

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

1. a) true
2. a) central limit theorem
3. b) modelling bounded count data
4. d) all of the mentioned
5. c) Poisson
6. b) false
7. b) hypothesis
8. a) 0
9. 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?

Answer: The normal distribution also known as gaussian distribution plots the random variables in a symmetrical bell-shaped curve. The bell shape is due to the fact that more data occurs near the mean, that is the frequency at the mean is higher.

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

To handle missing data the first thing is to identify whether the data is:

- i Missing at random
Here the data is missing in only some sub samples of data
- ii Missing completely at random
Here the data is completely missing in all observation
- iii Missing not at random
Here we know the reason why the data is missing and why the information is unknown

The imputation techniques recommended are:

- Mean, Median and Mode method
- Random Forest
- Multivariate Imputation by Chained Equation

12. What is A/B testing?

A/B testing is a simple experiment done by making 2 groups A and B. In the former we put half of the data that is unchanged, original. Whereas in the latter we put the data with some changes and then compare the two by framing a hypothesis. A null hypothesis which will give us the same values and an alternate hypothesis which is considered true and will show us the changes or difference in values of A and B

13. Is mean imputation of missing data acceptable practice?

Yes, mean imputation of missing data is an acceptable practice provided the no of missing data is less. Also mean imputation is not a feasible option

14. What is linear regression in statistics?

Linear regression is basically used to plot the relationship between two variables using a linear equation, of which one is the independent variable and the other being the dependent variable

15. What are the various branches of statistics?

There are three branches of statistics

- Data collection

This branch deals with the collection of data

- Descriptive Statistics

This branch deals with the representation of data either visually or numerically

- Inferential Statistics

This branch as the name suggests deals with making conclusions about the data
