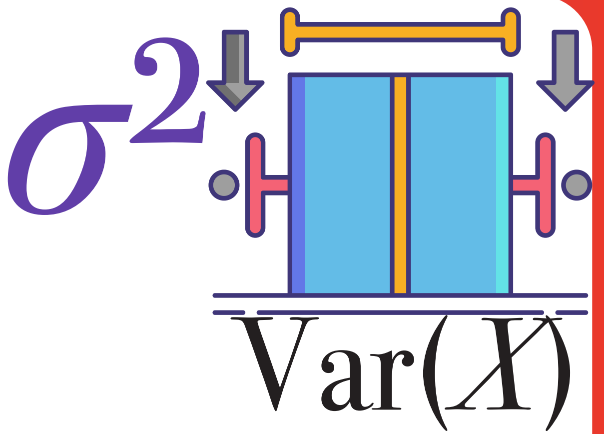
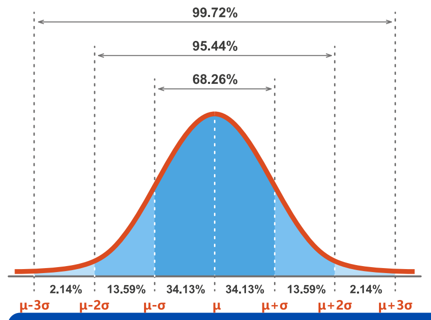


Day 3
Data Management and Analysis

Session 3
Measures of Dispersion



$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{n}$$

IQR

Session Outcome

After completing this session, researchers will be able to

- Understand the Concept of measures of dispersion and their Importance
- Understand Range, quartiles, interquartile range (IQR), mean absolute deviation (MAD), variance and standard deviation as measures of dispersion
- Explain the Formulas and Procedures to Calculate different measuress of dispersion

Session Outline

- Definition and Importance of Measure of Dispersion
- Range and Interquartile Range
- Mean Deviation
- Variance, Standard Deviation and Coefficient of Variation

Measure of Dispersion

Central tendency measures do not reveal the variability present in the data. Dispersion is a measure of variation. Dispersion is the scatteredness of the data series around it average. As per Bowley,

"Dispersion is a measure of the variation of the items".

Dispersion is the extent to which value in a distribution differ from the average of the distribution.

Objectives of Measuring Dispersion

- To determine the reliability of an average
- To compare the variability of two or more series
- For facilitating the use of other statistical measures
- Basis of statistical quality control

Measure of Dispersion

Properties of A Good Measure of Dispersion

- Easy to understand
- Simple to calculate
- Uniquely defined
- Based on all observations
- Not affected by extreme observations
- Capable of further algebraic treatment

Importance of Measure of Dispersion

1. **Identifying Performance Variability:** They reveal the spread of student performance, aiding in identifying those needing support or enrichment.
2. **Assessing Program Quality:** Dispersion analysis helps evaluate teaching effectiveness and curriculum alignment.

Importance of measures of central tendency:

Importance of Measure of Dispersion

3. Addressing Achievement Gaps: High dispersion highlights achievement disparities among demographic groups, guiding targeted interventions.

