Measures of Dispersion

Quantitative Aptitude & Business Statistics

Why Study Dispersion?

- An average, such as the mean or the median only locates the centre of the data.
- An average does not tell us anything about the spread of the data.

What is Dispersion

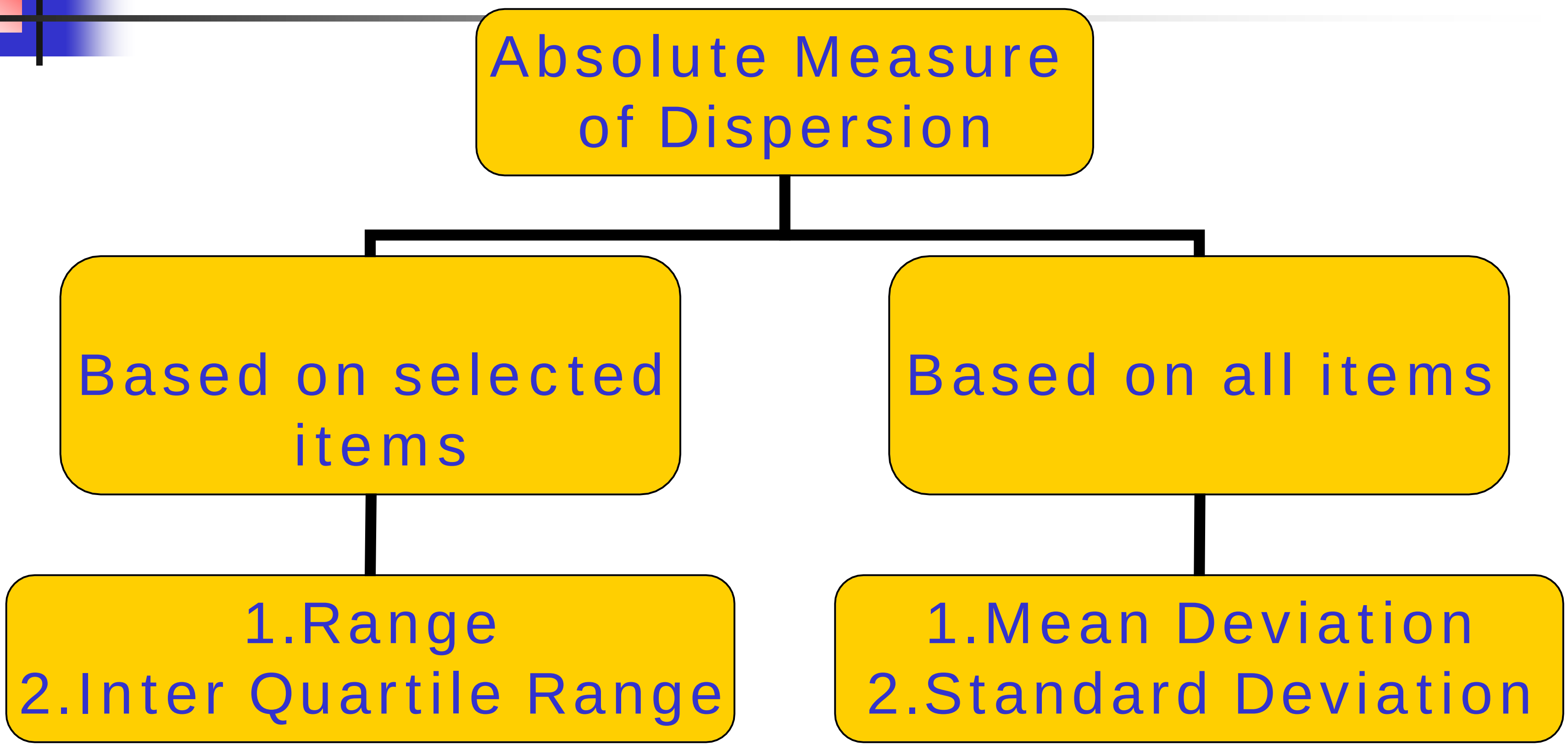
- Dispersion (also known as Scatter ,spread or variation) measures the items vary from some central value .
- It measures the degree of variation.

