

## Statistics – Worksheet\_1

Q1. A

Q2. A

Q3. B

Q4. D

Q5. C

Q6. A

Q7. B

Q8. A

Q9. C

Q10. Normal Distribution:

A: The normal distribution of data when plotted represents the bell curve has the symmetrical distribution of the data around three characteristics i.e. mean, median and mode. The distribution is symmetrical about the mean and the values can be measured by mean and standard deviation.

Q11. Handling missing data. Recommend imputation technique

A: There are several ways to handle missing data depending on the data.

1. Delete the missing values either deleting entire row or entire column if the column has too many values missing.
2. Imputing the values: (For continuous data)

Replace the missing data by arbitrary values by `fillna()` method.

Replace the missing value of column by mean or mode of the corresponding column.

Replace missing values with previous filled value – Forward Fill. (`method='ffill'`)

Replace missing values with next filled value – Backward Fill. (`method='bfill'`)

Replace missing values using interpolation. (`interpolate()`)

(For categorical data)

We can replace the missing value with most frequently value in column.

(`SimpleImputer(strategy='most_frequent')`) x

We can impute the missing value with 'missing' which is considered separate category Of that column

Recommended Imputation Technique:

For continuous data: impute the values with mean

For categorical data: impute the nan with most frequent

Q12. A/B Testing

A: A/B testing is an optimization technique in which an alternate prediction of dataset is concluded by making changes to specific variable or column to optimize the test result. A/B test then compares the result of original dataset and altered version of dataset and based on the comparison and conversion rate, chooses the statistically proven dataset.

Q13. Is mean imputation of missing data acceptable practice?

A: Yes, but it is not considered ideal practice as the result might be biased due to ignorance of correlation.

Q14. Linear Regression

A: Linear regression is a supervised machine learning statistical method used to predict the output for continuous variables.

Q15. Various Branches of statistics

A: Descriptive Statistics & Inferential Statistics