16-11-2012 Unit - 3 (1)

UNIT : III

~~S. VINOTHINI E/M~~
~~S. Sneha E/M~~
B. Srividhya

Measures of Dispersion - Range - Quartile deviation - Mean deviation about an average - Standard deviation and Coefficient of Variation - Individual, Discrete and Continuous type data.

2M

Dispersion is the measure of the Variation of the items. Dispersion is the degree of the variation of the Variables about a central Value.

Measure of dispersion is a measure of the deviations of the items of the series from its average.

Importance or significance of Measures of Dispersion:

1) To Judge the reliability of an average.
2) To decide consistency of performance, by comparing Variability of two or more series.
3) To devise a system of quality Control.
4) To study economic variables such as income, exports, imports, prices, wages etc.

It is useful in analysis of Time Series.

It facilitates the use of other statistical measures such as correlation, Regression, test of significance etc.

Requisites (or) characteristic (or) desirable properties of a good measure of Dispersion (or) Average.

1. It should be rigidly defined.
2. It should be based on all the items.
3. It should be simple to understand and easy to calculate.
4. It should be amenable (applyable to) to further mathematic treatment.
5. It should have [sampling Stability]

Absolute Measures : Absolute Measures it indicates the amount of Variation in a set of Values. They are [quoted] interms of the (units of observations)

They are (1) Range (2) Quartile deviation (or) Semi-inter Quartile Range (3) Mean deviation about Mean (or) Median (or) Mode (4) Standard deviation (5) Variance

Relative Measures : (Relative Measures are used to compare the variation in two (or) more sets). (They are free from the units of measurements of the sets. Comparing two sets by coefficient of Variation we can say which set is more ~~variable~~ Stable (or) less variable. or more consistent or more uniform.)

Relative Measures: (1) Coefficient of Range (2) Coefficient of Quartile deviation (3) Coefficient of Mean deviation and (4) Coefficient of Variation.

#12.

Range: Range is the difference between the largest value and the smallest value.

$$\text{Range} = L - S.$$

$$\text{Coefficient of Range} = \frac{L - S}{L + S}.$$

Merits: 1) It is easy to understand. 2) It is easy to calculate. 3) we can find it very ~~easily~~ Quickly.

Demerits: 1) It is not a reliable measure of dispersion, because it depends only on the two extreme values.

2) It is not suitable for mathematical treatment.

3) It is not based on all the items

4) It cannot be computed when the distribution has open-end classes.

Quartile deviation

The difference between the upper and lower quartiles i.e $Q_3 - Q_1$ is known as the inter quartile Range. and half of it i.e $\left(\frac{Q_3 - Q_1}{2}\right)$ is called the semi-inter quartile Range (or) the quartile deviation.

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

Quartiles: The three points which divide the series into four equal parts are called quartiles.
First Quartile (or) lower Quartile is Q_1.
Third Quartile (or) upper Quartile is Q_3.
The second Quartile Q_2 is nothing but Median.

Quartile deviation.
Note: Quartile deviation maybe defined as the "half the distance between the third and the first quartile".

Merits: (1) It is simple to understand and easy to compute.
2) It is not influenced by extreme Values
3) It can be found out with open end distribution
4) It is a best measure of dispersion than Range

Demerits: (1) It is not amenable to further

Mathematical treatment

2). It is affected by Sampling fluctations.

3). It is not a reliable measure of dispersion.

<u>Mean deviation</u>: The Arithmetic mean of all deviation from the measure of central tendency, considering all such deviations as positive.

Coefficient of Mean deviation $= \dfrac{\text{Mean deviation}}{\text{Mean or Median or Mode}}$

For Individual series,

Mean deviation from Mean $= \dfrac{\sum_i |x_i - \bar{x}|}{n}$

where $\bar{x} = \dfrac{\sum_i x_i}{n}$

For Discrete Series and Continuous

Mean deviation from Median $= \dfrac{\sum_i f_i |x_i - \bar{x}|}{\sum_i f_i}$

Where $\bar{x} = \dfrac{\sum_i f_i x_i}{\sum_i f_i}$

Mean deviation from Median $= \dfrac{\sum_i f_i |x_i - \text{Median}|}{\sum_i f_i}$

Mean deviation from Mode $= \dfrac{\sum_i f_i |x_i - \text{Mode}|}{\sum_i f_i}$

Merits: (i) It is based on all the observations Data

It is simple to understand and easy to calculate

It is not affected by the Extreme value ~~of extreme items~~ Values.