Significance of Measuring Dispersion

- To determine the reliability of an average.
- To facilitate comparison.
- To facilitate control.
- To facilitate the use of other statistical measures.

Properties of Good Measure of Dispersion

- Simple to understand and easy to calculate
- Rigidly defined
- Based on all items
- A meanable to algebraic treatment
- Sampling stability
- Not unduly affected by Extreme items.


Absolute Measure of Dispersion
Based on selected items
Based on all items
1.Range
2.Inter Quartile Range
1.Mean Deviation
2.Standard Deviation

Relative measures of Dispersion

Based on Selected items

1. **Coefficient of Range**
2. **Coefficient of QD**

Based on all items

1. **Coefficient of MD**
2. **Coefficient of SD & Coefficient of Variation**

- A small value for a measure of dispersion indicates that the data are clustered closely (the mean is therefore representative of the data).
- A large measure of dispersion indicates that the mean is not reliable (it is not representative of the data).

The Range

- The simplest measure of dispersion is the range.
- For ungrouped data, the range is the difference between the highest and lowest values in a set of data.

- RANGE = Highest Value - Lowest Value

- The range only takes into account the most extreme values.
- This may not be representative of the population.

The Range Example

- A sample of five accounting graduates revealed the following starting salaries: 22,000,28,000, 31,000, 23,000, 24,000.
- The range is 31,000 - 22,000 = 9,000.

Coefficient of Range

- Coefficient of Range is calculated as,
- Coefficient of Range =

$$\frac{L-S}{L+S}$$

$$=0.1698$$

- From the following data calculate Range and Coefficient of Range

Marks	**5**	**15**	**25**	**35**	**45**	**55**
No .of Students	**10**	**20**	**30**	**50**	**40**	**30**

- **Largest term (L)=55,**
- **Smallest term (S)=5**
- **Range=L-S=55-5=50**
- **Coefficient of Range**

$$= \frac{L - S}{L + S} = \frac{55 - 5}{55 + 5} = 0.833$$

- **From the following data ,calculate Range and Coefficient of Range**

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No .of Students	10	20	30	50	40	30

- Lower limit of lowest class (S)=0
- Upper limit of highest class (L)=60
- Coefficient of Range

$$= \frac{L - S}{L + S} = \frac{60 - 0}{60 + 0} = 1.000$$

Merits Of Range

- 1.Its easy to understand and easy to calculate.
- 2.It does not require any special knowledge.
- 3.It takes minimum time to calculate the value of Range.

Limitations of Range

- It does not take into account of all items of distribution.
- Only two extreme values are taken into consideration.
- It is affected by extreme values.

Limitations of Range

- It does not indicate the direction of variability.
- It does not present very accurate picture of the variability.

Uses of Range

- It facilitates to study quality control.
- It facilitates to study variations of prices on shares ,debentures, bonds and agricultural commodities.
- It facilitates weather forecasts.

Interquartile Range

- The interquartile range is used to overcome the problem of outlying observations.
- The interquartile range measures the range of the middle (50%) values only

- Inter quartile range = $Q_3 - Q_1$
- It is sometimes referred to as the quartile deviation or the semi-inter quartile range.

