Statistics worksheet

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

10. What do you understand by the term Normal Distribution?

- Normal Distribution is the proper term for a bell bottom curve
- In normal Distribution mean, median and mode are equal
- In Normal Distribution the mean is 0 and the standard deviation is 1.
- All Normal distribution are symmetrical, but all symmetrical are not normal distribution
- Normal Distribution model is motivated from the central limit theorem
-

11. How do you handle missing data? What imputation techniques do you recommend?

- Missing data is one of the crucial parts which need to handle by a data scientist, firstly when we have the data need to check whether the data is proper if it is not then we need to check any missing value
- If there is any missing value, we need to do two steps to handle the data. 1. Removing the data of the missing value (its not the efficient way), 2. Imputation method

- Imputation: It is the process of replacing missing values with substituted data
- Normal Imputation: if our data is numerical, then just used mean and median values to replace the missing data, if our data is categorical then we use mode value to replace the missing data
- Model Based Imputation: In model-based imputation we train our data and predict the missing value

- So, I suggest both normal based and model-based techniques.

12.What is A/B testing?

- A/B Testing is a split testing method used to compare two different versions of a webpage or a application to determine which one is better to perform. It is a randomized control experiment.

13.Is mean imputation of missing data acceptable practice?

- Mean imputation is the practice of replacing null values in a data set with the mean of the data. Mean imputation is generally bad practice because it doesn't take into account feature correlation and also it reduces the variance of the data and increases bias in the data and it gives less accurate model output.

14.What is linear regression in statistics?

Linear Regression is a linear approach to modelling the relationship between a scalar response And one or more explanatory response. There are two types of linear regression 1. simple linear regression and 2. Multiple linear regression

- Simple Linear Regression: It is a statistical method used to summarize and study the relationships between two continuous variables.one is independent variable and other one is dependent variable

- Multiple Linear Regression: It is a statical method used to summarize the relationship between dependent variable with more than two independent variables

15What are the various branches of statistics?

There are two types of branches 1. **Descriptive Statistics** 2. **Inferential Statistics**

- **Descriptive Statistics** we have two types 1. Measure of central Tendency and 2. Measure of variability
- **Inferential Statistics**: Different Types of Inferential Statistics they are

1. Regression analysis
2. Analysis of variance (ANOVA)
3. Analysis of covariance (ANCOVA)
4. Statistical significance (t-test)
5. Correlation analysis