STATISTICS WORKSHEET-1

* **MCQ** **answer type questions**

1. b) False
2. a) Central Limit Theorem
3. c) Modelling contingency tables
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

* **Subjective answer type questions**

1. **Normal Distribution:**
2. The normal distribution, also known as the Gaussian distribution, is a fundamental concept in statistics.
3. It describes a symmetric bell-shaped curve characterized by its mean (average) and standard deviation.
4. Many natural phenomena, such as heights, weights, and test scores, follow a normal distribution.
5. **Missing data** can occur due to various reasons (e.g., measurement errors).

* **Common imputation techniques include**:

1. **Mean Imputation**: Replacing missing values with the mean of the available data. Simple but may distort variability.
2. **Median Imputation**: Similar to mean imputation but uses the median.
3. **Multiple Imputation**: Generates multiple imputed datasets and combines results.
4. **Regression Imputation**: Predict missing values using regression models.
5. **K-Nearest Neighbours (KNN)**: Impute based on similar observations.
6. **Hot Deck Imputation**: Borrow values from similar cases.
7. **Time-Series Imputation**: For time-series data.
8. **Domain-Specific Methods**: Tailored to the context.
9. **A/B Testing**:
10. A/B testing is a method used to compare two
11. It involves randomly assigning users to different variants (A and B) and measuring their responses.
12. By comparing outcomes, we determine which variant performs better.
13. Commonly used in marketing, product development, and user experience optimization.
14. **Mean Imputation of Missing Data**, while mean imputation is simple, it has limitations:
15. Ignores relationships between variables.
16. Reduces variability.
17. May introduce bias if missingness is not random.
18. It's acceptable for small amounts of missing data or when other methods are impractical.
19. However, consider more advanced techniques for better results.
20. **Linear Regression**:
21. Linear regression is a statistical method used to model the relationship between a dependent variable (response) and one or more independent variables (predictors).
22. It assumes a linear relationship (straight line) between the variables.
23. The goal is to find the best-fitting line (regression equation) that minimizes the sum of squared differences between observed and predicted values.
24. Widely used for prediction, understanding associations, and identifying significant predictors.
25. **Branches of Statistics**:

* Statistics encompasses various branches:

1. **Descriptive Statistics**: Summarizing and visualizing data.
2. **Inferential Statistics**: Drawing conclusions about populations based on samples.
3. **Biostatistics**: Applied to biological and health sciences.
4. **Econometrics**: Applying statistical methods to economic data.
5. **Social Statistics**: Analysing social phenomena.
6. **Multivariate Statistics**: Dealing with multiple variables simultaneously.
7. **Bayesian Statistics**: Incorporating prior knowledge into analyses.
8. **Nonparametric Statistics**: Not assuming specific distributions.
9. **Time Series Analysis**: Studying data over time.
10. **Experimental Design**: Planning controlled experiments.
11. **Machine Learning**: Using algorithms for prediction. (Also known asStatistical Learning)