Statistics

Section – A (Only one correct option)

Ans -1 (A) True

Ans -2 (A) Central Limit Theorem

Ans -3 (C) Modeling contingency tables

Ans -4 (C) The Square of a standard normal random variable follows what is called chi-squared distribution

Ans -5 (C) Poisson

Ans -6 (B) False

Ans -7 (B) Hypothesis

Ans -8 (A) 0

Ans -9 (C) Outliers cannot conform to the regression relationship

Section – C (Subjective)

Ans -10 When data is distributed symmetrically i.e., most of the data revolves around measures of central tendency mean, median and mode and does not have any skewness and biasness, then the type of distribution is known as Normal Distribution. In normal distribution, the value of mean, median and mode are the same or equal. The rest of the data gets taped off symmetrically away from the centre toward either extreme points.

Ans -11 Missing data is handled initially by finding out the number of missing values in the whole data set column wise. Missing values are handled in two ways, one by deleting the missing values and another is to impute the missing values.

There are various methods for imputing the missing values, some of them are:-

1. Replacing the missing value with Mean, Median or Mode of the values of the particular column.
2. Replacing the missing value with the previous value of the column. This method is known as Forward fill.
3. Replacing the missing value with the next value of the column. This method is known as Backward fill.
4. Replacing the missing value with the most frequent value. This method is most often used in column with categorical data. However, it works with numerical data as well.
5. Replacing the missing value with functions from Sci-kit library. These functions are “SimpleImputer()”, “KNNImputer()”, and “IterativeImputer()”.

Ans -12 A/B testing also known bucket testing, is a type of testing in which we usually compare different versions of the same thing (could be any product, mail, web browser, mobile or desktop applications) in order to check the performance of those versions and identify which version is performing better in terms of user experience and sales and many more factors.

Ans -13 Mean imputation of missing value is not a good practice, as it increases the biasness of data, as replacing missing values with mean value, will not preserve relationship between variables, and since it ignores the correlation between them, the data distribution gets affected badly.

Ans -14 In statistics, there are two types of variables, one is known as dependent variable, and another is known as independent variables. In linear regression, usually there is only one dependent variable and its value is predicted on the basis of set of independent variables, they could be many. In linear regression, the value of dependent variable is predicted by all other independent variables linearly. By saying linearly, it means that data points are linearly separable on the basis of a linear equation that describes the correlation of the independent variables with the dependent variable. The linear equation is -> $Y = mX_i + b + e$, where “Y” is the dependent variable, “m” is the slope, “b” is the intercept and “e” is the error term in the regression line.

Ans -15 There are two branches of statistics:-

1. Descriptive Statistics – This type of statistics contains organisation, representation and description of data, with the help of tables, graphs and summary measures such as frequency, central tendency, dispersion etc.
2. Inferential Statistics – This type of statistics comes into play after the descriptive statistics of data gets done. This includes interpreting the meaning out from the representations done in descriptive statistics, in simple way, it is a way to draw conclusions from the data.