Statistics worksheet

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: A

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

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

Answer: Normal distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. Normal distribution is also called as Gaussian distribution.

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

Answer: 1. Delete the Data 2. Imputing Averages 3. Assign New Category 4. Certain Algorithms.

12. What is A/B testing?

Answer: A/B testing is a methodology for comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal. A/B testing is also known as split testing.

13. Is mean imputation of missing data acceptable practice?

14. What is linear regression in statistics?

Answer: Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called dependent variable. The variable you are using to predict the other variable's value is called independent variable.

15. What are the various branches of statistics?

Answer: There are two types of branches of statistics as Descriptive statistics and inferential statistics.

1. **Descriptive Statistics:** Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures.
2. **Inferential Statistics:** Inferential Statistics is a method that allows us to use information collected from a sample to make decisions, predictions or inferences from a population.