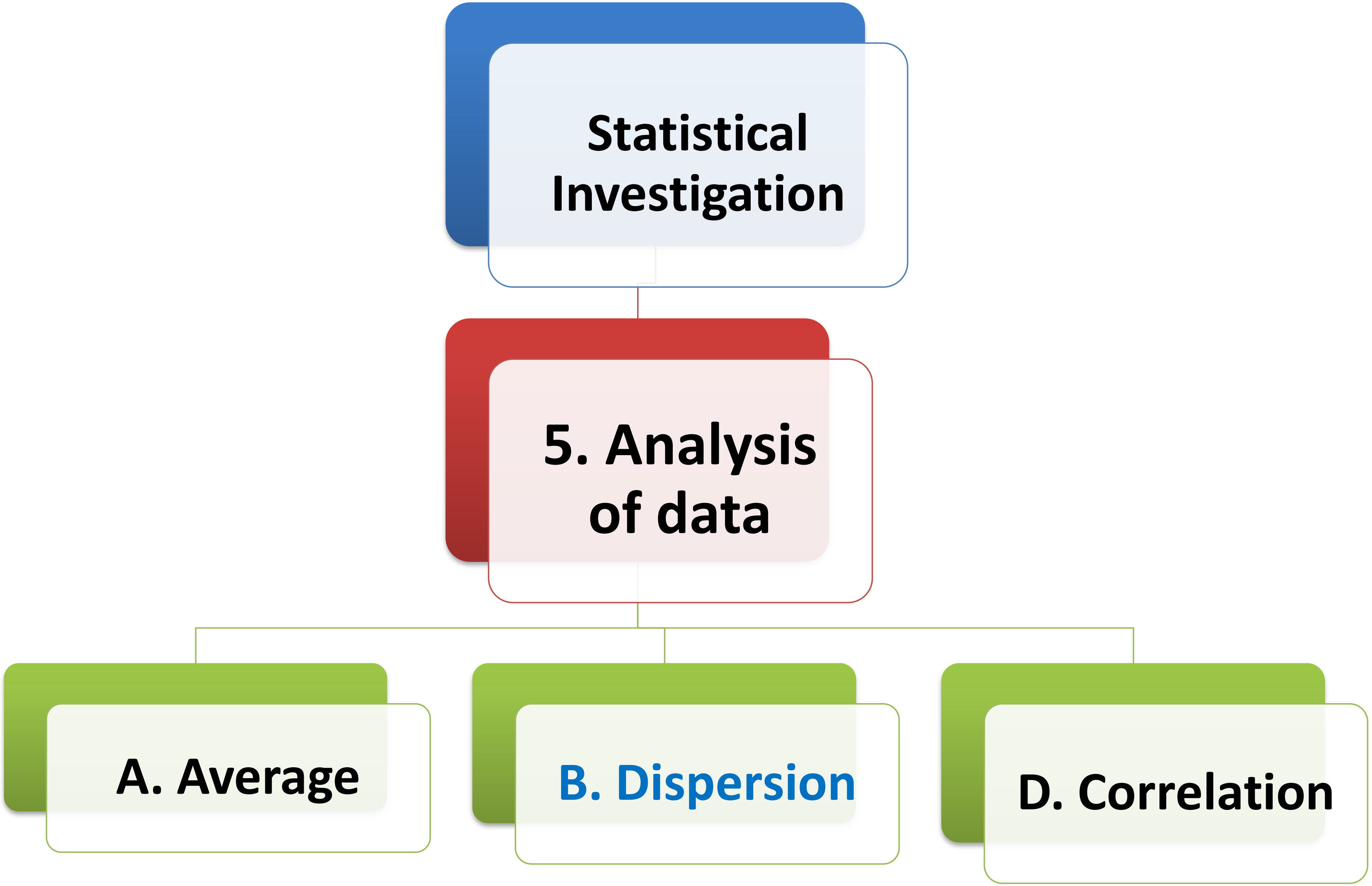
Measures of Dispersion (Deviation)

Mean Deviation, Standard Deviation and Error

Variance


Statistical Investigation
5. Analysis of data
A. Average
B. Dispersion
D. Correlation

Similarities? Differences?

