WORKSHEET

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)

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)

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)

mohammedsyedtaher1@gmail.com

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)

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans: c)

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

a) True

b) False

Ans:b)

mohammedsyedtaher1@gmail.com

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)

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)

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)

WORKSHEET

mohammedsyedtaher1@gmail.com

Q10and 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, also known as the Gaussian 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. In graph form, normal distribution will appear as a bell curve.

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

Ans: When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data.

- Imputation techniques: 1) Complete Case Analysis.2) Arbitrary Value Imputation, 3) Frequent Category Imputation.

12. What is A/B testing?

Ans: A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Ans: Although imputing missing values by using the mean is a popular imputation technique, there are serious problems with mean imputation. The variance of a mean-imputed variable is always biased downward from the variance of the un-imputed variable. This bias affects standard errors, confidence intervals, and other inferential statistics. Experts agree that mean imputation should be avoided when possible

mohammedsyedtaher1@gmail.com

14. What is linear regression in statistics?

Ans: 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 the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Ans: There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.

mohammedsyedtaher1@gmail.com