STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
a) True

b) False

Ans a) True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?
a) Central Limit Theorem

b) Central Mean Theorem

c) Centroid Limit Theorem

d) All of the mentioned

Ans a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

a) Modeling event/time data

b) Modeling bounded count data

c) Modeling contingency tables

d) All of the mentioned

Ans b) Modeling bounded count data

4. Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
c)
The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned

Ans d) All of the mentioned

5. _____random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

Ans b) False

7. 1. Which of the following testing is concerned with making decisions using data?
a) Probability

b) Hypothesis

c) Causal
d) None of the mentioned

Ans b) Hypothesis

8. 4. Normalized data are centered at____and have units equal to standard deviations of the original data.

a) 0

b) 5

c) 1

d) 10

Ans A) 0

9. Which of the following statement is incorrect with respect to outliers?

a) Outliers can have varying degrees of influence

b) Outliers can be the result of spurious or real processes

c) Outliers cannot conform to the regression relationship

d) None of the mentioned

Ans c) Outliers cannot conform to the regression relationship

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

Ans) Normal distribution is an important idea in statistics that looks like a bell-shaped curve. It is balanced around the middle, where the mean, median, and mode are all the same. The spread of the data is controlled by the standard deviation: a larger standard deviation makes the curve wider, and a smaller one makes it narrower. There's a helpful rule called the 68-95-99.7 rule, which says that about 68% of the data falls within one standard deviation of the mean, 95% within two, and 99.7% within three. This type of distribution is useful because many statistical tests assume that data follows this pattern, which helps in making predictions and decisions in areas like finance, healthcare, and social sciences.

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

1)Understand Why Data is Missing:

2)Imputation Techniques:

Mean/Median/Mode Imputation

Forward/Backward Fill:

K-Nearest Neighbors (KNN) Imputation:

Regression Imputation:

Multiple Imputation:

Interpolate:

Dropping Data:

In Overall check for Patterns ,Look at how much data is missing and why.

Choose the right method based on your data and the analysis you plan to do.

After filling in missing values, check how it affects your results.

In short, dealing with missing data involves understanding why it’s missing and choosing the right way to fill in those gaps, keeping in mind how it can change your analysis.

12. What is A/B testing?

Ans) A/B testing, or split testing, is a way to compare two versions of something, like a web page or app, to see which one is better. First, you decide what you want to measure, like clicks or sales. Then, you create two versions, making a small change, such as a different button color or headline. You randomly divide your audience into two groups: Group A sees the original version, and Group B sees the new version. After running the test for some time, you check which version performed better. If one is more effective, you use that version to improve your site or app. A/B testing helps you make smarter decisions based on real data, improves user experience, and can increase the number of people taking action, like buying something or signing up.

13. Is mean imputation of missing data acceptable practice?

Ans) Mean imputation is a common way to fill in missing data by replacing missing values with the average of the existing values. While it’s easy to do and keeps all your data, it has some downsides. One problem is that it reduces the variety in the data, which can lead to wrong conclusions. If the missing data isn't random, using the mean can introduce bias and

make the results less trustworthy. It also doesn't consider how different variables are related to each other. Mean imputation might be okay if only a small amount of data is missing and if the missing data is completely random. However, for better accuracy, it’s often better to use other methods like median imputation or K-Nearest Neighbors

14. What is linear regression in statistics?

Ans Linear regression is a statistical method used to understand how one variable affects another. It finds the best straight line that shows the relationship between a dependent variable (the one you want to predict) and one or more independent variables (the ones you use to make predictions). The basic formula is y=mx+by = mx + by=mx+b, where yyy is the dependent variable, xxx is the independent variable, mmm is the slope of the line, and bbb is where the line crosses the y-axis. There are two types: simple linear regression, which uses one independent variable, and multiple linear regression, which uses two or more. Linear regression works under certain assumptions, like having a linear relationship and consistent variance in the errors. It is commonly used in fields like economics and social sciences for tasks like forecasting and understanding relationships. The performance of the model is measured using metrics like R-squared and Mean Squared Error. Overall, linear regression is a useful tool for analyzing data and making predictions.

15. What are the various branches of statistics?

Ans Two main branches are Descriptive Statistics and Inferential Statistics otter types are as follows

Inferential Statistics
Bayesian Statistics
Regression Analysis
Multivariate Statistics
Statistical Learning Theory
Experimental Design
Time Series Analysis
Survival Analysis