4. Evaluating Assessment Tools: Dispersion indicates assessment reliability and validity, informing improvements.

5. Resource Allocation: Understanding performance variability guides resource allocation for tailored support.

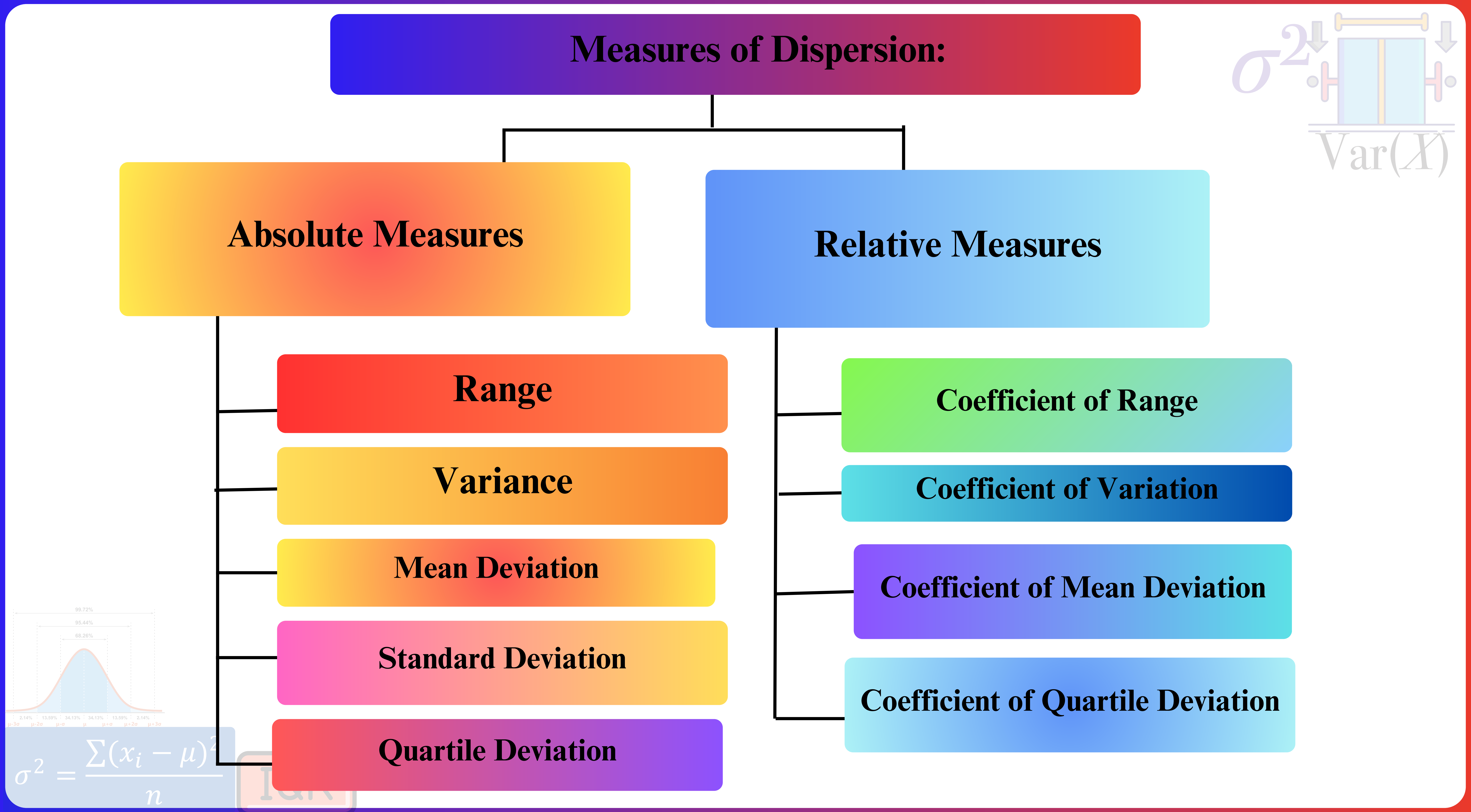
6. Monitoring Progress: Dispersion trends inform goal setting and instructional adjustments.

7. Comparative Analysis: Measures aid in benchmarking and comparative analysis for continuous improvement.

8. Research and Policy: They support evidence-based policymaking and research on educational phenomena.

Classifications of Measures of Dispersion:

- Measures of Dispersion
  - Absolute Measures
    - Range
    - Variance
    - Standard Deviation
    - Mean Deviation
    - Quartile Deviation
  - Relative Measures
    - Coefficient of Range
    - Coefficient of Variation
    - Coefficient of Mean Deviation
    - Coefficient of Quartile Deviation



Measures of Dispersion:
Absolute Measures
Relative Measures
Range
Variance
Mean Deviation
Standard Deviation
Quartile Deviation
Coefficient of Range
Coefficient of Variation
Coefficient of Mean Deviation
Coefficient of Quartile Deviation

Range

1. Range

Definition: Range in statistics is the difference between the highest and lowest values in a dataset. Range offers a straightforward measurement of the data's spread or variability.

