STATISTICS

(WORKSHEET-ANSWERKEY)

Q=1) (A)

Q=2) (A)

Q=3) (B)

Q=4) (D)

Q=5) (C)

Q=6) (B)

Q=7) (B)

Q=8) (A)

Q=9) (C)

Q=10) A normal distribution curve is symmetrical, bell Shaped curve defined by the mean and standard deviation of a data set. The normal curve is a probability distribution with a total area under the curve of 1.

Q=11) To handle missing data I suggest 3 techniques,

1.) Delete the record missing values.

Only if it is a huge dataset, delete the record.

2.) Create a separate modes to handle missing value.

By creating a separate model we can predict the output for the missing value. This method takes huge time to complete. When this method we can use is if the dataset is small otherwise no.

3.) Then using statistical methods mean, median or mode.

. By using the Average of the mean data(then replace the NAN/ missing value.)

. By using the median method (change the data to sorting format – then replacing the missing value).

. By using Mode (which ever the value is having more frequency. That value will be replacing to the NAN or missing data).

Q=12) Ans=) It's 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 one of the most prominent and widely used statistical tools.

Q=13) Ans=) it's a bad practice.

Mean imputation preserves the mean of the observed data. Leads to an underestimate of the standard deviation. Distorts relationship between variables by " pulling" estimates of the correlation toward zero.

Q=14) Ans=) Linear regression is a linear model that assumes a linear relationship between input variables (independent variables 'x') and output variables (dependent variables 'y') such that 'y' can be calculated from a linear combination of input variables (x)

Q=15) Ans=) There are two main branches of statistics:

1. Descriptive statistics.
2. Inferential statistics.

Statistics are used to describe or summarize the characteristics of a sample or data set, Such as a variable's mean, standard deviation, or frequency. Inferential statistics, in contrast, employs any number of techniques to relate

variables in a data set to one another, for example using correlation or regression analysis. These can then be used to estimate forecasts or infer casuality.