

## WEEKLY ASSIGNMENT TASK : 7 JANUARY

### Descriptive Statistics:

#### 1.What is the purpose of descriptive statistics?

Main purpose of descriptive statistics are: 1) providing basic information about variables in a dataset and 2) highlighting potential relationships between variables. Graphical/Pictorial Methods are measures of the three most common descriptive statistics that can be displayed graphically or pictorially. It is used to summarise data.

#### 2.Can you explain the difference between mean, median, and mode?

Mean is the measurement of central tendency that represents the average value of the dataset. Mean is calculated by adding all the values in the dataset and dividing by the total number of values, i.e.,

$$\text{Mean} = \text{Sum of all values} / \text{Number of Observations}$$

Median is the measure of central tendency that represents the middle value of the dataset when the data are arranged in order (either ascending or descending). Once you get the data, the first thing you have to do is to arrange the data either in ascending or descending order.

Case-1: When the number of terms is odd.

$$\text{Median} = ((n+1)/2)\text{th term}$$

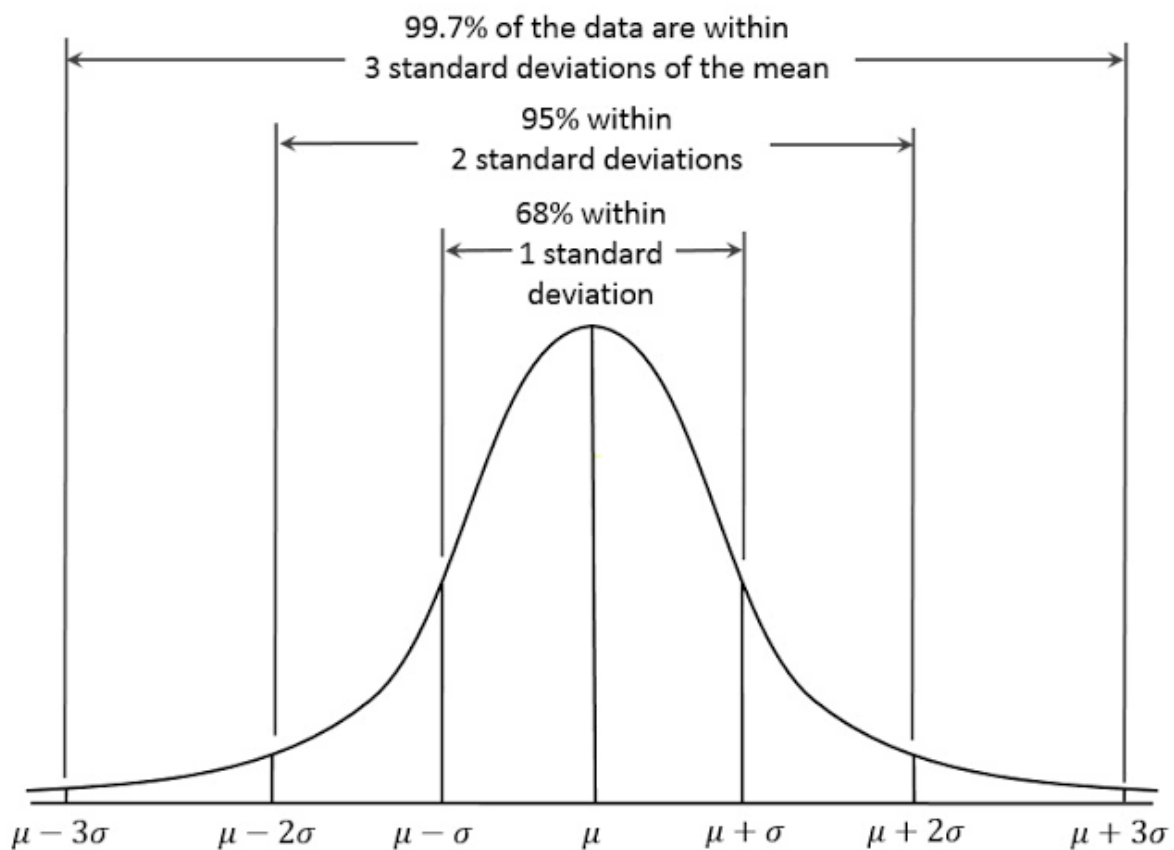
Case-2: When the number of terms is even

$$\text{Median} = [(n/2)\text{th term} + (n/2 + 1)\text{th term}] / 2$$

Mode is the most frequently occurring value. It may be possible that there doesn't exist any mode or there exists more than one mode.

#### 3.How do you interpret the standard deviation of a dataset?

A standard deviation is a measure of how spread out data is in relation to the mean. It indicates how much the values in a data set differ from the mean. A low standard deviation indicates that the values tend to be close to the mean, while a high standard deviation indicates that the values are spread out over a wider range



In larger datasets, points that are more than 2 or 3 standard deviations from the mean are considered outliers.