Range Formula:

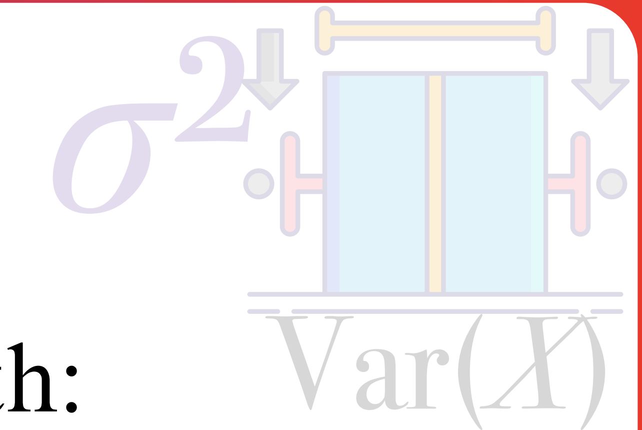
Range=L-S

Where, L=Largest Observation and S=Smallest Observation.

Coefficient of Range

The coefficient of range is given as

$$Coe.of\ R = \frac{L - S}{L + S}$$



Range

Example: Consider the following dataset of exam scores for a class tenth:

77 89 92 64 78 95 82

Find the Range of the above data.

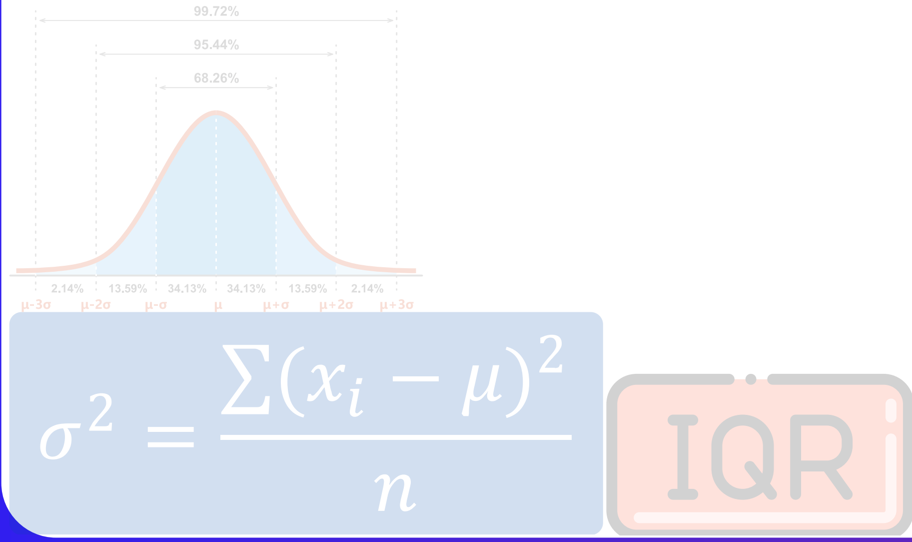
Range=Largest Value- Smallest Value

Range=95-64=31

So, the range of the exam scores in this dataset is 31.

And the coefficient of range is given by

$$Coe.\,of\ R = \frac{L - S}{L + S} = \frac{95 - 64}{95 + 64} = 0.194$$



Range

Range in Grouped Data:

Class Interval	Frequency
0-10	12
10-20	10
20-30	15
30-40	13
40-50	11

Solution:

Range =

Upper Limit of the Last Class Interval – Lower Limit of the First Class Interval

= 50-0 = 50

Range

Advantages of Range:

- The range is calculated easily since it is based on only two observations.
- It is easily understood
- It is rigidly defined.

