

ANS1. a) True

ANS2. a) Central Limit Theorem

ANS3. b) Modeling bounded count data

ANS4. a) The exponent of a normally distributed random variables follows what is called the log-normal distribution

ANS5. c) Poisson

ANS6. b) False

ANS7. b) Hypothesis

ANS8. a) 0

ANS9. c) Outliers cannot conform to the regression relationship

ANS10. Normal distribution is a statistical conception that describes a symmetric, bell-shaped probability distribution. In a normal distribution, data tends to cluster around a central point with the majority of the observations falling close to the average value. The distribution is characterized by its mean and standard deviation.

In a perfectly normal distribution

The mean, median, and mode are all equal and located at the center of the distribution.

The curve is symmetric, with half of the data points falling to the left and half to the right of the mean.

The standard deviation determines the spread of the distribution. The larger the standard deviation, the wider the curve.

Normal distribution is an abecedarian concept in statistics and is frequently used in colorful fields to model and dissect natural marvels, making it a handy tool in data analysis and hypothesis testing.

ANS11. Handling missing data is a crucial aspect of data analysis, and there are various techniques to address this issue. The choice of imputation method depends on the nature of the data and the specific characteristics of the missing values. Here are some common imputation techniques:

Mean imputation is a simple technique that involves replacing missing values with the mean of the observed values in the same variable.

Mode imputation is another simple technique that involves replacing missing values with the mode of the observed values in the same variable.

Regression imputation is a more complex technique that involves using regression models to predict missing values based on the observed values in other variables.

Stochastic imputation is a technique that involves generating random values from a probability distribution to replace missing values

Multiple imputation is a technique that involves generating multiple plausible imputed datasets and then analyzing each dataset separately to obtain a final result.

ANS12. A/B testing is a research methodology that involves comparing multiple versions of a single variable to determine which version is more effective. In an A/B test, two variants (A and B) are allocated in random order to various groups of users, and their outcomes are compared to determine which variant is the most effective. A/B testing is widely used in marketing to test the effectiveness of different versions of emails, web pages, product designs, apps, and more.

ANS13. Mean imputation is a technique for handling missing data that involves replacing missing values with the mean of the observed values in the same variable. While it is a popular technique due to its simplicity, it has several drawbacks. For example, mean imputation can lead to bias in multivariate estimates such as correlation or regression coefficients.

Despite these drawbacks, mean imputation can be a useful technique in certain situations. For example, mean imputation is a good choice when the data is missing at random and the variable is normally distributed¹. However, it is important to carefully consider the nature of the data and the research question before deciding on an imputation technique.

ANS14. Linear regression is a tool in statistics that helps to model the relationship between two variables by fitting a linear equation to the observed data. In other words, it is a technique used to predict the value of a dependent variable based on the value of one or more independent variables.

The simplest form of linear regression is simple linear regression, which involves only one independent variable and one dependent variable.

Linear regression can be used for both simple and multiple regression analysis. In simple linear regression, there is only one independent variable, while in multiple linear regression, there are two or more independent variables.

ANS15. Descriptive statistics: This branch of statistics deals with the presentation and collection of data. It uses different methods like mean median mode etc.

Inferential statistics: This branch of statistics involves drawing conclusions from the statistical analysis that has been performed using descriptive statistics. It gives us what is the relationship between the data.

Predictive statistics : This branch deals with the prediction can be drawn from the present data from using machine learning.

Prescriptive statistics : This branch uses for gives us suggestion resulted from the data.