

STATISTICS WORKSHEET-1

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

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

Ans- The normal distribution is a smooth distribution of data around the mean value where the standard deviation defines the width of this distribution. It is commonly referred to as the “bell curve.”

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

Ans- Missing data is one of the most common issues in data analytics and machine learning. How to handle missing data varies depending on the type of data and the type of analysis or model you’re using. Here are some of the most common ways to handle missing data:

- Mean, Median, or Mode Imputation
- Regression Imputation
- K-Nearest Neighbours Imputation

12. What is A/B testing?

Ans- A/B testing (also known as split testing) is a statistical approach to comparing two versions of the same product or webpage to see which version performs better. In A/B testing, a sample of users is split into two groups (A and B) and each group is exposed to a different version of a particular element (A or B). This could be a webpage, an advertisement, or a feature. The objective is to see which version results in better results (or higher performance) based on predetermined metrics.

13. Is mean imputation of missing data acceptable practice?

Ans- The process of replacing null values in a data collection with the data’s mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation.

14. What is linear regression in statistics?

Ans- Linear regression is the study of the relationship between one or more independent variable and a dependent variable by the fitting of a linear equation to observed data. The objective of linear regression is the finding of the best-fitting linear straight line (or a hyperplane for multiple independent variables) to minimize the sum of the squared difference between the observed and predicted values of the linear model.

15. What are the various branches of statistics?

Ans- a) Descriptive Statistics, also known as descriptive statistics, is the process of presenting and summarizing data. It includes central tendency measures (e.g. mean, median, mode), as well as dispersion measures (range, variance and standard deviation).

b) Inferential Statistics, or inferential statistics, refers to the process of making inferences or predictions about a population from a set of data. Examples of inferential statistics include hypothesis testing, confidence interval, and regression analysis.