

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) FalseAns: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 mentionedAns: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 mentionedAns: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 mentionedAns:d-all of the above
5. _____ random variables are used to model rates.
 - a) Empirical
 - b) Binomial
 - c) Poisson
 - d) All of the mentionedAns:c-poisson
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
 - a) True
 - b) FalseAns: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 mentionedAns: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) 10Ans: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

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?

Ans: The normal distribution, also known as the Gaussian or standard normal distribution, is the probability distribution that plots all of its values in a symmetrical fashion, and most of the results are situated around the probability's mean. Values are equally likely to plot either above or below the mean.

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

Ans: 1. Mean or Median Imputation

When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. However, there can be multiple reasons why this may not be the most feasible option:

There may not be enough observations with non-missing data to produce a reliable analysis

In predictive analytics, missing data can prevent the predictions for those observations which have missing data

2. Multivariate Imputation by Chained Equations (MICE)

MICE assumes that the missing data are Missing at Random (MAR). It imputes data on a variable-by-variable basis by specifying an imputation model per variable. MICE uses predictive mean matching (PMM) for continuous variables, logistic regressions for binary variables, bayesian polytomous regressions for factor variables, and proportional odds model for ordered variables to impute missing data.

To set up the data for MICE, it is important to note that the algorithm uses all the variables in the data for predictions. In this case, variables that may not be useful for predictions, like the ID variable, should be removed before implementing this algorithm.

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 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

13. Is mean imputation of missing data acceptable practice?

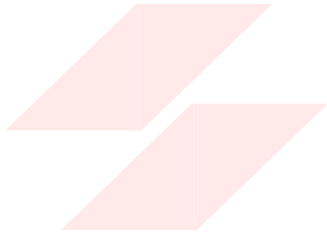
14. What is linear regression in statistics?

Ans: Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data.

15. What are the various branches of statistics?

Ans: Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

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.



FLIP ROBO