No. of workers f_i	$d_i = x_i - A = x_i - 900$	$u_i = \frac{x_i - 900}{20}$	f_iu_i
800	7	–100	–5	–35
820	14	–80	–4	–56
860	19	–40	–2	–38
900	25	0	0	0
920	20	20	1	20
980	10	80	4	40
1000	5	100	5	25
	$N = \Sigma f_i = 100$			$\Sigma f_i u_i = -44$

We have, $N = 100$, $\Sigma f_iu_i = -44$, $A = 900$ and $h = 20$

$\therefore$ Mean = $\bar{X} = A + h\left(\frac{1}{N}\Sigma f_iu_i\right) \Rightarrow \bar{X} = 900 + 20 \times \frac{-44}{100} = 900 - 8.8 = 891.2$

Hence, mean wage = Rs. 891.2

Ex.7 Find the mean of the following frequency distribution :

Class-interval :	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of workers f :	7	10	15	8	10

Sol. Calculation of Mean

Class-interval	Mid-values (x_i)	Frequency f_i	$d_i = x_i - 25$	$u_i = \frac{x_i - 25}{10}$	$f_i u_i$
0 – 10	5	7	–20	–2	–14
10 – 20	15	10	–10	–1	–10
20 – 30	25	15	0	0	0
30 – 40	35	8	10	1	8
40 – 50	45	10	20	2	20
		$N = \sum f_i = 50$			$\sum f_i u_i = 4$

We have, $A = 25$, $h = 10$, $N = 50$ and $\Sigma f_i u_i = 4$.

$\therefore$ Mean $= A + h\left\{\frac{1}{N}\Sigma f_i u_i\right\}$ $\Rightarrow$ mean $= 25 + 10 \times \frac{4}{50} = 25.8$

Ex.8 Find the mean marks of students from the following cumulative frequency distribution :

Marks	Number of students	Marks	Number of students
0 and above	80	60 and above	28
10 and above	77	70 and above	16
20 and above	72	80 and above	10
30 and above	65	90 and above	8
40 and above	55	100 and above	0
50 and above	43		

Sol. Here we have, the cumulative frequency distribution. So, first we convert it into an ordinary frequency distribution. We observe that there are 80 students getting marks greater than or equal to 0 and 77 students have secured 10 and more marks. Therefore, the number of students getting marks between 0 and 10 is 80 – 77 = 3.

Similarly, the number of students getting marks between 10 and 20 is 77 – 72 = 5 and so on. Thus, we obtain the following frequency distribution.

Marks	Number of students	Marks	Number of students
0 – 10	3	50 – 60	15
10 – 20	5	60 – 70	12
20 – 30	7	70 – 80	6
30 – 40	10	80 – 90	2
40 – 50	12	90 – 100	8

Now, we compute arithmetic mean by taking 55 as the assumed mean.

Computation of Mean

Marks	Mid-value (x_i)	Frequency (f_i)	$u_i = \frac{x_i - 55}{10}$	f_iu_i
0 – 10	5	3	–5	–15
10 – 20	15	5	–4	–20
20 – 30	25	7	–3	–21
30 – 40	35	10	–2	–20
40 – 50	45	12	–1	–12
50 – 60	55	15	0	0
60 – 70	65	12	1	12
70 – 80	75	6	2	12
80 – 90	85	2	3	6
90 – 100	95	8	4	32
Total		$\Sigma f_i = 80$		$\Sigma f_iu_i = -26$

We have,

$N = \Sigma f_i = 80$, $\Sigma f_iu_i = -26$, $A = 55$ and $h = 10$

$\therefore \ \overline{X} = A + h\left\{\frac{1}{N}\Sigma f_iu_i\right\}$

$\Rightarrow \overline{X} = 55 + 10 \times \frac{-26}{80} = 55 - 3.25 = 51.75$ Marks

GEOMETRIC MEAN

(i) In case of individual series

If $x_1, x_2, ... x_n$ are n positive values of a variate x, then their geometric mean G is defined as

$G = (x_1x_2......x_n)^{1/n}$ or $G = \text{anti log}\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$

(ii) In case of descrete series

If $f_1, f_2, ... f_n$ are frequencies of the values $x_1, x_2, ...x_n$ of the variate respectively, then $G = \left(x_1^{f_1}\,x_2^{f_2}.....x_n^{f_n}\right)^{1/N}$

$G = \text{anti log}\left(\frac{\Sigma f_i \log x_i}{N}\right)$ where $N = \sum_{i=1}^{n} f_i$

Note : If G_1, G_2 are geometric means of two series containing n_1, n_2 values respectively, then the geometric mean of their combined series G is given by

$\log G = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2}$

Solved Examples

Ex.9 Find the geometric mean of 1, 2, 2^2, ...2^n.

Sol. G.M. $= (1.2.2^2......2^n)^{1/n+1} = [2^{(1+2+..+n)}]^{1/(n+1)} = 2^{n/2}$

HARMONIC MEAN

If $x_1, x_2, ..., x_n$ are n non-zero values of a variate x, then their harmonic mean H is defined as follows-

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}}$$

If $f_1, f_2, ..., f_n$ are frequencies of the values $x_1, x_2, ..., x_n$ of the variate respectively, then

$$H = \frac{N}{\Sigma f_i\left(\frac{1}{x_i}\right)}$$

Note : If A, G, H are AM, GM, HM respectively of some obsevations, then $A \geq G \geq H$

Solved Examples

Ex.10 Find the harmonic mean of $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, ..., \frac{1}{17}$.

Sol. H.M. $= \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$ (For individual series)

$$= \frac{1}{\frac{1}{16}\sum_{i=1}^{16}(2 + 3 + ... + 17)} = \frac{2}{19}$$

MEDIAN

Median is the middle most or the central value of the variate in a set of observations when the observations are arranged either in ascending or in descending order of their magnitudes. It divides the arranged series in two equal parts.

