DESCRIPTIVE STATISTICS

1. What is statistics?
   - The word 'Statistics' is derived form Latin word 'Statis' which means a "political state". Clearly, statistics is closely linked with the administrative affairs of a state such as facts and figures regarding defense force, population, housing, food, financial resources etc.
   - American Heritage Dictionary defines statistics as "Mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference of numerical data".
   - The Merriam-Webster's Collegiate Dictionary definition is "A branch of mathematics dealing with the collection, analysis, interpretation and presentation of masses of numerical data".

2. Statistics
   - It is plural noun which describes a collection of numerical data such as employment statistics, accident statistics, population statistics, birth and expenditure, of exports and imports, etc.
   - As a singular noun, it is used to describe as a branch of applied mathematics, whose purpose is to provide methods of dealing with collections of data and extracting information from them in compact form by tabulating, summarizing and analyzing the numerical data or set of observations.
   - "by statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."

- This definition gives all the characteristics of statistics which are
  - Aggregate of facts
  - Affected by multiplicity of causes
  - Numerically expressed
  - Estimated according to reasonable standard of accuracy
  - Collected in a systematic manner
  - Collected for predetermined purpose
  - Placed in relation to each other

3. Descriptive statistics
   - Describe, organize or summarize data
   - E.g. average height of 1st year medical students

4. Inferential statistics
   - Generalizing to a population after having observed only a sample
   - E.g. average blood pressure of all the PPSP faculty

5. Measure of central tendency
   - Mean
     - Sum of the values of all items divided by the number of items in the series
     - It is sensitive to extreme value of the item
     - When repeated samples drawn from same population tend to have same mean
   - Median
     - Value of the item that divides the series of items when arranged in ascending or descending order of values
     - Insensitive to small number of extreme values of items
   - Mode
     - Most frequently occurring values of the items in a series

- Insensitive to small number of extreme values items
- Respective position of mean, median and mode
  - Positively skewed distribution (skewed to the right)
    - Mode → Median → Mean
  - Negatively skewed distribution (skewed to the left)
    - Mean → Median → Mode

6. Measure of variability or dispersion
   - Range
     - Difference between the lowest and highest value of the items in a series
   - Variance
     - Variance is a technical word for variability
     - It is measure of spread about the mean (spread from the mean)
     - Variance is the mean of the squared deviation from the mean
     - $S^2 = \Sigma(x_i - x_{mean})^2/n$
   - Standard deviation
     - Standard deviation is square root of variance
     - Remedies the problem of squaring the deviation in variance
     - $SD = \sqrt{[\Sigma(x_i - x_{mean})^2/n-1]}$
     - 68% of items fall within ± 1 SD
     - 95% of items fall within ± 2 SD
     - 99% 0f items fall within ± 3 SD
   - Special properties of the normal distribution
     - Its shape is such that it
       - Embraces 68.26% of the cases within 1 sd around the mean
       - Embraces 95.46% of the cases within 2 sd around the mean
       - Embraces 99.74% of the cases within 3 sd around the mean