Notes of Statistics

2nd Sem

Measures of Dispersion

The measures of dispersion shows the scatterings of data. It tells the variation of the data from one another and gives a clear idea about the distribution of the data. The measures of dispersion shows the homogeneity or the hetrogeneity of the distribution of the observations.

Characteristics of Measures of dispersion :-

- A measures of dispersion should be rigidly defined.
- It must be easy to calculate and understand.
- Not affected much by the fluctuations of observations.
- Based on all observations.

Types of Measures of Dispersion :- There are two types of measures of variation or Dispersion as follows -

1) **Absolute Measures of Dispersion** - The measures of dispersion which is expressed in terms of the units of the observations (e.g. Rupees, Meter, Years etc) is called absolute measure.

There are certain measures namely :-

(i) Range

(ii) Quartile Deviation

(iii) Mean Deviation

(iv) Standard Deviation

2) Relative Measures of Dispersion :- The relative measures of variation / dispersion are used to compare the distribution of two or more data sets which is is independent of unit of observations.

Common relative dispersion methods include :

(i) Coefficient of Range
(ii) coefficient of Quartile Deviation
(iii) coefficient of Mean Deviation
(iv) coefficient of Variation
(v) coefficient of Standard Deviation

Range :- Range is the most common and easily understandable measure of dispersion. Range is the difference between the highest and the lowest value in a distribution. It is denoted by R.

Thus, Range = Largest value − Smallest value

or $\boxed{R = L - S}$

where L = largest value & S = smallest value

Merits of Range :-

- Range is the simplest measures of dispersion.
- It is simple to understand and easy to calculate.
- It is rigidly defined.

Demerits of Range :-

- It is based on two extreme observations. Hence, get affected by fluctuations.
- The range provides no information about the structures of the series.
- It is not a reliable measure of dispersion.

Coefficient of Range :- To compare the series, the relative measure of variation is defined as –

$$\text{coefficient of Range (C.R.)} = \frac{\text{Largest value} - \text{smallest value}}{\text{Largest value} + \text{smallest value}}$$

or

$$\boxed{C.R. = \frac{L-S}{L+S}}$$

where L = Largest value

& S = smallest value

Methods of Calculation of Range :-

(i) Calculation of Range for Individual Series :-

In the individual series, range is calculated as the difference between the highest and lowest value of the series.

i.e. $\boxed{\text{Range } (R) = L - S}$

where L = largest value & S = smallest value

Example - Find out range and coefficient of range of following data -

10, 20, 30, 40, 80, 5, 100

Sol Here, L = 100 , S = 5

then R = 100 - 5 = 95 units

$$C.R. = \frac{L-S}{L+S} = \frac{100-5}{100+5} = \frac{95}{105} = 0.904$$

(ii) Calculation of Range for Discrete Series :-

Range of the discrete series is also determined as the difference between the largest and smallest value of the series. Frequency of the series is not taken into account.

So, $\boxed{\text{Range } (R) = L - S}$

Example - Find range and coefficient of range of the given data.

Data	5	10	25	30	50
Frequency	1	5	3	9	4

Sol - Here, Largest value $L = 50$
Smallest value $S = 5$

$\therefore$ Range $= L - S = 50 - 5 = 45$ units

and coefficient of Range $= \dfrac{L-S}{L+S} = \dfrac{50-5}{50+5} = \dfrac{45}{55} = 0.818$

(iii) Calculations of Range for Continuous Series :-

Range for a continuous series can be obtained by subtracting the lower limit of lowest class interval from the upper limit of the highest class interval.

Example - Find the range and its coefficient of the following frequency distribution.

Age (in years)	15-20	20-25	25-30	30-35
No. of Persons	10	15	20	7

Sol - Here, $L = 35$, $S = 15$

$\therefore$ $R = L - S = 35 - 15 = 20$ years

& $C.R. = \frac{L-S}{L+S} = \frac{35-15}{35+15} = \frac{20}{50} = 0.4$

Quartile Deviation :- Quartiles are the points which divide the data set into four equal parts. Q_1 gives the value of the item $\frac{1}{4}$th the way up the distribution and Q_3 are value of the item $\frac{3}{4}$th the way of distribution.

