Statistics-1

1) True (Ans--A)
2) Central Limit Theorem(Ans--A)
3) Modelling bounded count data (Ans--B)
4) All of the mentioned (Ans--D)
5) Poisson (Ans--C)
6) False (Ans--B)
7) Hypothesis (Ans--B)
8) 0 (Ans--A)
9) Outliers cannot conform to the regression relationship (Ans--C)
10) The normal distribution is the most common type of distribution assumed in technical stock market analysis and in other types of statistical analyses. The standard normal distribution has two parameters: the mean and the standard deviation. The normal distribution model is important in statistics and is key to the Central Limit Theorem (CLT). This theory states that averages calculated from independent, identically distributed random variables have approximately normal distributions, regardless of the type of distribution from which the variables are sampled.
11) A data scientist doesn't want to produce biased estimates that lead to invalid results. The concept of missing data is implied in the name: it's data that is not captured for a variable for the observation in question. Missing data reduces the statistical power of the analysis, which can distort the validity of the results.
    Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.
    a) Mean imputation
    b) Regression imputation
    c) Interpolation and extrapolation etc
12) A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics
13) The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation. mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.
14) Linear regression models the relationships between at least one explanatory variable and an outcome variable. These variables are known as the independent and dependent variables, respectively. When there is one independent variable (IV), the procedure is known as simple linear regression. When there are more IVs, statisticians refer to it as multiple regression.
15) Main two branch of Statistics are
    Descriptive statistics-: mostly focus on the central tendency, variability, and distribution of sample data. Central tendency means the estimate of the characteristics, a typical element of a sample or population, and includes descriptive statistics such as mean, median, and mode. Variability refers to a set of statistics that show how much difference there is among the elements of a sample or population along the characteristics measured, and includes metrics such as range, variance, and standard deviation.

Inferential Statistics-: Inferential statistics are tools that statisticians use to draw conclusions about the characteristics of a population, drawn from the characteristics of a sample, and to decide how certain they can be of the reliability of those conclusions. Based on the sample size and distribution statisticians can calculate the probability that statistics, which measure the central tendency, variability, distribution, and relationships between characteristics within a data sample, provide an accurate picture of the corresponding parameters of the whole population from which the sample is drawn.