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.Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
Ans :- 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
Ans ; - 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
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

Q10 and 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 also known as Gaussian Distribution. If we plot a random variable 'X' by using probability density function or Histogram 'a bell shaped' curve is formed. Hence this shows data is normally distributed and most of the data lies between first standard deviation. Hence normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest of the data is spread towards the right or left extreme.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans:- To perform data analysis when data is provided there might be missing values in the dataset which are written as NaN in numpy. It is important to fill these values otherwise whatever analysis is going to be done on data may produce wrong output
Some methods to handle missing data:

- Delete rows which contain missing values.
- Use of mean , mode , median of particular column to fill the missing data.
- Use of zero if it is good for data
- Replace the data with min or max data

12. What is A/B testing?
Ans :- A/B testing is a type of experiment in which traffic on web, app or any other platform is being split into A and B groups and analysis is done. It allows decision-makers to choose the best design or best model for a website by looking the analytics results obtained with two possible alternatives A and B

13. Is mean imputation of missing data acceptable practice?
Ans :- The process of replacing null values in a dataset with the data mean is known as mean imputation.
Mean imputation of missing data is not acceptable practice as it reduces accuracy of the model sometimes. Problems occur while using mean imputation includes ignorance of feature correlation, also it reduces variance of data.

14. What is linear regression in statistics?
Ans :- Linear regression is a statistical practice which provides relationship between dependent variable and independent variable using equation y = mx + c
Where,
m is slope of line
x is independent feature
y is dependent feature
c is intercept

Regression analysis is used to predict the value of y variable by using independent x variable.
Linear regression is well known algorithm in Machine learning

15. What are the various branches of statistics?
Ans :- Branches of Statistics:-

1. Descriptive Statistics
   The data is summarized through given observation
2. Inferential Statistics
   Used to interpret the meaning of Descriptive statistics