1. A) True
2. A) Central Limit Theorem
3. B) Modelling bounded count 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. C) Outliers cannot conform to the regression relationship

10. Normal distribution, also known as Gaussian distribution, is a continuous probability distribution that is symmetrical and bell-shaped. It is commonly used in statistics to model real-world phenomena, such as the distribution of heights or IQ scores in a population. The distribution is characterized by its mean (μ) and standard deviation (σ), and about 68% of the data falls within one standard deviation of the mean, about 95% of the data falls within two standard deviations, and about 99.7% of the data falls within three standard deviations.
11. Handling missing data is an important step in data analysis as missing data can lead to biased or inaccurate results. Imputation techniques are used to fill in the missing data points. Some commonly used imputation techniques are mean imputation, median imputation, mode imputation, regression imputation, and multiple imputation. The choice of imputation technique depends on the nature of the data and the reason for the missingness.
12. A/B testing, also known as split testing, is a statistical method used to compare two versions of a product, webpage, or marketing campaign. The objective is to determine which version performs better based on a chosen metric, such as click-through rate or conversion rate. The two versions, A and B, are randomly assigned to different groups of users, and their performance is measured and compared to determine the winner.
13. Mean imputation is a commonly used imputation technique where missing values are replaced with the mean value of the available data. While mean imputation is a simple and fast technique, it can lead to biased or inaccurate results as it assumes that the missing values are missing completely at random. Therefore, mean imputation should be used with caution and only when the missingness is expected to be non-systematic.
14. Linear regression is a statistical method used to model the relationship between two variables, where one variable is considered the dependent variable and the other variable is considered the independent variable. The goal is to find a linear equation that best predicts the value of the dependent variable based on the independent variable. The linear equation takes the form of Y = a + b, X, where Y is the dependent variable, X is the independent variable, a is the intercept, and b is the slope.
15. Statistics is a broad field that includes various branches, such as descriptive statistics, inferential statistics, probability theory, regression analysis, time series analysis, multivariate analysis, Bayesian statistics, and nonparametric statistics. Descriptive statistics involves summarizing and describing data using measures such as mean, median, mode, and standard deviation. Inferential statistics involves making inferences about a population based on a sample. Probability theory is the study of random events and the likelihood of their occurrence. Regression analysis is a statistical method used to model the relationship between variables. Time series analysis is the analysis of data collected over time. Multivariate analysis involves analysing multiple variables simultaneously. Bayesian statistics is a statistical approach that involves updating prior knowledge as new data is collected. Nonparametric statistics is a branch of statistics that does not rely on the assumption of a specific probability distribution.