(*) Descriptive statistics.

↓

Numerical summaries of data.

↓

Location	Shape
Measures of central tendency. (mean, median, mode, quartile, percentile.)	Variance, standard deviation, Range.

e.x (1) −10, 0, 10, 20, 30.

$$\text{mean} = \frac{-10+0+10+20+30}{5}$$

$$\boxed{\text{mean} = 10.}$$

Range: $30-(-10) = 40$.

variance

$$= \sigma^2 = \frac{(-10-10)^2+(0-10)^2+(10-10)^2+(20-10)^2+(30-10)^2}{5}$$

$$= \frac{400+100+0+100+400}{4}$$

$$= \frac{1000}{4} =$$

(2). 8, 9, 10, 11, 12.

$$\text{mean} = \frac{8+9+10+11+12}{5}$$

$$\boxed{\text{mean} = 10}$$

Range: $12-8 = 4$

$$\sigma^2 = \frac{(8-10)^2+(9-10)^2+(10-10)^2+(11-10)^2+(12-10)^2}{5}$$

$$= \frac{4+1+0+1+4}{4}$$

$$= 10/4 =$$

(1) is more disperce than (2).

✱ Measures of Central tendency ✱.

→ **Requisities for an ideal measure of central tendency [char. of a good average]:→**

According to proffessor Yule ; the following are the characteristic to be satisfied by an ideal measure of central tendency.

- It should be rigidly defined.
- It should be easy to understand and calculate.
- It should be based on all the observations
- It should be suitable for further mathematical treatment.
- It should be affected as little as possible by fluctuations of sampling.
- It should not be affected much by extreme values.

✱ Various measures of central tendency :→

Measures of central tendency gives us an ideal about the 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.

According to professor Bowley "Averages are statistical constants which enable us to comprehend in a single effort the significance of the whole."

The following are the five measure of central tendency which are commonly used in Practice.

<1> Arithmetic mean.
<2> Median
<3> Mode.
<4> Geometric mean
<5> Harmonic mean.

<1> ARITHMETIC MEAN :→

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

For ex: the arithmetic mean $\bar{x}$ of n-observation $x_1, x_2, x_3, \ldots x_n$ is given by:

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

→ In case of frequency distribution x_i/f_i; where f_i is the frequency of the variable x_i ;

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

$$= \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$\therefore \boxed{\bar{x} = \frac{\Sigma f_i x_i}{N}} \quad \text{where } N = \Sigma f_i \quad \text{---} \; (*)$$

→ In case of grouped or continuous freq. distribution ; x is taken as the mid-value of the corresponding class.

→ If the values of x and/or f are large; the calculation of mean by above formula (*) is quite tedious and time-consuming. In such a case the calculations can be reduced by using the step-deviation method which consists in taking the deviations (differenc of the given observations from any arbitrary value A.

Let $d = x - A$ then ;

$$\boxed{\bar{x} = A + \frac{\Sigma f_i d_i}{N}} ; \text{ where } \begin{cases} A = \text{Assumed mean} \\ d_i = x_i - A \\ N = \Sigma f_i \end{cases}$$

→ In case of grouped or continuous frequency distribution with class intervals of equal magnitude; the calculations are further simplified by taking:

$$d = \frac{x - A}{h} ;$$

where x is the mid value of the class and h is the common magnitude of the class intervals.

Here the arithmetic mean is:

$$\boxed{\bar{X} = A + h \cdot \frac{\Sigma f_i d_i}{N}}$$

where A = Assumed mean

$d_i = x_i - A / h$

h = width of class-interval

$N = \Sigma f_i$

Ex-7 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

Sol^n :→

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

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{2800}{100} = 28.$$

Ex-10 Find the average marks of students from the following table:

Marks	No. of students.
Above 0	80
Above 10	77
Above 20	72
Above 30	65
Above 40	55
Above 50	43
Above 60	28
Above 70	16
Above 80	10
Above 90	8
Above 100	0

Sol^n :→ Here Let $A = 55$; $h = 10$.

