

STATISTICS WORKSHEET-1

Q1) a) True

Q2) a) Central Limit Theorem

Q3) b) Modeling bounded count data

Q4) c) The square of a standard normal random variable follows what is called chi-squared distribution

Q5) c) Poisson

Q6) False.

Q7) b) Hypothesis. Hypothesis testing is concerned with making decisions using data.

Q8) a) 0

Q9) c) Outliers cannot conform to the regression relationship.

10. The Gaussian distribution, commonly known as the Normal distribution, is a symmetric probability distribution with a bell-shaped curve. It's a basic idea in statistics and probability theory. Most data points in a normal distribution cluster around the mean, with fewer data points moving away from the mean in either direction. Many natural occurrences in the actual world, such as height measures or IQ scores, have a normal distribution.

11. Different approaches are used to handle missing data in order to fill in the gaps in data sets. One such method is imputation, in which missing values are predicted based on the available data. Mean imputation, median imputation, and more sophisticated techniques like multiple imputation or k-nearest neighbours imputation are a few suggested imputation techniques. The type of data and its possible impact on analysis influence the method selection.

12. A/B testing, commonly referred to as split testing, compares two iterations of a website, app, or other digital content to see which one performs better in terms of a particular outcome, such as click-through rate or conversion rate. Users are randomly assigned to one of two versions (A or B), data on their interactions are gathered, and the version with the most favourable results is determined by analysis of the data.

13. Mean imputation includes substituting the mean of the available data for that variable for missing values. Although it is a straightforward method, it has the potential to bias results and understate variability by changing the distribution and relationships in the data. In general, it's thought to be a less preferable approach, particularly when there are a lot of missing data.

14. The relationship between a dependent variable and one or more independent variables can be modelled statistically using linear regression. It is predicated that a linear equation can roughly represent this relationship. Finding the line that minimises the difference between the observed data points and the projected values based on the linear equation is the best course of action. A lot of people utilise linear regression to make predictions, determine the direction and strength of correlations, and draw conclusions.

15. Various branches of statistics include:

- Descriptive Statistics: Involves summarizing and presenting data to describe its main features.

- Inferential Statistics: Deals with making inferences or predictions about a population based on a sample.

- Probability Theory: Studies the likelihood of different outcomes in uncertain situations.

- Biostatistics: Applies statistical methods to biological and medical data.

- Econometrics: Applies statistical methods to economics and financial data.

- Bayesian Statistics: Utilizes Bayes' theorem to update probabilities based on new information.

- Multivariate Statistics: Analyzes data with multiple variables simultaneously.

- Nonparametric Statistics: Involves methods that don't rely on specific distribution assumptions.

- Time Series Analysis: Focuses on analyzing data points collected over time to identify patterns or trends.