Demerits: i) It is not suitable for further algebraic treatment.

2). It is less reliable and non-mathematical because signs are ignored.

3) It cannot be computed for open-end classes

Standard deviation:

Standard deviation is the square-root of the arithmetic mean of the squares of all the deviations taken from the arithmetic mean.

~~For~~ It is denoted by the Greek Letter 'σ' (small sigma). Vari = $\sigma \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - (\bar{x})^2}$

For Individual Series:

$$\text{Standard deviation } '\sigma' = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$$

$$\text{Where } \bar{x} = \frac{\Sigma x_i}{n}$$

(or)

$$\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

For Discrete and Continuous distributions:

$$\text{Standard deviation } '\sigma' = \sqrt{\frac{\Sigma_i f_i (x_i - \bar{x})^2}{\Sigma_i f_i}}$$

$$\text{Where } \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

(or) Standard deviation 'σ' = $\sqrt{\frac{\Sigma_i f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}$

Comparison (or) Distinguish between Mean deviation and Standard deviation.

Mean Deviation	Standard deviation
1). Algebraic signs are ignored while calculating deviations.	Algebraic signs are taken into account.
2) It is not suitable for algebraic treatment.	It is suitable for algebraic treatment.
3) Deviations are calculated from mean, median or mode.	Deviations are calculated only from arithmetic mean.
4). It is simple to calculate	It is somewhat complex.
5). It Lacks Mathematical properties.	It has mathematical properties.

Mathematical Properties.

1. Standard deviation is not affected by change of origin (mean) but is affected by change of scale (Variance)

That is X, $X+A$, and $X-A$ where A is a constant have the same standard deviation. But X, kx, X/k where k is a constant have different standard deviations

2. Root Mean square deviation is greater than S.D

i.e $$\sqrt{\frac{\Sigma(X-A)^2}{n}} > \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$$

Merits : 1) It is rigidly defined.
(2) It is useful for furthur statistical analysis, such as skewness, correlation, regression etc.
3) It is based on all observations.
4) It is not very much affected by sampling fluctations.
5). It is possible to find Combined Standard deviation of two or more groups.

Demerits :

1). It gives more weight to extreme values.
2) It can-not be computed for open end distributions.
3) It is not easy to understand and it is difficult to calculate.

4. **Coefficient of Variation**. It is the percentage variation in the mean, Standard deviation being considered as the total variation in the mean.

$$\text{Coefficient of Variation} = \frac{\text{Standard deviation}}{\text{Mean}} \times 100.$$

Importance of Co-efficient of Variation

(i) It is used to Compare the variability of two or more Series.

(ii) The series for which Co-efficient of variation is more is said to be more Variable.

i) Find Quartile deviation and its coefficient.

Height in inches	50	51	52	53	54	55	56
Frequency	10	12	15	10	14	18	6

x_i	f_i	Cum fre
50	10	10
51	12	22
52	15	37
53	10	47
54	14	61
55	18	79
56	6	85
Total	85	

$$\frac{N}{4} = \frac{\Sigma f_i}{4} = \frac{85}{4} = 21.25$$

Cum frequency just greater than 21.25 is 37.

$\therefore Q_1 = 52$.

$$\frac{3N}{4} = \frac{3 \times \Sigma f_i}{4} = \frac{3 \times 85}{4} = 63.75$$

Cum frequency just greater than 63.75 is 79.

$\therefore Q_3 = 55$.

55 − 51

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{\cancel{79-52}}{2} = \frac{27}{2}$$

Q.D = 13.5

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

$= \frac{\cancel{79} - \cancel{52}}{\cancel{79} + \cancel{52}}$ $\frac{55-51}{55+51} = \frac{4}{106}$

$= \frac{27}{131}$

$= \cancel{0.2061}$. 0.0377

2). Calculate the Quartile deviation

Marks.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Students	5	7	10	16	11	7	3	2	2	0

Class	f	c f
0-10	5	5
10-20	7	12
20-30	10	22
30-40	16	38
40-50	11	49
50-60	7	56
60-70	3	59
70-80	2	61
80-90	2	63
90-100	6	63
Total		

$N/4 = \frac{63}{4} = 15.75$

Cumulative frequency just greater than 15.75 is 22.

∴ First Quartile class is 20-30.

$l = 20$; $m = 12$; $f = 10$; $c = 10$.

