

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)

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: d)

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?

Data are symmetrically distributed around mean, median, and mode or bell-shaped distribution. The form of a normal distribution is determined by its mean and standard deviation. Mean =0 and Standard deviation=1

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

Missing data can be handle by mean, mode &B with use of imputation techniques.

We can use Simple imputer, Knn Imputer & Iterative imputer for missing data handling.

12. What is A/B testing?

A/B testing (also known as bucket testing, split-run testing, or split testing) is a user experience research methodology. A/B tests consist of a randomized experiment that usually involves two variants (A and B), although the concept can be also extended to multiple variants of the same variable. 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 multiple versions of a single variable, for example by testing a subject's response to variant A against variant B, and determining which of the variants is more effective.

13. Is mean imputation of missing data acceptable practice?

It is not acceptable for handling all missing data.

14. What is linear regression in statistics?

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y = c + b \cdot x$, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

15. What are the various branches of statistics?

The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data.