(i) Median of an individual series :

Let $x_1, x_2, \ldots\ldots, x_n$ be n observations arranged in ascending or descending order. Then

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

Median (M) =

$$\frac{\left(\frac{n}{2}\right)^{th}\text{observation} + \left(\frac{n}{2}+1\right)^{th}\text{observation}}{2}$$ if n is even

(ii) Median of the discrete frequency distribution :

Find the cumulative frequency (C.F.)

Median (M) = Value of $\left(\frac{N+1}{2}\right)^{th}$ observation if N is odd

Median (M) =

$$\frac{\left(\frac{N}{2}\right)^{th}\text{observation} + \left(\frac{N}{2}+1\right)^{th}\text{observation}}{2}$$ if N is even

where $N = \sum_{i=1}^{n} f_i$

(iii) Median of continuous frequency distribution :

Let the number of observation be N. Prepare the cumulative frequency table. Find the median class i.e. the class in which the observation whose cumulative frequency is equal to or just greater than $\frac{N}{2}$ lies.

The median value is given by the formula :

$$\text{Median (M)} = \ell + \left[\frac{\left(\frac{N}{2}\right) - c}{f}\right] \times h$$ where

N = total frequency = Σf_i

ℓ = lower limit of median class

f = frequency of the median class

c = cumulative frequency of the class preceding the median class

h = class interval (width) of the median class

MERITS AND DEMERITS OF MEDIAN

The following are some merits and demerits of median :

Merits :

(i) It is easy to compute and understand.

(ii) It is well defined an ideal average should be

(iii) It can also be computed in case of frequency distribution with open ended classes.

(iv) It is not affected by extreme values.

(v) It can be determined graphically.

(vi) It is proper average for qualitative data where items are not measured but are scored.

Demerits :

(i) For computing median data needs to be arranged in ascending or descending order.

(ii) It is not based on all the observations of the data.

(iii) It cannot be given further algebraic treatment.

(iv) It is affected by fluctuations of sampling.

(v) It is not accurate when the data is not large.

(vi) In some cases median is determined approximately as the mid-point of two observations whereas for mean this does not happen.

Solved Examples

Ex.11 Find the median of the following marks obtained by 100 students

x:	0-10	10-20	20-30	30-40	40-50
f:	8	30	40	12	10

Sol.

Class (x)	Number of students (f)	Cumulative frequency (cf)
0-10	8	8
10-20	30	38
20-30	40	78
30-40	12	90
40-50	10	100

$\because \frac{N+1}{2} = 50.5$ which lies in the value 78 of cf. hence class of this cf i.e. 20–30 is the median class.

So $\ell = 20$, f = 40, F = 38, N = 100, h = 10

$\therefore$ median $= \ell + \frac{\frac{1}{2}N - F}{f} \times h$

$= 20 + \frac{50 - 38}{40} \times 10 - 23$

Ex.12 Find the median of observations 4, 6, 9, 4, 2, 8, 10

Sol. Values in ascending order are 2, 4, 4, 6, 8, 9, 10

here n = 7 so $\frac{n+1}{2} = 4$

so median = 4^{th} observaiton = 6

Ex.13 Obtain the median for the following frequency distribution :

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

Sol.

x	f	cf
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
	N = 120	

Here, N = 120 $\Rightarrow$ $\frac{N}{2} = 60$

We find that the cummulative frequency just greater than $\frac{N}{2}$ i.e., 60 is 65 and the value of x corresponding to 65 is 5. Therefore, Median = 5.

Ex.14 Calculate the median from the following distribution :

Class	Frequency
5-10	5
10-15	6
15-20	15
20-25	10
25-30	5
30-35	4
35-40	2
40-45	2

Sol.

Class	Frequency	Cumulative Frequency
5 – 10	5	5
10 – 15	6	11
15 – 20	15	26
20 – 25	10	36
25 – 30	5	41
30 – 35	4	45
35 – 40	2	47
40 – 45	2	49
	N = 49	

We have, N = 49

$\therefore \frac{N}{2} = \frac{49}{2} = 24.5$

The cummulative frequency just greater than $\frac{N}{2}$ is 26 and the corresponding class is 15-20.
Thus 15-20 is the median class such that $\ell = 15$, $f = 15$, F = 11 and h = 5

$\therefore$ Median $= \ell + \frac{\frac{N}{2} - F}{f} \times h = 15 + \frac{24.5 - 11}{15} \times 5$

$= 15 + \frac{13.5}{3} = 19.5$

MODE

In a frequency distribution the mode is that value of the variate which has the maximum frequency.

Method for determining mode :

(i) For ungrouped distribution :

In case of ungrouped distribution, values are first arranged in ascending or descending order and their respective frequencies are written against them. Then by observing their frequencies, the value containing maximum frequency is determined to be the mode.

(ii) For grouped distribution :

In this case, first we find the modal class (containing the maximum frequency), then the exact value of the mode is found by during the following formula ;

$$\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times h$$

where l = lower limit of the modal class

f = frequency of the modal class

f_1 = frequency of the class preceding the modal class

f_2 = frequency of the class following the modal class

h = width of the modal class

MERITS, DEMERITS OF MODE :

The following are some merits and demerits of mode :

Merits :

(i) It is readily comprehensible and easy to compute. In some case it can be computed merely by inspection.

(ii) It is not affected by extreme values. It can be obtained even if the extreme values are not known.

(iii) Mode can be determined in distributions with open classes.

(iv) Mode can be located on graph also.

Demerits :

(i) It is ill-defined. It is not always possible to find a clearly defined mode. In some cases, we may come across distributions with two modes. Such distributions are called bimodal. If a distribution has more than two modes, it is said to be multimodal.

(ii) It is not based upon all the observation.

(iii) Mode can be calculated by various formulae as such the value may differ from one to other. Therefore, it is not rigidly defined.