Exercise

- The number of complaints received by the manager of a supermarket was recorded for each of the last 10 working days.
- 21, 15, 18, 5, 10, 17, 21, 19, 25 & 28

Interquartile Range Example

Sorted data

- 5, 10, 15, 17, 18, 19, 21, 21, 25 & 28

$$Q_1 = \frac{N+1}{4}$$

$$Q_1 = \frac{11}{4}$$

$$Q_1 = 2.75$$

$$= 2 item + 0.75(15 - 10)$$

$$= 10 + 3.75 = 13.75$$

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

$$Q_3 = \frac{33}{4}$$

$$Q_3 = 8.25 = 8 + 0.25(9th item - 8th item)$$

$$= 21 + 0.25(25 - 21)$$

$$= 21 + 0.25(4) = 23$$

Inter quartile range = 23 – 13.75 = 9.25

- Co-efficient of Quartile Deviation =

$$= \frac{Q3 - Q1}{Q3 + Q1}$$

$$= 0.1979$$

- **From the following data ,calculate Inter Quartile Range and Coefficient of Quartile Deviation**

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No .of Students	10	20	30	50	40	30

Calculation of Cumulative frequencies

Marks	No. of Students	Cumulative Frequencies
0-10	10	10
10-20	20	30 cf
20-30	30f	60Q1 Class
30-40	50	110 cf
40-50	40 f	150Q3Class
50-60	30	180
	180	

L1

L3

- Q_1= Size of N/4th item = Size of 180/4th item = 45th item
- There fore Q_1 lies in the Class 20-30

$$Q_1 = L + \left(\frac{\frac{N}{4} - c.f}{f} \right) \times C$$

$$Q_1 = 20 + \left(\frac{\frac{180}{4} - 30}{30} \right) \times 10$$

$$= 20 + \left(\frac{15}{30} \right) \times 10$$

$$= 25$$

- Q_3= Size of 3.N/4th item = Size of 3.180/4th item = 135th item
- There fore Q_3 lies in the Class 40-50

$$Q_3 = L + \left(\frac{3.\frac{N}{4} - c.f}{f} \right) \times C$$

$$Q3 = 40 + \left(\frac{3.\frac{180}{4} - 110}{40} \right) \times 10$$

$$= 40 + \left(\frac{25}{10} \right) \times 10$$

$$= 46.25$$

- Inter Quartile Range $=Q_3-Q_1$

  $=46.25-25=21.25$

- Coefficient of Quartile Deviation

$$=\frac{Q_3-Q_1}{Q_3+Q_1} \quad =0.2982$$

Merits Of Quartile Deviation

- Its easy to understand and easy to calculate.
- It is least affected by extreme values.
- It can be used in open-end frequency distribution.

Limitations of Quartile Deviation

- It is not suited to algebraic treatment
- It is very much affected by sampling fluctuations
- The method of Dispersion is not based on all the items of the series .
- It ignores the 50% of the distribution.

Mean Deviation

- The mean deviation takes into consideration all of the values
- Mean Deviation: The arithmetic mean of the absolute values of the deviations from the arithmetic mean

$$MD = \frac{\sum \left| x - \bar{x} \right|}{n}$$

Where: X = the value of each observation $\bar{X}$ = the arithmetic mean of the values

n = the number of observations

|| = the absolute value (the signs of the deviations are disregarded)

Mean Deviation Example

- **The weights of a sample of crates containing books for the bookstore are (in kgs.) 103, 97, 101, 106, 103.**
- $\bar{X}$ **= 510/5 = 102 kgs.**
- $\sum |x-\bar{x}|$ **= 1+5+1+4+1=12**
- **MD = 12/5 = 2.4**
- **Typically, the weights of the crates are 2.4 kgs. from the mean weight of 102 kgs.**

Frequency Distribution Mean Deviation

- If the data are in the form of a frequency distribution, the mean deviation can be calculated using the following formula:

$$MD = \frac{\sum f\,|x - \bar{x}|}{\sum f}$$

Where *f* = the frequency of an observation x

n = Σ*f* = the sum of the frequencies

Frequency Distribution MD Example

Number of outstanding accounts	Frequency	fx	|x-x|	f|x-x|
0	1	0	2	2
1	9	9	1	9
2	7	14	0	0
3	3	9	1	3
4	4	16	2	8
Total:	24	Σfx = 48		Σ f|x-x| = 22

$$\bar{x} = \frac{\sum fx}{\sum f}$$

mean = 48/24 = 2

$$MD = \frac{\sum f\,|x - \bar{x}|}{\sum f}$$

$$MD = 22/24 = 0.92$$

Calculate Mean Deviation and Coefficient of Mean Deviation

Marks (X)	No. of Students	Cumulative Frequencies
0-10	10	10
10-20	20	30
20-30	30	60 cf
30-40	50 f	110
40-50	40	150
50-60	30	180
	180	