	S1	S2	S3	S4	S5	Total	Mean
A	16	16	16	16	16	80	16
B	14	15	16	17	18	80	16
C	3	8	17	24	28	80	16

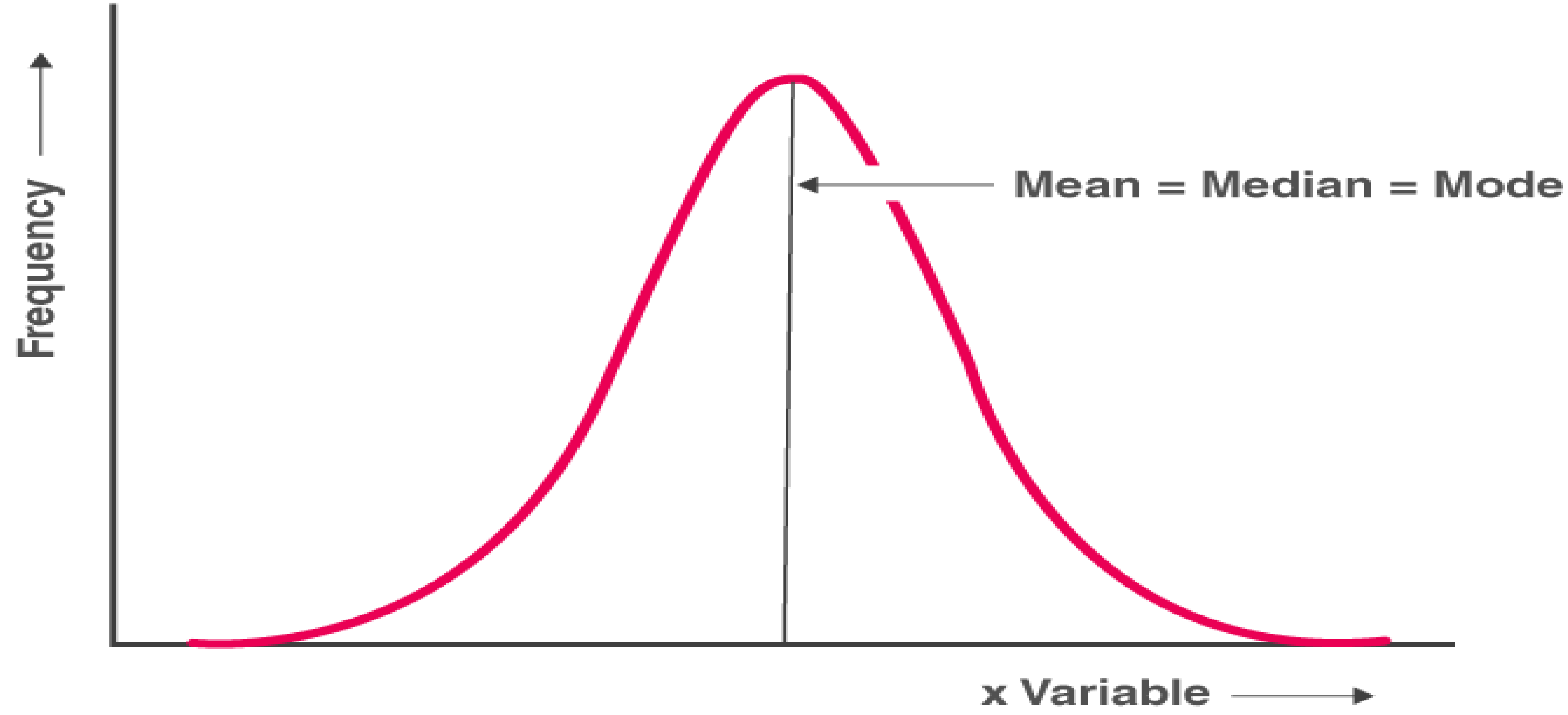
Normal Distribution

- The area under a normal curve has a *normal distribution* (Gaussian distribution)
- Properties of a normal distribution
  - It is symmetric about its mean
  - The highest point is at its mean
  - The height of the curve decreases as one moves away from the mean in either direction, approaching, but never reaching zero