(iv) It is affected to a greater extent by fluctuations of sampling.

USES OF MODE :

Mode is used by the manufacturers of ready-made garments, shoes and accessories in common use etc. The readymade garment manufacturers made those sizes more which are used by most of the persons than other sizes. Similarly, the makers of shoes will make that size maximum which the majority people use and others in less quantity.

RELATIONSHIP BETWEEN MEAN, MODE AND MEDIAN :

(i) In symmetrical distribution, Mean = Mode = Median

(ii) In skew (moderately asymmetrical) distribution, Mode = 3 Median – 2 Mean

Graphical Representation of Data :

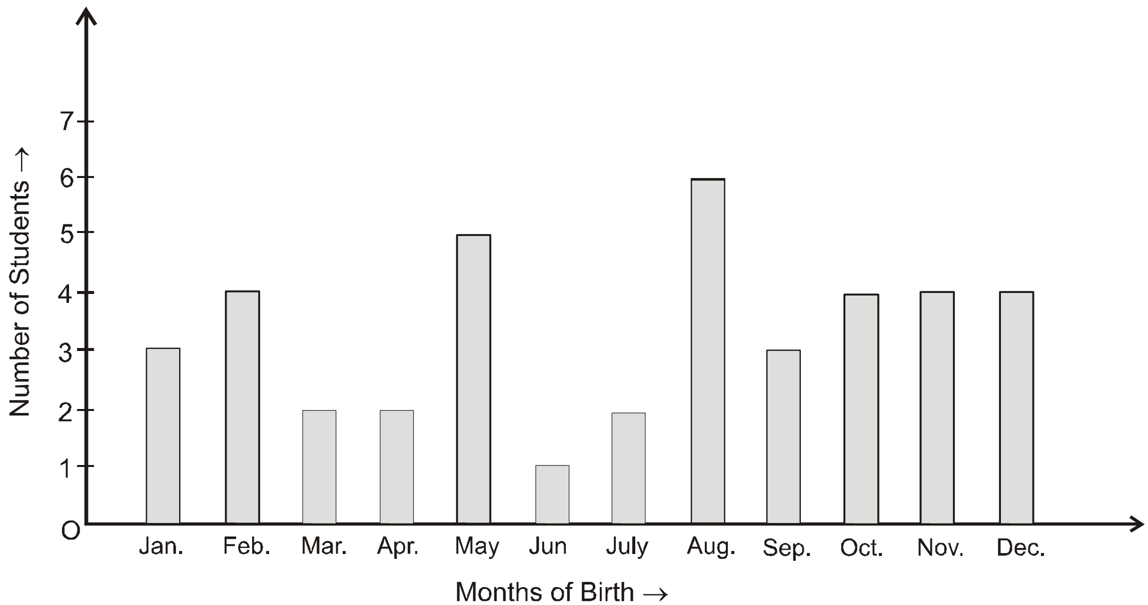
Usually comparisons among the individual items are best shown by means of graphs. The representation then becomes easier to understand than the actual data. We shall study the following graphical representations in this section

(A) Bar graphs

(B) Histograms of uniform width and of varying widths

(C) Frequency polygons

(A) Bar graphs : 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, 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. For example, 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 :



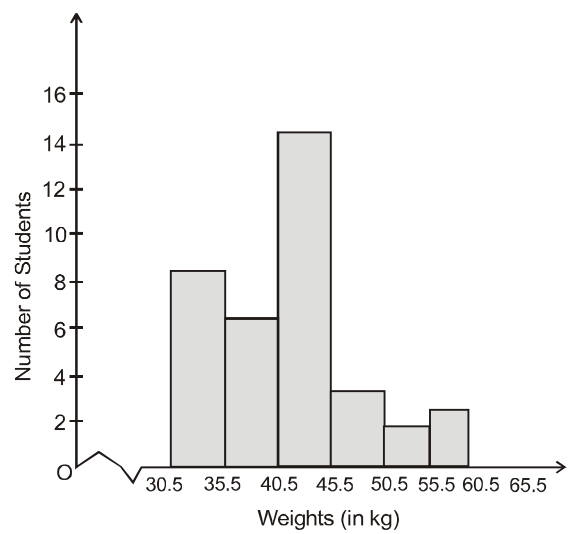
The variable here is the 'month of birth' and the value of the variable is the 'Number of students born'.

The maximum number of students were born in the month of August.

(B) Histogram : This is form of representation like the bar graph, but it is used for continuous class intervals. For instance, consider the frequency distribution table representing the weights of 36 students of a class :

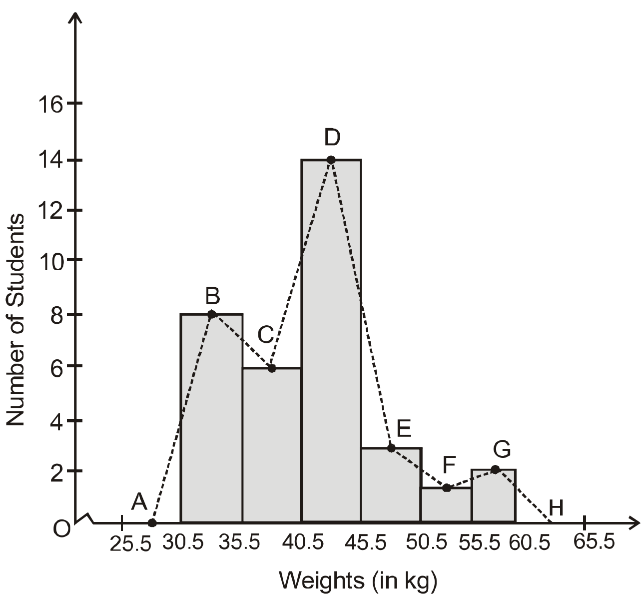
Weight (in kg)	Number of students
30.5-35.5	9
35.5-40.5	6
40.5-45.5	15
45.5-50.5	3
50.5-55.5	1
55.5-60.5	2
Total	**36**

Let us represent the data given above graphically as follows :



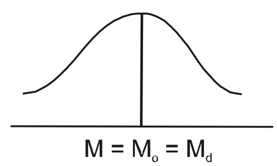
Since there are no gaps in between consecutive rectangles, the resultant graph appears like a solid figure. This is called a histogram. Unlike a bar graph, the width of the bar plays a significant role in its construction. Here, in fact areas of the rectangles erected are proportional to the corresponding frequencies since the widths of the rectangles are all equal.

