1. a) True
2. a) Central Limit Theorem
3. b) Modelling 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. The normal distribution, also known as the Gaussian distribution, is a probability distribution that is often used to model natural phenomena such as measurements of physical characteristics or errors in experimental data.
    It is characterized by a bell-shaped curve that is symmetric around its mean. The mean, median, and mode of a normal distribution are all equal, and the distribution is completely defined by two parameters: its mean (μ) and standard deviation (σ).
    In a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, 95% falls within two standard deviations, and 99.7% falls within three standard deviations. This makes the normal distribution a useful tool for statistical analysis and hypothesis testing.
11. Some common imputation techniques for handling missing data:
    1. Mean or Median Imputation: This technique involves replacing missing values with the mean or median value of the non-missing data. It is a simple and fast method, but it can lead to biased estimates and can distort the variability of the data.
    2. Multiple Imputation: This technique involves creating several imputed datasets, each of which is analysed separately, and then combining the results. It is a more complex method that can handle missing data with greater accuracy and precision, but it requires more computational resources and can be time-consuming.
    3. Regression Imputation: This technique involves using regression analysis to estimate the missing values based on the relationship between the missing variable and other variables in the dataset. It can be a powerful technique, but it assumes that there is a linear relationship between the missing variable and the other variables in the dataset.
    4. K-Nearest Neighbour Imputation: This technique involves finding the K-nearest observations to the observation with missing values and using their values to impute the missing values. It can be effective when there is a clear clustering of values in the data, but it can be sensitive to the choice of K.
12. A/B testing, also known as split testing, is a statistical method used to compare two versions of a product or service to determine which one performs better. It involves randomly assigning participants to two groups, where one group is shown the original version of the product or service (the control group) and the other group is shown a modified version (the treatment group). The purpose of A/B testing is to identify which version of the product or service leads to a higher conversion rate or better performance in some other metric of interest. Examples of metrics that could be measured include click-through rates, engagement rates, or sales revenue.
13. One problem with mean imputation is that it assumes that the missing values are missing completely at random (MCAR), meaning that the probability of a value being missing is unrelated to the value itself or any other variables in the dataset. However, in practice, missing data is often missing not at random (MNAR), meaning that the probability of a value being missing depends on the value itself or other variables in the dataset. In such cases,

mean imputation can lead to biased estimates of the mean and standard deviation of the data.

Another problem with mean imputation is that it can reduce the variability of the data and underestimate the standard error of estimates, which can lead to overconfidence in statistical tests and hypothesis tests. This is because mean imputation reduces the amount of variation in the data, which can lead to inflated statistical significance.

In summary, mean imputation is a simple and easy-to-use method for handling missing data, but it can lead to biased results and distorted variability. As such, it should be used with caution and only in situations where the assumption of MCAR is reasonable. If the assumption of MCAR is not reasonable, more advanced imputation techniques such as multiple imputation or regression imputation may be more appropriate.

14. Linear regression is a statistical technique used to model the relationship between a dependent variable (also known as the response variable) and one or more independent variables (also known as predictors or explanatory variables). The objective of linear regression is to find the best linear relationship between the dependent variable and the independent variable(s) that can be used to predict the value of the dependent variable for new observations.
15. Statistics is a broad field that encompasses a wide range of subfields, some of which are listed below:
    1. Descriptive Statistics: This branch of statistics involves summarizing and describing the main features of data, such as measures of central tendency and dispersion.
    2. Inferential Statistics: This branch of statistics involves making inferences about a population based on a sample of data. It includes techniques such as hypothesis testing, confidence intervals, and regression analysis.
    3. Probability Theory: Probability theory is the study of the likelihood of events occurring. It is used in many areas of statistics, including hypothesis testing and regression analysis.
    4. Bayesian Statistics: Bayesian statistics is a branch of statistics that uses prior knowledge or beliefs about a parameter to update the likelihood of the parameter given the data.
    5. Biostatistics: Biostatistics is the application of statistical methods to biological and medical data, including clinical trials, epidemiological studies, and genetics.
    6. Econometrics: Econometrics is the application of statistical methods to economic data, including regression analysis, time-series analysis, and panel data analysis.
    7. Social Statistics: Social statistics is the application of statistical methods to social science data, including surveys, opinion polls, and demographic data.
    8. Machine Learning: Machine learning is a branch of artificial intelligence that uses statistical methods to develop algorithms that can learn from data and make predictions or decisions.