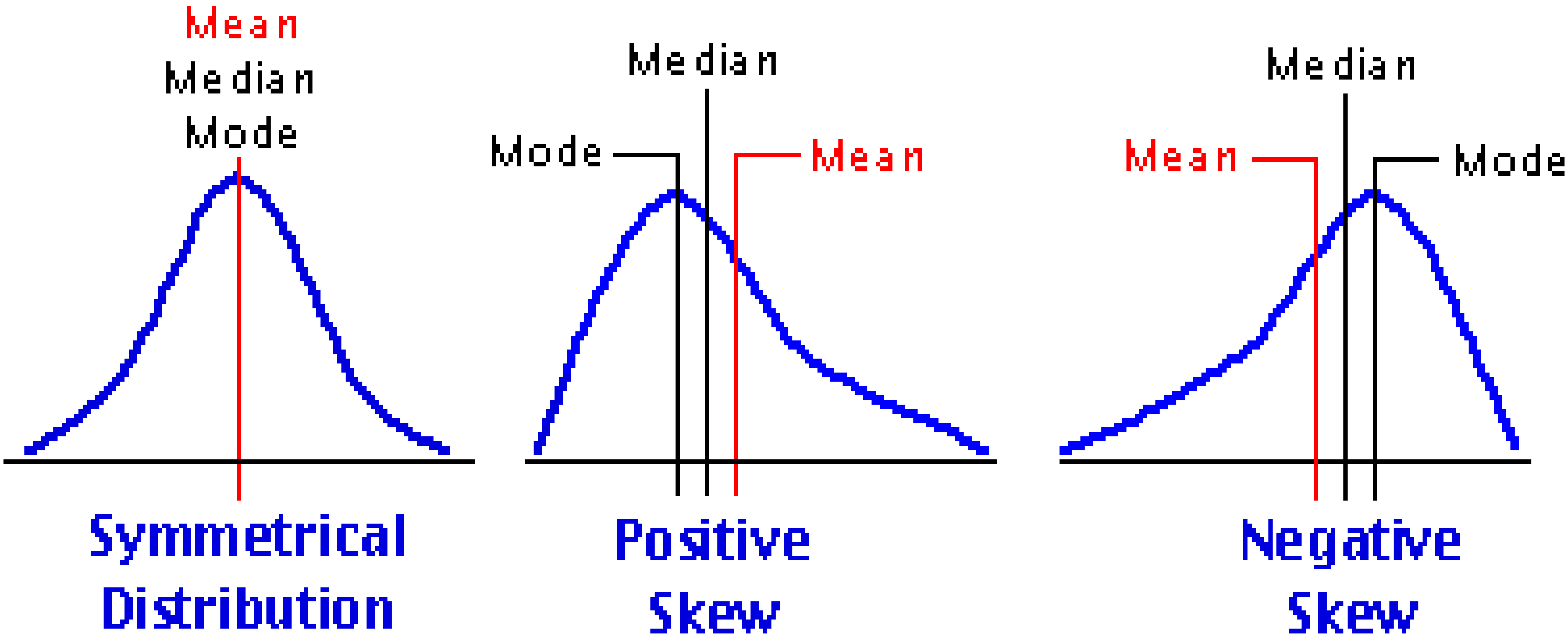
Skewed Distribution

- The data are not distributed symmetrically in skewed distributions
  - Consequently, the mean, median, and mode are not equal and are in different positions
  - Scores are clustered at one end of the distribution
  - A small number of extreme values are located in the limits of the opposite end
- Skew is always toward the direction of the longer tail
  - Positive if skewed to the right
  - Negative if to the left

