1. a) True
2. a) Central Limit Theorem
3. b) Modeling bounded count data
4. c) The square of a standard normal random variable follows what is called chi-squared distribution
5. c) Poisson
6. a) True
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. The normal distribution, or Gaussian distribution, is a continuous probability distribution that is symmetrical and bell-shaped. It is defined by its mean (μ) and standard deviation (σ).
    The curve is symmetrical around the mean. Most data points are near the mean, with fewer as you move away. Empirical Rule is bout 68% of data lies within one standard deviation of the mean, 95% within two, and 99.7% within three. It is analyzing for real world data.

11. Handling missing data is crucial for analysis. Here are some common techniques:

    Remove

        Data:Remove Rows/Columns: Only if minimal data is missing or not important.

    Imputation Techniques:

        Mean/Median/Mode Imputation: Replace with mean, median, or mode.

        Forward/Backward Fill: Use previous or next value (good for time series).

        KNN Imputation: Use nearest neighbors' values (more accurate, but slower).

        MICE: Iteratively predict missing values using other features.

        Regression Imputation: Predict missing values using a regression model.

    Advanced Techniques:

        Deep Learning Models: For complex data.

        Multiple Imputation: Create multiple datasets with different imputed values and combine results.

    Recommendation: Choose based on data type and amount of missingness; always check the impact of imputation on results

12. A/B testing compares two versions (A and B) to see which performs better. Here's the process:

    - **Create Two Versions**: Change one element between them (like a headline or button).
    - **Split Audience**: Randomly divide users into two groups; one sees A, the other sees B.
    - **Measure Results**: Track metrics like clicks or conversions for both versions.
    - **Analyze**: Determine if differences are statistically significant.

    A/B testing helps in making data-driven decisions by identifying which version leads to better outcomes, thereby optimizing user experience and business goals.

13. Mean imputation, where missing values are replaced with the mean of the available data, is a simple and commonly used technique. However, it has some limitations and should be used with caution:
    Keeps the dataset size consistent, which can be useful in some analyses.

    Consider the Data Context: Use mean imputation cautiously; it can impact your analysis.

    Explore Alternatives: Look into methods like multiple imputation, K-nearest neighbors (KNN) imputation, or model-based approaches, which better preserve relationships in the data.
    Assess Impact: Check how imputation affects your results, especially in predictive modeling or inferential analysis.

14. Linear regression is a statistical method used to model and analyze the relationship between a dependent variable and one or more independent variables. The key idea is to fit a linear equation (a straight line) to observed data.

    $Y=\beta0+\beta1X+\epsilon$

    - YYY: Dependent variable
    - XXX: Independent variable(s)
    - β0\beta_0β0: Intercept
    - β1\beta_1β1: Slope (how much Y changes with X)
    - ϵ\epsilonϵ: Error term

    It's used for predicting values and understanding relationships between variables.

15. Statistics has two branches
    a. Descriptive statistics
    b. Inferential statistics

Descriptive Statistics: This branch focuses on summarizing and describing the features of a dataset. It includes measures like mean, median, mode, standard deviation, and visualizations like charts and graphs. Descriptive statistics help in understanding the basic characteristics of the data.

Inferential Statistics: This branch deals with making predictions or inferences about a population based on a sample. It involves hypothesis testing, confidence intervals, regression analysis, and other techniques to draw conclusions from data and make decisions under uncertainty.

.