STATISTICS

1. A. True
2. A. Central limit theorem
3. A. Modeling even/time data
4. C. The square of a standard normal random variable follows what is called chi-squared distribution
5. C. Poisson
6. B. False
7. B. Hypothesis
8. A. 0
9. A. Outliers can have varying degrees of influence
10. Normal distribution represents a specific pattern of data distribution. The data tends to cluster symmetrically around a central mean value, creating a characteristic bell-shaped curve. The majority of data points are closer to the mean, and the further you move from the mean, the fewer data points you encounter.
11. Replace missing values with the mean value of the non-missing data in that column. If the amount of missing data is small and doesn't significantly affect the analysis, you might choose to simply remove the rows or samples with missing values.This is a simple method, but it might distort the data's distribution and variability. for categorical data, you can replace missing values with the mode of the non-missing data in that category. In time series data, missing values can be filled by propagating the last known value forward fill or using the next available value backward fill.
12. A/B testing is a method used in marketing, product development, and other fields to compare two versions of a webpage, application, or other variable to determine which performs better. It involves dividing a target audience into two groups: Group A experiences the original version (control group), while Group B experiences a modified version (experimental group) that includes a single change or variation. By analyzing the performance metrics of each group, such as click-through rates, conversion rates, or user engagement, you can determine which version is more effective in achieving your goals. A/B testing helps in making data-driven decisions to optimize and improve various aspects of a product or process.
13. Mean imputation is a quick and easy approach, it's not without its limitations. Its appropriateness depends on the nature of the data, the amount of missing data, and the goals of your analysis. In cases where missing data is substantial and/or not missing at random, more advanced imputation techniques or careful consideration of the impact of imputation on the analysis might be necessary. It's generally recommended to explore multiple imputation techniques or domain-specific approaches to handle missing data effectively.

14. Linear regression is used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between these variables, where changes in the independent variables correspond to proportional changes in the dependent variable. The goal of linear regression is to find the best-fitting line that minimizes the difference between the actual observed data points and the predicted values by the model. This line is determined by estimating coefficients for each independent variable, which represent the slope and intercept of the line.
15. A. Descriptive Statistics:- Involves methods for summarizing and describing data, including measures of central tendency (mean, median, mode), measures of dispersion (range, variance, standard deviation), and graphical representations.

    B. Inferential Statistics:- Deals with making predictions or inferences about populations based on sample data. It includes techniques like hypothesis testing, confidence intervals, and regression analysis.

    C. Biostatistics:- Applies statistical methods to biological and medical data, aiding in medical research, clinical trials, epidemiology, and public health studies.

    D. Econometrics:- Applies statistical methods to economic data, helping economists model and analyze economic relationships, forecast trends, and test economic theories.