$Q_1 = l + \left\{\frac{N/4 - m}{f}\right\} \times c$

$= 20 + \left\{\frac{15.75 - 12}{10}\right\} \times 10$

$Q_1 = 20 + 3.75 = 23.75$.

$$\frac{3N}{4} = \frac{3 \times 63}{4} = \frac{189}{4} = 47.25$$

Cumulative frequency just greater than 47.25 is,

∴ Third Quartile class is 40 – 50.

$l = 40$; $f = 11$; $m = 38$; $C = 10$.

$$Q_3 = l + \left\{\frac{3N/4 - m}{f}\right\} \times C$$

$$= 40 + \left\{\frac{47.25 - 38}{11}\right\} \times 10$$

$$= 40 + \left\{\frac{9.25}{11}\right\} \times 10$$

$$= 40 + \left(\frac{92.5}{11}\right)$$

$$= 40 + 8.4091$$

$$= 48.4091$$

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Quartile deviation $= \dfrac{Q_3 - Q_1}{2}$

$$= \frac{48.4091 - 23.75}{2}$$

$$= \frac{24.6591}{2}$$

Q.D $= 12.3296$

3) Find Mean deviation and its coefficient

~~21~~ 23 25 28 30 32 38 39 46 48

x_i	$\lvert x_i - \bar{x} \rvert$
21	12
23	10
25	8
28	5
30	3
32	1
38	5
39	6
46	13
48	15
Total 330	78

Arithmetic Mean $\bar{x} = \frac{\Sigma x_i}{n}$

$\bar{x} = \frac{330}{10}$

$\boxed{\bar{x} = 33}$

Mean deviation from mean $= \frac{\Sigma \lvert x_i - \bar{x} \rvert}{n}$

$= \frac{78}{10}$

$\boxed{M.D = 7.8}$

Coefficient of Mean deviation $= \frac{M.D}{Mean}$

$= \frac{7.8}{33}$

$= 0.2364$

4) Find Mean deviation and its coefficient.

Marks	10	15	20	25	30
No: of Students	2	4	6	8	5

Mean $\bar{x} = \frac{\sum_i f_i x_i}{\sum_i f_i}$

Mean deviation $= \frac{\sum_i f_i \lvert x_i - \bar{x} \rvert}{\sum_i f_i}$

x_i	f_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
10	2	20	12	24
15	4	60	7	28
20	6	120	2	12
25	8	200	-3	24
30	5	150	8	40
	25	550		128

$$\bar{x} = \frac{\sum_i f_i x_i}{\sum_i f_i}$$

$$= \frac{550}{25}$$

$$\boxed{\bar{X} = 22}$$

Mean deviation from mean $= \dfrac{\sum_i f_i \lvert x_i - \bar{x} \rvert}{\sum_i f_i}$

$$= \frac{128}{25}$$

$$\boxed{M.D = 5.12}$$

Coefficient of Mean deviation $= \dfrac{\text{Mean deviation}}{\text{Mean}}$

$$= \frac{5.12}{22}$$

$$= 0.2327$$

5) Find Mean deviation from Mean and Median

Class	0-6	6-12	12-18	18-24	24-30
frequency	8	10	12	9	5.

Class	f_i	x_i	$f_i x_i$	cum fre.	$\lvert x_i-\bar{x}\rvert$	$f_i\lvert x_i-\bar{x}\rvert$	$\lvert x_i-Mdn\rvert$	$f_i\lvert x_i-Mdn\rvert$
0-6	8	3	24	8	11.0455	88.3640	11	88
6-12	10	9	90	18	5.0455	50.4550	5	50
12-18	12	15	180	30	0.9545	11.4540	1	12
18-24	9	21	189	39	6.9545	62.5905	7	63
24-30	5	27	135	44	12.9545	64.7725	13	65
Total.	44		618			277.6360		278

Arithmetic Mean $\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{618}{44} = 14.0455$

Mean deviation from Mean $= \frac{\Sigma f_i\lvert x_i-\bar{x}\rvert}{\Sigma f_i} = \frac{276.636}{44} = 6.2872$

$\frac{N}{2} = \frac{44}{2} = 22.$

Cumulative frequency Just greater than 22 is 30.

$\therefore$ Median class = 12-18.

$l = 12$; $f = 12$; $m = 18$; $c = 6$.

