

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

- a) True b) False

Answer: 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

Answer: 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

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.

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?

Answer: 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.

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

Answer: The simplest imputation method is replacing missing values with the mean or median values of the dataset at large, or some similar summary statistic. This has the advantage of being the simplest possible approach, and one that doesn't introduce any undue bias into the dataset.

12. What is A/B testing ?

Answer: A/B testing (also called split testing) is comparing two versions of a web page to figure out the better performing variation.

13. Is mean imputation of missing data acceptable practice?

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

14. What is linear regression in statistics?

Answer: Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value.

15. What are the various branches of statistics?

Answer: The various branches of statistics were descriptive statistics and inferential statistics.





