Statistics

1. a) True
2. a) Central Limit Theorem
3. b) Modelling bounded count data
4. d) All of the above mentioned
5. c) Poisson
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10.

Normal Distribution:

Normal distribution is also known as Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are frequent in occurrence than data far from the mean. The normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3. Normal distributions are symmetrical, but not all symmetrical distributions are normal.

11.

How do you handle missing data?

The best way to handle missing data is to develop contingency plans to minimise the damage.

Best techniques to handle missing data:

Use detention methods to eliminate missing data: The detention methods only work for certain datasets where participants have missing fields. There are several deleting methods-two common ones include Listwise Deletion and Pairwise Deletion. It means deleting any participants or data entries with missing values.

Use regression analysis to symmetrically eliminate data: Regression is useful for handling missing data because it can be used to predict the null value using other information from the dataset.

Average Imputation:

Uses the average value of the responses from the other entries to fill out missing values. However, a word of caution when using this method-it can artificially reduce the variability of the dataset.

Common-point imputation:

Is when the data scientists utilise the middle point or the most chosen value. For example, on a five-point scale, the substitute value will be 3. Something to keep in mind when utilising this method is the three types of middle values: mean, median and mode, which is valid for numerical data (it should be noted that for non-numerical data only the median and mean are relevant.

12.

A/B Testing:
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.
Mean imputation is typically considered terrible practice since it ignores feature correlation.
14.
In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).
15.
Branches of Statistics:
The two main branches of statistics are descriptive statistics and inferential statistics. Both are employed in scientific analysis of data, and both are equally important for the student of statistics.
Descriptive Statistics
Descriptive Statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing

experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

Inferential Statistics

Inferential Statistics as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important, and this aspect is dealt with in inferential statistics.