Normal Distribution



Skewed Distribution



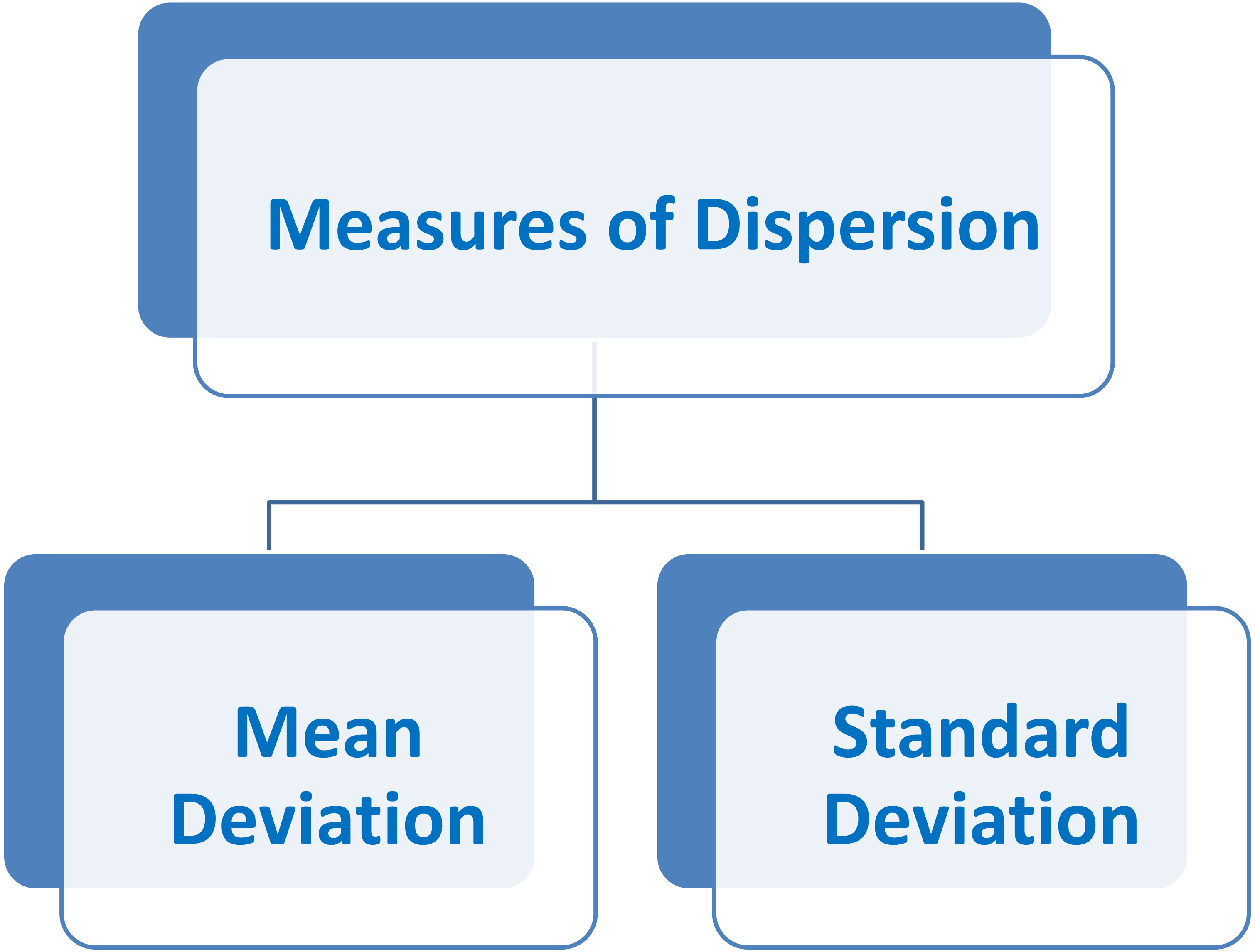
Measures of Dispersion

- It is the degree of the scatter
- It is the measure of the variations of the items
- It is the degree to which numerical data tend to spread about an average value

Measures of Dispersion

Objectives –

1. To determined the reliability of an average
2. To serve as a basis for the control of variability
3. To compare two or more series with regard to their variability
4. To facilitate the use of other statistical measures


Measures of Dispersion
Mean Deviation
Standard Deviation

Merits and Demerits of Mean Deviation

Merits	Demerits
• Readily understood • Based on all observations • Less affected by extreme items • Better measure for comparison • Flexible because can also be calculated by Median	• Not used for further mathematical treatments • Can not be calculated in open end classes • May not give accurate results

Mean Deviation/ Average Deviation

- It is based on all the items of a distribution
- It is an average amount of scatter of items in a distribution

$$\frac{\Sigma(x - \bar{x})}{n} \quad \text{or} \quad \frac{\Sigma(d\bar{x})}{n}$$

Standard Deviation

- It is introduced by Karl Pearson in 1893
- It is positive square root of the arithmetic mean of the squares of the deviation of the given observations
- Represented by Greek letter sigma 'σ'

1. Direct method – Deviation taken from actual mean
2. Short cut method – Deviation taken from assumed mean

Direct method

1. Find out the actual mean ($\bar{x}$)
2. Find out deviation from mean ($x - \bar{x}$)
3. Square the deviation of each value and take the total of squared deviations ($\Sigma(x-\bar{x})^2$)
4. Divide the total by the number of observations $\frac{\Sigma(x-\bar{x})^2}{N}$
5. Find the square root of the product