Median $= l + \left\{\frac{\frac{N}{4} - m}{f}\right\} \times c$

$= 12 + \left\{\frac{22-18}{12}\right\} \times 6 = 12 + \frac{4}{2} = 12 + 2 = 14$

Mean deviation from Median $= \frac{\Sigma f_i\lvert x_i - Mdn\rvert}{\Sigma f_i} = \frac{278}{44} = 6.3182$

6. Find Standard deviation.

5 8 7 11 9 10 8 2 4 6

x_i	x_i^2
5	25
8	64
7	49
11	121
9	81
10	100
8	64
2	4
4	16
6	36
Total 70	560

Standard deviation $\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$

$$= \sqrt{\frac{560}{10} - \left(\frac{70}{10}\right)^2}$$

$$= \sqrt{56 - (7)^2}$$

$$= \sqrt{56 - 49}$$

$$= \sqrt{7}$$

$$= 2.6458$$

Standard deviation $\sigma = 2.6458$.

Q(7) Find Standard deviation.

x	56	63	70	77	84	91	98
f	3	6	14	16	13	6	2

If $d = \frac{x - A}{c}$; then Standardard deviation $\sigma = c \times \sqrt{\frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2}$

x_i	f_i	$d_i = \frac{x_i - A}{c}$ ($A=77$, $c=7$)	$f_i d_i$	$f_i d_i^2$
56	3	-3	-9	27
63	6	-2	-12	24
70	14	-1	-14	14
77	16	0	0	0
84	13	1	13	13
91	6	2	12	24
98	2	3	6	18
Total	60		9	107

Standard deviation,

$$\sigma = c\sqrt{\frac{\sum f_i d_i^2}{\sum f_i} - \left\{\frac{\sum f_i d_i}{\sum f_i}\right\}^2}$$

$$= 7\times\sqrt{\frac{208}{60} - \left\{\frac{-26}{60}\right\}^2}$$

$$= 7\times\sqrt{3.4667 - (0.4333)^2}$$

$$= 7\times\sqrt{3.4667 - 0.1877}$$

$$= 7\times\sqrt{3.279}$$

$$= 7\times 1.8108$$

$$\sigma = 12.6756$$

Standard deviation $\sigma = c\times\sqrt{\frac{\sum f_i d_i^2}{\sum f_i} - \left\{\frac{\sum f_i d_i}{\sum f_i}\right\}^2}$

$$\sigma = 7\times\sqrt{\frac{107}{60} - \left\{\frac{9}{60}\right\}^2}$$

$$= 7\times\sqrt{1.7833 - (0.15)^2}$$

$$= 7\times\sqrt{1.7833 - 0.0225}$$

$$= 7\times\sqrt{1.7608}$$

$$= 7\times 1.32695$$

$$\sigma = 9.28865$$

~~(8)~~ Find Mean and Standard deviation.

Expenditure: (less than) Below Rs 5, Below Rs 10, Below Rs 15, Below Rs 20, Below Rs 25

No: of Students: 6, 16, 28, 38, 46.

Class	f	Mid x_i	$d_i = \frac{x_i - A}{c}$ $A = 12.5; c = 5$	$f_i d_i$	$f_i d_i^2$
0-5	6	2.5	-2	-12	24
5-10	10	7.5	-1	-10	10
10-15	12	12.5	0	0	0
15-20	10	17.5	1	10	10
20-25	8	22.5	2	16	32
Total	46			4	76

Arithmetic Mean $\bar{x} = A + \left\{\frac{\Sigma f_i d_i}{\Sigma f_i}\right\} \times c$

$$\bar{x} = 12.5 + \left\{\frac{4}{46}\right\} \times 5$$

$$= 12.5 + \left(\frac{20}{46}\right)$$

$$= 12.5 + 0.4349$$

$$\bar{x} = 12.9348.$$

Standard deviation $\sigma = c \times \sqrt{\frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2}$

$$\sigma = 5 \times \sqrt{\frac{76}{46} - \left(\frac{4}{46}\right)^2}$$

$$= 5 \times \sqrt{1.6522 - (0.087)^2}$$

$$= 5 \times \sqrt{1.6522 - 0.00757}$$

$$= 5 \times \sqrt{1.64463}$$

$$= 5 \times 1.2824$$

$$= 6.412$$

$$\sigma = 6.412.$$

Calculate the Mean, and Standard deviation and Coefficient of Variation.

