STATISTICS WORKSEEHT- 1

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

a) True

b) False

Ans:- a. True

02. 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 increase?

a) Central Limit Theorem

b) Central Mean Theorem

c) Centroid Limit Theorem

d) All of the mentioned

Ans:- a) Central Limit Theorem

03. 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

04. 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

05._____ random variables are used to model rates.

a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Ans:- c) Poisson

06. Usually replacing the standard error by its estimated value does not changes the CLT.

a) True
b) False

Ans:- b) False

07. 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

Q 8. Normalized data are entered at _____ and have all units equal to standard deviations of the original data.

a) 0
b) 5
c) 1
d) 10

Ans:- a) 0

Q 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 process.
c) Outliers cannot conform to the regression relationship
d) None of the mentioned

Ans:- c) Outliers cannot conform to the regression relationship

Q 10. What do you understand by the term normal distribution?

Ans:- A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme. In graph form normal data will

appear as a bell curve and the curve is symmetric at the center. In normal distribution mean median and mode all are equal.

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

Ans:- We handle missing data by filling up those missing values if its necessary using through various imputation technique or otherwise we can remove those values if it is useless for our dataset.

Various imputation techniques we can use is

1. **Mean/median/mode imputation**
2. **Assigning unique values**
3. **Predicting missing values Substitution**
4. **Regression imputation**
5. **Delete rows with the missing values**

Q 12. What is A/B testing?

Ans:- A/B testing is basically a form of hypothesis testing. It is a two-sample hypothesis testing to compare two versions. This testing use to improve and optimize user experience and marketing.

Q 13. Is mean imputation of missing data acceptable practice?

Ans:- Mean imputation cannot works proper every time. We cannot implement this technique every time because if we wants to fill missing values it takes mean of every value and fill it with those values which may

be good or bad for example a failure student who did not answer even a single question in exam and its score section is missing and we fill those values with the mean data and that student gets good marks…so It is not so good practice.

Q 14. What is linear regression in statistics?

Ans : -Linear regression is a statistical technique which is used for predictive analysis. This technique will identify the strength of the impact that the indepdendent variable makes on dependent variable. In statistics, linear regression is a linear approach for modeling the relationship between a scaler response and one or more explanatory variables also known as dependent and independent variable

Q 15. What are the various branches of statistics?

Ans:- There are two main branches of statistics. Both of these are used in scientific data analysis and are equally significant

1. **Descriptive statistics:- it deals with preservation and collection of data from samples by making use of index for getting mean or standard deviation. The methods used in descriptive statistic are displaying, analysing and describing the data.**
2. **Inferential statistics:- It involves drawing the right conclusion from the statistical analysis that has been performed using descriptive spastics.**