

1. **A- True**
2. **A -Central Limit Theorem**
3. **B -Modeling bounded count data**
4. **D- All of the mentioned**
5. **C- Poisson**
6. **B- False**
7. **B- Hypothesis**
8. **A- 0**
9. **C- Outliers cannot conform to the regression relationship**

## 10. What do you understand by the term Normal Distribution?

Normal distribution, also known as Gaussian distribution, is a continuous probability distribution that is symmetrical around its mean, indicating that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, normal distribution will appear as a bell curve. The properties of a normal distribution include its bell-shaped curve, the mean, median, and mode all being equal, and the distribution being defined by its mean ( $\mu$ ) and standard deviation ( $\sigma$ ). It is widely used in statistics because many real-world variables are approximately normally distributed.

## 11. How do you handle missing data? What imputation techniques do you recommend?

Handling missing data is crucial for accurate statistical analysis and can be done using several methods:

1. **Deletion Methods:**
  - **Listwise Deletion:** Remove any observations with missing values.
  - **Pairwise Deletion:** Use available data without deleting entire observations.
2. **Imputation Methods:**
  - **Mean/Median Imputation:** Replace missing values with the mean or median of the column.
  - **Mode Imputation:** Replace missing values with the mode of the column (useful for categorical data).
  - **K-Nearest Neighbors (KNN) Imputation:** Use the average of the nearest neighbors' values to impute missing data.
  - **Multiple Imputation:** Use multiple models to estimate missing values and combine the results.
  - **Predictive Modeling:** Use regression or machine learning models to predict and replace missing values.

The choice of method depends on the nature of the data and the extent of the missing values. For smaller datasets, mean or median imputation may be sufficient, while for larger or more complex datasets, multiple imputation or predictive modeling may be more appropriate.

## 12. What is A/B testing?

A/B testing, also known as split testing, is a method used to compare two versions of a webpage, product, or marketing campaign to determine which one performs better. By randomly splitting the audience into two groups (A and B), one group receives the original version (control) and the other receives the new version (variant). The performance of each

version is then measured using a predefined metric (e.g., click-through rate, conversion rate). Statistical analysis is used to determine if any observed differences in performance are statistically significant, helping to make data-driven decisions.

### **13. Is mean imputation of missing data an acceptable practice?**

Mean imputation is a simple and commonly used method to handle missing data, where missing values are replaced with the mean of the observed values for that variable. While it is easy to implement and can be effective for datasets with a small proportion of missing data, it has limitations. Mean imputation can distort the variance and covariance of the data, leading to biased estimates and invalid statistical inferences. It does not account for the uncertainty of the missing values and can underestimate variability. Therefore, while it may be acceptable in some cases, more sophisticated methods like multiple imputation or model-based approaches are often recommended for better accuracy and reliability.

### **14. What is linear regression in statistics?**

Linear regression is a statistical method used to model and analyze the relationship between a dependent variable and one or more independent variables. The goal is to find the best-fitting line (regression line) that predicts the dependent variable based on the values of the independent variables.

### **15. What are the various branches of statistics?**

Statistics is broadly divided into two main branches:

1. **Descriptive Statistics:** This branch involves summarizing and describing the features of a dataset. Common techniques include measures of central tendency (mean, median, mode), measures of variability (range, variance, standard deviation), and graphical representations (histograms, pie charts, box plots).
2. **Inferential Statistics:** This branch focuses on making inferences about a population based on a sample. It includes hypothesis testing, confidence intervals, regression analysis, and analysis of variance (ANOVA).