MEASURES OF DISPERSION

Introduction:

The measures of central tendency serve to locate the center of the distribution, but they do not reveal how the items are spread out on either side of the center. This characteristic of a frequency distribution is commonly referred to as dispersion. In a series all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion. Small dispersion indicates the high uniformity of the items, while large dispersion indicates less uniformity.

CHARACTERISTICS OF A GOOD MEASURE OF DISPERSION:

An ideal measure of dispersion is expected to possess the following properties

1. It should be rigidly defined
2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy to calculate.

ABSOLUTE AND RELATIVE MEASURES:

There are two kinds of measures of dispersion, namely

1. An Absolute measure of dispersion
2. Relative measure of dispersion.

An Absolute measure of dispersion indicates the amount of variation in a set of values in terms of units of observations. On the other hand relative measures of dispersion are free from the units of measurements of the observations. They are pure numbers. They are used to compare the variation in two or more sets, which are having different units of measurements of observations.

The various absolute and relative measures of dispersion are listed below.

Absolute measure Relative measure

1. Range	1. Coefficient of Range
2. Quartile deviation	2. Coefficient of Quartile deviation
3. Mean deviation	3. Coefficient of Mean deviation
4. Standard deviation	4. Coefficient of variation

RANGE:

Definition: This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

In symbols, Range = L – S.

Where L = Largest value.

S = Smallest value.

In individual observations and discrete series, L and S are easily identified. In continuous series, the following method is followed. L = Upper boundary of the highest class

S = Lower boundary of the lowest class.

Co-efficient of Range:

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$

1.Find the value of range and its co-efficient for the following data:

8,10, 5,9,12,11

Solution:

L = 12

S = 5

Range = L-S = 12-5 = 7

Co-efficient of Range = $\frac{L-S}{L+S} = \frac{12-5}{12+5}$ = 7/17 = 0.4118

2. Calculate the range and its co-efficient from the following distribution:

Size	60-62	63-65	66-68	69-71	72-74
Numer	5	18	42	27	8

Size	Number f	True Class Interval
60-62	5	59.5-62.5
63-65	18	62.5-65.5
66-68	42	65.5-68.5
69-71	27	68.5-71.5
72-74	8	71.5-74.5

Solution:

L = 74.5

S = 59.5

Range = L-S = 74.5-59.5 = 15

Co-efficient of Range = $\frac{L-S}{L+S} = \frac{74.5-59.5}{74.5+59.5}$ = 15/134 = 0.1119

Merits:

1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, etc., range is most widely used.

Demerits

1. It is very much affected by the extreme items.

2. It is based on only two extreme observations.

3. It cannot be calculated from open-end class intervals.

4. It is not suitable for mathematical treatment.

5. It is a very rarely used measure.

QUARTILE DEVIATION (Q.D.)

Definition:

Quartile Deviation is half of the diference between the first and the third quartiles. Hence it is also known as Semi Inter Quartile Range.

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

$$\text{Co-efficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

1. Find the Quartile Deviation and its coefficient for the following data:

391, 384, 591, 407, 672, 522, 777, 733, 1490, 2488

Solution:

Ascending Order: 384, 391, 407, 522, 591, 672, 733, 777, 1490, 2488

Position of $Q_1 = \frac{N+1}{4} = \frac{10+1}{4}$ = 11/4 = 2.75

$Q_1 = 2^{nd}$ value + 0.75(3^{rd} value – 2^{nd} value)

= 391 +0.75(407-391)

= 391 + 0.75(16)

= 391 + 12

Q_1 = 403

Position of $Q_3 = 3(\frac{N+1}{4})$ = 3(2.75) = 8.25

$Q_3 = 8^{th}$ value +0.25(9^{th} value – 8^{th} value)

= 777 + 0.25(1490-777)

= 777 + 0.25(713)

= 777 + 178.25

Q_3 = 955.25

$$Q.D. = \frac{Q_3 - Q_1}{2} = \frac{955.25 - 403}{2} = \frac{552.25}{2} = 276.13$$

$$\text{Co-efficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{955.25 - 403}{955.25 + 403} = 0.4066$$