L

Median Class

- Calculation of Mean deviations from Median $\frac{N}{2} = \frac{180}{2}$

$= 90^{th}$ item

Median in Class = 30-40, L=30, c.f=60, f=50 and c=10

$$M = L + \left(\frac{\frac{N}{2} - c.f}{f} \right) \times c$$

$$= 30 + \left(\frac{\frac{180}{2} - 30}{50} \right) \times 10$$

$$= 30 + 6 = 36$$

Calculation of Mean Deviation of Median

- Mean Deviation from Median

$$= \frac{\sum f \left| x - M \right|}{N}$$

$$= \frac{2040}{180} = 11.333$$

- Coefficient of Mean Deviation

$$= \frac{\text{Mean Deviation}}{\text{Median}}$$

$$= \frac{11.333}{36} = 0.3148$$

Merits of Mean Deviation

- It is easy to understand
- It is based on all items of the series
- It is less affected by extreme values
- It is useful small samples when no detailed analysis is required.

Limitations of Mean Deviation

- It lacks properties such that (+) and(-)signs which are not taken into consideration.
- It is not suitable for mathematical treatment.

- It may not give accurate results when the degree of variability in a series is very high.

Standard Deviation

- Standard deviation is the most commonly used measure of dispersion
- Similar to the mean deviation, the standard deviation takes into account the value of every observation

- The values of the mean deviation and the standard deviation should be relatively similar.

Standard Deviation

- The standard deviation uses the squares of the residuals
- Steps;
  - Find the sum of the squares of the residuals
  - Find the mean
  - Then take the square root of the mean

$$\sigma = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{N}}$$

- From the following data calculate Standard Deviation
- 5,15,25,35,45 and 55

$$\bar{X} = \frac{\sum X}{N}$$

$$\bar{X} = \frac{180}{6} = 30$$

$$\sigma = \sqrt{\frac{\sum (X - \overline{X})^2}{N}}$$

$$= \sqrt{\frac{1750}{6}} = 17.078$$

Frequency Distribution SD

- **If the data are in the form of a frequency distribution the standard deviation is given by**

$$\sigma = \sqrt{\frac{\sum f.(X - \overline{X})^2}{N}}$$

$$= \sqrt{\frac{\sum f.x^2}{N} - \left(\frac{\sum f.x}{N}\right)^2}$$

- **From the following data ,calculate Standard Deviation .**

Marks (X)	5	15	25	35	45	55
No. of Students (f)	10	20	30	50	40	30

- **Mean =** $$\overline{X} = \frac{\sum f.x}{N} = \frac{6300}{180} = 35$$

Marks (X)	No. of Students (f)	(X-35)=x	fx^2
5	10	-30	9000
15	20	-20	8000
25	30	-10	3000
35	50	0	0
45	40	10	4000
55	30	20	12000
	N=180		$\sum fx^2 = 36000$

$$\sigma = \sqrt{\frac{\sum f.x^2}{N} - \left(\frac{\sum f.x}{N}\right)^2}$$

$$= \sqrt{\frac{\sum f.x^2}{N} - \left(\overline{X}\right)^2}$$

$$= \mathbf{14.142}$$

Properties of Standard Deviation

- Independent of change of origin
- Not independent of change of Scale.
- Fixed Relationship among measures of Dispersion.