Marks.	Mid values (x)	Frequency (f)	$u = \frac{x-A}{h}$	$f_i u_i$
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
		$N = 80$		$\Sigma f_i u_i = -26$

$$\bar{x} = A + h \cdot \frac{\Sigma f_i u_i}{N}$$

$$= 55 + \frac{10(-26)}{80}$$

$$= 55 - 3.25$$

$$= 51.75 \text{ marks.}$$

Ex-11 Find the mean from the following data.

Marks	No. of students	Marks	No. of students.
below 10	5	below 60	60
below 20	9	below 70	70
below 30	17	below 80	78
below 40	29	below 90	83
below 50	45	below 100	85

Solⁿ:→ Here $h = 10$; $A = 55$

Marks.	Mid value (x)	frequency (f)	$u = \frac{x-A}{h}$	$f_i u_i$
0-10	5	5	-5	-25
10-20	15	4	-4	-16
20-30	25	8	-3	-24
30-40	35	12	-2	-24
40-50	45	16	-1	-16
50-60	55	15	0	0
60-70	65	10	1	10
70-80	75	8	2	16
80-90	85	5	3	15
90-100	95	2	4	8
		$N = 85$		$\Sigma f_i u_i = -56$

$$\bar{X} = A + h \cdot \frac{\sum f_i u_i}{N}$$

$$= 55 + \frac{10(-56)}{85}$$

$$= 55 - \frac{112}{17}$$

$$= 55 - 6.59$$

$$= 48.41 \text{ marks.}$$

* PROPERTIES OF ARITHMETIC MEAN :

PROP :- 1 :→

The algebraic sum of the deviations of the given set of observations from their Arithmetic mean is zero.

→ We know $\bar{x} = \frac{\Sigma fx}{N}$ $\Rightarrow$ $\Sigma fx = N \cdot \bar{x}$ ---- (*)

Now

$$\Sigma f(x-\bar{x}) = \Sigma (fx - f\bar{x})$$
$$= \Sigma fx - \Sigma f\bar{x}$$
$$= \Sigma fx - \bar{x}\,\Sigma f$$
$$= \Sigma fx - \bar{x}.N$$
$$= \Sigma fx - \Sigma fx \quad (\because \text{from } (*))$$
$$= 0$$

Hence the results.

PROP : 2 :→

The sum of the squares of deviations of the given set of observations is minimum when taken from the arithmetic mean.

ie For a given frequency distribution ; the sum $S = \Sigma f(x-A)^2$; which represents the sum of the squares of deviations of given observations from any arbitrary value 'A' is minimum when $A = \bar{x}$

→ Let the frequency distribution be x_i/f_i
Let s be the sum of the squares of the deviations of the given values from any arbitrary point A (say)

$$\therefore \; S = \Sigma f(x-A)^2 \quad \text{------} \quad (*)$$

We have to show that : S is minimum when $A = \bar{x}$

Now s is minimum when

$$\frac{ds}{dA} = 0 \quad \text{and} \quad \frac{d^2s}{dA^2} > 0$$

differentiate eqn $(*)$; w.r.t. A ;

$$\frac{ds}{dA} = \Sigma f \; 2(x-A)(-1) = -2\Sigma f(x-A)$$

$$\text{------} \quad (**)$$

$$\frac{ds}{dA} = 0 \Rightarrow \Sigma f(x-A) = 0$$

$$\Rightarrow \Sigma fx - \Sigma fA = 0$$

$$\Rightarrow \Sigma fx - A\Sigma f = 0$$

$$\Rightarrow A = \frac{\Sigma fx}{\Sigma f} = \bar{x}$$

Also; diff. eqn $(**)$; w.r.t. A ;

$$\frac{d^2s}{dA^2} = -2\Sigma f(-1) = 2\Sigma f = 2 \cdot N > 0$$

since total frequency is always positive.

Hence; $A = \bar{x}$ provides a minimum of S.