2. Calculate Q.D. and its co-efficient for the following data:

Height (incm.): 28, 32, 18, 16, 42, 12, 39

Solution:

Ascending Order: 12, 16, 18, 28, 32,39,42

N = 7 (Odd)

Position of $Q_1 = \frac{N+1}{4} = \frac{7+1}{4}$ = 8/4 = 2

$Q_1 = 2^{nd}$ value = 16

Position of $Q_3 = 3(\frac{N+1}{4}) = 3(2) = 6$

$Q_3 = 6^{th}$ value = 39

Q.D. = $\frac{Q_3 - Q_1}{2}$ = $\frac{39-16}{2}$ = 23/2 = 11.5

Co-efficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ = 39-16/39+16 = 23/55 = 0.4182

3. Weekly wages of a labourer are given below. Calculate Q.D. and its co-efficient.

Weekly Wages (Rs.) X	No of Weeks f	c.f
100	5	5
200	8	13
400	21	34
500	12	46
600	6	52
	$\sum f = 52$	

Solution:

Position of $Q_1 = \frac{\sum f + 1}{4}$ = 52+1/4 =53/4 = 13.25

= 13^{th} value + 0.25(14^{th} value – 13^{th} value)

Q_1 = 200+0.25(400-200)

= 200+0.25(200)

= 200+50 = 250

Position of $Q_3 = 3(\frac{\sum f + 1}{4}) = 3(13.25) = 39.75$

$Q_3 = 39^{th}$ value + 0.75(40^{th} value – 39^{th} value)

= 500 +0.75(500-500) = 500

Q.D. = $\frac{Q_3 - Q_1}{2}$ = 500-250/2 = 250/2 = 125

Co-efficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{500-250}{500+250} = \frac{250}{750} = 0.3333$

4.Calculate Quartile Deviation and co-efficient of Q.D. from the following data.

Life Expectancy (in months)	79	43	40	34	42	41
No of Patients	4	7	15	8	7	2

Ascending Order:

Life Expectancy (in months) X	No of Patients f	c.f
34	8	8
40	15	23
41	2	25
42	7	32
43	7	39

79	4	43
	$\sum f$ = 43	

Solution:

Position of Q_1 = $\frac{\sum f + 1}{4}$ = 43+1/4 = 44/4 = 11

Q_1 = 40

Position of Q_3 = 3($\frac{\sum f + 1}{4}$) = 3(11) = 33

Q_3 = 43

Q.D. = $\frac{Q_3 - Q_1}{2}$ = 43-40/2 = 3/2 = 1.5

Co-efficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ = 43-40/43+40 = 3/83 = 0.0361

5.For the data given below here, calculate the Q.D. and co-efficient of Q.D.

X	f	True class Intervals	c.f
351-500	48	350.5-500.5	48
501-650	189	500.5-650.5	237
651-800	88	650.5-800.5	325
801-950	47	800.5-950.5	372
951-1100	28	950.5-1100.5	400
	$\sum f$ = 400		

$\frac{\sum f}{4} = \frac{400}{4} = 100$

Q_1 Class = 500.5-650.5

L_1 = 500.5

p.c.f_1 = 48

f_1 = 189

i_1 = 150

$Q_1 = L_1 + \left[\frac{\sum f/4 - p.c.f_1}{f_1}\right] i_1 = 500.5 + \left[\frac{100-48}{189}\right] 150 = 500.5 + \left[\frac{52}{189}\right] 150 =$

