

STATISTICS WORKSHEET-1

Answers to the questions are in bold:

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

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

3. Which of the following is incorrect with respect to use of Poisson distribution?

b) Modeling bounded count data

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.

c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

b) False

7. 1. Which of the following testing is concerned with making decisions using data?

b) Hypothesis

8. 4. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

a) 0

9. Which of the following statement is incorrect with respect to outliers?

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?

Ans : It is probability distribution, it is symmetric. It is perfectly symmetrical around its centre. There is one peak in normal distribution. It is also called as bell curve. In a normal distribution the mean is zero and the standard deviation is 1.

Eg. Height, weight, etc.

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

Ans : We can handle missing data by imputation(replacing missing data) or deleting or removing the data. We can use below imputation techniques :

- a. Deletion , Pairwise deletion, Listwise deletion**
- b. Dropping rows, dropping complete columns**

12. What is A/B testing?

Ans : It is also known as split testing. It is the process of comparing two versions or two variants to see which one performs better based on a given metric.

Eg. Two WebPages/websites can be compared.

13. Is mean imputation of missing data acceptable practice?

Ans : Mean imputation ignores feature correlation and so it is typically considered as terrible practice.

Mean imputation decreases variance of our data while increasing bias. As a result if variance is reduced then the model is less accurate.

14. What is linear regression in statistics?

Ans : It is used to predict the value of a variable based on the value of another variable. The variable which is predicted is called dependant variable and the variable which is used to predict another variable's value is called independent variable.

Simply it estimates the relationship between one independent variable and one dependent variable using a straight line.

15. What are the various branches of statistics?

Ans : The two main branches of statistics is

- a. Descriptive – Data is typically arranged and displayed in tables or graphs summarizing details such as histograms ,bar or scatter plots.
Types : Mean, median , mode, range, variance, standard deviation.**
- b. Inferential – This is used to draw conclusions and inferences i.e to make valid generalizations from samples. It is used to describe the meaning of collected data.
Types: Hypothesis testing, z-score.**