Measure of dispersion

Dispersion

The extent (limit) to which the values are spread out from an average is called dispersion.

Measure of dispersion

Any formula used to measure the dispersion is called a Measure of dispersion.

Types of Measure of dispersion

There are two main types of Measure of dispersion.

1) Absolute Measure of dispersion
2) Relative Measure of dispersion

Absolute Measure of dispersion

It is the dispersion which measures the variation present among the values in terms of the same unit as the unit of the data.

Types

i) Range ii) Quartile deviation
iii) Mean deviation
iv) Standard deviation
v) Variance

Relative Measure of dispersion

It is the dispersion which measures the variation present in the data relative to their average. It is expressed in the form of ratios, Cofficient or percentage. It is independent of the unit of the data.

Types

i) Coefficient of Range
ii) Coefficient of Quartile deviation
iii) Coefficient of Mean deviation
iv) Coefficient of Standard deviation
v) Coefficient of Variance or variation

Range

Un-group data

The difference between the largest value and the smallest value in the data is called Range. It is denoted by R

R=largest value-smallest value where X_0=smallest value
$R=X_m-X_0$ X_m=largest value

Discrete frequency distribution

The difference between the largest value of the variable X and smallest value of the variable X it is called Range. It is denoted by R

R=largest value-smallest value where X_0=smallest value of the variable X
$R=X_m-X_0$ X_m= largest value of the variable X

Continuous data or Grouped data

The difference between the upper class boundary of the highest class and the lower class boundary of the lowest class it is called Range. It is denoted by R

R=largest value-smallest value where X_0=lower class boundary
$R=X_m-X_0$ X_m= upper class boundary

Coefficient of Range

It is relative measure of dispersion and is based on the value of range. It is also called range coefficient of dispersion. It is defined as

$$Coefficient\ of\ range = \frac{X_m - X_0}{X_m + X_0}$$

The range X_m-X_0 is standardized by the total X_m+X_0

Advantages or merits of Range

1) It easy to calculate
2) It is suitable if the data is homogenous
3) It is useful in small sample inquires
4) It is easy to interpret

Disadvantages or demerits of Range

1) It is highly rough measure of dispersion
2) It gives no idea between the two extreme values
3) It is not based on all the values
4) It is not capable mathematical treatment

Example: 4.1: The marks obtained by 9 students are given below.
45, 32, 37,46,39,36,41,48,36.

Solution:

X_0= Lowest marks =32
X_m=highest marks = 48

Range=X_m-X_0=48-32=16

$$Coefficient\ of\ range = \frac{X_m - X_0}{X_m + X_0} = \frac{48-32}{48+32} = \frac{16}{80} = 0.2$$

Example 4.2: Find the range and coefficient of range from the following discrete frequency distribution.

X	10	11	12	13	14	15	16	17	18	19	20	21
f	9	36	75	105	116	107	88	66	45	30	18	5

Solution:

X_0= Smallest value of the variable X =10
X_m= Smallest value of the variable X = 21

Range=X_m-X_0=21-10=11

$$Coefficient\ of\ range = \frac{X_m - X_0}{X_m + X_0} = \frac{21-10}{21+10} = \frac{11}{31} = 0.355$$

Example 4.3: Find the range and coefficient of range from the following frequency distribution.

Classes	5-9	10-14	15-19	20-24	25-29	30-34
f	9	36	75	105	116	107

Solution:

C.b	4.5-9.5	9.5-14.5	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5
f	9	36	75	105	116	107

*C.b=class boundary

X_0= lowest class boundary of the highest class=4.5
X_m= upper class boundary of the highest class = 34.5

Range=X_m-X_0=34.5-4.5=30

$$Coefficient\ of\ range = \frac{X_m - X_0}{X_m + X_0} = \frac{34.5-4.5}{34.5+4.5} = \frac{30}{39} = 0.769$$

Quartile deviation

The half of the difference between the upper quartile and lower quartile is called quartile deviation. It is also called semi inter quartile range. It is denoted by Q.D or S.I.Q.R