(C) Frequency Polygon : There is yet another visual way of representing quantitative data and its frequencies. Consider the histogram represented by figure shown above. Let us join the mid-points of the upper sides of the adjacent rectangles of this histogram by means of line segments. Let us call these mid-points B, C, D, E, F and G. When joined by line segments, we obtain the figure BCDEFG. To complete the polygon, we assume that there is a class interval with frequency zero before 30.5 – 35.5 and one after 55.5 – 60.5 and their mid-points are A and H respectively. ABCDEFGH is the given frequency polygon corresponding to the given data

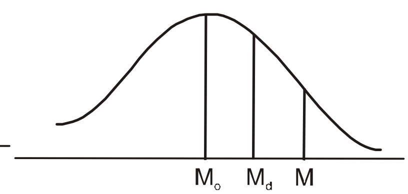


Althrough there exists no class preceding the lowest class and no class succeeding the highest class, addition of the two class intervals with zero frequency enables us to make the area of the frequency polygon the same as the area of the histogram. Frequency polygons can also be drawn independently without drawing histogram. Frequency polygons are used when the data is continuous and very large. It is very useful for comparing two different sets of data of the same nature, for example, comparing the performance of two different sections of the same class.

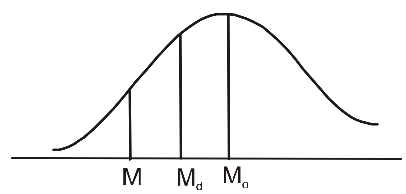
Skewness : We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data. The term 'skewness' refers to lack of symmetry. We can define skewness of a distribution as the tendancy of a distribution to depart from symmetry. In a symmetrical distribution we have Mean = Median – Mode. When the distribution is not symmetrical, it is called asymmetrical or skewed.In a skewed distribution Mean ≠ Median ≠ Mode. In positively skewed distribution we have Mean > Median > Mode. In Negatively skewed distribution, we have Mean < Median < Mode.



Symmetrical Distribution



Positively skewed Distribution



Negatively skewed Distribution

Solved Examples

Ex.15 Find mode of data 2, 4, 6, 8, 8, 12, 17, 6, 8, 9.

Sol. 8 occurs maximum number of times so mode = 8

Ex.16 Compute the mode for the following frequency distribution :

Size of items	Frequency
0 - 4	5
4 - 8	7
8 - 12	9
12 - 16	17
16 - 20	12
20 - 24	10
24 - 28	6
28 - 32	3
32 - 36	1
36 - 40	0

Sol. Here, the maximum frequency is 17 and the corresponding class is 12-16 So 12-16 is the modal class.

We have, $\ell = 12$, $h = 4$, $f = 17$, $f_1 = 9$ and $f_2 = 12$

$$\therefore \text{ Mode} = \ell + \frac{f - f_1}{2f - f_1 - f_2} \times h$$

$$\Rightarrow \text{ Mode} = 12 + \frac{17 - 9}{34 - 9 - 12} \times 4$$

$$\Rightarrow \text{ Mode} = 12 + \frac{8}{13} \times 4 = 12 + \frac{32}{13} = 12 + 10.66$$

$= 32.66$

MEASURES OF DISPERSION

The dispersion of statistical distribution is the measure of the deviation of its values about the average (central) value. It gives an idea of scatteredness of different values from the mean.

The following measures of dispersion are commonly used :

(i) Range (ii) Quartile deviation
(iii) Mean deviation (iv) Standard deviation

(i) Range

The difference between the greatest and the least values of variate of a distribution, is called the range of that distribution. If the distribution is continuous grouped distribution, then its

Range = upper limit of the maximum class - lower limit of the minimum class.

Also

the coefficient of the range

$$= \frac{\text{difference of extreme values}}{\text{sum of extreme values}}$$

(ii) Quartile deviation

Quartile deviation $Q = \frac{Q_3 - Q_1}{2}$

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Note : If the distribution is symmetrical, then

$Q = M - Q_1 = Q_3 - M$ where M is the median.

(iii) Mean deviation

The mean deviation of a distribution is the arithmetic mean of the absolute values of the deviation of different values of the variate from a statistical average (mean, median or mode) of the distribution.

If $x_1, x_2 \ldots\ldots x_n$ are n observation then mean deviation about a point A is given by

Mean deviation $= \frac{1}{n}\sum |x_i - A|$

In case of discrete series

Mean deviation $= \frac{1}{N}\sum_{i=1}^{n} f_i\,|x_i - A|$, where $N = \sum_{i=1}^{n} f_i$

Coefficient of mean deviation $= \frac{\text{mean deviation}}{A}$

Note : Mean deviation is least when taken from the median.

Solved Examples

Ex.17 Find the mean deviation of number 3, 4, 5, 6, 7.