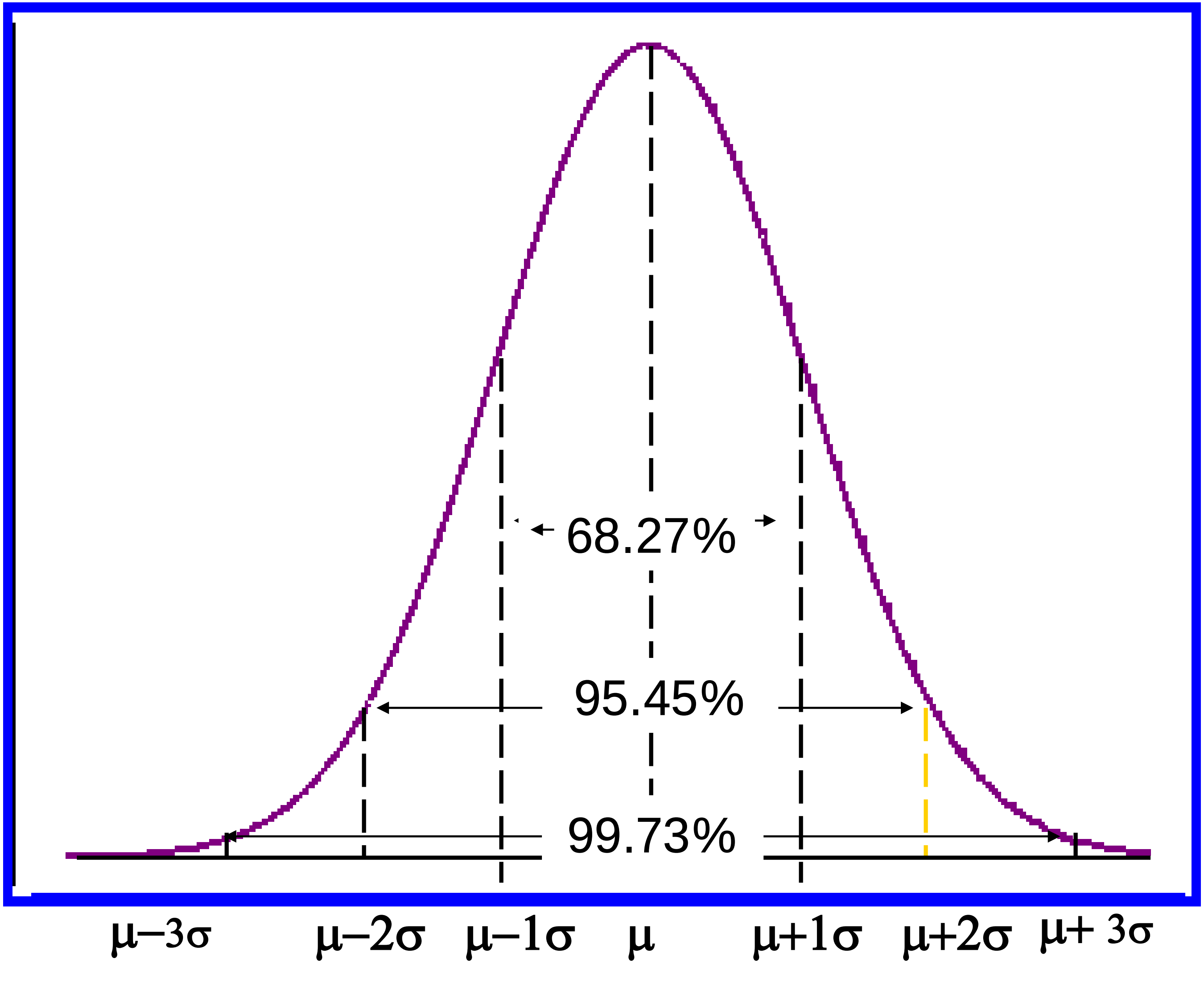
- **In a normal distribution there is fixed relationship**

$$QD = \frac{2}{3}\sigma$$

$$MD = \frac{4}{5}\sigma$$

Thus SD is never less than QD and MD

Bell - Shaped Curve showing the relationship between σ and μ.



