

STATISTICS WORKSHEET-1

1. TRUE
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C

10. **The normal distribution**, also known as the Gaussian distribution, is the most important probability distribution is statistics for independent, random variables. The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions. While the normal distribution is symmetrical, not all symmetrical distributions are normal.
11. Missing data can be handled through various imputation techniques like:
 - (i) Mean imputation
 - (ii) Median imputation
 - (iii) Mode imputation
 - (iv) Random imputation
 - (v) Multiple imputation

I would recommend multiple imputation technique because we can generate missing values from the dataset many times.

12. **A/B testing** - An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.
13. Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not preserve relationships between variables such as correlations.
14. **Linear regression** - Linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. **Branches of statistics** - There are two types of statistics:-

- (i) Descriptive statistics: if data can be described without any statistical tools then it is called descriptive statistics. Example – marks in class, height of student.
- (ii) Inferential statistics: if data is too big then we use inferential statistics. Take a few samples from different data and we find the average. This is called inferential statistics. The average is then applicable to all the data from where we have selected our samples.