STATISTICS WORKSHEET-1

Ans:-1 A

Ans:-2. A

Ans:-3 B

Ans:-4 D

Ans:-5 C

Ans:-6 B

Ans:-7 B

Ans:-8. A

Ans:-9. C

Ans:-10 When the prediction take up a continuous value and are not discrete, we assume that those values are sampled from a normal distribution or GAUSSIAN distribution and representing a symmetric bell shaped curve.

Ans:-11 There are a lot of techniques to treat missing value. I am trying to think what is the best way to organize some of the most commonly used methods, if you use SAS to implement it -

- **Ignore the records with missing values.**

Many tools ignore records with missing values. When the percentage of records with missing values is small, we could ignore those records.

- **Substitute a value such as mean**.

When the percentage is large and also when it makes sense to do something to avoid bias modeling results, substituting a value (e.g. mean, median) is a commonly used way. But this method could cause bias distribution and variance. That's where the following imputation methods come in.

- **Predict missing values.**

Depending on the type of the imputed variable (i.e. continuous, ordinal, nominal) and missing data pattern (i.e. monotone, non-monotone), below are a few commonly used models. If you plan to do it in SAS, there are SAS codes that you can write to identify the missing data pattern.

  - Logistic Regression
  - Discriminant Regression
  - Markov Chain Monte Carlo (MCMC)

Ans:-12. A/B testing is hypothesis testing ,an objective method of making decision or inference from sample data (evidence).

1. Sample data used to choose between two choices i.e. Hypothesis or statements about a population.
2. We typically do this by comparing what we have observed to what we expected if one of the statements(null hypothesis) was true.

    a. Null hypothesis
    b. Alternative hypothesis

Ans:-13. Mean imputation reduces the variance of the imputed variables.

MI shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.

MI does not preserve relationships between variables such as correlations.

Ans:-14. Linear Regression uses one independent variable to explain or predict the outcome of the other

dependent variable using a straight line. L.R whenever we predict the label on continuous values. Linear regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions.

Ans:-15 Branches of Statistics

There are two main branches of statistics
1.Descriptive Statistics:- one procedure is used to summarize, organize and make sense of a set of scores or observation.
a. Measure of center
:-Mean
:-Median
:-Mode
b. Measure of Dispersion
:-Range
:-Percentile
:-variance
:-standard deviation
:-Skewness
:-Kurtosis
2.Inferential Statistics :-Procedure used that allow researchers to infer or generalize observation made with sample to the larger population from which they were selected.