1. a) True
2. a) Central Limit Theorem
3. c) Modeling contingency tables
4.
5. c) Poisson
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. Normal distribution generally refers to the set of data where MEAN=MEDIAN=MODE. In other words it can be called symmetrical distribution of data.
11. For handling missing data we can do the following steps:
i) If the column having the missing data is not so important for the dataset we can drop it using pandas 'dropna' method.
ii) If the number of missing data is too high we cannot drop it, instead we can fill it with some data like (mean/median/mode of the respective row/column).
iii) We can use SimpleImputer from sklearn library to impute mean/median in place of missing values.
iv) We can use '.fillna' method to fill the missing values with certain values.
12.
13. Normally mean imputation is not a good practice as it may not give the accurate result. For example, certain data set is missing the marks of a student who usually
scores less in the exams, here if we are considering the results of mean imputation then it will definitely give us a value which may be too high as per the actual
performance of the student. So, it is not acceptable to use mean imputing technique.
14. Linear Regression in statistics can be defined as a technique used to calculate the value of variable(y) based on the values of another variable(x). Suppose say
there is houseprice estimation; variable(x) is having some values like Crime rate near the area, Hospital availablity, School facilities available near the area,
if the area is located in remote place etc., and we use these values of (x) to predict the house_price(y). Here we need to predict the house price depending on the
values provided by (x).
15. Statistics is divided as Descriptive Statistics and inferential ststistics. Descriptive statistics can be defined as properties of a population or a sample of population.
Inferential Statistics can be defined as the techniques through which we test the data collected through hypothesis testing, anova and draw cocnlusions for the given data.