Disadvantages of Range:

- The range is not based on all observations and hence does not depict the variability of all observations.
- The amount of range is affected by the extreme values.
- It is affected much by sampling fluctuations.
- It is not calculated from frequency distribution with open-end classes.

Range

Uses of Range:

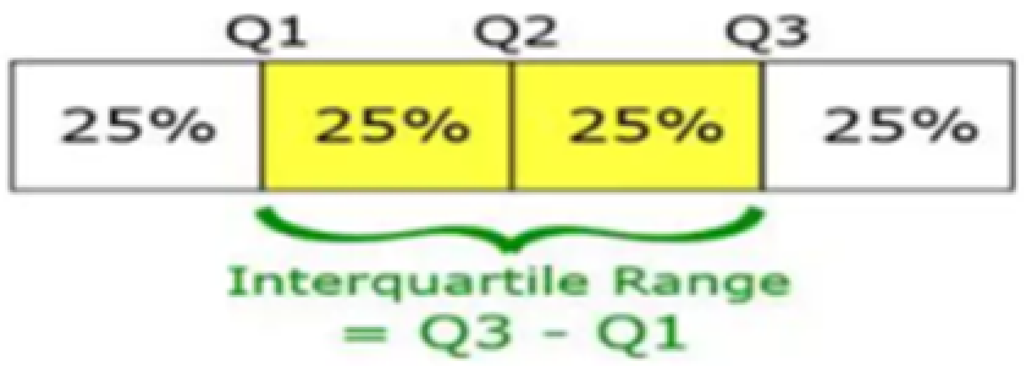
- Assessing student performance variability.
- Informing curriculum development and differentiation.
- Establishing benchmarks and setting goals.
- Identifying achievement gaps.
- Designing assessments.
- Allocating resources effectively.
- Communicating student progress to parents.
- Conducting research and program evaluation.

Interquartile Range

2. Interquartile Range

Definition: The interquartile range (IQR) is a measure of statistical dispersion, specifically a measure of variability in a dataset. It's calculated as the difference between the upper quartile (Q3) and the lower quartile (Q1) in a dataset.

observation divides the data into four equal part denoted as Q_1, Q_2, Q_3



Q_1 = First Quartile $=\left(\frac{n+1}{4}\right)$ *th ordered observation*

Q_2 = Second Quartile = $2*\left(\frac{n+1}{4}\right)$ *th orderedobservation* =Median

Q_3 = Third Quartile = $3*\left(\frac{n+1}{4}\right)$ *th orderedobservation*

Interquartile Range

Example: Let's say we have the following dataset:

5 6 8 10 12 15 18 20 22 25 5 6 8
10 12 15 18 20 22 25

Solution:

To calculate the interquartile range (IQR), we first need to find the first quartile (Q1) and the third quartile (Q3).

Arrange the dataset in ascending order:

5 6 8 10 12 15 18 20 22 25

Find the median (Q2):

Interquartile Range

Since we have 10 numbers, the median (Q_2) is the average of the 5th and 6th numbers:

$$Q_2 = \frac{12+15}{2} = 13.5$$

Find the median of the lower half (Q_1):

$$Q_1 = \frac{6+8}{2} = 7$$

Find the median of the upper half (Q_3):

$$Q_3 = \frac{18+20}{2} = 19$$

Calculate the interquartile range (IQR): IQR=Q3-Q1=19-7=12

So, the interquartile range (IQR) for this dataset is 12.

Mean Deviation

3. Mean Deviation

Definition: The mean Deviation of a series can be defined as the arithmetic average of the deviations of various items from a measure of central tendency (mean, median, or mode). Mean Deviation is also known as the First Moment of Dispersion or Average Deviation.