Q_1 = 500.5+41.27 = 541.77

$= 3\left(\frac{\sum f}{4}\right) = 3(100) = 300$

Q_3 Class = 650.5-800.5

L_3 = 650.5

p.c.f_3 = 237

f_3 = 88

i_3 = 150

$$Q_3 = L_3 + \left[\frac{3(\sum f/4) - p.c. f_3}{f_3}\right] i_3 = \ 650.5 + \left[\frac{300-237}{88}\right]150 \ = \ 650.5 + \left[\frac{63}{88}\right]150 =$$

$$Q_3 = 650.5+107.39 = 757.89$$

$$\text{Q.D.} = \ \frac{Q_3 - Q_1}{2} \ = \ \frac{757.89 - 541.77}{2} = 216.12/2 = 108.06$$

$$\text{Co-efficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{757.89 - 541.77}{757.89 + 541.77} = \frac{216.12}{1299.66} = 0.1663$$

Merits:

1. It is Simple to understand and easy to calculate
2. It is not affected by extreme values.
3. It can be calculated form data with open end classes also.

Demerits:

1. It is not based on all the items. It is based on two positional values Q1 and Q3 and ignores the extreme 50% of the items.
2. It is not amenable to further mathematical treatment.
3. It is affected by sampling fluctuations.

MEAN DEVIATION:

The range and quartile deviation are not based on all observations. They are positional measures of dispersion. They do not show any scatter of the observations of an average. The mean deviation is a measure of dispersion based on all items in a distribution.

Definition: Mean deviation is the arithmetic mean of the deviations of a series computed from any measure of central tendency; i.e., the mean, median or mode, all the deviations is taken as positive i.e., signs are ignored.

Coefficient of mean deviation:

Mean deviation calculated by any measure of central tendency is an absolute measure. For the purpose of comparing variation among different series, a relative mean deviation is required. The relative mean deviation is obtained by dividing the mean deviation by the average used for calculating the mean deviation.

$$\text{Coefficient of mean deviation:} = \frac{\text{Mean deviation}}{\text{Mean or Median or Mode}}$$

If the result is desired in percentage, the coefficient of mean

$$\text{Coefficient of mean deviation:} = \frac{\text{Mean deviation}}{\text{Mean or Median or Mode}} \times \ 100$$

Calculate Mean Deviation about Mean for the numbers given below:

1,2,3,4,5,

X	$\lvert X - \overline{X} \rvert$ 3
1	2
2	1
3	0
4	1
5	2
$\sum X = 15$	$\sum \lvert X - \overline{X} \rvert = 6$

$$\overline{X} = \frac{\sum X}{N} = 15/5 = 3$$

$$\text{M.D. about Mean} = \frac{\sum \lvert X - \overline{X} \rvert}{N} = 6/5 = 1.2$$

Co-efficient of M.D. about Mean = <u>Mean Deviation about Mean</u>

Mean

= 1.2/3 = 0.4000

Discrete Series:

$$\text{Mean Deviation about Mean} = \frac{\sum f \lvert X - \overline{X} \rvert}{\sum f}$$

Calculate the M.D. from Mean for thefollowing data:

X	f	fx	$\lvert X - \overline{X} \rvert$ 6	f$\lvert X - \overline{X} \rvert$
2	1		4	4
4	4		2	8
6	6		0	0
8	4		2	8
10	1		4	4
	$\sum f = 16$	$\sum fx = 96$	$\sum \lvert X - \overline{X} \rvert$	$\sum f \lvert X - \overline{X} \rvert$ =24

$$\overline{X} = \frac{\sum fx}{\sum f} = 96/16 = 6$$

$$\text{M. D about Mean} = \frac{\sum f \lvert X - \overline{X} \rvert}{\sum f} = 24/16 = 1.5$$

Co-efficient of M.D. about Mean =

<u>M.D. about Mean</u>

Mean

= 1.5/6 = 0.2500

. Calculate the M.D. from Mean for thefollowing data:

Marks	No of Students f	m	fm	$\lvert m - \overline{X} \rvert$ $\overline{X} = 33.4$	f$\lvert m - \overline{X} \rvert$

0-10	6	5	30	28.4	170.4
10-20	5	15	75	18.4	92
20-30	8	25	200	8.4	67.2
30-40	15	35	525	1.6	24
40-50	7	45	315	11.6	81.2
50-60	6	55	330	21.6	129.6
60-70	3	65	195	31.6	94.8
	$\sum f = 50$		$\sum fm = 1670$		$\sum f\lvert m-\overline{X}\rvert = 659.2$

