STATISTICS

1. The central limit theorem states that the sampling distribution of the mean approaches a normal distribution, as the sample size increases. This fact holds especially true for sample sizes over 30.
   As sample size increases, the sample mean and standard deviation will be closer in value to the population mean μ and standard deviation σ.
   The central limit theorem tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach normality as the sample size (N) increases.

2. Sampling means selecting a group (a sample) from a population from which we will collect data for our research. Samplings is an important aspect of a research study as the results of the study majorly depend on the sampling technique used.
   There are two types of sampling methods:
   - Probability sampling: involves random selection, allowing you to make strong statistical inferences about the whole group.
   - Non-Probability sampling: involves non-random selection based on convenience or other criteria, allowing you to easily collect data

3.

BASIS FOR COMPARISON	TYPE I ERROR	TYPE II ERROR
Meaning	Type I error refers to non-acceptance of hypothesis which ought to be accepted.	Type II error is the acceptance of hypothesis which ought to be rejected.
Equivalent to	False positive	False negative
What is it?	It is incorrect rejection of true null hypothesis.	It is incorrect acceptance of false null hypothesis.
Represents	A false hit	A miss
Probability of committing error	Equals the level of significance.	Equals the power of test.

4. A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

5. Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are 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 is a tool used in financial modeling to analyze how the different values of a set of independent variables affect a specific dependent variable under certain specific conditions. In general, sensitivity analysis is used in a wide range of fields, ranging from biology and geography to economics and engineering.

Mathematically, the dependent output formula is represented as,
$Z = X^2 + Y^2$

Firstly, the analyst is required to design the basic formula, which will act as the output formula. For instance, say NPV formula can be taken as the output formula.

Next, the analyst needs to identify which are the variables that are required to be sensitized as they are key to the output formula. In the NPV formula in excel, the cost of capital and the initial investment can be the independent variables.

Next, determine the probable range of the independent variables.

Next, open an excel sheet and then put the range of one of the independent variable along the rows and the other set along with the columns.

8. Hypothesis testing concerns on how to use a random sample to judge if it is evidence that supports or not the hypothesis. Hypothesis testing is formulated in terms of two hypotheses: H0: the null hypothesis; • H1: the alternate hypothesis.

Null hypothesis (H0): The null hypothesis here is what currently stated to be true about the population. In our case it will be the average height of students in the batch is 100.

H0 : μ = 100

Alternate hypothesis (H1): The alternate hypothesis is always what is being claimed. "In our case, Tedd believes(Claims) that the actual value has changed". He doesn't know whether the average has gone up or down, but he believes that it has changed and is not 100 anymore.

H1: $\mu \neq 100$

Always remember that an alternate hypothesis is always written with a $\neq$ or < or > sign. Please refer the below table for more clarity.

H$_0$	H$_1$
Equal (=)	Not equal to ($\neq$) or greater than (>) or less than (<)
greater than or equal to ($\geq$)	less than (<)
less than or equal to ($\leq$)	more than (>)

So if the alternate hypothesis is written with a $\neq$ sign that means that we are going to perform a 2-tailed test because chances are it could be more than 100 or less than 100 which makes it 2-tailed.

9. Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

10. The interquartile range is a measure of where the "middle fifty" is in a data set. Where a range is a measure of where the beginning and end are in a set, an interquartile range is a measure of where the bulk of the values lie. That's why it's preferred over many other measures of spread when reporting things like school performance or SAT scores.
    The interquartile range formula is the first quartile subtracted from the third quartile:
    $$IQR = Q_3 - Q_1.$$

11. A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side.

12. Using visualizations:

You can use software to visualize your data with a box plot, or a box-and-whisker plot, so you can see the data distribution at a glance. This type of chart highlights minimum and maximum values (the range), the median, and the interquartile range for your data.

Many computer programs highlight an outlier on a chart with an asterisk, and these will lie outside the bounds of the graph.

13.   The p-value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P-values are used in hypothesis testing to help decide whether to reject the null hypothesis.

14.   The binomial distribution is given by the formula:

$P(X= x) = {}^{n}C_{x}p^{x}q^{n-x}$, where $= 0, 1, 2, 3, \ldots$ $P(X = 6)$ $= 105/512$.

Hence, the probability of getting exactly 6 heads is 105/512.

15.   ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.