Formula for Mean Deviation

For ungrouped data:-

Mean deviation (MD) = $\frac{\Sigma|x-\bar{x}|}{n}$

For grouped data:-

Mean deviation (MD) = $\frac{\Sigma f * |x-\bar{x}|}{\Sigma f}$ where $\bar{x}$= mean

f=frequency

Mean Deviation

Advantages

- Simple to Compute
- Easy to Understand
- Less Affected by Extreme Values
- It has A Fixed Value

Disadvantages

- Not useful for Algebric Calculation
- Not Appplicable for Open End Class Series
- Ignores Negative Signs
- Difficult to Determine in case of Fractiond

Coefficient of Mean Deviation

Coefficient of Mean Deviation

Mean Deviation is an absolute measure of dispersion. In order to transform it into a relative measure, it is divided by the average, from which it has been calculated. It is known as the Coefficient of Mean Deviation.

Coefficient of Mean Deviation from Mean

$$(MD_{\bar{X}}) = \frac{MD_{\bar{X}}}{\bar{x}}$$

Example :

Calculate mean deviation from mean and coefficient of mean deviation.

Classes	20-40	40-80	80-100	100-120	120-140
Frequency	3	6	20	12	9

Solution:

Coefficient of Mean Deviation

Classes(X)	Frequency(f)	Mid-point(m)	fm	$\lvert D\rvert = \lvert m-\bar{X}\rvert$	$f\lvert D\rvert$
20-40	3	30	90	64.8	194.4
40-80	6	60	360	34.8	208.8
80-100	20	90	1800	4.8	96
100-120	12	110	1320	15.2	182.4
120-140	9	130	1170	35.2	316.8
			$\sum fm = 4740$		$\sum f\lvert D\rvert = 998.4$

Mean $(\bar{X}) = \frac{\sum fm}{\sum f} = \frac{4740}{50} = 94.8$

Mean Deviation from Mean $(MD_{\bar{X}}) = \frac{\sum f|D|}{N} = \frac{998.4}{50} = 19.968$

Coefficient of Mean Deviation from Mean $= \frac{MD_{\bar{X}}}{\bar{X}} = \frac{19.968}{94.8} = 0.210$

Uses of Mean Deviation:

- Mean deviation is frequently used in studying the distribution of personal wealth of a nation.
- It is Used in analysis related to forecasting business cycle.

Variance

Variance: Variance is defined as, “The measure of how far the set of data is dispersed from their mean value”. Variance is represented with the symbol σ2. In other words, we can also say that the variance is the average of the squared difference from the mean.

Variance Formula

There are two formulas for Variance, that are:

- Population Variance
- Sample Variance

The formula for Population Variance:

The mathematical formula to find the variance of the given data is,

Variance

$$\sigma^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{N}$$

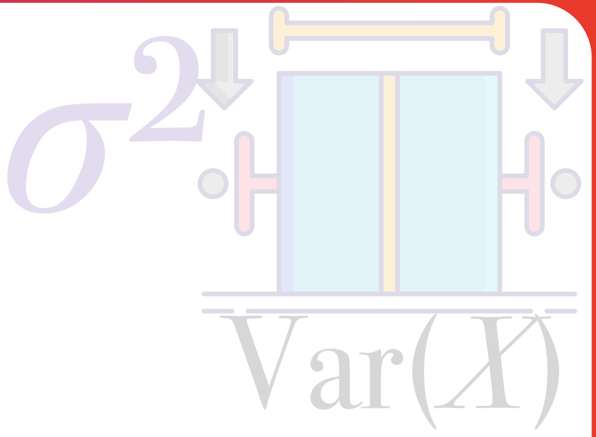
Where,

- σ^2 is the variance of the Population,
- N is the Number of Observation in the Population,
- X_i is the i^{th} observation in the Population, and
- μ is the mean of the Population.

