

STATISTICS ANSWER SHEET

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. Normal distribution is a probability distribution used for data analytics. It is a bell shaped curve which is symmetric around the curve. It helps us to check for outliers, data lying outside the range of -3 to +3 are outliers.
11. First we load the dataset and check for NaN values. We can check it through heatmap also, if there are any white spaces in between. We can either find the mean of that null value column and replace it with the null value or else we can drop that row, provided it does not have much data.
12. A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.
13. Yes, mean imputation of missing data is acceptable, but at the same time it is not a good solution because it may affect the relationship among the variables, which can be important points.
14. Linear regression is the simplest and most extensively used statistical technique for predictive modelling analysis. It is a way to explain the relationship between a dependent variable (target) and one or more explanatory variables(predictors) using a straight line. There are two types of linear regression - Simple and Multiple.
15. There are basically 4 branches of statistics :
 - Mathematical or theoretical statistics : it helps in forming the experimental and statistical distribution.

- Statistical method or function : it helps in the collection, tabulation and interpretation of data.
- Descriptive statistics : it helps in summarizing and organizing any dataset characteristics.
- Inferential statistics : it helps in finding the conclusion regarding the population after analysis on the sample drawn from it.