- **Minimum sum of Squares; The Sum of Squares of Deviations of items in the series from their arithmetic mean is minimum.**
- **Standard Deviation of n natural numbers**

$$= \sqrt{\frac{N^2 - 1}{12}}$$

- Combined standard deviation

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

- Where σ_{12} =Combined standard Deviation of two groups
- σ_1 =Standard Deviation of first group
- N1=No. of items of First group
- N2=No. of items of Second group
- σ_2 = Standard deviation of Second group

$$d_1 = \overline{X}_1 - \overline{X}_{12}$$

$$d_2 = \overline{X}_2 - \overline{X}_{12}$$

Where $\overline{X}_{12}$ is the combined mean of two groups

Merits of Standard Deviation

- It is based on all the items of the distribution.
- it is a meanable to algebraic treatment since actual + or – signs deviations are taken into consideration.
- It is least affected by fluctuations of sampling

Merits of Standard Deviation

- It facilitates the calculation of combined standard Deviation and Coefficient of Variation ,which is used to compare the variability of two or more distributions
- It facilitates the other statistical calculations like skewness ,correlation.
- it provides a unit of measurement for the normal distribution.

Limitations of Standard Deviation

- It can't be used for comparing the variability of two or more series of observations given in different units. A coefficient of Standard deviation is to be calculated for this purpose.
- It is difficult to compute and compared

Limitations of Standard Deviation

- It is very much affected by the extreme values.
- The standard deviation can not be computed for a distribution with open-end classes.

Variance

- Variance is the arithmetic mean of the squares of deviations of all the items of the distributions from arithmetic mean .In other words, variance is the square of the Standard deviation= σ^2

- Variance= $\sigma = \sqrt{variance}$

Interpretation of Variance

- Smaller the variance ,greater the uniformity in population.
- Larger the variance ,greater the variability

The Coefficient of Variation

- The coefficient of variation is a measure of relative variability
  - It is used to measure the changes that have taken place in a population over time

- To compare the variability of two populations that are expressed in different units of measurement
- It is expressed as a percentage

- Formula: $$CV = \frac{\sigma}{\bar{X}} \times 100$$

Where:

$\bar{X}$ = mean

σ = standard deviation

- 1. Dispersion measures

(a) The scatter ness of a set of observations

(b) The concentration of set of observations

(c) The Peaked ness of distribution

(d) None

- 1. Dispersion measures

(a) The scatter ness of a set of observations

(b) The concentration of set of observations

(c) The Peaked ness of distribution

(d) None

- 2. Which one is an absolute measure of Dispersion?

  (a) Range

  (b) Mean Deviation

  (c) Quartile Deviation

  (d) all these measures

- 2. Which one is an absolute measure of Dispersion?

  (a) Range

  (b) Mean Deviation

  (c) Quartile Deviation

  (d) all these measures

- 3.Which measures of Dispersion is not affected by the presence of extreme observations

(a) deviation

(b) Quartile Deviation

(C) Mean Deviation

(d) Range

- 3.Which measures of Dispersion is not affected by the presence of extreme observations

(a) deviation

(b) Quartile Deviation

(C) Mean Deviation

(d) Range

- 3.Which measures of Dispersion is not affected by the presence of extreme observations

(a) deviation

(b) Quartile Deviation

(C) Mean Deviation

(d) Range

- 4.which measures of Dispersion is based on all the items of observations

(a) Mean Deviation

(b) Standard Deviation

(C) Quartile Deviation

(d) a and b but not c

- 4.which measures of Dispersion is based on all the items of observations

(a) Mean Deviation

(b) Standard Deviation

(C) Quartile Deviation

(d) a and b but not c

- 5.Standard Deviation is
- (a) absolute measure
- (b) relative measure
- (c) both
- (d) none

- 5.Standard Deviation is
- (a) absolute measure
- (b) relative measure
- (c) both
- (d) none

