STATISTICS WORKSHEET-1

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. What do you understand by the term Normal Distribution?

Answer: The Normal Distribution, also known as the Gaussian distribution, is a continuous probability distribution that is symmetric about its mean. It's characterized by its bell-shaped curve. Key properties include:

- 68% of data falls within one standard deviation of the mean

- 95% within two standard deviations

- 99.7% within three standard deviations

It's crucial in statistics due to the Central Limit Theorem and its applicability to many natural phenomena.

11. How do you handle missing data? What imputation techniques do you recommend?

Answer: Missing data can significantly impact analysis. Common techniques include:

- Mean/median imputation: Replace missing values with the mean or median of the variable.

- Multiple imputation: Create multiple plausible imputed datasets and combine results.

- Regression imputation: Predict missing values using other variables.

- Hot deck imputation: Replace missing values with observed values from similar respondents.

- Maximum likelihood estimation: Estimate parameters directly from the incomplete data.

12. What is A/B testing?

Answer: A/B testing is an experimental approach to compare two versions (A and B) of a single variable. Process:

1. Form a hypothesis

2. Create two versions (control and variation)

3. Randomly split your sample group

4. Collect and analyse data

5. Draw conclusions

13. Is mean imputation of missing data acceptable practice?

Answer: Mean imputation is generally not acceptable because:

- It reduces variability in the data

- It weakens covariance and correlation estimates

- It biases estimates of variances and covariances

- It assumes missing completely at random (MCAR)

14. What is linear regression in statistics?

Answer: Linear regression models the relationship between a dependent variable (Y) and one or more independent variables (X) using a linear equation:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$

Where:

- β_0 is the y-intercept

- $\beta_1, \beta_2, ..., \beta_n$ are the coefficients

- ε is the error term

It's used for prediction and understanding variable relationships.

15. What are the various branches of statistics?

Answer: Descriptive statistics: Summarizing and describing data

- Inferential statistics: Drawing conclusions from samples about populations

- Bayesian statistics: Incorporating prior knowledge into statistical inference

- Nonparametric statistics: Methods not relying on data belonging to any particular distribution
- Multivariate statistics: Analyzing multiple variables simultaneously
- Time series analysis: Analyzing data points collected over time
- Spatial statistics: Analyzing spatial data and patterns

Each branch has its own set of methods and applications in various fields.