STATISTICS WORKSHEET-1

Answers[1-15]

1. a
2. a
3. b
4. d
5. c
6. b
7. b
8. a
9. c
10. Normal Distribution is also known as the 'Gaussian Distribution'. It is a probability distribution that is symmetric about the mean, showing the data near the mean are more frequent in occurrence than data far from the mean.In other word it describes how the values of a variable are distributed
11. We can use different methods to handle missing data points, such as dropping missing values, imputing them using machine learning, or treating missing values as a separate category.
    Imputation Techniques:
    1. Mean/Median Imputation
    2. Random Sample Imputation
    3. Multivariate imputation technique or Multiple Imputation
    4. Pattern Substitution
    5. Maximum Likelihood estimation
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. A/B testing is a type of split testing and is commonly used to drive improvements to specific variables or elements by measuring user or audience engagement. The approach is commonly used to optimise marketing campaigns or digital assets like websites. In A/B testing a specific variable is altered such as a title, image, or element layout. A sample of the audience is shown the control version and the altered version in a 50/50 split. Half traffic will interact with the original version, the other half will interact with the newer version. Engagement or the completion of a defined goal is the metric that is compared between the versions after a set period of time.
13. Mean imputation is typically considered terrible practice since it ignores feature correlation. Mean imputation also has some drawbacks, such as reducing the variance and standard deviation of the data, ignoring the distribution and correlation of the data, and potentially creating unrealistic values.
14. Linear regression is a type of statistical analysis used to predict the relationship between two variables. It assumes a linear relationship between the independent variable and the dependent variable and aims to find the best-fitting line that describes the relationship. The line is determined by minimizing the sum of the squared differences between the predicted values and the actual values.
    In statistics, a regression model is linear when all terms in the model are one of the following:

1.The constant
2.A parameter multiplied by an independent variable (IV)
That is:

Dependent variable = constant + parameter * IV + … + parameter * IV

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

15. Branches of statistic:
    1. **Descriptive Statistics**
    Descriptive statistics mostly focus on the central tendency, variability, and distribution of sample data. Central tendency means the estimate of the characteristics, a typical element of a sample or population. It includes descriptive statistics such as mean, median, and mode.
    2. **Inferential Statistics**
    Inferential statistics are tools that statisticians use to draw conclusions about the characteristics of a population, drawn from the characteristics of a sample, and to determine how certain they can be of the reliability of those conclusions. Based on the sample size and distribution, statisticians can calculate the probability that statistics, which measure the central tendency, variability, distribution, and relationships between characteristics within a data sample, provide an accurate picture of the corresponding parameters of the whole population from which the sample is drawn.