

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 variable 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. 10. Usually replacing the standard error by its estimated value does change the CLT.

- a) True
- b) False

Ans: - b) False

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

Ans: - b) Hypothesis

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

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

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Ans:

Normal distribution, also known as the Gaussian distribution. It is a probability distribution function which is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a 'Bell Curve'. In a normal distribution the mean is '0' and the standard deviation is '1'. It has zero skewness and a kurtosis of 3. These are symmetrical, but not all those symmetrical distributions are normal.

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

Ans:

In order to handle the Missing data we use Fill NaN method and some imputer techniques.

Some of the imputer techniques which I have used so far are:

1.Simple Imputer 2.KNN Imputer 3. Iterative Imputer

12. What is A/B testing?

Ans:

A/B testing is also known as bucket testing or split-run testing). It is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective. A/B tests are useful for understanding user engagement and satisfaction of online features, such as a new feature or product. Large social media sites like LinkedIn, Facebook, and Instagram use A/B testing to make user experiences more successful and as a way to streamline their services.

13. Is mean imputation of missing data acceptable practice?

Ans:

It Is a Bad practice in general, mean imputation preserves the mean of the observed data Which Leads to an underestimate of the standard deviation and Distorts relationships between variables by "pulling" estimates of the correlation towards zero.

14. What is linear regression in statistics?

Ans:

In statistics, linear regression is a linear approach to model the relationship between a scalar response and one or more explanatory variables which are also known as dependent and independent variables. The case of one explanatory variable is called simple linear regression, for more than one, the process is called multiple linear regression.

15. What are the various branches of statistics?

Ans:

There are two types of statistical branches they are:

Descriptive Statistics:

In this type of statistics, the data is summarised through the given observations. The summarisation is one from a sample of population using parameters such as the mean or standard deviation.

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.

Descriptive statistics are also categorised into four different categories:

Measure of frequency

Measure of dispersion

Measure of central tendency

Measure of position

The frequency measurement displays the number of times a particular data occurs. Range, Variance, Standard Deviation are measures of dispersion. It identifies the spread of data. Central tendencies are the mean, median and mode of the data. And the measure of position describes the percentile and quartile ranks.

Inferential Statistics

This type of statistics is used to interpret the meaning of Descriptive statistics. That means once the data has been collected, analysed and summarised then we use these stats to describe the meaning of the collected data. Or we can say, it is used to draw conclusions from the data that depends on random variations such as observational errors, sampling variation, etc.

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.