The formula for the population standard deviation is:

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

x: Random variable

$\mu$ : Population mean

N: Total number of values

4. Describe the concept of skewness in statistics.

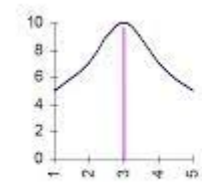
Skewness is a measurement of the distortion of symmetrical distribution or asymmetry in a data set. Skewness is demonstrated on a bell curve when data points are not distributed symmetrically to the left and right sides of the median on a bell curve. If the bell curve is shifted to the left or the right, it is said to be skewed.

Positive Skewness

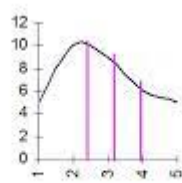
A distribution is positively skewed when its tail is more pronounced on the right side than it is on the left. Since the distribution is positive, the assumption is that its value is positive.

### Negative Skewness

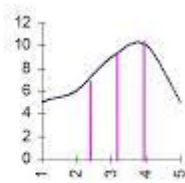
Negative skewness, on the other hand, occurs when the tail is more pronounced on the left rather than the right side. Contrary to the positive skew, most of the values are found on the right side of the mean when it comes to negative skewness.



Mean = Median = Mode



Mode > Med > Mean



Mean < Med < Mode

### **Inferential Statistics:**

#### **5.What is the main goal of inferential statistics?**

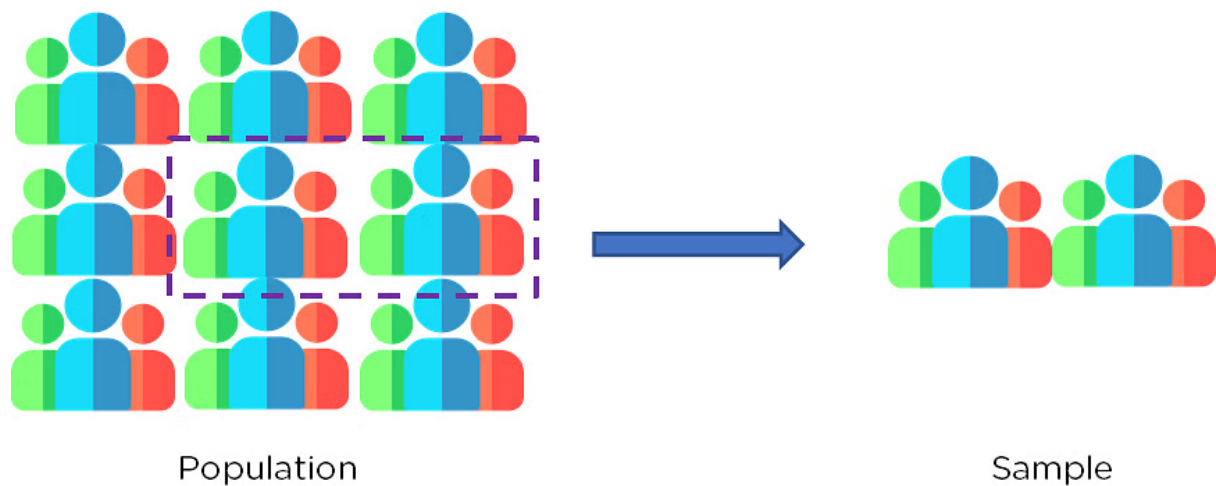
Inferential statistics is a branch of statistics that uses data from a sample to make inferences about a larger population. It involves using analytical tools to draw conclusions about a population by examining random samples.

The goal of inferential statistics is to make generalizations about a population. It involves making predictions, generalizations, and conclusions about a population based on the analysis of a sample of data.

#### **6.Explain the difference between a population and a sample.**

population is the entire set of items from which you draw data for a statistical study. It can be a group of individuals, a set of items, etc. It makes up the data pool for a study.

A sample is defined as a smaller and more manageable representation of a larger group. A subset of a larger population that contains characteristics of that population. A sample is used in statistical testing when the population size is too large for all members or observations to be included in the test.



### 7. What is a confidence interval, and how is it useful in inferential statistics?

A confidence interval displays the probability that a parameter will fall between a pair of values around the mean. Confidence intervals measure the degree of uncertainty or certainty in a sampling method.

The confidence interval shows the range of values you expect the true estimate to fall between if you redo the study many times. Confidence intervals can provide important information to statistical significance of studies especially when a  $p$ -value is borderline (i.e., it is equal to the critical  $p$ -value). If the null value (the value that indicates no difference and is usually zero or one) is included in the confidence interval, then the result is not statistically significant.

### 8. Define $p$ -value

The  $p$  value, or probability value, tells you how likely it is that your data could have occurred under the null hypothesis. The  $p$  value tells you how often you would expect to see a test statistic as extreme or more extreme than the one calculated by your statistical test if the null hypothesis of that test was true. The  $p$  value gets smaller as the test statistic calculated from your data gets further away from the range of test statistics predicted by the null hypothesis. The  $p$  value is a proportion: if your  $p$  value is 0.05, that means that 5% of the time you would see a test statistic at least as extreme as the one you found if the null hypothesis was true.  $P$ -values are calculated from the null distribution of the test statistic. They tell you how often a test statistic is expected to occur under the null hypothesis of the statistical test, based on where it falls in the null distribution.