BIOSTATISTICS

- <u>STATISTICS</u> : is the science of compiling, classifying and tabulating numerical data & expressing the results in a mathematical or graphical form.

- <u>BIOSTATISTICS</u> : is that branch of statistics concerned with mathematical facts & data related to biological events.

MEASURES OF CENTRAL TENDENCY

- A single estimate of a series of data that summarizes the data - *the measure of central tendency.*

- **Objective:**

- To condense the entire mass of data.

- To facilitate comparison.

PROPERTIES

- Should be easy to understand and compute.

- Should be based on each and every item in the series.

- Should not be affected by extreme observations.

- Should be capable of further statistical computations.

- Should have sampling stability.

- The most common measures of central tendeny that are used in dental sciences are :

✓ **Arithmetic mean** – mathematical estimate.

✓ **Median** – positional estimate.

✓ **Mode** – based on frequency.

Arithmetic mean

- Simplest measure of central tendency.

- **Ungrouped data**:

  ✓ Mean $=\dfrac{\text{Sum of all the observations in the data}}{\text{Number of observations in the data}}$

- **Grouped data**:

  ✓ Mean $=\dfrac{\text{Sum of all the variables multiplied by the corresponding frequency in the data}}{\text{Total frequency}}$

MEDIAN:-

- *Middle value in a distribution such that one half of the units in the distribution have a value smaller than or equal to the median and one half have a value greater than or equal to the median.*

- All the observations are arranged in the order of the magnitude.

- Middle value is selected as the median.

- Odd number of observations : $(n+1)/2$.

- Even number of observations: mean of the middle two values is taken as the mean.

MODE

- *The mode or the modal value is that value in a series of observations that occurs with the greatest frequency.*

- When mode is ill defined, it can be calculated using the relation

✓ **Mode = 3 median – 2 mean**

- Most commonly used: arithmetic mean.

- Extreme values in the series : median.

- To know the value that has high influence in the series: mode.

Measures of dispersion

- Dispersion is the **degree of spread** or **variation of the variable** about a central value.

- **Measures of dispersion used**:

- To determine the reliability of an average.

- To serve as basis for control of variability.

- To compare two or more series in relation to their variability.

- Facilitate further statistical analysis.

RANGE

- It is the simplest method, Defined as the difference between the value of the smallest item and the value of the largest item.

- This measure gives no information about the values that lie between the extremes values.

- Subject to fluctuations from sample to sample.

MEAN DEVIATION

- It is the average of the deviations from the arithmetic mean.

- $M.D = \dfrac{\sum X - X_i}{n}$, where $\sum$ (**sigma**) is the sum of, **X** is the arithmetic mean, **Xi** is the value of each observation in the data, **n** is the number of observation in the data.

STANDARD DEVIATION(SD)

- Most important and widely used.

- Also known as root mean square deviation, because it is the square root of the mean of the squared deviations from the arithmetic mean.

- Greater the standard deviation, greater will be the magnitude of dispersion from the mean.

- A small SD means a higher degree of uniformity of the observations.

CALCULATION

- **For ungrouped data:**
- Calculate the mean(X) of the series.
- Take the deviations (d) of the items from the mean by : **d=X$_i$ – X,** where X$_i$ is the value of each observation.
- Square the deviations (d^2) and obtain the total ($\sum$ d^2)
- Divide the $\sum$ d^2 by the total number of observations i.e., (n-1) and obtain the square root. This gives the standard deviation.
- Symbolically, standard deviation is given by:

$$SD = \sqrt{\sum d^2 / (n-1)}$$

- **For grouped data with single units for class intervals:**

- $S = \sqrt{\sum(X_i - X) \times f_i / (N - 1)}$

- Where,

  ✓ X_i is the individual observation in the class interval

  ✓ f_i is the corresponding frequency

  ✓   X is the mean

  ✓ N is the total of all frequencies

- **<u>For grouped data with a range for the class interval</u>:**

$$S = \sqrt{\sum(X_i - X) \times f_i / (N - 1)}$$

Where,

    X_i is the midpoint of the class interval

    f_i is the corresponding frequency

    X is the mean

    N is the total of all frequencies

COEFFICIENT OF VARIATION(C.V.)

