

Statistics worksheet

Q1. what do you understand by term normal distribution?

Ans. normal distribution is always symmetrical in curve. The mean median and mode are always equal in normal distribution. It always has only one peak which is the highest point of the distribution. Half values fall below the mean and half values above the mean. It is bell in shaped.

Q2. How do you handle missing data? What imputation technique do you recommend?

Ans. there are different ways to handle missing data. The methods depend on the type of missing data. If the data is missing completely at random(MCAR) then mean, median and mode imputation is effective. If the data is missing at random (MAR) then multiple imputation is used. If its not missing at random (MNAR) then maximum likelihood estimation method is reliable. There is also a simple method which is deleting the rows and columns. You can delete the rows and columns of missing values but it leads to loss of data. KNN k-nearest neighbour also used where it finds the similar data points. This is known as the best method to find the missing values.

Q3. what is A/B testing?

Ans. it's a testing method for two websites or products to know which one performs better. It is mostly used by designers to show the new features and changes impact. It is also known as the split testing.

Q4. Is mean imputation of missing data acceptable practice?

Ans. No, mean imputation of missing data is not acceptable practice for handling missing data. It does not show the relationship between variables. It also lead to wrong estimates of variance and co-variance.

Q5. what is linear regression in statistics?

Ans. it shows the relationship between two quantitative variables. One dependent variable and other is independent variable. For example, the hight and weight of a person has linear relationship with each other. As the height of a person increases the weight of a person also increases.

Q6. What are the various branches of statistics?

Ans. The two main branches of statistics are descriptive statistics and inferential statistics.

The **descriptive statistics** is known as the first part of statistics which deals with the collection and preservation of data. Collection of samples, population and then summarize them in graphs and tables is the part of descriptive statistics. Here we organize the data for future predictions or make it ready for inferential statistics.

Inferential statistics is called the 2nd part of statistics. Which is used for conclusions and to make predictions depending upon the summarized data. The regression models and test hypotheses are the examples of inferential statistics.

Macq's statistics worksheet.

Q1.option A

Q2. Option A

Q3. Option B

Q4. Option A

Q5. Option C

Q6. Option B

Q7. Option B

Q8. Option A

Q9. Option C

ML WORKSHEET QUESTIONS AND ANSWERS.

Q1. Explain the term regularization?

Ans. Regularization is used to make the model easier to understand. It makes the model less sensitive. It is used to solve the problem of overfitting. Which improves the model's reliability. The data augmentation one type of regularization is used to modify the training data . it increases the number of samples and expands the model to a greater extent. In the same way the early stopping is another regularization technique. in training model its just reduce the number of iterations. Regularization is also used in linear regression models.

Q2. Which particular algorithms are used for regularization?

Ans. Lasso (L1), Ridge (L2), Dropout, Ensembling, Elastic net.

Q3. Explain the term error present in linear regression equation?

Ans. the equation is

$$Y=mx+b$$

Where y is dependent variable

X=independent variable

m=slope

b=estimated intercept

The term error in the regression model represents the uncertainty . when the actual value of y differs from the expected value then the value of b does not remain the zero. Which means some factors influences the value of y. this error counts the variables which are not included in the model. The certainty of the line depends on the value of the "b". A larger error value mean the line is less certain. This error is also known as the disturbance or remainder term.

MCQS of ML.

Q1. Ans option A

Q2. Ans option A

Q3. Ans option B

Q4. Ans option C

Q5. Ans option C

Q6. Ans option B

Q7. Ans option D

Q8. Ans option B

Q9. Ans option C

Q10. Ans option B

Q11. Ans option B

Q12. Ans option A and B

PYTHON MCQ'S

Q1. Ans option C

Q2. Ans option B

Q3. Ans option C

Q4. Ans option A

Q5. Ans option D

Q6. Ans option C

Q7. Ans option A

Q8. Ans option C

Q9. Ans option A and C

Q10. Ans option A and B