Sol. Mean deviation $= \frac{1}{5}\sum_{i=1}^{5} |x_i - \bar{x}|$

$= \frac{1}{5}\{|3 - 5| + |4 - 5| + |5 - 5| + |6 - 5| + |7 - 5|\}$

$= \frac{1}{5}\{2 + 1 + 1 + 2\} = 1.2$

Ex.18 Find the mean deviation about mean from the following data :

x_i :	3	9	17	23	27
f_i :	8	10	12	9	5

Sol. Calculation of mean deviation about mean.

x_i	f_i	$f_i x_i$	$\|x_i - 15\|$	$f_i\|x_i - 15\|$
3	8	24	12	96
9	10	90	6	60
17	12	204	2	24
23	9	207	8	72
27	5	135	12	60
	$N = \Sigma f_i = 44$	$\Sigma f_i x_i = 660$		$\Sigma f_i\|x_i - 15\| = 312$

Mean $= \bar{X} = \frac{1}{N}(\Sigma f_i\, x_i) = \frac{660}{44} = 15$

Mean deviation = M.D.

$= \frac{1}{N}\Sigma f_i\,|x_i - 15| = \frac{312}{44} = 7.09$

Ex.19 Calculate mean deviation about median for the following data 3, 9, 5, 3, 12, 10, 18, 4, 7, 19, 21.

Sol. Data in ascending order is 3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21

Median = $\frac{n+1}{2}$ th value = 6^{th} value = 9

Mean deviation about median = $\frac{\sum_{i=1}^{11}|x_i - \text{median}|}{11}$

$= \frac{58}{11}$

Ex.20 Find mean deviation from mean

x	5	7	9	10	12	15
f	8	6	2	2	2	6

Sol.

x	f	fx	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
5	8	40	– 4	4	32
7	6	42	– 2	2	12
9	2	18	0	0	0
10	2	20	1	1	2
12	2	24	3	3	6
15	6	90	6	6	36
	N = 26	$\sum fx = 234$			$\sum f\lvert x-\bar{x}\rvert = 88$

$\bar{x} = \frac{\sum fx}{\sum f} = 9$

Now, M.D. $(\bar{x}) = \frac{\sum f|x-\bar{x}|}{\sum f} = \frac{88}{26} = 3.38$

Ex.21 Find the mean deviation about the median of the following frequency distribution :

Class	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30
Frequency	8	10	12	9	5

Sol. Calculation of mean deviation about the median

Class	Mid values (x_i)	Frequency (f_i)	Cumulative Frequency (c.f.)	$\lvert x_i - 14\rvert$	$f_i\lvert x_i - 14\rvert$
0 – 6	3	8	8	11	88
6 – 12	9	10	18	5	50
12 – 18	15	12	30	1	12
18 – 24	21	9	39	7	63
24 – 30	27	5	44	13	65
			$N = \sum f_i = 44$		$\sum f_i\lvert x_i - 14\rvert = 278$

Here N = 44, so $\frac{N}{2}$ = 22 and the cumulative frequency just greater than $\frac{N}{2}$ is 30. Thus 12-18 is the median class.

Now Median = $\ell + \frac{N/2 - F}{f} \times h$, where ℓ = 12, h = 6, f = 12, F = 18

or Median = $12 + \frac{22-18}{12} \times 6 = 12 + \frac{4\times 6}{12} = 14$

Mean deviation about median

$= \frac{1}{N}\sum f_i|x_i - 14| = \frac{278}{44} = 6.318$

Ex.22 Find the mean deviation from the mean for the following data :

Classes	Frequencies
10 - 20	2
20 - 30	3
30 - 40	8
40 - 50	14
50 - 60	8
60 - 70	3
70 - 80	2

Sol. We prepare the table as follows : Computation of mean deviation from mean

Classes	Mid – values (x_i)	frequencies f_i	$f_i\, x_i$	$\lvert x_i - \bar{X}\rvert = \lvert x_i - 45\rvert$	$f_i \lvert x_i - \bar{X}\rvert$
10 – 20	15	2	30	30	60
20 – 30	25	3	75	20	60
30 – 40	35	8	280	10	80
40 – 50	45	14	630	0	0
50 – 60	55	8	440	10	80
60 – 70	65	3	195	20	60
70 – 80	75	2	150	30	60
		$n = \sum f_i = 40$	$\sum f_i x_i = 1800$		$\sum f_i \lvert x_i - \bar{X}\rvert = 400$

We have, N = 40 and $\sum f_i x_i = 1800$

$$\therefore \quad \bar{X} = \frac{\sum f_i x_i}{N} = \frac{1800}{40} = 45$$

Now $\quad \Sigma f_i \lvert x_i - \bar{X}\rvert = 400$ and $N = \Sigma f_i = 40$

$$\therefore \quad \text{M.D.} = \frac{1}{N}\Sigma f_i \lvert x_i - \bar{X}\rvert = \frac{400}{40} = 10$$

LIMITATIONS OF MEAN DEVIATION :

Following are some limitations or demerits of mean deviation.

(i) In a frequency distribution the sum of absolute values of deviations from the mean is always more than the sum of the deviations from median. Therefore, mean deviation about mean is not very scientific Thus, in many cases, mean deviation may give unsatisfactory results.

(ii) In a distribution, where the degree of variability is very high, the median is not a representative central value. Thus, the mean deviation about median calculated for such series can not be fully relied.

(iii) In the computation of mean deviation we use absolute values of deviations. Therefore, it cannot be subjected to further algebraic treatment.

VARIANCE AND STANDARD DEVIATION :

The variance of a variate x is the arithmetic mean of the squares of all deviations of x from the arithmetic mean of the observations and is denoted by var(x) or σ^2 .

