STATISTICS WORKSHEET-1

Answer 1. a) True

Answer 2. a) Central Limit Theorem

Answer 3. b) Modeling bounded count data

Answer 4. c) The square of a standard normal random variable follows what is called chi-squared distribution

Answer 5. c) Poisson

Answer 6. b) False

Answer 7. b) Hypothesis

Answer 8. a) 0

Answer 9. c) Outliers cannot conform to the regression relationship

Answer 10. The normal distribution, also known as the Gaussian distribution, is a crucial concept in statistics. It forms a bell-shaped curve where data clusters around the mean, displaying symmetry. Many real-world phenomena, like heights and errors, follow this pattern due to the central limit theorem. It's characterized by mean (μ) and standard deviation (σ) parameters that determine its center and spread.

Answer 11. As missing values can lead to inaccurate results. There are several techniques for imputing missing data, and the choice of technique depends on the nature of the data, the amount of missing data, and the specific goals of your analysis. I recommend Regression Imputation When there's a clear relationship between the variable with missing data and other variables, regression imputation can work well. It's useful for situations where you can predict the missing value based on other features.

Answer 12. A/B testing is a statistical approach to compare two versions of something, often used in marketing or research. Groups A and B experience different versions, and their outcomes are compared. Statistical analysis helps determine if differences are significant, aiding decisions on which version is more effective.

Answer 13. Mean imputation is straightforward but risky. It assumes randomness in missing data and can introduce bias, affecting results. It's acceptable when data is missing at random, but advanced methods like regression imputation or multiple imputation are more reliable for capturing data relationships.

Answer 14. Linear regression is a statistical technique for modeling relationships between a dependent variable and one or more independent variables. It assumes a linear connection, aiming to find the best-fitting line that minimizes the difference between observed and predicted values. It's used for prediction, understanding variable impact, and identifying trends.

Answer 15. Statistics encompasses several branches. Descriptive statistics summarizes data. Inferential statistics draws conclusions from samples. Probability deals with uncertainty. Biostatistics analyzes biological data. Econometrics applies statistics to economics. Social statistics studies social trends. Multivariate stats handles multiple variables. Time series analysis examines sequential data. Experimental design plans controlled studies. Statistical machine learning merges statistics with machine learning for predictive modeling, each branch serving distinct analytical goals