WORKSHEET

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
Answer-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
Answer-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
Answer-A

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
Answer-D

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned
Answer-C

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
Answer-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
Answer-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
Answer-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
Answer-C

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

Answer

Normal distribution is also known as Gaussian distribution. It is a continuous distribution in nature just as binomial distribution. Normal distribution is an arrangement of a dataset in which most values data are in the middle of the range and the rest spread on either side. The graphical representations of normal distribution are in bell-shape curve in which mean, median and mode are equals.

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

Answer

Missing data can be delt in many ways. The very first option is to remove list in which missing data is present but it will affect the entire analysis. So, we try to avoid it. Another method is “Imputation” in which it develops reasonable guesses for the missing data. The most common method used in imputation is “mean, median and mode”. For categorical data we use mode method (most occurring method) and for continuous data we use mean and mode method.
There are some other methods are also used for handling the missing through imputation in which we use KNN imputer technique, iterative imputer technique etc.
In my recommendation we should use simple imputer technique (mean, median and mode) because it easy to understand and very relatable.

12. What is A/B testing?

Answer

A/B testing is also known as Split testing or bucket testing. It is a user experience research methodology. A/B testing consist of a randomized experiment whit two variants A and B. It includes application of statistical hypothesis testing or two sample hypothesis testing as used in the filed of statistics.

13. Is mean imputation of missing data acceptable practice?

Answer

Yes, but it can only be acceptable to the continuous data.

14. What is linear regression in statistics?

Answer

Linear regression is a method use to analyze data and predict the value of a variable 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 (value of variable) is independent variable. Example- Experience and Salary (in which salary is dependent variable and experience is independent variable)

15. What are the various branches of statistics?

Answer

There are four main branches of statistics.

- Mathematical statistics- It helps in forming the experimental and statistical distribution
- Statistical method or function- It helps in collection, tabulation, interpretation of the data. It helps in analyzing in data and returns insight.
- Descriptive statistics- In the descriptive statistics each data is analyzed but we could not use it when there is a lot of data in the sample. It can be used only when data is small.
- Inferential statistics- In this we analyze very big data sample by taking some random data from the sample and then analyze it and the result will be inferable for the whole sample. For example- in covid-19 pandemic time we used to read/watch the covid positivity rate of India or of different states in which medical staff only examined few random people and on the basis of that result they declared the positivity rate for whole state. In this few random people’s sample is treated as descriptive and the whole population is treated as inferential statistics.