- A relative measure of dispersion.

- To compare two or more series of data with either different units of measurement or marked difference in mean.

- **C.V.= (Sx100)/ X**

- Where, C.V. is the coefficient of variation

- S is the standard deviation

- X is the mean

- Higher the C.V. greater is the variation in the series of data
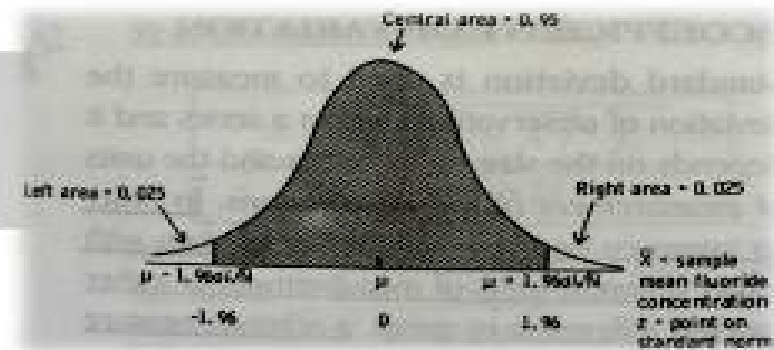
NORMAL DISTRIBUTION CURVE

- Gaussian curve

- Half of the observations lie above and half below the mean

  - Normal or Gaussian distribution
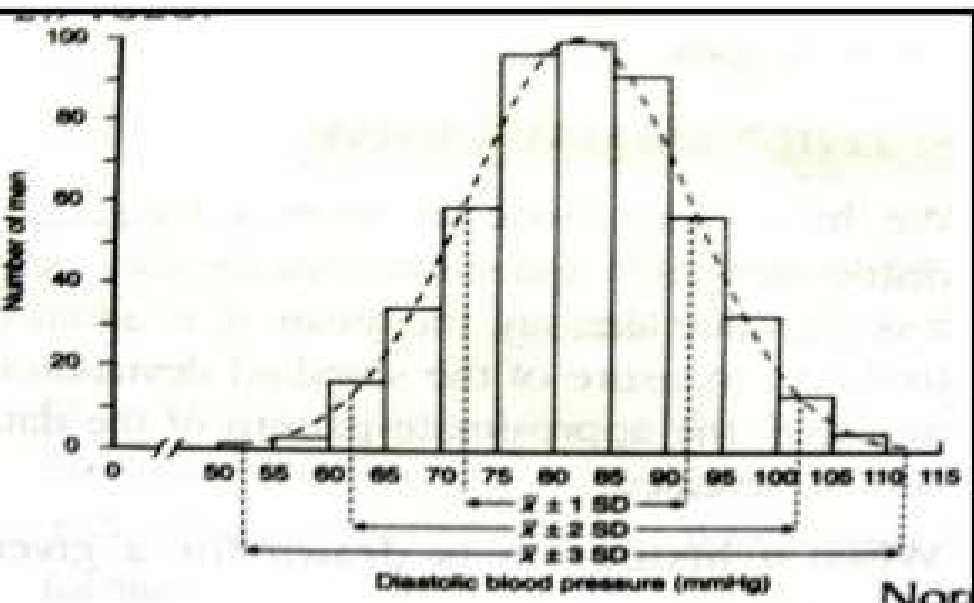
Properties

- Bell shaped.

- Symmetrical about the midpoint.

- Total area of the curve is 1. Its mean zero & standard deviation 1.

- Height of curve is maximum at the mean and all three measures of central tendency coincide.

- Maximum number of observations is at the value of the variable corresponding to the mean, numbers of observations gradually decreases on either side with few observations at extreme points.

- Area under the curve between any two points can be found out in terms of a relationship between the mean and the standard deviation as follows:

✓ Mean ± 1 SD covers 68.3% of the observations

✓ Mean ± 2 SD covers 95.4% of the observations

✓ Mean ± 3 SD covers 99.7% of the observations

- These limits on either side of mean are called confidence limits.

- Forms the basis for various tests of significance .