The positive square root of the variance of a variate x is known as standard deviation

i.e. standard deviation (S.D.) = $\sqrt{\text{var}(x)} = \sqrt{\sigma^2} = \sigma$

(i) Variance of Individual observations :

If $x_1, x_2, ..., x_n$ are n values of a variable x, then by defintion

$$\text{var}(x) = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$$

If the values of variable x are large, the calculation of variance from the above formulae is quite tedious and time consuming. In that case, we take deviation from an arbitrary point A (say) then

$$\text{var}(x) = \frac{1}{n}\left[\sum_{i=1}^{n}(d_i - \bar{d})^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^2, \text{ where } d_i = x_i - A$$

$$\text{Also var}(x) = \frac{h^2}{n}\left[\sum_{i=1}^{n}(u_i - \bar{u})^2\right]$$

$$= h^2\left[\frac{1}{n}\sum_{i=1}^{n} u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} u_i\right)^2\right] \quad \text{where } u_i = \frac{x_i - A}{h}$$

(ii) Variance of discrete frequency distribution :

If $x_1, x_2, ..., x_n$ are n observation with frequencies $f_1, f_2, ..., f_n$, then

$$\text{var}(x) = \frac{1}{N}\left\{\sum_{i=1}^{n} f_i(x_i - \bar{x})^2\right\} = \left(\frac{1}{N}\sum_{i=1}^{n} f_i x_i^2\right) - \bar{x}^2$$

where $N = \sum_{i=1}^{n} f_i$

If the value of x or f are large, we take the deviations of the values of variable x from an arbitrary point A. (say)

$\therefore\ d_i = x_i - A$; i = 1, 2, ..., N

$$\therefore\ \text{Var}(x) = \frac{1}{N}\left[\sum_{i=1}^{N} f_i(d_i - \bar{d})^2\right]$$

$$= \frac{1}{N}\left(\sum_{i=1}^{n} f_i d_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{n} f_i d_i\right)^2 \quad \text{where } N = \sum_{i=1}^{n} f_i$$

Sometimes $d_i = x_i - A$ are divisibly by a common number h(say)

then $u_i = \frac{x_i - A}{h}$, i = 1, 2, ..., N

then $\text{var}(x) = \frac{h^2}{N}\left[\sum_{i=1}^{N} f_i(u_i - \bar{u})^2\right]$

$$= h^2\left[\frac{1}{N}\sum_{i=1}^{n} f_i u_i^2 - \left(\frac{1}{N}\sum_{i=1}^{n} f_i u_i\right)^2\right] \quad \text{where } N = \sum_{i=1}^{n} f_i$$

(iii) Variance of a grouped or continuous frequency distribution : In a grouped or continuous frequency distribution, any of the formulae discussed in discrete frequency distribution can be used.

Note : Variance is independent of change of origin but dependent on change of scale. Adding or subtracting a positive number from each observation of a group does not affect the variance. If each observation is multiplied by a constant h then variance of the resulting group becomes h^2 times the original variance.

Note : While calculating S.D., the deviations are to be taken about arithmetic mean only.

Note : Coefficient of standard deviation $= \frac{\text{S. D.}}{\text{Mean}} = \frac{\sigma}{\bar{x}}$

COEFFICIENT OF VARIATION :

The mean deviation and standard deviation have the same units in which the data is given. Whenever we want to compare the variability of two series with data expressed in different units, we require measure of dispersion which is independent of the units. This measure is coefficient of variation (C.V.)

$$\text{C.V.} = \frac{\text{S.D}}{\text{Mean}} \times 100 = \frac{\sigma}{\bar{x}} \times 100$$

The series having greater C.V. is said to be more variable and less consistent than the other.

STANDARD DEVIATION OF COMBINED GROUP :

Let $\bar{x}_1, \bar{x}_2$ are A.M. and σ_1, σ_2 are S.D. of two groups having number of observations as n_1 and n_2 respectively then combined standard deviation σ of all the observations taken togther is given by

$$\sigma = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}} \quad \text{where } d_1 = \bar{x}_1 - \bar{x},$$

$d_2 = \bar{x}_2 - \bar{x}$ and $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Solved Examples

Ex.23 Find the mean and variance of first n natural numbers.

Sol. $\bar{x} = \frac{\sum x}{n} = \frac{1+2+3+.....+n}{n} = \frac{n+1}{2}$

$$\text{Variance} = \frac{\sum x^2}{n} - (\bar{x})^2$$

$$= \frac{1^2 + 2^2 + 3^2 + ... + n^2}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12}$$

Ex.24 Find the variance and standard deviation of the following frequency distribution :

Variable(x_i)	2	4	6	8	10	12	14	16
Frequency(f_i)	4	4	5	15	8	5	4	5

So. Calculation of variance and standard deviation

Variable x_i	Frequency f_i	$f_i x_i$	$x_i = \overline{X}$ $= x_i - 9$	$(x_i - \overline{X})^2$	$f_i(x_i - \overline{X})^2$
2	4	8	– 7	49	196
4	4	16	– 5	25	100
6	5	30	– 3	9	45
8	15	120	– 1	1	15
10	8	80	1	1	8
12	5	60	3	9	45
14	4	56	5	25	100
16	5	80	7	49	245
	$N = \sum f_i = 50$	$\sum f_i x_i = 450$			$\sum f_i(x_i - \overline{X})^2$ $= 754$

Here $N = 50, \Sigma f_i x_i = 450$

$$\therefore \overline{X} = \frac{1}{N}(\Sigma f_i x_i) = \frac{450}{50} = 9$$

We have $\Sigma f_i(x_i - \overline{X})^2 = 754$

$$\therefore \text{Var}(X) = \frac{1}{N}\left[\sum f_i(x_i - \overline{X})^2\right] = \frac{754}{50} = 15.08$$

$$\text{S.D.} = \sqrt{\text{Var}(X)} = \sqrt{15.08} = 3.88$$

Ex.25 Find the standard deviation (S.D.) of first n natural numbers.

Sol. Standard deviation

$$(\sigma) = \sqrt{\frac{1}{n}\Sigma x_i^2 - \left(\frac{1}{n}\Sigma x_i\right)^2} \quad \text{(individual series)}$$

$$= \sqrt{\frac{1}{n}\frac{n(n+1)(2n+1)}{6} - \frac{1}{n^2}\frac{n^2(n+1)^2}{4}} = \sqrt{\frac{n^2-1}{12}}$$

Ex.26 If $\sum_{i=1}^{18}(x_i - 8) = 9$ and $\sum_{i=1}^{18}(x_i - 8)^2 = 45$ then find the standard deviation of $x_1, x_2 ...... x_{18}$.