Quartile deviation or semi-inter-quartile deviation is defined as -

$$\boxed{\text{Quartile deviation } (Q.D.) = \frac{Q_3 - Q_1}{2}}$$

Coefficient of Quartile Deviation :- The relative measures of quartile deviation are determine as -

$$\boxed{\text{coefficient of Quartile Deviation } (C.Q.D) = \frac{Q_3 - Q_1}{Q_3 + Q_3}}$$

Advantages of Quartile Deviation :-

- It is simple to understand and easy to compute.
- It is not influenced by the extreme values.
- It can be found out with open-end distribution.

Disadvantages of Quartile Deviation :-

- It ignores the 25% of the items and the last 25% of the items.
- It is positional average, hence it is not used to further mathematical treatement.
- It gives only a rough measure, that means not a reliable measures of dispersion.

Methods of Calculations of Quartile Deviation :-

(i) Computation of Quartile Deviation for Individual Series -

For individual series, Q_1 and Q_3 are calculated by using following formula -

$$Q_1 = \left(\frac{N+1}{4}\right) \text{ th Value}$$

&

$$Q_3 = 3\frac{(N+1)}{4} \text{ th Value}$$

First arrange the given data in ascending or decending order.

where, $N =$ Number of observations or data.

So, $$\boxed{Q.D = \frac{Q_3 - Q_1}{2}}$$

Example - From the following data, find Q.D.

10, 20, 30, 40, 50, 60, 70

Sol The values are already arranged in ascending order. Here N = 7

$$Q_1 = \frac{N+1}{4} \text{ th item} = \frac{7+1}{4} = 2\text{nd value} = 20$$

$$Q_3 = \frac{3(N+1)}{4} \text{ th item} = \frac{3(7+1)}{4} \text{ th value} = 6\text{th value} = 60$$

$$\therefore \quad Q.D = \frac{Q_3 - Q_1}{2} = \frac{60-20}{2} = 20 \text{ units}$$

and. $$C.Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{60-20}{60+20} = \frac{40}{80} = 0.5$$

(ii) Computation of Q.D. for Discrete Series :-

To compute quartile deviation we follow the following steps -

1) First we find out cumulative frequency (c.f.) of given data.

2) Now calculate ~~total~~ frequency which is denoted by N i.e. $N = \Sigma f$

3) Apply the formula for Q_1 and Q_3 as -

$$Q_1 = \frac{(N+1)}{4} \text{ th value}$$

& $$Q_3 = \frac{3(N+1)}{4} \text{ th value}$$

Choose the cumulative frequency which is closed to the value of Q_1 and Q_3 and corresponding data is called Q_1 and Q_3.

Finally,

$$\boxed{Q.D. = \frac{Q_3 - Q_1}{2}}$$

Example - Find out quartile deviation and its coefficient -

X	5	15	20	50	70	82
f	8	12	20	10	7	3

Sol

X	f	C.f.
5	8	8
15	12	20
20	20	40
50	10	50
70	7	57
82	3	60

$N = \Sigma f = 60$

Now,

$Q_1 = \frac{N+1}{4}$ th value $= \frac{60+1}{4}$ th value $= 15.25$ th value
$= 15$

& $Q_3 = \frac{3(N+1)}{4}$ th value $= \frac{3(60+1)}{4}$ th value $= 45.75$ th value
$= 50$

Scanned with CamScanner

So,
$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{50-15}{2} = 17.5 \text{ unit}$$

&
$$C.Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{50-15}{50+15} = 0.538$$

(iii) Calculation of Quartile Deviation for Continuous series –

It includes the following steps –

1) Calculate cumulative frequency of each class interval and also determine total frequency which is expressed by $N = \Sigma f$ = total frequeny.

2) Apply initial formula to determine Q_1 and Q_3 as –

$$Q_1 = \frac{N}{4}\text{th value} \quad \& \quad Q_3 = \frac{3N}{4}\text{th value}$$

3) select the cumulative frequency which is near to the calculated value of Q_1 and Q_3 and choose the corresponding class interval.