Formula for Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Variance

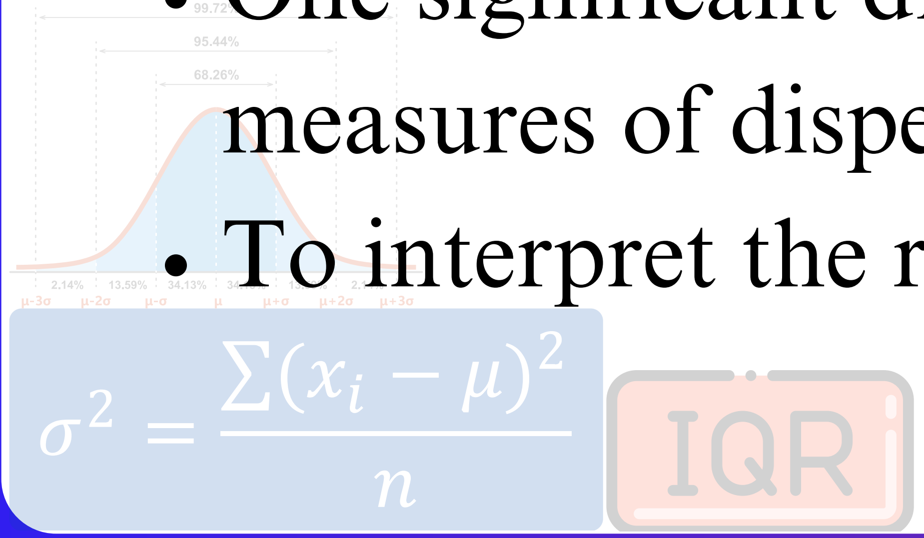


Where,

- s^2 is the variance of the Sample,
- n is the Number of data Point
- X_i is the values of the data
- $\bar{X}$ is the mean of X_i

Drawback of Variance

- One significant drawback of variance is that it is not as intuitively interpretable as other measures of dispersion, such as the range or standard deviation.
- To interpret the result, we use the standard deviation.