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

Short cut method

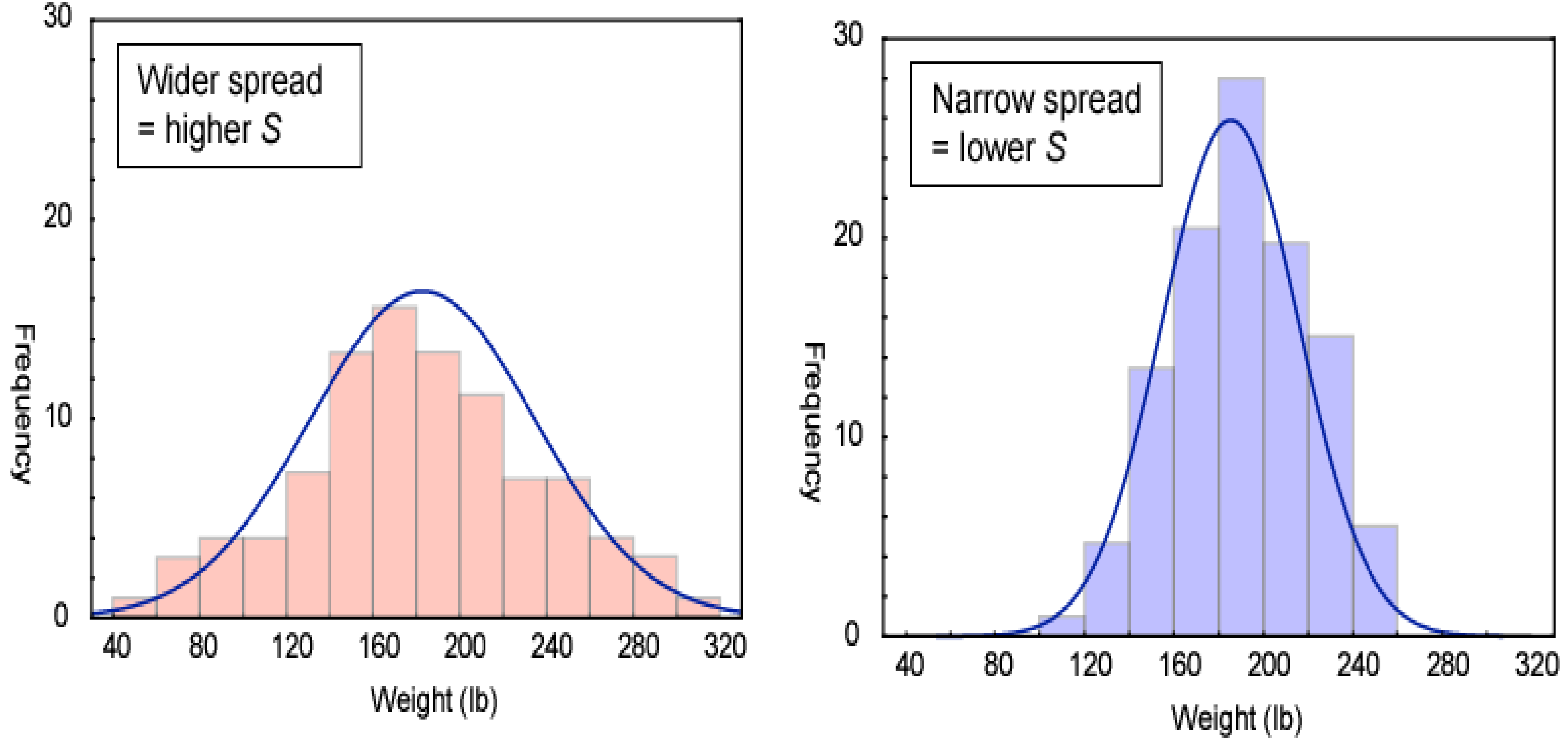
1. Take the assumed mean (A)
2. Find deviation from assumed mean (x-A=d)
3. Square the deviation of each value (d^2) and take the total of squared deviations ($\Sigma\ d^2$)
4. Apply the formula

$$\sqrt{\frac{\Sigma d2}{N}} - \left[\frac{\Sigma d}{N}\right]2$$

Standard Deviation

Merits	Demerits
• Rigidly defined and based on all observations • Most important for authenticity of data and widely used • Possible for further algebraic treatments • Squaring the deviations make all of them positive	• Not easy to understand • Affected by the values of every items in the series

Wide spread results in higher SDs
narrow spread in lower SDs



Standard Error

It is a statistical constant which measures the dispersion of the sample means around the total population mean

$$\frac{SD}{\sqrt{N}}$$

Variance

It is the square of the standard deviation

$$\text{Variance} = SD^2$$

$$SD^2 = \text{Variance}$$

$$SD = \sqrt{\text{Variance}}$$

Coefficient of Variation

It was first developed by Karl Pearson

- The standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated
- But the deviation from two different units can not be compared (weight in g and length in cm)
- Therefore, the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison
- This is known as Coefficient of variation (CV)

- **If the coefficient of variation is greater, then it is said to the group is more variable, less stable, less uniform or less homogenous**

$$\text{Coefficient of variation} = \frac{Satndard\ Deviation}{Mean} \times 100$$

- **It is most useful in comparing the variability of several different samples each with different mean**

Ex . Air pollution study at different sites and at different time intervals