- 6.Coefficient of standard deviation is
- (a) SD/mean
- (b) SD/Median
- (c) SD/Mode
- (d) Mean/SD

- 6.Coefficient of standard deviation is
- (a) SD/mean
- (b) SD/Median
- (c) SD/Mode
- (d) Mean/SD

- **7.Coefficient of Quartile Deviation is calculated by formula**
- **(a)(Q3-Q1)/4**
- **(b) (Q3-Q1)/2**
- **(c) (Q3-Q1)/(Q3+Q1)**
- **(d) (Q3+Q1)/(Q3-Q1**

- **7.Coefficient of Quartile Deviation is calculated by formula**
- **(a)(Q3-Q1)/4**
- **(b) (Q3-Q1)/2**
- **(c) (Q3-Q1)/(Q3+Q1)**
- **(d) (Q3+Q1)/(Q3-Q1**

- **8.The standard Deviation of 5,8,5,5,5,8and 8 is**
- **(a) 4**
- **(b) 6**
- **(C) 3**
- **(d) 0**

- **8.The standard Deviation of 5,8,5,5,5,8and 8 is**
- **(a) 4**
- **(b) 6**
- **(C) 3**
- **(d) 0**

- 9.If all the observations are increased by 10,then

(a) SD would be increased by 10

(b)Mean deviation would be increased by 10

(c) Quartile Deviation would be increased by 10

(d) all these remain unchanged

- 9.If all the observations are increased by 10,then

(a) SD would be increased by 10

(b)Mean deviation would be increased by 10

(c) Quartile Deviation would be increased by 10

(d) all these remain unchanged

- 10.For any two numbers SD is always
- (a) Twice the range
- (b) Half of the range
- © Square of range
- (d) none of these

- 10.For any two numbers SD is always
- (a) Twice the range
- (b) Half of the range
- © Square of range
- (d) none of these

- 11Mean deviation is minimum when deviations are taken about
- (a) Arithmetic Mean
- (b) Geometric Mean
- © Harmonic Mean
- (d) Median

- 11.Mean deviation is minimum when deviations are taken about
- (a) Arithmetic Mean
- (b) Geometric Mean
- © Harmonic Mean
- (d) Median

- 12.Root mean square deviation is
- (a) Standard Deviation
- (b) Quartile Deviation
- © both
- (d) none

- 12.Root mean square deviation from mean is
- (a) Standard Deviation
- (b) Quartile Deviation
- © both
- (d) none

- **13.Standard Deviation is**
- (a) Smaller than mean deviation about mean
- (b) Smaller than mean deviation about median
- © Larger than mean deviation about mean
- (d) none of these

- **13.Standard Deviation is**
- (a) Smaller than mean deviation about mean
- (b) Smaller than mean deviation about median
- © Larger than mean deviation about mean
- (d) none of these

- 14.Is least affected by sampling fluctuations
- a) Standard Deviation
- (b) Quartile Deviation
- © both
- (d) none

- 14.Is least affected by sampling fluctuations
- a) Standard Deviation
- (b) Quartile Deviation
- © both
- (d) none

- 15. Coefficient of variation of two series is 60% and 80% respectively. Their standard deviations are 20 and 16 respectively, what is their A.M
- A) 15 and 20
- B) **33.3 and 20**
- C) 33.3 and 15
- D) 12 and 12.8

- 15. Coefficient of variation of two series is 60% and 80% respectively. Their standard deviations are 20 and 16 respectively, what is their A.M
- A) 15 and 20
- B) **33.3 and 20**
- C) 33.3 and 15
- D) 12 and 12.8

- 16. For the numbers 1, 2, 3, 4, 5, 6, 7 standard deviation is:
- A) 3
- B) 4
- C) **2**
- D) None of these

- 16. For the numbers 1, 2, 3, 4, 5, 6, 7 standard deviation is:
- A) 3
- B) 4
- C) **2**
- D) None of these

THE END

Measures of Dispersion