Standard Deviation

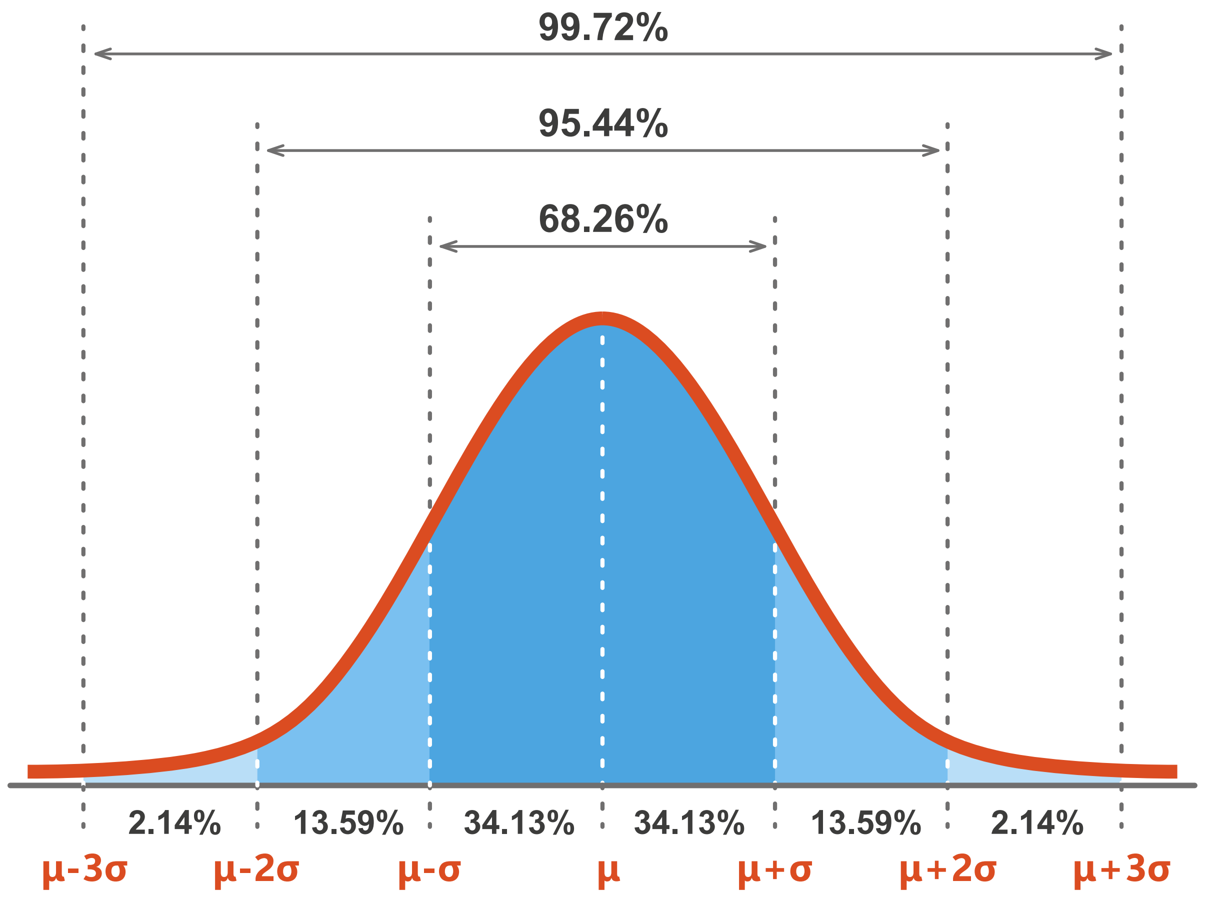
Standard Deviation

Definition: How far our given set of data varies along with the mean of the data is measured in standard deviation. Thus, we define standard deviation as the

“spread of the statistical data from the mean or average position”.

We denote the standard deviation of the data using the symbol σ.

We can also define the standard deviation as the square root of the variance.



Standard Deviation

Population Standard Deviation

The mathematical formula to find the standard deviation of the given data is,

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{N}}$$

Where,

- σ is the Standard Deviation of the Population,
- N is the Number of Observation in the Population,
- X_i is the i^{th} observation in the Population, and
- μ is the mean of the Population.

Standard Deviation

Sample Standard Deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Where,

- σ is the Standard Deviation of the Population,
- N is the Number of Observation in the Population,
- X_i is the i^{th} observation in the Population, and
- μ is the mean of the Population.

Standard Deviation

Sample Standard Deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Where,

- s is the standard deviation of sample
- n is the number of data point
- X_i is the values of the data
- $\bar{X}$ is the mean of X_i

Standard Deviation

Advantages of Standard Deviation

- It is rigidly defined
- Based on all the items in the series
- It is less affected by the sampling fluctuation
- It is suitable for Algebraic operation and statistical analysis

Disadvantages of Standard Deviation

- It is a complex method
- It is difficult to understand
- It is affected by the extreme values

Standard Deviation

Uses of Standard Deviation:

- It is a good measure to study the formation of observations in a distribution
- The relative measure corresponding to standard deviation is used to compare the dispersion of two or distribution
- The standard deviation of an estimate is used to have an idea of the precision of the estimate.

