1. In probability theory, the central limit theorem (CLT) states that the distribution of a sample variable approximates a normal distribution (i.e., a “bell curve”) as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population's actual distribution shape. Put another way, CLT is a statistical premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population. The Central Limit Theorem is important for statistics because it allows us to safely assume that the sampling distribution of the mean will be normal in most cases. This means that we can take advantage of statistical techniques that assume a normal distribution.
2. Sampling is a process in statistical analysis where researchers take a predetermined number of observations from a larger population. The method of sampling depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

   Random Sampling: a random sample may include choosing the names of 25 employees out of a hat in a company of 250 employees. The population is all 250 employees, and the sample is random because each employee has an equal chance of being chosen.

   Judgment Sampling: when the samples are chosen from the population based on some conditions.
3. A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.
4. A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution. The normal distribution is also known as a Gaussian distribution or probability bell curve. It is symmetric about the mean and indicates that values near the mean occur more frequently than the values that are farther away from the mean. Regardless of its exact shape, a normal distribution bell curve is always symmetrical about the mean. A symmetrical distribution means that a vertical dividing line drawn through the maximum/mean value will produce two mirror images on either side of the line, in which half the population is less than the mean and half is greater. However, the reverse is not always true; that is, not all symmetrical distributions are normal. In the bell curve, the peak is always in the middle, and the mean, mode and median are all the same.

5. Covariance is a measure of how two variables change together. The terms covariance vs correlation is very similar to each other in probability theory and statistics. Both the terms describe the extent to which a random variable or a set of random variables can deviate from the expected value. It is calculated as the covariance of the two variables divided by the product of their standard deviations. Covariance can be positive, negative, or zero. A positive covariance means that the two variables tend to increase or decrease together. A negative covariance means that the two variables tend to move in opposite directions. A zero covariance means that the two variables are not related. Correlation can only be between -1 and 1. A correlation of -1 means that the two variables are perfectly negatively correlated, which means that as one variable increases, the other decreases. A correlation of 1 means that the two variables are perfectly positively correlated, which means that as one variable increases, the other also increases. A correlation of 0 means that the two variables are not related.
6. Univariate statistics summarize only one variable at a time.
   Bivariate statistics compare two variables.
   Multivariate statistics compare more than two variables.
7. Sensitivity analysis determines how different values of an independent variable affect a particular dependent variable under a given set of assumptions. In other words, sensitivity analyses study how various sources of uncertainty in a mathematical model contribute to the model's overall uncertainty. This technique is used within specific boundaries that depend on one or more input variables.
   Sensitivity = Percentage change in output / Percentage change in input
8. Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test. It is used to estimate the relationship between 2 statistical variables.

   The Null Hypothesis is the assumption that the event will not occur. A null hypothesis has no bearing on the study's outcome unless it is rejected. H0 is the symbol for it.

   The Alternate Hypothesis is the logical opposite of the null hypothesis. The acceptance of the alternative hypothesis follows the rejection of the null hypothesis. H1 is the symbol for it.

   H0 is the null hypothesis for two tailed test and H1 is the alternate hypothesis for example if the average mark of the group is 90 marks but the teacher thinks that it has decreased to 80 so in this case 90 is the NULL Hypothesis and 80 is the alternate hypothesis.

9. Quantitative data are measures of values or counts and are expressed as numbers. Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are data about categorical variables (e.g. what type).
10. The range is calculated by subtracting the lowest value from the highest value. The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.
11. A bell curve is a common type of distribution for a variable, also known as the normal distribution. The highest point on the curve, or the top of the bell, represents the most probable event in a series of data (its mean, mode, and median in this case), while all other possible occurrences are symmetrically distributed around the mean, creating a downward-sloping curve on each side of the peak. The width of the bell curve is described by its standard deviation.
12. The interquartile range (IQR) tells you the range of the middle half of your dataset. You can use the IQR to create "fences" around your data and then define outliers as any values that fall outside those fences.

    METHODOLOGY:

    Sort your data from low to high

    Identify the first quartile (Q1), the median, and the third quartile (Q3).

    Calculate your IQR = Q3 – Q1

    Calculate your upper fence = Q3 + (1.5 * IQR)

    Calculate your lower fence = Q1 – (1.5 * IQR)

    Use your fences to highlight any outliers, all values that fall outside your fences.
13. The p value, or probability value, tells you how likely it is that your data could have occurred under the null hypothesis. It does this by calculating the likelihood of your test statistic, which is the number calculated by a statistical test using your data. 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.
14. $P(x:n,p) = {}^{n}C_{x}\, p^{x}\, (1-p)^{n-x}$

    $P(x:n,p) = {}^{n}C_{x}\, p^{x}\, (q)^{n-x}$
15. Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups.

The outcome of ANOVA is the ‘F statistic’. This ratio shows the difference between the within group variance and the between group variance, which ultimately produces a figure which allows a conclusion that the null hypothesis is supported or rejected. If there is a significant difference between the groups, the null hypothesis is not supported, and the F-ratio will be larger. ANOVA is used in medical fields to test the effectiveness of various medicines , used in financial department for the calculation of the budgets etc.