Statistics

Q1. True

Q2. Central limit Theorem

Q3. Modelling Bounded count data

Q4. All of the mentioned

Q5. Poisson

Q6. False

Q7. Hypothesis

Q8. 0

Q9. Outliers cannot conform to the regression relationship

Q10. A 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. The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distribution are similarly unlikely.

There are 3 types of distributions:

1. Symmetric Distribution
2. Skewed left
3. Skewed Right

Q11. Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical program will make the decision for you. Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea. Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analyzing the entire data set as if the imputed values were the true observed values.

The following are some of the most prevalent methods:

1. Mean imputation: Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks. Almost all of the methods described below are superior to mean imputation.
2. Regression imputation: The result of regressing the missing variable on other factors to get a predicted value. As a result, instead of utilizing the mean, you're relying on the anticipated value, which is influenced by other factors. This keeps the associations between the variables in the imputation model, but not the variability around the anticipated values.

Q12. A/B testing is a statistical way of comparing two or more versions such as version A & version B to determine not only which version performs better but also to understand if it difference between two versions is statistically significant. 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. This is the way businesses are run these days and they have to take a data-driven approach.
In A/B testing, A refers to 'control' or the original testing variable whereas B refers to 'variation' or a new version of the original testing variable. A common dilemma that companies face is that they think they understand the customer but in reality, customers would behave much differently than you would think consciously or subconsciously. Users don't often even know why they make the choice they make, they just do. but when running an experiment on an A/B test, you might find out otherwise and the results can often be very humbling, and customers can behave much differently than you would think so it's best to conduct tests rather than relying on intuition. We do A/B testing with the help of a Hypothesis.
Hypothesis is making a guess (not a wild guess) based on assumptions without scientific proof or explaining the situation based on reasonable assumptions.

Null Hypothesis H_0 Decisions always lead to the status quo. Current status/assumption doesn't change

Alternate Hypothesis H_1 Decisions lead to the opposite of Null Hypothesis H_0.

Q13. Mean imputation is the replacement of a missing observation with the mean of the non-missing observations for that variable. Mean imputation is typically considered terrible practice since it ignores feature correlation.
Problems associated with using Mean imputation are: -

1. Mean imputation does not preserve the relationships among variables.
2. Mean Imputation Leads to An Underestimate of Standard Errors

Q14. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least squares" method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable / features) from Y (independent variable / label). Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study. You can also use linear regression to provide better insights by uncovering patterns and relationships that your business colleagues might have previously seen and thought they already understood. For example, performing an analysis of sales and purchase data can help you uncover specific purchasing patterns on particular days or at certain times. Insights gathered from regression analysis can help business leaders anticipate times when their company's products will be in high demand.

Examples of linear-regression success: -

1. Evaluating trends and sales estimates
2. Analyze pricing elasticity
3. Assess risk in an insurance company

Q15. Data is a collection of values and Statistics is the branch of mathematics that deals with data. There are three real branches of statistics: -

1. Data collection: - Data collection is all about how the actual data is collected.
2. Descriptive Statistics: - It is the part of statistics that deals with presenting the data we have. This can take two basic forms – presenting aspects of the data either visually (via graphs, charts, etc.) or numerically (via averages and so on).

3. Inferential statistics: - This is quite a wide area. Inferential statistics is the aspect that deals with making conclusions about the data.