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

Q_1=lower quartile

Q_3=upper quartile

Coefficient of Quartile deviation

The relative measure of quartile deviation is called coefficient of quartile deviation or quartile coefficient of dispersion. It is defined as Ratio between Q_3-Q_1 and Q_3+Q_1

$$Coefficien\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Which is a pure number and is used for comparing the variation in two or more sets of data

Advantages or merits of Q.D

1) It is easy to calculate
2) It is simple to understands
3) It is not affected by extreme values
4) It is superior to range
5) It is useful badly skewed distributions
6) I is not affected by the dispersion of the individual values of the variable
7) It is specially useful for measuring variation in case of open end distribution

Disadvantages or demerits of Q.D

1) It is not based on all the values

2) Q.D is same for all sets of values having same quartiles
3) It is amenable mathematical treatment
4) The Q.D is not widely used as other measure of dispersion
5) It is ignore first 25% and last 25% of the observations in the data
6) Its value is much affected by sampling fluctuations

Because of these limitations quartile deviation is not a useful measure in statistical inference.

Exampe.4.4: Following are the marks obtained by 20 students, Calculate lower quartile Q_1 and upper quartile Q_3 also Q.D and coefficient of Quartile. (Un-group data)
38,60,41,40,29,40,51,56,40,56,39,40,54,40,53,54,37,53,45,50.

Solution: First we arrange the data in ascending order
Arrange
29,37,38,39,40,40,40,40,40,41,45,50,51,53,53,54,54,56,56,60

Q_1=The value of $(\frac{n+1}{4})th$ item = (21/4) th=5.25th value

=5th value+0.25(6th value-5th value) =40+0.25(40-40) = 41+0.25(0) =40+0=40Ans

Q_3=The value of $3(\frac{n+1}{4})th$ item =3 (21/4)th=3(5.25)th value=15.75th value

=15th value+0.75(16th value-15thvalue)

=53+0.75(5453) =53+0.75(1)=53+0.75=53.75Ans

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{53.75 - 40}{2} = 6.875$$

$$Coefficien\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{53.75 - 40}{53.75 + 40} = \frac{13.75}{93.75} = 0.147$$

Example 4.5: Find the range and coefficient of range from the following discrete frequency distribution.

X	10	11	12	13	14	15	16	17	18	19	20	21
f	9	36	75	105	116	107	88	66	45	30	18	5

Solution:

X	10	11	12	13	14	15	16	17	18	19	20	21
f	9	36	75	105	116	107	88	66	45	30	18	5
c.f	9	45	120	225	341	448	536	602	647	677	695	700

*Q_1 and Q_3 *C.f=Cumulative frequency

Q_1=$(\frac{n+1}{4})th$ item = (701/4) th=175.25th value=13

$Q_3 = 3(\frac{n+1}{4})th$ item =3 (701/4) th=3(175.25)th value=525.75th value=16

=15th value+0.75(16th value-15thvalue)

=53+0.75(5453) =53+0.75(1)=53+0.75=53.75Ans

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{16-13}{2} = 1.5$$

$$Coefficien\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{16-13}{16+13} = \frac{3}{29} = 0.103$$

Example 4.6: Find the semi inters quartile range and coefficient of Q.D from the following frequency distribution.

Classes	5-9	10-14	15-19	20-24	25-29	30-34
f	9	36	75	105	116	107

Solution:

C.b	4.5-9.5	9.5-14.5	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5
f	9	36	75	105	116	107
c.f	9	45	120	225	341	448

*C.b=class boundary *C.f=Cumulative frequency

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$ Lower quartile $(\frac{448}{4})th = 112th$

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right) = 14.5 + \frac{5}{75}(112-45) = 18.97$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$ $(\frac{3*448}{4})th = (3*112)th = 336th$

$$= 24.5 + \frac{5}{116}(336-225) = 29.28$$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{29.28-18.97}{2} = 5.157$$

$$Coefficien\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{29.28-18.97}{29.28+18.97} = \frac{10.314}{48.25} = 0.214$$

Mean deviation

It is defined as the mean of the deviations of the values taken from their averages (mean, median, and mode) without algebraic sign. It is denoted by M.D

