

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?

Ans.10) Normal Distribution is a proper term of a probability bell curve. In Normal Distribution the mean is 0 and the Standard deviation is 1.

Ans.11) One way of handling missing values is the deletion of the rows or columns having null values. If any columns have more than half of the values as null then you can drop the entire column. In the same way rows can also be dropped. (1) Mean Imputation. (2) Substitution (3) Hot deck Imputation (4) Cold deck imputation (5) Regression imputation.

Ans.12) A/B testing is also known as Split testing, refers to a randomized experimentation process wherein two or more versions of a variable are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

Ans.13) Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

Ans.14) Linear Regression:- It is used to predict the value of variables based on the value of another variable. The variable we want to predict is called dependent variable and the variable we are using to predict the other variables value is called Independent variable.

Ans.15) There are 3 real branches of statistics:

- 1) Data Collection
- 2) Descriptive statistics
- 3) Inferential statistics.