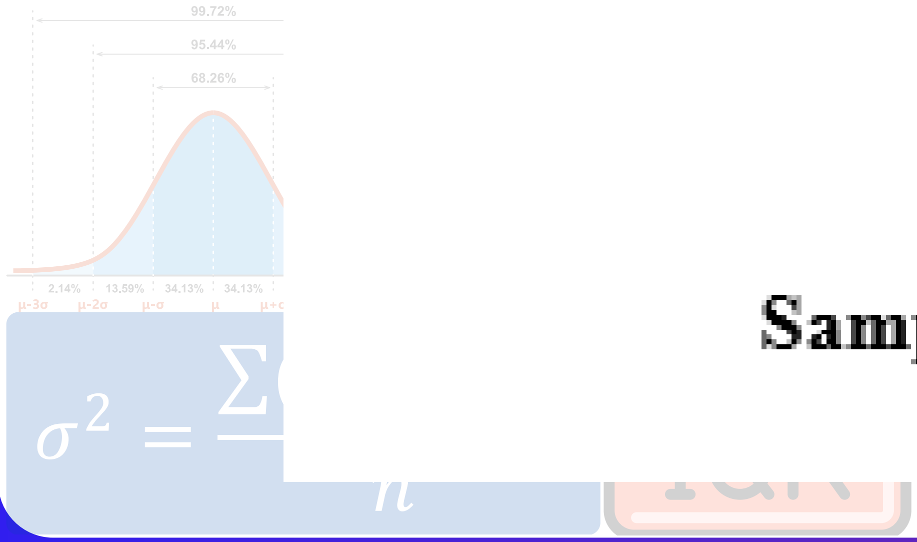
Coefficient of variation

Coefficient of variation

The coefficient of variation (CV) is a statistical measure used to quantify the relative variability or dispersion of a dataset compared to its mean. It is particularly useful for comparing the variability of datasets that have different units or scales.

Coefficient of Variation Formulas

	Coefficient of Variation	Standard Deviation
Population	$\frac{\sigma}{\mu} \times 100$	$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{N}}$
Sample	$\frac{s}{\mu} \times 100$	$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$



Coefficient of variation

Example: Let's consider a hypothetical scenario where we have the population data of math test scores for 10 students in a primary school. We will use this data to calculate the population variance, standard deviation, and coefficient of standard deviation.

Student ID	Math Score (x_i)
1	85
2	90
3	78
4	92
5	80
6	88
7	85
8	82
9	90
10	86

Coefficient of variation

Example: Let's consider a hypothetical scenario where we have the population data of math test scores for 10 students in a primary school. We will use this data to calculate the population variance, standard deviation, and coefficient of standard deviation.

Solution: Now, let's calculate the variance, standard deviation, and coefficient of standard deviation for these math scores.

Student ID	Math Score (x_i)
1	85
2	90
3	78
4	92
5	80
6	88
7	85
8	82
9	90
10	86
	$\sum_{i=1}^{n} x_i = 856$

Coefficient of variation

1. Calculate the Mean (Average):

$$\text{Mean } \mu = \frac{\sum_{i=1}^{n} x_i}{N} = \frac{856}{10} = 85.6$$

2. Calculate Population Variance

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{N}$$

$$= \frac{(85-85.6)^2+(90-85.6)^2+\ldots+(86-85.6)^2}{10}$$

$$= 31.664$$

3. Calculate Population Standard Deviation:

$$\text{Population Standard Deviation } \sigma = \sqrt{31.664} \approx 5.63$$

4.Calculate Coefficient of Population Standard Deviation:

$$\text{Coefficient of Population Standard Deviation} = \frac{\textit{Population Standard Deviation}}{\textit{Mean}}$$

$$\text{Coefficient of Population Standard Deviation} = \frac{5.63}{85.6} \approx 0.0658$$

Coefficient of variation

Interpretation:

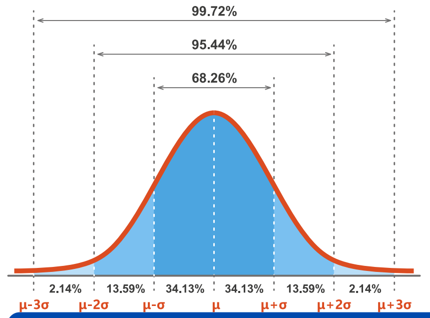
- The variance of the math test scores is approximately 31.664.
- The standard deviation of the math test scores is approximately 5.63.
- The coefficient of standard deviation indicates that the standard deviation is approximately 6.58% of the mean score.

Uses of Coefficient of Variation:

In the educational sector, the coefficient of variation (CV) is used to:

1. Ensure assessment consistency.
2. Evaluate teaching effectiveness.
3. Identify student progress.
4. Control quality in assessment tools.
5. Monitor educational programs.

Thank You



$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{n}$$

IQR