

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) Modelling 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 the above 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. It doesn't change

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 can conform to the regression relationship

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

Answer – The Normal Distribution is a continuous probability distribution where the values are distributed in a symmetrical manner around the mean, i.e left side can be superimposed on the right side and vice-versa. Moreover there is no skewness and it's shape is similar to a bell hence a bell shaped curve. In case of a normal distribution, Mean=Median=Mode.

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

Answer – Missing data can be handled in two ways:

1. By dropping the missing values
2. By using the imputation techniques

We can use several imputation techniques like mean, median, mode, KNN according to the data set.

12. What is A/B testing?

Answer – A/B testing is basically a way to compare two versions of something to figure out which one performs better. It is most often associated with websites and applications.

13. Is mean imputation of missing data acceptable practice?

Answer – Yes, mean imputation of missing data acceptable practise in some cases but it is not the best technique as it is sensitive to outliers. In mean imputation we replace the missing value with the mean of all data formed within a specific cell or class.

14. What is linear regression in statistics?

Answer – Linear Regression tends to establish a relationship between a dependent variable (Y) and one or more independent variable (X) by finding the best fit of the straight line. The equation of the linear model is $Y = mX + c$ where m is the slope and c is the intercept.

15. What are the various branches of statistics?

Answer – There are two main branches of statistics :

1. Descriptive Statistics – It deals with the collection and presentation of the data.
2. Inferential Statistics – It is used to draw right conclusions from the data collected about the population.