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

Answer : 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

Answer : 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

Answer: 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

Answer : d) All of the mentioned

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

Answer : c) Poisson

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

Answer : b) False

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) Hypothesis

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)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) Outliers cannot conform to the regression relationship

WORKSHEET

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

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

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

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

Filling in missing data is also called data imputation.

According to Han and Kamber, missing values can directly be imputed by:

- ➢ Ignore the tuple: usually done when class label is missing (when doing classification)—not effective when the % of missing values per attribute varies considerably
- ➢ Fill in the missing value manually: tedious + infeasible
- ➢ Fill in it automatically with a global constant
- ➢ Fill with the attribute mean
- ➢ Fill with the attribute mean for all samples belonging to the same class
- ➢ Fill with the most probable value: inference-based such as Bayesian formula or decision tree

 12. What is A/B testing?

A/B testing is a method of 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.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is typically considered terrible practice since it ignores feature correlation

14. What is linear regression in statistics?

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?

➢ Inferential Statistic.
➢ Descriptive Statistic