STATISTICS WORKSHEET-1

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

a) True

b) False

ANSWER : OPTION 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

ANSWER : OPTION 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

ANSWER : OPTION 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

ANSWER : OPTION C (The square of a standard normal random variable follows what is called chi-squared distribution)

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

ANSWER : OPTION C (Poisson)

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

a) True

b) False

ANSWER : OPTION B (False)

7. Which of the following testing is concerned with making decisions using data?

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

ANSWER : OPTION B (Hypothesis)

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

a) 0

b) 5

c) 1

d) 10

ANSWER : OPTION 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

ANSWER : OPTION C (Outliers cannot conform to the regression relationship)

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

ANSWER : Normal Distribution shows us how data is distributed around average or mean of data. It is continues probability distribution and having bell shaped curve.

1. **For normal distribution the curve is symmetry about its mean or average.**
2. **For normal distribution Mean = Median = Mode.**
3. **For normal distribution smaller is the standard deviation narrower and taller is the curve and vice versa.**
4. **For normal distribution 68% data covered for 1 standard deviation,95% data covered for 2 standard deviation and 97.5% data covered for 3 standard deviation.**

Normal distribution used in, Statistical Analysis, Quality Control, Finance and Economics and many more fields.

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

ANSWER : For handling missing data we can use multiple ways to handle depends on type of data we have in dataset.

1. **NaNs will check for NaNs, if its less in count and it will not affect for anylysis will drop NaNs.**
2. **Replacing value with mean/median or mode.**
3. **Replacing missing value with the weighted average of similer observation this can used for small dataset only.**
4. **Replacing missing values with observed values from randomly selected records.**
5. **Replacing missing values using linear regression to predict the missing values.**
6. **Replacing missing values using various imputation technics.**

Imutation Techiques

1. **Knn imputation : This technique is used by estimating based on values by their nearest neighbours. This technique is used in multivariate data.**

   Steps in knn imputation – 1.identify neighbours
   2.Finding weighted average
   3.Imputation.

12. What is A/B testing?

ANSWER : A/B testing refers to a process of experimentation randomly wherein two or more versions of a variable like web page, page element, etc. are shown to different segments of website visitors or users at the same time to determine which version leaves the maximum impact and drive business metrics. We can perform the A/B test on the Site Pages, Flow, and Elements, Business Model, Backend Functionality and Algorithms, New Products or Services.

Advantages of A/B Testing :

1. **Increase Revenue and Conversions**
2. **Rapid Iteration**
3. **Learn what works**
4. **Uses Actual Site Visitors**
5. **Data-Driven Decision Making**

Steps Involved in A/B Testing :

1. **Select the type of test**
2. **The metric we can take from Bounce Rate, Exit Rate, Engagement metrics here.**
3. **Creation of the hypothesis you need to test.**
4. **Set the variation pattern**
5. **Creat the pattern and structure.**
6. **Activate for an experiment basis for a shorter time and analyze the improvement**
7. **Calculate the Results using Metrics.**
8. **Publish the variation**
9. **Repeat.**

13. Is mean imputation of missing data acceptable practice?

ANSWER :

Mean imputation will be acceptable practice if,

1. **Missing values are completely random in nature.**
2. **If other imputation methods are not intensively working.**
3. **If Mean imputation will not affect trueness of dataset.**
4. **If mean imputation will not affect variability of dataset.**
5. **If it doesn't affect relationship.**

14. What is linear regression in statistics?

ANSWER:

Linear regression is one of the most basic types of regression in supervised machine learning. The linear regression model consists of a predictor variable and a dependent variable related linearly to each other.

This can be expressed in straight line equation y = mx+c

Where,

1. **Dependent Variable = y this is variable which want to predict**
2. **Independent variable = x = this is feature or input variable**
3. **Constant Value = c**

Linear Regression used predicting future values of a dependant variable based on new values of the independent variable.

15. What are the various branches of statistics?

ANSWER :

Below are the branches of statistics

1. **Descriptive Statistics**
2. **Inferential Statistics**
3. **Multivariate Statistics**
4. **Biostatistics**
5. **Environmental Statistics**
6. **Educational Statistics**