Sol. Let $d_i = x_i - 8$ but

$$\sigma_x^2 = \sigma_d^2 = \frac{1}{18}\sum d_i^2 - \left(\frac{1}{18}\sum d_i\right)^2$$

$$= \frac{1}{18} \times 45 - \left(\frac{9}{18}\right)^2 = \frac{5}{2} - \frac{1}{4} = \frac{9}{4}$$

Therefore $\sigma_x = 3/2$.

Ex.27 Find the coefficient of variance of first n natural numbers.

Sol. Coefficient of variance $= \frac{\sigma}{\overline{x}} \times 100$

$$= \sqrt{\frac{n^2-1}{12}} \times 100 \times \frac{1}{\frac{(n+1)}{2}} \quad = 100\sqrt{\frac{(n-1)}{3(n+1)}}$$

Ex.28 Determine the variance of the following distribution

x:	0-2	2-4	4-6	6-8	8-10	10-12
f:	2	7	12	19	9	1

Sol.

Class	**Mid value**	**f**	$u_i = \frac{x_i - 7}{2}$	**fu_i**	**fu_i^2**
0–2	1	2	–3	–6	18
2–4	3	7	–2	–14	28
4–6	5	12	–1	–12	12
6–8	7	19	0	0	0
8–10	9	9	1	9	9
10–12	11	1	2	2	4
Sum		50	---	–21	71

$$\sigma^2 = h^2\left[\frac{\Sigma f_i u_1^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2\right]$$

$$= 4\left[\frac{71}{50} - \left(\frac{-21}{50}\right)^2\right] = 4\ [1.42 - 0.1764] = 4.97$$

Ex.29 Calculate the mean and standard deviation for the following distribution :

Marks	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of Students	3	6	13	15	14	5	4

Sol. Calculation of Standard deviation

Class interval	Frequency f_i	Mid – values x_i	$u_i = \frac{x_i - 55}{10}$	$f_i u_i$	u_i^2	$f_i u_i^2$
20 – 30	3	25	– 3	– 9	9	27
30 – 40	6	35	– 2	– 12	4	24
40 – 50	13	45	– 1	– 13	1	13
50 – 60	15	55	0	0	0	0
60 – 70	14	65	1	14	1	14
70 – 80	5	75	2	10	4	20
80 – 90	4	85	3	12	9	36
	$N = \sum f_i = 60$			$\sum f_i u_i = 2$		$\sum f_i u_i^2 = 134$

Here $N = 60$, $\Sigma f_i u_i = 2$, $\Sigma f_i u_i^2 = 134$ and $h = 10$

$$\therefore \text{ Mean } = \bar{x} = A + h\left(\frac{1}{N}\sum f_i u_i\right)$$

$$\Rightarrow \bar{x} = 500 + 10\left(\frac{2}{60}\right) = 55.333$$

$$\text{Var}(X) = h^2\left[\frac{1}{N}\sum f_i u_i^2 - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$$

$$= 100\left[\frac{134}{60} - \left(\frac{2}{60}\right)^2\right] = 222.9$$

$$\therefore \text{ S.D.} = \sqrt{\text{Var}(X)} = \sqrt{222.9} = 14.94$$

Ex.30 Calculate the mean and standard deviation for the following data :

Wages upto (in Rs.)	No. of workers
15	12
30	30
45	65
60	107
75	157
90	202
105	222
120	230

Sol. We are given the cummulative frequency distribution. So first we will prepare the frequency distribution as given below :

Class Interval	Cummlative frequency	Mid – values	Frequency	$u_i = \frac{x_i - 67.5}{15}$	$f_i u_i$	$f_i u_i^2$
0 – 15	12	7.5	12	– 4	– 48	192
15 – 30	30	22.5	18	– 3	– 54	162
30 – 45	65	37.5	35	– 2	– 70	140
45 – 60	107	52.5	42	– 1	– 42	42
60 – 75	157	67.5	50	0	0	0
75 – 90	202	82.5	45	1	45	45
90 – 105	222	97.5	20	2	40	80
105 – 120	230	112.5	8	3	24	72
			$\sum f_i = 230$		$\sum f_i u_i = -105$	$\sum f_i u_i^2 = 733$

Here $A = 67.5, h = 15, N = 230, \Sigma f_i u_i = -105$ and $\Sigma f_i u_i^2 = 733$

$$\therefore \text{ Mean} = A + h\left(\frac{1}{N}\sum f_i u_i\right) = 67.5 + 15\left(\frac{-105}{230}\right)$$

$= 67.5 - 6.85 = 60.65$ and Var(X)

$$= h^2\left[\frac{1}{N}\sum f_i u_i^2 - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$$

$$\Rightarrow \text{Var}(X) = 225\left[\frac{733}{230} - \left(\frac{-105}{230}\right)^2\right]$$

$= 225[3.18 - 0.2025] = 669.9375$

$$\therefore \text{ S.D.} = \sqrt{\text{Var}(X)} = \sqrt{669.9375} = 25.883$$

Ex.31 Suppose that samples of polythene bags from two manufactures, A and B are tested by a prospective buyer for bursting presure, with the following results :

Bursting presure	**Number of bags manufactured by manufactured**	
in kg	**A**	**B**
5 -10	2	9
10 -15	9	11
15 - 20	29	18
20 -25	54	32
25 -30	11	27
30 -35	5	13

Which set of the bag has the highest average brusting pressure ? Which has more uniform pressure ?

Sol. For determining the set of bags having higher average bursting pressure, we compute mean and for finding out set of bags having more uniform pressure we compute coefficient of variation.

