Module - 01

Skewness

* **Measures of central Tendency (OR Average):-**

According to professor Bowley, averages are "statistical constants which enable us to comprehend in a single effort the significance of the whole. They give us an idea about concentration of the values in the central part of the distribution. An average of a statistical series is the value of the variable which is representative of the entire distribution.

The Following are five measures of central tendency that are in common use.

i) Arithmetic mean or simple mean
ii) median
iii) mode
iv) Geometric mean
v) Harmonic mean

* **Requisites for an Ideal measure of central Tendency:-**

According to professor yule, the Following are the characteristics to be satisfied by an ideal measure of central tendency.

i) It should be rigidly defined.
ii) It should be readily comprehensible and easy to calculate.
iii) It should be based on all the observations.

iv) It should be suitable for further mathematical treatment. By this we mean that if we are given the averages and sizes of a number of series, we should be able to calculate the average of the composite series obtained on combining the given series.

v) It should be affected as little as possible by fluctuations of sampling.

* Arithmetic mean :-

Arithmetic mean of a set of observation is their sum divided by the number of observations.

The arithmetic mean denoted by $\bar{x}$ of n observations $x_1, x_2, x_3, \ldots x_n$ is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

$$\therefore \boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

For e.g. consider the height in cm of 5 students in a class

140, 153, 173, 170, 156

Let, $x_1 = 140$, $x_2 = 153$, $x_3 = 173$, $x_4 = 170$, $x_5 = 156$

Here $n = 5$

$$\therefore \text{Arithmetic mean } (\bar{x}) = \frac{\sum_{i=1}^{5} x_i}{n}$$

$$\therefore \quad \bar{x} = \frac{140+153+173+170+156}{5} = \frac{792}{5}$$

$$\therefore \quad \boxed{\bar{x} = 158.4}$$

* **Arithmetic mean for frequency distribution:-**

In case of the frequency distribution, $x_i | f_i$ where $i = 1, 2, 3, \ldots, n$ where f_i is the frequency of the variable x_i. Then arithmetic mean is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

$$\therefore \quad \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \longrightarrow (1)$$

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

$$\therefore (1) \Rightarrow \quad \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

$$\therefore \quad \boxed{\bar{x} = \frac{1}{N} \sum_{i=1}^{n} f_i x_i}$$

* **Note:-** In case of grouped or continuous frequency distribution, x is taken as the mid-point of the corresponding class.

Que:- Find the arithmetic mean of the following frequency distribution.

x	1	2	3	4	5	6	7
F	5	9	12	17	14	10	6

Answer:- The computation of mean is given in the following table.

x	F	Fx
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
Total	$N = \Sigma F = 73$	$\Sigma Fx = 299$

From above table,

$N = 73$ and $\Sigma Fx = 299$

$\therefore$ Arithmetic mean $(\bar{x}) = \frac{1}{N} \sum_{i=1}^{7} f_i x_i$

$\therefore \bar{x} = \frac{1}{73} \times 299 = \frac{299}{73}$

$\therefore \boxed{\bar{x} = 4.09}$

Que] calculate the arithmetic mean of the marks from the following table.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	12	18	27	20	17	6

Answer:- The calculation of mean is given in the following table.

Marks	No. of students (F)	midpoint (x)	Fx
0-10	12	5	60
10-20	18	15	270
20-30	27	25	675
30-40	20	35	700
40-50	17	45	765
50-60	6	55	330
Total	$N = \Sigma F = 100$		$\Sigma Fx = 2800$

From above table,

$N = \Sigma F = 100$

$\Sigma Fx = 2800$

$\therefore$ Arithmetic mean $(\bar{x}) = \frac{1}{N} \Sigma Fx$

$= \frac{1}{100} \times 2800$

$\boxed{\bar{x} = 28}$

Que 7 calculate the mean of the following frequency distribution-

Class Interval	0-8	8-16	16-24	24-32	32-40	40-48
Frequency	8	7	16	24	15	7

Answer:- The calculation of mean is given in following table.

Class Interval	Frequency (F)	midpoint (x)	Fx
0-8	8	4	32
8-16	7	12	84
16-24	16	20	320
24-32	24	28	672
32-40	15	36	540
40-48	7	44	308
Total	$N = \Sigma F = 77$		$\Sigma Fx = 1956$

From above table,

$N = \Sigma F = 77$

$\Sigma Fx = 1956$

$\therefore$ Arithmetic mean $(\bar{x}) = \frac{1}{N} \Sigma Fx$

$= \frac{1}{77} \times 1956$

$= \frac{1956}{77}$

$\boxed{\bar{x} = 24.40}$

* **median:-** Median of a distribution is the value of the variable which divides it into two equal parts. It is the value which exceeds and is exceeded by the same number of observations. That means, it is the value such that the number of observations above it is equal to the number of observations below it. The median is thus a potential average.

① In case of ungrouped data, if the number of observations is odd then median is the middle value after the values have been arranged in ascending or descending order of magnitude.