Mean:

$$\overline{X} = \frac{\sum fm}{\sum f} = 1670/50 = 33.4$$

$$\text{M.D. about Mean} = \frac{\sum f\lvert m-\overline{X}\rvert}{\sum f} = 659.2/50 = 13.18$$

$$\text{Co-efficient of M.D. about Mean} = \frac{\text{M.D. about Mean}}{\text{Mean}}$$

$$= 13.18/33.4 = 0.3946$$

Merits:

1. It is simple to understand and easy to compute.
2. It is rigidly defined.
3. It is based on all items of the series.
4. It is not much affected by the fluctuations of sampling.
5. It is less affected by the extreme items.
6. It is flexible, because it can be calculated from any average.
7. It is a better measure of comparison.

Demerits:

1. It is not a very accurate measure of dispersion.
2. It is not suitable for further mathematical calculation.
3. It is rarely used. It is not as popular as standard deviation.
4. Algebraic positive and negative signs are ignored. It is mathematically unsound and illogical.

STANDARD DEVIATION

Definition:

Standard deviation is the root mean square deviation from their arithmetic mean. It is denoted by ‘σ’. Variance is denoted by ‘σ^2’. The corresponding relative measure is Coefficient of Variation.

$$\text{Coefficient of variation} = \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \text{ X } 100$$

$$\text{Coefficient of variation} = \frac{\text{S.D}}{\text{A.M}} \text{ X } 100$$

$$\text{Coefficient of variation(C.V.)} = \frac{\sigma}{\overline{X}} X100$$

Standard Deviation – Formulae

1. Individual Series $\sigma = \sqrt{\frac{\sum X^2}{N} - \left[\frac{\sum X}{N}\right]^2}$
2. Discrete Series $\sigma = \sqrt{\frac{\sum fX^2}{\sum f} - \left[\frac{\sum fX}{\sum f}\right]^2}$
3. Continuous Series $\sigma = \sqrt{\frac{\sum fm^2}{\sum f} - \left[\frac{\sum fm}{\sum f}\right]^2}$

Individual Series:

1.Calculate S.D and Coefficient of variation for the data given below.

S.No	Marks X	X^2
1	5	25
2	10	100
3	20	400
4	25	625
5	40	1600
6	42	1764
7	45	2025
8	48	2304
9	70	4900
10	80	6400
	$\sum X = 385$	$\sum X^2 = 20143$

$\overline{X} = \frac{\sum X}{N}$ = 385/10 = 38.5

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \left[\frac{\sum X}{N}\right]^2}$$

$$= \sqrt{\frac{20143}{10} - (38.5)^2} = \sqrt{2014.3 - 1482.25}$$

$$= \sqrt{532.05} = 23.07$$

C.V = $\frac{\sigma}{\overline{X}} X100$ = 23.07/38.5 X 100 = 59.31

2. Calculate S.D and Coefficient of variation for the data given below.

Roll No	Marks X	X^2

101	10	100
102	30	900
103	20	400
104	25	625
105	15	225
	$\sum X = 100$	$\sum X^2 = 2250$

$$\overline{X} = \frac{\sum X}{N} = 100/5 = 20$$

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \left[\frac{\sum X}{N}\right]^2}$$

$$= \sqrt{\frac{2250}{5} - (20)^2} = \sqrt{450 - 400} = \sqrt{50} = 7.07$$

$$C.V = \frac{\sigma}{\overline{X}} X100 = 7.07/20 \text{ X}100 = 35.35$$

Discrete Series:

3.. Calculate S.D and Coefficient of variation.

