STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.
(Correct Answer are marked in RED)

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

Answer: A

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: A

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: B

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: D

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Answer: C

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True b) False

Answer: B

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

Answer: B

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

Answer: 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: C

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

10. What do you understand by the term Normal Distribution?
Answer: The normal distribution is also known Gauss distribution. a normal distribution is an important class of Statistical Distribution with a wide range of uses. the normal distribution is applied in most Machine Learning Algorithms. two main parameter of normal distribution are - mean and standard deviation.

11. How do you handle missing data? What imputation techniques do you recommend?
Answer: Missing data is an internal part of the process. a data analyst, pours a lot of resources, time, and energy into making sure the data set is accurate as possible.
The best techniques to handle missing data are:

- Use deletion methods to eliminate the missing data
- Use regression analysis to systematically eliminate the data
- we can use data imputation techniques
- Keeping things under the control

12. What is A/B testing?
Answer: it is one of the popular controlled experiment used to optimize web marketing strategies. It allows decision-makers to choose the best design or mockups for a website by looking at the analytics results obtained with two possible alternatives A and B.

13. Is mean imputation of missing data acceptable practice?
Answer: The process of replacing the null values in a data collection with the data mean is known as mean imputation.
Mean imputation is considered a terrible practice since it ignores feature correlation.
Mean imputation decrease the variance of the data while increasing the bias. As a result due to the reduced variance, the model is less accurate and the confidence interval is narrow.

14. What is linear regression in statistics?
Answer: Linear regression analyses are used to predict the value of a variable based on the value of another or different variable. The variable we want to predict is known as the dependent variable. The variable we are using to predict the other variable value is known as independent variable.
Linear regression is one of the well-known algorithms in statistics and machine learning. the many different names by which linear regression is known.

15. What are the various branches of statistics?
Answer: Branches of Statistics:
1. Descriptive Statistics.
It is the first part of statistics that deals with the collection of data.

2. Inferential Statistics
These are the techniques that enable statisticians to use the collected data from the sample to conclude, bring decisions, or to predict.