Answers for Assignment 3(Statistics)

Ans1- a) True

Ans2- a) Central Limit Theorem

Ans3- c) Modeling contingency tables

Ans4- d) All of the above

Ans5- c) Poisson

Ans6- b) False

Ans7- b) Hypothesis

Ans8- a) Zero

Ans9- c) Outliers cannot conform to the regression relationship

Ans10- A normal distribution is a bell-shaped frequency distribution curve of a continuous random variable. The mean of a normal distribution is Zero and standard deviation is 1. It is also known as Gaussian distribution. Normal distributions are symmetrical, but not all symmetrical distribution are normal.

Ans11-There are several techniques for handling missing data.

1) Delete the sample with any missing data elements.
2) Impute the value of the missing data
3) Remove a variable that has a high incidence of missing data.

Few imputation techniques for handling missing data are as follows.

One of the most common techniques is **mean imputation** where the mean value of a variable is used in place of the missing data value for that same variable. Another technique is **K-Nearest Neighbour (KNN)** imputation where the missing values are replaced by the average of the K nearest neighbors[1]. **Multiple Imputation by Chained Equation (MICE)** is another technique that can be used to impute missing data.

Ans12- A/B testing is a user experience research methodology that involves comparing two or more versions of a variable to see which one performs better in terms of a desired outcome. A/B testing is also known as split testing. This testing is widely used by marketers, designers and developers.

Ans13- Yes, mean imputation of missing data is acceptable as mean imputation is a common method for handling missing data. It involves replacing missing values with the mean of the observed values. This method preserves the mean of the observed data and keeps the sample size up to the full sample size.

Ans14- Linear regression is a statistical method that establishes a linear relationship between two variables based on a line of best fit. It is the most common form of regression analysis. Linear regression model use a straight line to describe the relationship between variables.

Ans15- There are basically four branches or division in which statistics is divided.

1) Mathematical or Theoretical statistics

2) Statistical method or function
3) Descriptive statistics
4) Inferential statistics

However, the two main branches of the statistics are Descriptive and Inferential.