

## **Statistics**

1. Bernoulli random variables take (only) the values 1 and 0.

a) True b) False

Ans)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

Ans)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

Ans)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

Ans)D

5. \_\_\_\_\_ random variables are used to model rates.

a) Empirical b) Binomial c) Poisson d) All of the mentioned

Ans)C

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

a) True b) False

Ans)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

Ans)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

Ans)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

Ans)D

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

Ans) The normal distribution is described by the mean ( $\mu$ ) and the standard deviation ( $\sigma$ ).

The normal distribution is often referred to as a 'bell curve' because of its shape:

- Most of the values are around the centre ( $\mu$ )
- The median and mean are equal
- It has only one mode
- It is symmetric, meaning it decreases the same amount on the left and the right of the centre

The area under the curve of the normal distribution represents probabilities for the data.

The area under the whole curve is equal to 1, or 100%

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

Ans)To, handle missing data we use various imputers like

KNN imputer - knn finds relation with other column and try to impute data acc to relation.

Iterative Imputer - it will treat other continuous features as train data, train on it and treat null column as label and predict null entry.

Simple Imputer- it makes mean of the features and fills the missing value it.

12. What is A/B testing?

Ans) **A/B testing** (also known as **bucket testing** or **split-run testing**) 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.

13. Is mean imputation of missing data acceptable practice?

No, it's not ,it totally depends upon domain knowledge, sometimes we have to remove the missing data and sometimes we use mean or various other imputers.

14. What is linear regression in statistics?

Ans) **Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:**

- One variable, denoted  $x$ , is regarded as the predictor, explanatory, or independent variable.
- The other variable, denoted  $y$ , is regarded as the response, outcome, or dependent variable.

15. What are the various branches of statistics?

Ans) The two main branches of statistics are **descriptive statistics** and **inferential statistics**

**Descriptive statistics** deals with the presentation and collection of data. This is usually the first part of a statistical analysis.

**Inferential statistics**, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.