X	f	fx	x^2	f x^2
6	7	42	36	252
9	12	108	81	972
12	13	156	144	1872
15	10	150	225	2250
18	8	144	324	2592
	$\sum f = 50$	$\sum fx = 600$		$\sum fX^2 = 7938$

$$\overline{X} = \frac{\sum fx}{\sum f} = 600/50 = 12$$

$$\sigma = \sqrt{\frac{\sum fX^2}{\sum f} - \left[\frac{\sum fX}{\sum f}\right]^2}$$

$$= \sqrt{\frac{7938}{50} - 12^2} = \sqrt{158.76 - 144}$$

$$= \sqrt{14.76} = 3.84$$

$$C.V = \frac{\sigma}{\overline{X}} X100 = 3.84/12\text{X } 100 = 32$$

4.Calculate S.D and Coefficient of variation.

No of goals Scored in a match x	No of Matches f	fx	x^2	f x^2
0	1	0	0	0
1	2	2	1	2
2	4	8	4	16
3	3	9	9	27
4	0	0	16	0
5	2	10	25	50
	$\sum f = 12$	$\sum fx = 29$		$\sum fm^2 = 95$

$$\overline{X} = \frac{\sum fx}{\sum f} = 29/12 = 2.42$$

$$\sigma = \sqrt{\frac{\sum fX^2}{\sum f} - \left[\frac{\sum fX}{\sum f}\right]^2}$$

$$= \sqrt{\frac{95}{12} - (2.42)^2}$$

$$= \sqrt{7.9167 - 5.8564}$$

$$= \sqrt{2.0603} = 1.44$$

$$C.V = \frac{\sigma}{\overline{X}} X100 = 1.44/2.42 \text{ X } 100 = 59.50$$

Continuous Series

5.. Find the S.D and Coefficient of Variation:

Class Interval	Frequency f	m	fm	m^2	f m^2

0-10	2	5	10	25	50
10-20	5	15	75	225	1125
20-30	9	25	225	625	5625
30-40	3	35	105	1225	3675
40-50	1	45	45	2025	2025
	$\sum f = 20$		$\sum fm = 460$		$\sum fm^2 = 12500$

$\overline{X} = \frac{\sum fm}{\sum f}$ = 460/20 = 23

$\sigma = \sqrt{\frac{\sum fm^2}{\sum f} - \left[\frac{\sum fm}{\sum f}\right]^2}$

$= \sqrt{\frac{12500}{20} - (23)^2}$

$= \sqrt{625 - 529}$

$= \sqrt{625 - 529} = \sqrt{96} = 9.80$

C.V = $\frac{\sigma}{\overline{X}} X100$ = 9.80/23 X 100 = 42.61

6.The following data were obtained while observing the life span of a few neon lights of a company. Calculate S.D. and C.V.

Life Span (Years)	No. of Neon lights f	m	fm	m^2	f m^2
4-6	10	5	50	25	250
6-8	17	7	119	49	833
8-10	32	9	288	81	2592
10-12	21	11	231	121	2541
12-14	20	13	260	169	3380
	$\sum f = 100$		$\sum fm = 948$		$\sum fm^2 = 9596$

$\overline{X} = \frac{\sum fm}{\sum f}$ = 948/100 = 9.48

$\sigma = \sqrt{\frac{\sum fm^2}{\sum f} - \left[\frac{\sum fm}{\sum f}\right]^2}$

$= \sqrt{\frac{9596}{100} - (9.48)^2}$

$= \sqrt{95.96 - 89.87}$

$= \sqrt{6.09} = 2.47$

C.V = $\frac{\sigma}{\overline{X}} X100$ = 2.47/9.48X100 = 26.05

CO- EFFICIENT OF VARIATION

1. The means ans standard deviation values for the number of runs of two players A and B are 55; 65 and 4.2; 7.8 respectively. Who is the more consistent player?

Given: Player A Player B

	Player A	Player B
Mean($\overline{X}$)	55	65
S.D.(σ)	4.2	7.8

Coefficient of variation of Player A = $\frac{\sigma}{\overline{X}} X100 = \frac{4.2}{55}$ X 100 = 7.64

Coefficient of variation of Player B = $\frac{\sigma}{\overline{X}} X100 = \frac{7.8}{65}$ X 100 = 12

Coefficient of variation of player A is less. Therefore Player A is the more consistent player.