

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: *The Normal Distribution is defined by the probability density function for a continuous random variable in a system. Let us say, $f(x)$ is the probability density function and x is the random variable. Hence, it defines a function which is integrated between the range or interval $(x \text{ to } x + dx)$, giving the probability of random variable X , by considering the values between x and $x+dx$.*

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

Ans: *Imputation is the process of replacing missing values with substituted data. It is done as a pre-processing step. If the data is numerical, we can use mean and median values to replace else if the data is categorical, we can use mode which is a frequently occurring value. If the data is numerical so we can use the mean value. The missing values will be represented as NaN, If the missing data is not a Number*

12. What is A/B testing?

Ans: *A/B testing is an efficient and effective way to gauge your audience's response to a design or content idea because it doesn't disturb your users' experience or send out disruptive feedback surveys. Just try something new and let the results speak for themselves*

13. Is mean imputation of missing data acceptable practice?

Ans: *Mean imputation is the practice of replacing null values in a data set with the mean of the data. Mean imputation is generally bad practice because it doesn't take into account feature correlation.*

14. What is linear regression in statistics?

Ans: *Linear regression strives to show the relationship between two variables by applying a linear equation to observed data. One variable is supposed to be an independent variable, and the other is to be a dependent variable. For example, the weight of the person is linearly related to his height.*

15. What are the various branches of statistics?

Ans: *a) Descriptive statistics*

b) Inferential statistics

Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures. For example, the collection of people in a city using the internet or using Television.

Inferential Statistics is a method that allows us to use information collected from a sample to make decisions, predictions or inferences from a population. It grants us permission to give statements that goes beyond the available data or information. For example, deriving estimates from hypothetical research.