

STATISTICS WORKSHEET-1

(Answers are shown in red colour)

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

Subjective Answers

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

Answer- Normal Distribution also called Gaussian distribution. It is continuous distribution in nature. We cannot use it in case of discrete data. Every event is independent from one another. Mean, Median, Mode all line up in such a way that the centre of distribution is mean. It is easily identifiable by Bell Shaped curve.

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

Answer- We can handle missing data with the help of imputer, Imputers which I learnt and recommend is Simple indicator with indicator, Knn imputer, iterative imputer.

12. What is A/B testing?

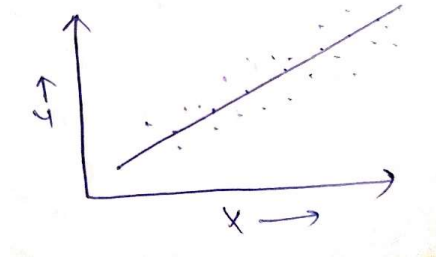
Answer- A/B testing is basically used to compare 2 different products and this comparison is made through user input like whether the user is clicking the product1 or whether the user is clicking on product 2. It is heavily used in E-comm industry, Netflix, etc. It is totally based on user behaviour.

13. Is mean imputation of missing data acceptable practice?

Answer- No, It is not acceptable. Because it ignores feature correlation, It decreases the variance of our data while increasing bias and model becomes less accurate.

14. What is linear regression in statistics?

Answer- In linear regression dependent and independent variables are present. But that dependent variable is continuous in nature. OR The relationship between dependent variable and independent variable is in linear in nature.



15. What are the various branches of statistics?

Descriptive – If we are able to describe any data then we can say it as descriptive statistics. Eg. Marks of students, Avg. weight or height of population etc.

Inferential- If our data is too big and it is too difficult to explain, but at the same time we have to pick data from individual batch and calculate the avg. And this calculated avg. value we use to explain the dataset.