Manufacturer A :

Computation of mean and standard deviation

Bursting pressure	Mid – values x_i	f_i	$u_i = \frac{x_i - 17.5}{5}$	$f_i u_i$	$f_i u_i^2$
5 – 10	7.5	2	– 2	– 4	8
10 – 15	12.5	9	– 1	– 9	9
15 – 20	17.5	29	0	0	0
20 – 25	22.5	54	1	54	54
25 – 30	27.5	11	2	22	44
30 – 35	32.5	5	3	15	45
		$N = \sum f_i = 110$	$\sum u_i = 3$	$\sum f_i u_i = 78$	$\sum f_i u_i^2 = 160$

$$\overline{X}_A = a + h\left(\frac{\sum f_i u_i}{N}\right)$$

$$\Rightarrow \overline{X}_A = 17.5 + 5 \times \frac{78}{110}$$

$[\because\ h = 5, a = 17.5]$

$$\Rightarrow \overline{X}_A = 17.5 + 3.5 = 21$$

$$\sigma_A^2 = h^2\left[\frac{1}{N}\sum f_i u_i^2 - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$$

$$\Rightarrow \sigma_A^2 = 25\left[\frac{160}{110} - \left(\frac{78}{110}\right)^2\right]$$

$$\Rightarrow \sigma_A^2 = 25\left(\frac{17600 - 6084}{110 \times 110}\right) = 23.79$$

$$\Rightarrow \sigma_A = \sqrt{23.79} = 4.87$$

$$\therefore \text{ Coefficient of variation} = \frac{\sigma_A}{\overline{X}_A} \times 100$$

$$= \frac{4.87}{21} \times 100 = 23.19$$

Manufacurer B :

Bursting pressure	Mid – value x_i	f_i	$u_i = \frac{x_i - 17.5}{5}$	$f_i u_i$	$f_i u_i^2$
5 – 10	7.5	9	– 2	– 18	36
10 – 15	12.5	11	– 1	– 11	11
15 – 20	17.5	18	0	0	0
20 – 25	22.5	32	1	32	32
25 – 30	27.5	27	2	54	108
30 – 35	32.5	13	3	39	117
		$N = \sum f_i = 110$	$\sum u_i = 3$	$\sum f_i u_i = 96$	$\sum f_i u_i^2 = 304$

$$\overline{X}_B = a + h\left(\frac{\sum f_i u_i}{N}\right)$$

$$\Rightarrow \overline{X}_B = 17.5 + 5 \times \frac{96}{110} = 17.5 + 4.36 = 21.81$$

$$\sigma_B^2 = h^2\left[\frac{1}{N}\left(\sum f_i u_i^2\right) - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$$

$$\Rightarrow \sigma_B^2 = 25\left[\frac{304}{110} - \left(\frac{96}{110}\right)^2\right]$$

$$\Rightarrow \sigma_B^2 = 25\left(\frac{33440 - 9216}{110 \times 110}\right) = 50.04 \quad \sigma_B = 7.07$$

$$\therefore \text{ Coefficient of variation} = \frac{\sigma_B}{\overline{X}_B} \times 100$$

$$= \frac{7.07}{21.81} \times 100 = 32.41$$

We observe that the average brusting pressure is higher for manufacturer B. So, bags manufactured by B have higher bursting pressure.

The coefficient of variation is less for manufacturer A. So bags manufactured by A have more uniform pressure.

(i) Mathematical Properties of Variance

* If all values of the variate in a distribution are added (subtracted) by the same quantity (say λ), then the variance of the distribution remains unchanged. Hence

  Var $(X + \lambda)$ = Var (X)

* If all values of the variate in a distribution are multiplied by a constant number k, then the variance of the distribution is multiplied by k^2. Hence

  Var (kX) = k^2 var (X)

* From above results (i) and (ii) it is obvious that

  Var (aX + b) = a^2 Var (X)

* For a continuous distribution standard deviation is not less than the mean deviation with respect to AM.

* If AM's of two series containing n_1, n_2 values are m_1, m_2 and their variance's are σ_1^2, σ_2^2 respectively, then the variacne σ^2 of their combined series is given by

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2}(m_1 - m_2)^2$$

* $\sigma = \frac{3}{2}Q$ (approx.) mean deviation = $\frac{4}{5}\sigma$ (approx.).

(ii) Mean and Variance of Binomial Distribution

If the frequencies of the values 0, 1, 2,..... n of a variate are represented by the following coefficients of a binomial.

$q^n, {}^nc_1q^{n-1}p, {}^nc_2q^{n-2}p^2, \ldots\ldots, p^n$

where p is the probability of the success of the experiment (variate), q is the probability of its failure and p + q =1 i.e., distribution is a binomial distribution, then

$P(x = r) = {}^nc_r\, q^{n-r}\, p^r$

mean $\overline{x} = \Sigma p_i x_i = np$

variance $\sigma^2 = npq = \overline{x}\,q$

Solved Examples

Ex.32 If frequencies of the values 0, 1, 2, ..., n of a variate are represented by the following binomial coefficients $q^n, {}^nc_1q^{n-1}p, {}^nc_2q^{n-2}p^2, \ldots\ldots, p^n$ where p + q = 1, then find the mean of this series.

Sol. Here

$\Sigma f_i = q^n + {}^nc_1q^{n-1}p + {}^nc_2q^{n-2}p^2 + \ldots\ldots + p^n$

$= (q + p)^n = 1$ .....(1) and

$\Sigma f_i x_i = 0.\, q^n + 1.\, {}^nc_1q^{n-1}p + 2.{}^nc_2q^{n-2}p^2 + \ldots\ldots + n.p^n$

$$= np\left[q^{n-1} + (n-1)q^{n-2}p + \frac{(n-1)(n-2)}{1.2}q^{n-3}p^2 + \ldots\ldots + p^{n-1}\right]$$

$= np\,(q + p)^{n-1} = np \qquad [\because\ p + q = 1]$ .....(2)

$$\therefore \text{ mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{np}{1} = np$$