4) Now, finally apply the formula –

$$Q_1 = L_1 + \frac{\frac{N}{4} - C}{f}(L_2 - L_1)$$

and

$$Q_3 = L_1 + \frac{\frac{3N}{4} - C}{f}(L_2 - L_1)$$

where,
L_1 = lower limit of selected class
L_2 = upper limit of selected class
f = frequency of class

and C = cumulative frequency of preceding class

Thus,

$$\boxed{Q.D = \frac{Q_3 - Q_1}{2}}$$

Example - Calculate Q.D. from the following data -

Marks	0-20	20-40	40-60	60-80	80-100
No. of students	7	10	30	42	11

Sol

Marks (X)	No of students (f)	c.f
0-20	7	7
20-40	10	17
40-60	30	47
60-80	42	89
80-100	11	100
	$N = \Sigma f = 100$	

$Q_1 = \frac{N}{4}$ th value $= \frac{100}{4}$ th value $=$ 25 th value

40-60 class interval are selected.

Here, $L_1 = 40$, $L_2 = 60$, $f = 30$, $c = 17$

$$\therefore \quad Q_1 = L_1 + \frac{\frac{N}{4} - c}{f}(L_2 - L_1)$$

$$= 40 + \frac{25 - 17}{30}(60 - 40)$$

$$= 40 + 5.33 = 45.33$$

$Q_3 = \frac{3N}{4}$ th value $= \frac{3 \times 100}{4}$ th value $=$ 75 th value

So, 60 − 80 class interval are selected.

Here, $L_1 = 60$, $L_2 = 80$, $f = 42$, $c = 47$

$\therefore$

$$Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f}(L_2 - L_1)$$

$$= 60 + \frac{75 - 47}{42}(80 - 60)$$

$$= 60 + 13.33 = 73.33$$

Thus,

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$= \frac{73.33 - 45.33}{2} = 14 \text{ units}$$

& $$C.Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{73.33 - 45.33}{73.33 + 45.33} = 0.235$$

Mean Deviation :— Mean deviation is also known as 'first moment of dispersion.'

Mean deviation of a series is the arithmetic average of the deviations of various items from the median or mean of the series.

It is denoted by M.D.

Note - Median is preferred since the sum of the deviations from the median is less than that from the mean.

Formula for Calculating Mean Deviation

(i) $$M.D.\ (\text{about mean}) = \frac{\Sigma|X-\overline{X}|}{N} \quad \text{or} \quad \frac{\Sigma|dx|}{N}$$

(ii) $$M.D.\ (\text{about median}) = \frac{\Sigma|X-M|}{N} \quad \text{or} \quad \frac{\Sigma|dx|}{N}$$

where, $\overline{X}$ = mean, M = median and N = number of observation

Coefficient of Mean Deviation :-

(i) $$\text{coefficient of M.D. from mean} = \frac{M.D.\ (\text{about mean})}{\text{mean}}$$

(ii) $$\text{coefficient of M.D. from median} = \frac{M.D.\ (\text{about median})}{\text{median}}$$

Advantages of Mean Deviation :-

- It is based on all the observations and thus it is definitely a better measure of dispersion.
- Mean deviation is rigidly defined.
- It provides a minimum value when the deviations are taken from the medium.

Disadvantage of Mean Deviation :-

- Not easily understandable.
- Its calculation is not easy and time-consuming.
- Ignorance of negative sign creates artificiality and become useless for further mathematical treatment.

Methods of Calculations of Mean Deviation :-

1) Computation of Mean Deviation - Individual Series :-

The process of computing M.D. involves the following steps-

i) Calculate the average (mean, median) of the series.
ii) Take the deviations (or difference) of each data items from average (ignore -ve sign) which is denoted by $|dx|$
(iii) Compute the total of these deviations i.e. $\Sigma|dx|$
(iv) Now apply the formula -

$$\boxed{\text{Mean Deviation (M.D.)} = \frac{\Sigma|dx|}{N}}$$

where N is the number of observations.