For e.g. consider the following observation-

7, 3, 5, 9, 8, 13, 12

First arrange the data in ascending order.

∴ 3, 5, 7, 8, 9, 12, 13

The middle term is 8.

∴ median (Md) = 8

② In case of even number of observations, there are two middle terms and median is obtain by taking arithmetic mean of middle terms.

For e.g. 8, 6, 10, 12, 5, 9

First arrange the data in ascending order.

∴ 5, 6, $\underline{8, 9}$, 10, 12

The middle terms are 8 and 9.

$$\therefore \text{median} = \frac{8+9}{2} = \frac{17}{2} = 8.5$$

* Median for Discrete Frequency Distribution –

In case of discrete frequency distribution median is obtained by considering the cumulative frequencies.

The steps for calculating median are given below –

1. Find $\frac{1}{2}N$, where $N = \Sigma F$
2. See the cumulative frequency just greater than $\frac{1}{2}N$.
3. The corresponding value of x is median.

Que] 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

Answer:- The computation of median is

x	F	C.F.
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
Total	$N = \Sigma F = 120$	

Here, $N = 120$

consider,

$\frac{1}{2}N = \frac{1}{2} \times 120 = 60$

$\therefore$ The cumulative frequency just greater than 60 is 65 and it's corresponding x value is 5.

$\therefore$ $\boxed{\text{median}(md) = 5}$

* median For continuous Frequency Distribution:-

In case of continuous Frequency Distribution, we perform the following steps-

① Determine the cumulative frequency of each class interval.

② calculate $\frac{1}{2}N$, where $N=\Sigma F$

③ Find the cumulative Frequency (CF) just greater than $\frac{1}{2}N$.

④ The class corresponding to c.F. just greater than $\frac{1}{2}N$ is called median class.

⑤ The median is obtained by the formula-

$$\boxed{\text{median}(md) = l + \frac{h}{F}\left(\frac{N}{2} - c\right)}$$

where,
l is the lower limit of the median class.
h is the magnitude of the median class.
F is the Frequency of the median class.
c is the c.F. of the class preceding (i.e. above) the median class.

* Note:- ① h = common magnitude

h = upper limit of class − lower limit of class

② value of median is always between lower limit and upper limit of class interval.

Que] Find the median salary of the following distribution -

Salary (in Rs.)	2000-3000	3000-4000	4000-5000	5000-6000	6000-7000
No. of employees	3	5	20	10	5

Answer:- The computation of median is

Salary (in Rs.)	No. of employees (F)	C.F.
2000-3000	3	3
3000-4000	5	8
4000-5000	20	28
5000-6000	10	38
6000-7000	5	43
Total	$N = \Sigma F = 43$	

Here, $N = 43$

$\therefore \frac{1}{2}N = \frac{1}{2} \times 43 = 21.5$

$\therefore$ cumulative frequency just greater than 21.5 is 28 and it's corresponding class interval is 4000-5000.

$\therefore$ 4000-5000 is median class.

$\therefore$ median (md) $= l + \frac{h}{F}\left(\frac{N}{2} - c\right) \rightarrow (1)$

Here, $l = 4000$, $h = 5000 - 4000 = 1000$

$F = 20$, $\frac{N}{2} = 21.5$, $c = 8$

(1) $\Rightarrow$ median (md) $= 4000 + \frac{1000}{20}(21.5 - 8)$

$= 4000 + 50(13.5)$

$= 4000 + 675 = 4,675$

$\therefore$ $\boxed{\text{median (md)} = 4,675}$

Que] Find the median from the following data.

Work-hours	3-6	7-10	11-14	15-18
No. of employees	150	580	900	1370

Answer:- In a given data, the frequency of class interval is not continuous. We make it continuous by taking boundary limit.

Work-Hours	Boundaries	No. of employees (F)	C.F.
3-6	2.5-6.5	150	150
7-10	6.5-10.5	580	730
11-14	10.5-14.5	900	1630
15-18	14.5-18.5	1370	3000
Total		$N = \Sigma F = 3000$	

Here, $N = 3000 \Rightarrow \frac{1}{2}N = \frac{1}{2} \times 3000 = 1500$

$\therefore$ c.f. just greater than 1500 is 1630 and it's corresponding class interval (Here take Boundaries) 10.5-14.5 is median class.

$\therefore$ $\text{median}(md) = l + \frac{h}{F}\left(\frac{N}{2} - c\right) \rightarrow (1)$

Here $l = 10.5$, $h = 14.5 - 10.5 = 4$, $F = 900$,

$\frac{N}{2} = 1500$, $c = 730$

$\therefore (1) \Rightarrow$ $\text{median}(md) = 10.5 + \frac{4}{900}(1500 - 730)$

$= 10.5 + \frac{4}{900}(770)$

$= 10.5 + \frac{3080}{900}$

$= 10.5 + 3.42$

$= 13.92$

$\therefore$ $\boxed{\text{median}(md) = 13.92}$