Ans1) a) True

Ans 2) a) Central Limit Theorem

Ans3 ) b) Modeling bounded count data

Ans 4) d) All of the mentioned

Ans 5) c) Poisson

Ans 6) a) True

Ans 7) b) Hypothesis

Ans 8) a) 0

Ans 9) c) Outliers cannot conform to the regression relationship

ANS 10) The normal distribution is most widely used for all types of distributions. As normal distribution approximates many natural phenomena. It has developed into a standard of reference for many probability problems. Many things actually are normally distributed or very close of it. The normal distribution is easy to work with mathematically. In many of practical cases the method developed using normal theory work quiet well even when the distribution is not normal.

There is a strong connection between the size of a sample N and the extent to which a sampling distribution approaches the normal form. Many distribution based on Large N can be approximated by the normal distribution even though the population distribution itself is definitely not normal.

ANS 11) As we know the data is available in the messy form, scattered and missing form. This may leads to problems during the analysis, decision making to the data scientist. It also impacts the statistical power of analysis which impacts the result validity. There are two types of data missing. Missing at Random(MAR) and Missing Not at Random(MNAR). Missingno is a simple Python library that presents a series of visualizations to recognize the behavior and distribution of missing data inside a pandas data frame. It can be in the form of a barplot, matrix plot, heatmap, or a dendrogram.

ANS 12) A/B Testing is the process of experiment where two or more types of varients of a page are shown to the user at randomly and statistics analysis is used to determine which variant is performing better and accurately for the given performance goal.

ANS13) The process of replacing null values in a data collection with the data's mean is known as mean imputation. The mean imputation of missing data is not an acceptable practice as 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.

ANS14) Linear Regression : It is also called as Supervised Learning. Linear regression is the simplest and most extensively used statistical technique for predictive modelling analysis. It is a way to explain the relationship between a

dependent variable (target) and one or more explanatory variables(predictors) using a straight line. There are two types of linear regression - Simple and Multiple.