It is given as **Un-group data**

$$M.D = \frac{\sum|X_i - Mean|}{n}$$ Mean deviation from mean

$$M.D = \frac{\sum|X_i - Median|}{n}$$ Mean deviation from median

$$M.D = \frac{\sum|X_i - Mode|}{n}$$ Mean deviation from mode

Group data

$$M.D = \frac{\sum f|X_i - Mean|}{\sum f}$$ Mean deviation from mean

$$M.D = \frac{\sum f|X_i - Median|}{\sum f}$$ Mean deviation from median

$$M.D = \frac{\sum f|X_i - Mode|}{\sum f}$$ Mean deviation from mode

Note

Without algebraic sign mean absolute value of the deviation i.e. $|X_i - Mean|$ or $|X_i - Median|$ .The absolute value of the positive number in the itself where as the absolute value of the negative number is the number without its minus sign i.e. $|X_i| = X_i$, $|-X_i| = X_i$

Advantages or merits of M.D

1) There is no confusion in its definition
2) It is based on all the values
3) It is easy to calculate
4) It is easy to understand
5) It gives weight to the observations according to their size
6) It is less affected by extreme values

Disadvantages or demerits of M.D

1) It has a mathematical flaw of ignoring signs
2) It has no further mathematical treatment
3) It is affected by extreme values
4) It is not generally used in social sciences

5) Coefficient of Mean deviation

6) Ratio between mean deviation and average used in its calculation is called coefficient of mean deviation. It is denoted by
7) $Coefficient\ of\ M.D\ or\ mean\ coeffcient\ of\ dispersion \quad = \frac{M.D_{\bar{X}}}{\bar{X}}$
8) $Coefficient\ of\ M.D\ or\ median\ coeffcient\ of\ dispersion \quad = \frac{M.D_{\tilde{X}}}{\tilde{X}}$

9) $Coefficient\ of\ M.D\ or\ \bmod e\ coeffcient\ of\ dispersion = \frac{M.D_{\hat{X}}}{\hat{X}}$

10) **Example 4.7:** Calculate the Mean Deviation and Co-efficient of M.D from i) Mean
11) ii) Median iii) Mode. From the following data
12) 32, 45,37,46,39,36,41,48 and 36
13) **Solution:** First we arrange the observations 32, 36, 36,37,39,41,45,46,48
14) Mean= $\bar{x} = \frac{\sum x}{n}$ =40
15) Median= 39
16) Mode=36
17)

X	$\lvert X - Mean \rvert$	$\lvert X - Median \rvert$	$\lvert X - Mode \rvert$
32	8	7	4
36	4	3	0
36	4	3	0
37	3	2	1
39	1	0	3
41	1	2	5
45	5	6	9
46	6	7	10
48	8	9	12
Total	$\sum \lvert X - Mean \rvert = 40$	$\sum \lvert X - Median \rvert = 39$	$\sum \lvert X - Mode \rvert = 44$

18) $M.D = \frac{\sum \lvert X_i - Mean \rvert}{n}$ =4.44 Mean deviation from mean

19) $M.D = \frac{\sum \lvert X_i - Median \rvert}{n}$ =4.33 Mean deviation from median

20) $M.D = \frac{\sum \lvert X_i - Mode \rvert}{n}$ =4.8 Mean deviation from mode

21) $Coefficient\ of\ M.D_{\bar{x}}\ or\ mean\ coeffcient\ of\ dispersion = \frac{M.D_{\bar{X}}}{\bar{X}}$ =0.111

22) $Coefficient\ of\ M.D_{\tilde{x}}\ or\ median\ coeffcient\ of\ dispersion = \frac{M.D_{\tilde{X}}}{\tilde{X}}$ =0.1111

23) $Coefficient\ of\ M.D_{\hat{x}}\ or\ \bmod e\ coeffcient\ of\ dispersion = \frac{M.D_{\hat{X}}}{\hat{X}}$ =0.136

24) **Example 4.8**: Calculate the Mean Deviation and Co-efficient of M.D from i) Mean
25) ii) Median iii) Mode. From the following frequency distribution.
26)

Classes	5-9	10-14	15-19	20-24	25-29	30-34
f	9	36	75	105	116	107

27)
28) Solution:
29)