PROP: 3 :→ [Mean of combined series]

If $\bar{x}_1$ is the arithmetic mean of n_1 observations and $\bar{x}_2$ is the arithmetic mean of n_2 other observations then Prove that the combined arithmetic mean of these (n_1+n_2) observations is

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

→ We know that If $\bar{x}$ is the mean of n observations then,

$$\bar{x} = \frac{\Sigma x}{n} \Rightarrow \Sigma x = n.\bar{x}$$

i.e Sum of n-observations $=(n)$(arithmetic mean). ——— (*)

If $\bar{x}_1$ is the mean of n_1 observations and $\bar{x}_2$ is the mean of n_2 other observations; then by using (*) we get,

The sum of n_1 observations $= n_1 \cdot \bar{x}_1$

The sum of n_2 observations $= n_2 \cdot \bar{x}_2$

$\therefore$ The sum of (n_1+n_2) observations $= n_1\bar{x}_1 + n_2\bar{x}_2$

$\therefore$ The mean $\bar{x}$ of the (n_1+n_2) observations is

$$\bar{x} = \frac{\text{sum of } (n_1+n_2) \text{ observations}}{n_1+n_2}$$

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

Generalization :→ In general ; If $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_k$ are the arithmetic means of k-series with $n_1, n_2, n_3, \ldots n_k$ observations respectively then the mean $\bar{x}$ of the combined series of size $(n_1+n_2+\ldots+n_k)$ is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1+n_2+n_3+\ldots+n_k}$$

Ex-22 The mean wage of 200 workers in a factory is Rs. 50 The mean wage of 75 workers of the first shift is Rs. 60. Find the mean wage of the rest.

Solⁿ :→

Here no. of workers in first shift $(n_1) = 75$

" " " other " $(n_2) = 200 - 75$
$= 125$

Mean wage of workers in first shift $(\bar{x}_1)$ = Rs. 60
" " " all the workers $(\bar{x})$ = Rs. 50

Now :

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

$$\therefore \quad 50 = \frac{(75 \times 60) + (125 . \bar{x}_2)}{200}$$

$$\Rightarrow \quad 10000 - 4500 = 125 . \bar{x}_2$$

$$\Rightarrow \quad \bar{x}_2 = \frac{5500}{125}$$

$$\Rightarrow \quad \boxed{\bar{x}_2 = 44 \text{ Rs.}}$$

⟨2⟩ MEDIAN :→

The Median is the central value of the variable which divides the group in two equal parts ; one part comprising all the values greater and the other all values less than median.

Thus Median of a distribution may defined as that value of the variable which exceeds and is exceeded by the same number of observations. i.e it is the value such that the number of observations above it is equal to the number of observations below it.
Thus, arithmetic mean is based on all the items of the distribution while the median is only positional average i.e its value depends on the position occupied by a value in the frequency distribution :

Merits & De-Merits of Median :

(*). Merits :→

⟨1⟩ It is rigidly defined.
⟨2⟩ It is easily understood and easy to calculate.
⟨3⟩. It can be calculated for distribution with open - end classes.

<4> Median is the only average to be used while dealing with qualitative characteristic which cannot be measured quantitatively but can still be arranged in Ascending or descending order of magnitude

for ex. To find the average beauty, average honesty, average intelligency etc.

<5> It is not affected by extreme values.
<6> The value of the median can be located graphically.

* De - Merits :→

<1> In case of even number of observations for an ungrouped data; Median cannot be determined exactly. We merely estimate it as the arithmetic mean of the two middle terms.
<2> Median, being a positional average; is not based on all the items of the distribution
<3> The median does not lend itself to algebraic treatment.
<4> Median is not suitable for further mathematical treatment i.e given the sizes and the median values of different groups we cannot compute the median of the combined group.
<5> As compared with mean; it is affected much by fluctuations of sampling.

Calculation of Median :

* For ungrouped data :→

If the number of observations n is odd then the median is the middle value after the observations have been arranged in ascending or descending order of magnitude.

When n is even; there are two middle values $n/2$ and $(n+1)/2$ then the median is the arithmetic mean of these middle values after they are arranged in ascending or descending order.

* For discrete frequency distribution :→

In case of discrete frequency distribution Median is obtained by considering the cumulative frequencies the steps for calculating the median are given below.

