STATISTICS WORKSHEET-1

Ans1.(A)

Ans2. (A)

Ans3. (A)

Ans4. (A)

Ans5. (C)

Ans6. (B)

Ans7. (B)

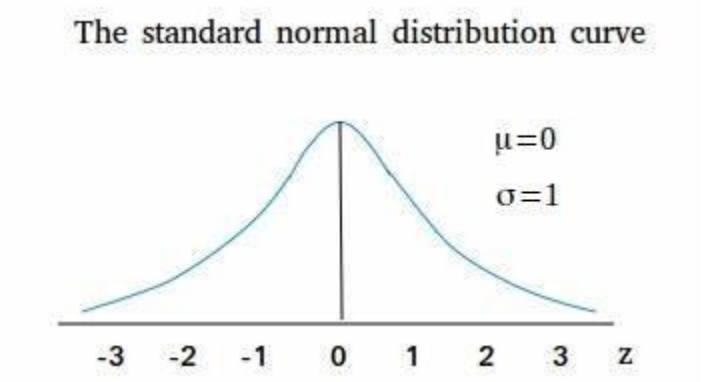
Ans8. (A)

Ans9. (C)

Ans10.

Normal Distribution: -

The normal distribution is the most widely known and used of all distributions. Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference of many probability problems.



Ans 11.

Multiple imputation is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing

data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of the right values.

Using the imputed data, the process is repeated to make multiple imputed data sets. Each set is then analyzed using the standard analytical procedures, and the multiple analysis results are combined to produce an overall result.

The various imputations incorporate natural variability into the missing values, which creates a valid statistical inference. Multiple imputations can produce statistically valid results even when there is a small sample size or a large amount of missing data.

Ans12.

AB Testing: -

An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

Ans 13.

Although imputing missing values by using the mean is a popular imputation technique, there are serious problems with mean imputation. The variance of a mean-imputed variable is always biased downward from the variance of the un-imputed variable. This bias affects standard errors, confidence intervals, and other inferential statistics.

Ans14.

Linear Regression: -

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

(1) Does a set of predictor variables do a good job in predicting an outcome (dependent) variable?

(2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable?

These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula **y = c + b*x**, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Types of Linear Regression:-

1) Simple linear regression
2) Multiple linear regression
3) Logistic regression
4) Ordinal regression
5) Multinomial regression
6) Discriminant analysis

Ans15.

Types of statistics: -

There are two main branches of statistics:

1) Descriptive Statistics
2) Inferential Statistics

Descriptive Statistics: - It is the first phase of statistics. It involves any kind of data processing designed to the collection, organization, presentation and analyzing the important features of the data without attempting to infer/ conclude anything that goes beyond the known data.
In short, descriptive statistics describes the nature or characteristics of the observed data without making conclusion or generalization.
Inferential Statistics: - It concerns with drawing conclusions or predictions about a population from the analysis of a random sample drawn from that conclusion.
It includes method like,

- Point estimation
- Interval estimation
- Hypothesis testing