Example - Find the mean deviation of the following data-

10, 20, 30, 40, 50

Sol.

mean $\overline{X} = \frac{\Sigma X}{N} = \frac{150}{5} = 30$

& median $M = \frac{N+1}{2}$ th value $= \frac{5+1}{2}$ th value

$= 3$rd value $= 30$

Table for M.D. about mean and median

X	$\lvert X-\overline{X}\rvert$	$\lvert X-M\rvert$
10	20	20
20	10	10
30	0	0
40	10	10
50	20	20
	$\Sigma\lvert X-\overline{X}\rvert = 60$	$\Sigma\lvert X-M\rvert = 60$

M.D. (about mean) $= \frac{\Sigma|X-\overline{X}|}{N} = \frac{60}{5} = 12$

Coefficient of M.D. (about mean) $= \frac{M.D}{mean} = \frac{12}{30} = 0.4$

M.D. (about median) $= \frac{\Sigma|X-M|}{N} = \frac{60}{5} = 12$

coefficient of M.D. (about median) $= \frac{M.D.}{median} = \frac{12}{30} = 0.4$

2) Computation of Mean Deviation – Discrete Series –

It involves the following steps –

(i) Find out an average (mean, median)

ii) Find out the deviation of each data items from the central tendency (ignore -ve sign) & denote them by $|dx|$

iii) Multiply each deviation by its corresponding frequency i.e. $f \cdot |dx|$ and find out the total $\Sigma f \cdot |dx|$ & total frequency (Σf)

iv) Now apply the formula for mean deviation

$$\boxed{M.D. = \frac{\Sigma f \cdot |dx|}{\Sigma f}}$$

Example - Compute mean deviation & its coefficient from following data.

marks	5	10	15	20	25
frequency	6	7	8	11	8

Sol

Mean deviation about mean

marks (X)	f	X·f	$\lvert dx\rvert = \lvert X-\bar{X}\rvert$	$f \cdot \lvert dx\rvert$
5	6	30	11	66
10	7	70	6	42
15	8	120	1	8
20	11	220	4	44
25	8	200	9	72
	$\Sigma f = 40$	$\Sigma f \cdot x = 640$		$\Sigma f \cdot \lvert dx\rvert = 232$

$\therefore \quad \overline{X} = \frac{\Sigma f \cdot x}{\Sigma f} = \frac{640}{40} = 16$

Mean deviation (M.D.) $= \frac{\Sigma f \cdot |dx|}{\Sigma f} = \frac{232}{40} = 5.8$ marks

coefficient of M.D. (about mean) $= \frac{M.D.}{mean} = \frac{5.8}{16} = 0.363$

Mean deviation about median

marks(x)	frequency(f)	c.f.	$\lvert dx \rvert = \lvert x - M \rvert$	$f \cdot \lvert dx \rvert$
5	6	6	10	60
10	7	13	5	35
$\boxed{15}$	$\boxed{8}$	$\boxed{21}$	0	0
20	11	32	5	55
25	8	40	10	80
	$N = \Sigma f = 40$			$\Sigma f \cdot \lvert dx \rvert = 230$

median (M) $= \frac{N+1}{2}$ th value $= \frac{40+1}{2}$ th value $= 20.5$th value

$= 15$

$M.D. = \frac{\Sigma f \cdot |dx|}{\Sigma f} = \frac{230}{40} = 5.75$ marks

coefficient of M.D. (about median) $= \frac{M.D.}{median} = \frac{5.75}{15}$

$= 0.383$

3) ## Calculation of Mean deviation - Continuous Series :-

It uses the following steps-

i) Find mid value for each class interval.

ii) Compute average value (mean or median).

iii) Determine deviation of each mid value from their average value. i.e. $|dx|$ and multiply these deviation by the corresponding frequency $(f \cdot |dx|)$

iv) Calculate sum of product and also total frequency which is denoted by $\Sigma f \cdot |dx|$ and Σf respectively and apply the following formula-

$$\boxed{\text{Mean Deviation } (M.D.) = \frac{\Sigma f \cdot |dx|}{\Sigma f}}$$

Example - From the following data, find out mean deviation and coefficient of dispersion.