<1> Find $N/2$; where $N = \Sigma f_i$

<2> See the (less than) cumulative frequencies just greater than $N/2$ (i.e $(N+1)/2$)

<3> The corresponding value of x is median.

(*) · For a grouped (continuous) frequency distribution

The Median is given by the formula.

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

Median formula

where; l = lower limit of the median class; where median class is the class corresponding to cumulative - frequencies just greater than $N/2$

h = The width of median class.

f = The frequency of the median class.

N = Σf ; the total frequency.

c = Cumulative frequency of the class preceding the median class.

Ex-6 :→ According to the census of 1981; following are the population figures; in thousands of 10 cities :

2000, 1180, 1785, 1500, 560, 782, 1200, 385, 1123, 222.

Find the median.

Soln :→ Arranging data in Ascending order.

222, 385, 560, 782, 1123, 1180, 1200, 1500, 1785, 2000

Here $n = 10$ (even)

$\therefore$ Median $=$ A.M. of size of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n+1}{2}\right)^{th}$ item.

$=$ A.M of 5^{th} and 6^{th} items.

$$= \frac{1123 + 1180}{2}$$

$=$ 1151.5 thousands.

Ex-7 :→ 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 :→ The commutative frequency table 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

Here $N = \Sigma f = 120$

Odd data is given so, using formula for median for ODD data :

$$\therefore \frac{N+1}{2} = 60.5$$

Cumulative frequency just greater than $\left(\frac{N+1}{2}\right)$ is 65 and the value of x correspon-ding to c.f. 65 is 5.

$\therefore$ Median = 5.

Ex-8 Find the median from the following data:

Marks	No. of students.		Marks.	No. of students.
0 – 10	2		40–50	35
10 – 20	18		50 – 60	20
20 – 30	30		60 – 70	6
30 – 40	45		70 – 80	3.

Sol^n: Cumulative frequency table is.

Marks.	No. of students (f)	C.F.	
0 – 10	2	2	
10 – 20	18	20	
20 – 30	30	50	c
L 30 – 40	45 f	95	
40 – 50	35	130	
50 – 60	20	150	
60 – 70	6	156	
70 – 80	3	159.	

Here $N = \Sigma f = 159$.

∴ Median class in the class corresponding to c.f. just greater than $N/2 = 79.5$

∴ 30 - 40 is the midian class.

Here ; l = lower limit of midian class = 30
h = width of midian class = 40 - 30 = 10
f = frequency of midian class = 45
c = c.f. of the class preceding the median class = 50.

$$\therefore \text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$$= 30 + \frac{10}{45}[79.5 - 50]$$

$$= 30 + 6.56$$

$$= 36.56.$$

Ex-9. The following table shows the age distribution of persons in a particular region.

Age (years)	No. of persons (in thousands)	Age (years)	No. of persons (in thousands.)
Below 10	2	Below 50	14
" 20	5	" 60	15
" 30	9	" 70	15.5
" 40	12	70 and over	15.6

Find the median age.

Sol:→

Age (Years)	No. of Persons. (f)	C.f.
0 - 10	2	2
10 - 20	5 - 2 = 3	5
$\boxed{20 - 30}$	$\boxed{9 - 5 = 4}$	$\boxed{9}$
30 - 40	12 - 9 = 3	12
40 - 50	14 - 12 = 2	14
50 - 60	15 - 14 = 1	15
60 - 70	15.5 - 15 = 0.5	15.5
70 and over	15.6 - 15.5 = 0.1	15.6

Here $N/2 = \frac{\Sigma f}{2} = \frac{15.6}{2} = 7.8$

Cumulative freq. greater than 7.8 is 9.

∴ corr. class 20 - 30 is median class.

$$\therefore \text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$$= 20 + \frac{10}{4}\left[7.8 - 5\right]$$

$$= 20 + 7$$

$$= 27.$$

Ex-10 Calculate Mean and Median from the following:

Class intervals	frequency	Class intervals	frequency
6.5 - 7.5	5	10.5 - 11.5	32
7.5 - 8.5	12	11.5 - 12.5	6
8.5 - 9.5	25	12.5 - 13.5	1
9.5 - 10.5	48.		