C.b	4.5-9.5	9.5-14.5	14.5-19.5	19.5-24.5	24.5-29.5	29.5-34.5
f	9	36	75	$105=f_1$	$116=f_m$	$107=f_2$
c.f	9	45	120	225	341	448

30) *C.b=class boundary *C.f=Cumulative frequency

31) $\bar{x} = \dfrac{\sum fx}{\sum f} = \dfrac{10636}{448} = 23.74$

32) $Median = \tilde{X} = l + \dfrac{h}{f}\left(\dfrac{n}{2} - c\right) = 19.5 + \dfrac{5}{105}(224 - 120) = 24.45 \qquad (\dfrac{448}{2})th = 224th$

33) $\hat{X} = l + \dfrac{f_m - f_1}{f_m - f_1 + f_m - f_2} \times h = 24.5 + \dfrac{116 - 107}{116 - 107 + 116 - 105} \times 5 = 26.75$

x	f	$f\lvert X - Mean\rvert$	$f\lvert X - Median\rvert$	$f\lvert X - Mode\rvert$
7	9	150.66	157.05	177.75
12	36	422.64	448.2	531
17	75	505.5	558.75	731.25
22	105	182.7	257.25	498.75
27	116	378.16	295.8	29
32	107	883.82	807.85	561.75
	$\sum f$ 448	$\sum f\lvert X - Mean\rvert = 2523.48$	$\sum f\lvert X - Median\rvert = 2524.9$	$\sum f\lvert X - Mode\rvert = 2529.5$

34)

35) $M.D = \dfrac{\sum f\lvert X_i - Mean\rvert}{\sum f}$ Mean deviation from mean=5.633

36) $M.D = \dfrac{\sum f\lvert X_i - Median\rvert}{\sum f}$ Mean deviation from median=5.634

37) $M.D = \dfrac{\sum f\lvert X_i - Mode\rvert}{\sum f}$ Mean deviation from mode=5.646

38)

39) $Coefficient\ of\ M.D_{\bar{x}}\ or\ mean\ coeffcient\ of\ dispersion = \dfrac{M.D_{\bar{X}}}{\bar{X}}$ =0.237

40) $Coefficient\ of\ M.D_{\tilde{x}}\ or\ median\ coeffcient\ of\ dispersion = \dfrac{M.D_{\tilde{X}}}{\tilde{X}}$ =0.230

41) $Coefficient\ of\ M.D_{\hat{x}}\ or\ \mathrm{mod}\,e\ coeffcient\ of\ dispersion = \dfrac{M.D_{\hat{X}}}{\hat{X}}$ =0.211

42) Standard deviation

43) It is defined as the positive square root of the mean of the squared deviations of the values from their mean. It is denoted by S or δ

44) Calculation methods

45) Simple method

46) $$S=\sqrt{\frac{\sum\left(X-\overline{X}\right)^2}{n}}=\sqrt{\left(\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2\right)}$$

47) Or

48) $$S=\sqrt{\frac{\sum f\left(X-\overline{X}\right)^2}{\sum f}}=\sqrt{\left(\frac{\sum fx^2}{\sum f}-\left(\frac{\sum fx}{\sum f}\right)^2\right)}$$

Advantages or merits of S.D

1) It is rigidly defined
2) It is based on all the observations
3) It is capable of mathematical treatment
4) It is stable in repeated sampling experiments
5) It will be large if the observations are distant from the mean and small if they are close to mean
6) It is possible to calculate combined standard deviation of two or more groups which is not possible with any other measure of dispersion

Disadvantages or demerits of S.D

1) It is affected by extreme values
2) It is not an unbiased estimate of population standard deviation

Variance

It is defined as the arithmetic mean of the squared deviation taken from their mean. It is denoted by $S^2 \; or \, \sigma^2$

It is given as

Simple method or direct method

$$S^2=\frac{\sum\left(x-\bar{x}\right)^2}{n}=\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2$$

$$S^2=\frac{\sum f\left(x-\bar{x}\right)^2}{\sum f}=\frac{\sum fx^2}{\sum f}-\left(\frac{\sum fx}{\sum f}\right)^2$$

Properties of variance

1) Variance of constant is equal to zero.