size	0-2	2-4	4-6	6-8	8-10
frequency	3	5	6	4	2

Sol Mean Deviation about mean

X	f	mid value x	$f \cdot x$	$\lvert dx \rvert = \lvert x - \overline{X} \rvert$	$f \cdot \lvert dx \rvert$
0-2	3	1	3	3.7	11.1
2-4	5	3	15	1.7	8.5
4-6	6	5	30	0.3	1.8
6-8	4	7	28	2.3	9.2
8-10	2	9	18	4.3	8.6
	$\Sigma f = 20$		$\Sigma f \cdot x = 94$		$\Sigma f \cdot \lvert dx \rvert = 39.2$

mean $\overline{X} = \dfrac{\Sigma f \cdot x}{\Sigma f} = \dfrac{94}{20} = 4.7$

Mean deviation (about mean) $= \dfrac{\Sigma f \cdot \lvert dx \rvert}{\Sigma f} = \dfrac{39.2}{20} = 1.96$

coefficient of mean deviation $= \dfrac{\text{M.D. (about mean)}}{\text{mean}} = \dfrac{1.96}{4.7} = 0.417$

Mean deviation about median

X	f	x	c.f.	$\lvert dx \rvert = \lvert X - M \rvert$	$f \cdot \lvert dx \rvert$
0-2	3	1	3	3.7	11.1
2-4	5	3	8	1.7	8.5
4-6	6	5	14	0.3	1.8
6-8	4	7	18	2.3	9.2
8-10	2	9	20	4.3	8.6
	$N = \Sigma f = 20$				$\Sigma f \cdot \lvert dx \rvert = 39.2$

Median no. (m) = $\frac{N}{2}$ th value = $\frac{20}{2}$ th value = 10 th value

This lies in class interval 4–6. So.

Here, $L_1 = 4$, $L_2 = 6$, $f = 6$, $c = 8$

$\therefore$ median $M = L_1 + \frac{m-c}{f}(L_2 - L_1)$

$= 4 + \frac{10-8}{6}(6-4)$

$= 4 + 0.7 = 4.7$

$\therefore$ Mean deviation (about median) $= \frac{\sum f|dx|}{\sum f} = \frac{39.2}{20} = 1.96$

& coefficient of mean deviation $= \frac{\text{M.D. (about median)}}{\text{median}}$

$= \frac{1.96}{4.7} = 0.417$

Standard Deviation :– Standard deviation is also known as root mean square deviation for the reason that it is the square root of the mean of the squared deviation from the arithemetic mean. It is denoted by the the small greek letter sigma (σ) or simply s.d.

i.e. $$\boxed{\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}}$$

coefficient of standard deviation :- It is defined as-

$$\text{coefficient of s.d (C.S.D.)} = \frac{\text{Standard deviation}}{\text{mean}} = \frac{\sigma}{\bar{x}}$$

Advantage of standard Deviation :-

- It is based on all the observations and is rigidly defined.
- It is used to algebric treatment and processed many mathematical properties.
- It is less affected by fluctuations of data.
- It is possible to calculate the combined standard deviation of two or more groups.

Disadvantage of standard Deviation :-

- It is not easy to understand and calculate.
- It is time consuming process.

