

STATISTICS ANSWERS – WORKSHEET 1

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. It 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.
11. There are 2 primary ways of handling missing values:
 - a. Deleting the Missing values
 - b. Imputing the Missing Values

KNN Imputer and Iterative Imputer techniques are most commonly used imputation techniques

12. A/B testing is a marketing strategy that pits two different versions of a website, ad, email, popup, or landing page against each other to see which is most effective.
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 is a data analysis technique that predicts the value of unknown data by using another related and known data value. It mathematically models the unknown or dependent variable and the known or independent variable as a linear equation.
15. There are two branches of statistics, DESCRIPTIVE STATISTICS and INFERENCE STATISTICS.

DESCRIPTIVE STATISTICS

The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

INFERENCE STATISTICS

The branch of statistics that analyzes sample data to draw conclusions about a population.