

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
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
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
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
5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
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
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
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

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?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

10.

The normal distribution, commonly known as the Gaussian distribution, is a continuous probability distribution that, in statistics, can be used to describe a wide variety of natural occurrences. In a normal distribution, the data are symmetrically distributed around the mean, with the majority of the data falling close to the mean and a smaller percentage of the data falling farther away from the mean. The mean and standard deviation, which determine the distribution's shape and size, are the two factors that define the normal distribution. Several statistical tests and methods make the assumption that the data is normally distributed, which enables the use of certain statistical inferences and calculations.

11.

Imputing or removing missing values are 2 techniques for handling missing data. If the missing data are not entirely random, removing them could result in biased results. When there are missing values in the data, they are estimated using methods like mean imputation, regression imputation, or multiple imputation. The type and quantity of the missing data, as well as the research topic, influence the imputation strategy selection. In addition to reporting the patterns of missing data and the imputation techniques used in the study, it is crucial to assess how sensitive the results are to various imputation techniques.

12. A/B testing is a statistical technique that compares two iterations of a good or service to see which one works better. Users are randomly assigned to version A or version B, and their behaviour is then observed.

13.

Mean imputation is a popular technique for dealing with missing data, in which the mean of the available data is used to fill in for the missing values. While being a straightforward and practical procedure, mean imputation has a number of drawbacks and can add bias and mistakes into the data.

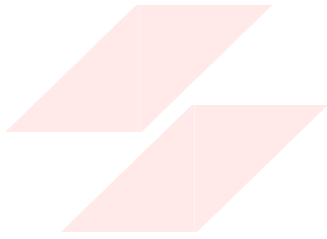
The missing data must be assumed to be missing at random, which may not always be the case when using mean imputation. Moreover, the data's variability is decreased, which might improve the precision of statistical studies. Mean imputation might also not be acceptable for variables with extreme values or non-normal distributions.

Because of this, it is important to utilise mean imputation cautiously and to be aware of its limitations. Depending on the type and quantity of missing data as well as the research objective, other imputation strategies, such as regression imputation or multiple imputation, may be more suitable.

14.

A dependent variable and one or more independent variables are modelled using the statistical technique of linear regression. Simple linear regression models relationships using only one independent variable and a straight line, whereas multiple linear regression models relationships using two or more independent variables and a hyperplane. Finding the line or hyperplane that fits the data the best and minimises the sum of the squared differences between the observed and predicted values is the objective of linear regression. In order to produce predictions, determine the direction and strength of the relationship between variables, and test relationship-related hypotheses, linear regression is frequently utilised.

15. Descriptive statistics, inferential statistics, biostatistics, econometrics, Bayesian statistics, machine learning, time series analysis, and multivariate analysis are the primary subfields of statistics.



FLIP ROBO