Marks	No: of Students
More than 0	100
" 10	90
" 20	75
" 30	50
" 40	25
" 50	15
" 60	5
" 70	0

$A = 35$

class	f	Mid pt: x_i	$d_i = \frac{x_i - A}{c}$, $c=10$	$f_i d_i$	$f_i d_i^2$
0–10	10	5	−3	−30	90
10–20	15	15	−2	−30	60
20–30	25	25	−1	−25	25
30–40	25	35	0	0	0
40–50	10	45	1	10	10
50–60	10	55	2	20	40
60–70	5	65	3	15	45
Total	100			−40	270

Mean $\bar{x} = A + \left\{\frac{\Sigma f_i d_i}{\Sigma f_i}\right\} \times c.$

$= 35 + \left\{\frac{-40}{100}\right\} \times 10$

$= 35 - 4 = 31.$

$\boxed{\bar{x} = 31.}$

Standard deviation. $\sigma = c \times \sqrt{\frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2}$

$\sigma = 10 \times \sqrt{\frac{270}{100} - \left(\frac{-40}{100}\right)^2}$

$= 10 \times \sqrt{2.7 - (0.4)^2}$

$= 10 \times \sqrt{2.7 - 0.16}$

$= 10 \times \sqrt{2.54}$

$= 10 \times 1.59374$

$\boxed{\sigma = 15.9374.}$

Coefficient of Variation $= \left\{\frac{\text{Standard deviation}}{\text{Mean}}\right\} \times 100$

$= \left\{\frac{15.9374}{31}\right\} \times 100$

$= \frac{1593.74}{31} = 51.41\%$

$\boxed{\text{Coefficient of Variation} = 51.41\%}$

Inter relationship between Q.D, M.D & S.D.

For a discrete distribution (Standard deviation is not less than Mean deviation.)

Proof: Let (x_i, f_i) be a discrete distribution

We have to prove, Standard deviation $\not<$ Mean deviation.

Standard deviation $\not<$ Mean deviation from Mean.

$$\Rightarrow (S.D)^2 \not< (M.D \text{ from Mean})^2$$

$$\Rightarrow (S.D)^2 \geq (M.D \text{ from mean})^2$$

$$\Rightarrow \frac{1}{N} \Sigma f_i (x_i - \bar{x})^2 \geq \left[\frac{1}{N} \Sigma f_i |x_i - \bar{x}|\right]^2$$

Let $|x_i - \bar{x}| = z_i$

$$\frac{1}{N} \Sigma f_i z_i^2 \geq \left\{\frac{1}{N} \Sigma f_i z_i\right\}^2$$

$V = \Sigma x^2 - (\Sigma x)^2$

$$\Rightarrow \frac{1}{N} \Sigma f_i z_i^2 - \left\{\frac{1}{N} \Sigma f_i z_i\right\}^2 \geq 0$$

i.e $Var(z) \geq 0$ $\because$ Always variance is positive.

$\therefore$ Standard deviation $\geq$ Mean deviation

(or) $S.D \not< M.D.$

i.e S.D is not less than M.D.

(or)

Standard deviation greater than M.D

$$Q.D : M.D : S.D = \frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma$$

$$= 10 : 12 : 15.$$

7.

Let X denotes the runs of A.
Let Y denotes the runs of B.

	X	X^2	Y	Y^2
	100	10000	70	4900
	60	3600	65	4225
	80	6400	80	6400
	20	400	85	7225
	30	900	60	3600
	90	8100	65	4225
	100	10000	75	5625
	5	25	85	7225
	15	225	60	3600
	50	2500	55	3025
Total.	550	42150	700	50050

Mean for A ; $\bar{X} = \frac{\Sigma X_i}{n}$

$$= \frac{550}{10} = 55$$

Stadandard deviation for A; $\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$

$$= \sqrt{\frac{42150}{10} - (55)^2}$$

$$= \sqrt{4215 - 3025}$$

$$= \sqrt{1190}$$

$\sigma_x = 34.4964.$

Mean for B; $\bar{y} = \frac{\Sigma y_i}{n}$

$$= \frac{700}{10}$$

$$= 70.$$

Standard deviation for B, $\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2}$

$$\sigma_y = \sqrt{\frac{50050}{10} - (70)^2}$$

$$= \sqrt{5005 - 4900}$$

$$= \sqrt{105}$$

$$= 10.247.$$

Coefficient of Variation for A; $= \frac{\sigma_x}{\bar{x}} \times 100$

$$= \frac{34.4964}{55} \times 100$$

$$= \frac{3449.64}{55}$$

$$= 62.72\%$$

Coefficient of Variation for B $= \frac{\sigma_y}{\bar{y}} \times 100$

$$= \frac{10.247}{70} \times 100$$

$$= \frac{1024.7}{70}$$

$$= 14.64\%$$

Coefficient of Variation for B is less than Coefficient of Variation for A.

$\therefore$