Methods of Calculation of Standard Deviation :-

1) Calculation of Standard Deviation - Individual Series

The following are the steps -

i) Find out actual mean of the series ($\overline{X}$)

ii) Determine the deviation of each value from the mean i.e. $dx = X - \overline{X}$ & square the deviations (dx^2)

iii) Find out sum of squared deviation which is expressed by Σdx^2 and apply the formula -

$$\boxed{\sigma = \sqrt{\frac{\Sigma dx^2}{N}}}$$

Example - Calculate standard deviation and its coefficient -

X - 10, 20, 30, 40, 50, 60

Sol

$$\text{mean } (\overline{X}) = \frac{10+20+30+40+50+60}{6}$$

$$= \frac{210}{6} = 35$$

X	$dx = X - \overline{X}$	dx^2
10	-25	625
20	-15	225
30	-5	25
40	5	25
50	15	225
60	25	625

$\Sigma dx^2 = 1350$

So,

$$\sigma = \sqrt{\frac{\Sigma dx^2}{N}}$$

$$= \sqrt{\frac{1350}{6}}$$

$$= 15$$

coefficient of s.d = $\frac{\sigma}{\bar{X}} = \frac{15}{35} = 0.4285$

2) <u>Calculation of Standard Deviation - Discrete Series :-</u>

The process of computing S.D. involves the following steps —

i) Calculate arithematic mean of the series ($\bar{X}$).

ii) Determine the deviation of each data from the mean ($dx = X - \bar{X}$) and square the deviations (dx^2).

iii) Multiply squared deviation by the respective frequency we get $f.dx^2$.

iv) Find sum of product & also total frequency i.e. $\Sigma f.dx^2$ and Σf respectively. Now apply the formula -

$$\boxed{\sigma = \sqrt{\frac{\Sigma f.dx^2}{\Sigma f}}}$$

<u>Example</u> - From the following data, find out standard deviation and coefficient of dispersion -

marks	10	20	30	40	50	60
No of student	8	12	20	10	7	3

X	f	$f \cdot X$	$dx = X - \overline{X}$	d_x^2	$f \cdot d_x^2$
10	8	80	−20.8	432.64	3461.12
20	12	240	−10.8	116.64	1399.68
30	20	600	−0.8	0.64	12.80
40	10	400	9.2	84.64	846.40
50	7	350	19.2	368.64	2580.48
60	3	180	29.2	852.64	2557.92
	$N = \Sigma f = 60$	$\Sigma f \cdot X = 1850$			$\Sigma f \cdot d_x^2 = 10858.40$

$$\overline{X} = \frac{\Sigma f \cdot X}{\Sigma f} = \frac{1850}{60} = 30.8$$

$$\sigma = \sqrt{\frac{\Sigma f \cdot d_x^2}{\Sigma f}} = \sqrt{\frac{10858.40}{60}} = 13.45$$

coefficient of s.d. $= \frac{\sigma}{\overline{X}} = \frac{13.45}{30.8} = 0.436$

3) Calculations of Standard Deviation – Continuous Series :–

Steps of calculation of standard deviation in Continuous series is as follows –

i) Find out the mid value of each group or class.

ii) Calculate mean of the series ($\overline{X}$)

iii) Determine the deviation of each value from the mean i.e. $dx = x - \overline{X}$ & square the deviation (d_x^2).

iv) Multiply squared deviation by the respective frequency we get $f \cdot dx^2$.

v) Find sum of product & also total frequency i.e. $\Sigma f \cdot dx^2$ and Σf respectively. Apply the formula –

$$\boxed{\sigma = \sqrt{\frac{\Sigma f \cdot dx^2}{\Sigma f}}}$$

Example – Find out standard deviation –

size (x)	0–20	20–40	40–60	60–80	80–100
frequency (f)	7	10	30	42	11

Sol

X	f	x	$dx = x - \overline{X}$	$f \cdot x$	dx^2	$f \cdot dx^2$
0–20	7	10	–48	70	2304	16128
20–40	10	30	–28	300	784	7840
40–60	30	50	–8	1500	64	1920
60–80	42	70	12	2940	144	6048
80–100	11	90	32	990	1024	11264
	$\Sigma f = 100$			$\Sigma f x = 5800$		$\Sigma f \cdot dx^2 = 4320$

$\therefore \quad \overline{X} = \frac{\Sigma f \cdot x}{\Sigma f} = \frac{5800}{100} = 58$

$$\sigma = \sqrt{\frac{\Sigma f \cdot dx^2}{\Sigma f}} = \sqrt{\frac{43200}{100}} = \sqrt{432} \approx 20.8$$