   Var (a) =0

2) The Variance of variable X is independent of origin. It remains unchanged when a Constant is added or Subtracted from each value of the variable X.

$$Var(X\pm a)=\frac{1}{N}\sum\left(X-\mu\right)^2=Var(X)$$

3) When all the values of variable X are multiplied or divided by a constant. The Variance of these values is multiplied or divided by square of constant.

$$Var(aX) = a^2 Var(X)$$

$$Var(\frac{1}{a}X) = \frac{1}{a^2} Var(X)$$

4) If two sets of data consisting of n_1, n_2 values have means $\overline{X}_1, \overline{X}_2$ and variances $S_1^{\ 2}, S_2^{\ 2}$ respectively then the combined variance of n_1, n_2 values is

$$S^2 = \frac{n_1 S_1^{\ 2} + n_2 S_2^{\ 2}}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2}\left(\overline{X}_1 - \overline{X}_2\right)^2$$

5) The variance of the sum or difference of two independent random variables is equal to the sum of their respective variances

i) Var(x+y) =Var(x) +Var(y)

ii) Var(x+y) =Var(x) +Var(y)

6) If k subgroups of the data consisting of N_1, N_2… N_k ($\sum N_i = N$) observations have respective means $\mu_1,\ \mu_2, \mu_3, \ldots, \mu_k$ and variances $\sigma^2{}_1, \sigma^2{}_2, \sigma^2{}_3, \ldots, \sigma^2{}_k$,than the variance σ^2 of the combined observations is given as

$$\sigma^2 = \frac{1}{N}\sum N_i\left(\sigma^2{}_i + D^2{}_i\right) \qquad \text{i=1,2,3,…,k}$$

Where $D_i = \mu_i - \mu$ and μ is combined mean

Coefficient of Variance or variation

The relative measure of dispersion of variance is called the Coefficient of variation. The coefficient of variation expressed the standard deviation as percentage in terms of arithmetic mean. It is denoted by C.V

$$C.V = \frac{S}{\overline{\overline{X}}} \times 100$$

Uses of coefficient of variation

1) It is used to compare the dispersion of two or more set of data
2) It is also used for the consistence performance. The smallest the coefficient of variation the more consistence is the performance

   Example.4.9: Find the mean S.D and Co-efficient of S.D and Co-efficient of variation from following ungroup data by Direct, short-cut and coding method.

   5, 10,15,20,25

   Solution:

X	X^2
5 10 15 20 25	25 100 225 400 625
$\sum x$ =75	$\sum x^2$ =1375

i) $\bar{x} = \frac{\sum x}{n} = \frac{75}{5} = 15$ $\qquad S^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{1375}{5} - (\frac{75}{5})^2 = 50$

$S.D = \sqrt{Var(x) = \sqrt{50}} = 7.071$

Co-efficient of S.d

$$Co-efficient\,of\ S.D = \frac{S}{\overline{\overline{X}}} = \frac{7.071}{15} = 0.4714$$

Co-efficient of variation

$$Co-efficient\,of\ \text{var}\,iation = \frac{S}{\overline{\overline{X}}} \times 100 = \frac{7.071}{15} \times 100 = 47.14\%$$

Example.4.10: Find S.D and Co-efficient of S.D and Co-efficient of variation from the following data by 1) Direct ii) Short-cut iii) Coding method

Classes	10-15	15-20	20-25	25-30	30-35
f	2	4	6	8	3

Solution:

x	f	fx	fx^2
12.5	2	25	312.5
17.5	4	70	1225
22.5	6	135	3037.5
27.5	8	220	6050
32.5	3	97.5	3168.75
Total	23	547.5	13793.75

$$\bar{x} = \frac{\sum fx}{\sum f} = 23.80 \quad S^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2 = \frac{13793.75}{23} - \left(\frac{547.5}{23}\right)^2 = 599.728 - 566.65 = 33.08$$

Co-efficient of S.d

$$Co-efficient\,of\ S.D = \frac{S}{\overline{\overline{X}}} = \frac{5.7515.}{23.80} = 0.2417$$

Co-efficient of variation

$$Co-efficient\,of\ \text{var}\,iation = \frac{S}{\overline{\overline{X}}} \times 100 = \frac{5.7515}{23.80} \times 100 = 24.17\%$$