

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

Bernoulli random variable are a type of discrete random variable that can only take on the value of 1 or 0

Option A

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

The theorem that states that the distribution of averages of independent and identically distributed (iid) variables, properly normalized, becomes that of a standard normal as the sample size increases is known as the **Central Limit Theorem**.

option A

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

The Poisson distribution is a discrete probability distribution that gives the probability of an event happening a certain number of times within a given interval of time or space. It is commonly used to model count data that is unbounded and where the events occur randomly and independently.

OPTION B

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

NO IDEA

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

Poisson random variables are used to model rates of events that occur randomly and independently in a given interval of time or space .

Option c

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

The **central limit theorem (CLT)** states that the distribution of the sample mean of a large number of independent and identically distributed random variables converges to a normal distribution. The **standard error** is the standard deviation of the sample mean, which measures how much the sample mean varies from the true population mean.

Option B

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

Hypothesis testing is a statistical method that allows researchers to make inferences about a population parameter based on sample data.

Option B

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

Normalized data are centered at zero and have units equal to standard deviations of the original data. This means that the mean of the normalized data is zero and the standard deviation is one. Normalizing data is a process of transforming data values to a common scale, often to make them comparable or reduce variability.

Option A

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

No idea

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 ?

No idea

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

No idea

12. What is A/B testing?

A/B testing is a method of comparing two or more versions of a variable, such as a web page, an email, or a product design, to see which one performs better in terms of a desired outcome, such as conversions, clicks, or sales. [A/B testing is also known as split testing or bucket testing, and it is a form of randomized controlled experiment](#)

13. Is mean imputation of missing data acceptable practice?

No idea

14. What is linear regression in statistics?

Linear regression is a statistical model that estimates the linear relationship between a dependent variable and one or more independent variables. [It can be used to describe how a change in one variable affects another variable, or to predict the value of a variable based on the values of other variables](#)

The simplest form of linear regression is simple linear regression, which involves only one independent variable and one dependent variable.

15. What are the various branches of statistics?

Statistics is the science of collecting, organizing, analyzing, and interpreting data to make decisions or draw conclusions. [There are two main branches of statistics descriptive statistics and inferential statistics](#)

Descriptive statistics summarizes the data using numerical measures, tables, graphs, and charts. Descriptive statistics helps to describe the basic features of the data and to explore its patterns.

Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population. Inferential statistics uses probability theory, hypothesis testing, confidence intervals, and regression analysis to draw conclusions and generalize the results.
