

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?

Normal distribution for Gaussian distribution is a type of continuous probability distribution for real random variable which showing that data near the mean are more frequent in occurrence than data far from the mean. In normal distribution the ideal condition is null mean and standard deviation of unity. The spread of data gives a bell shaped curve (mean=0, standard deviation=1)

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

Data missing is a common problem when handling huge number of data in statistics. There are different types of missing data's.

- MCAR(Missing Completely At Random): here the missing is related to observed data not the specific values or responses
- MNAR(Missing Not At Random): in this case the missing data prefer particular structure.

There are different imputation methods are there to handling this data missing like

- Substituting mean or median
- Deletion technique
- Regression imputation
- LOCF(last observed carried forward) etc.

But for handling the missing data I prefer multiple imputation techniques than single imputation technique because here missing values are always replaced with a set of suitable values.

12. What is A/B testing?

A/B test also known as bucket test or split run testing , which always allow a decision maker to prefer one depending the effectiveness from two alternatives like A or B, here comparing two versions of a single variable by testing the efficiency of each and helping to prefer the suitable one.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is not a preferable solution for handling missing data because it affects the data variables in different ways.

- It affects the natural relationship between variables
- Lose the variance of variables
- The process of shrinking the standard error will lead to the rejection of null hypothesis

14. What is linear regression in statistics?

Linear regression is a linear approach for modelling in which assumes a linear relationship between the input variables(x) and the single output variable (y). Simple linear regression is useful for finding relationship between two continuous variables. Here one is predictor or independent variable and other is the response or dependent variable. Simply this analysis is used to predict the value of a variable based on the value of another variable.

15. What are the various branches of statistics?

Statistics is a branch in mathematics deals with collecting and analysing numerical data in large quantities, it deals with the process of collection, analysis, interpretation and presentation of masses of numerical data. Statistics have two different branches,

- Descriptive statistics: Branch focuses on collecting, summarizing, and presenting a set of data.
- Inferential statistics: Branch focus on the analysis of sample data to draw conclusions about a population.