

STATISTICS WORKSHEET-3

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is the correct formula for total variation?

Ans. b) Total Variation = Residual Variation + Regression Variation

- a) Total Variation = Residual Variation – Regression Variation b) Total Variation = Residual Variation + Regression Variation
c) Total Variation = Residual Variation * Regression Variation d) All of the mentioned

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.

Ans. c) binomial

- a) random b) direct c) binomial d) none of the mentioned

3. How many outcomes are possible with Bernoulli trial?

Ans. a) 2 ##### b) 3 c) 4 d) None of the mentioned

4. If H_0 is true and we reject it is called

Ans. a) Type-I error #### b) Type-II error c) Standard error d) Sampling error

5. Level of significance is also called:

Ans. b) Size of the test

- a) Power of the test b) Size of the test c) Level of confidence d) Confidence coefficient

6. The chance of rejecting a true hypothesis decreases when sample size is:

Ans. b) Increase

- a) Decrease b) Increase c) Both of them d) None

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

Ans. b) Hypothesis

- a) Probability b) Hypothesis c) Causal d) None of the mentioned

8. What is the purpose of multiple testing in statistical inference?

Ans. d) All of the mentioned

- a) Minimize errors b) Minimize false positives c) Minimize false negatives d) All of the mentioned

9. Normalized data are centered at ___ and have units equal to standard deviations of the original data

Ans. a) 0

- a) 0 b) 5 c) 1 d) 10

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What is Bayes' Theorem?

Ans. Bayes' Theorem describes the [probability](#) of an [event](#), based on prior knowledge of conditions that might be related to the event.

11. What is z-score?

Ans. A Z-score is a statistical measure that tells you how much an observation differs from the mean. Z-score uses standard deviation to indicate difference between a dataset's mean and individual observation

12. What is t-test?

Ans. T Test is a statistical significance test that is used to compare the means of two groups and determine if the difference in means is statistically significant

13. What is percentile?

Ans. A percentile is a measure used in statistics indicating the value *below which* a given percentage of observations in a group of observations fall

14. What is ANOVA?

Ans. Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means.

15. How can ANOVA help?

Ans.

MACHINE LEARNING ASSIGNMENT – 3

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is an application of clustering?

Ans. d. All of the above

a. Biological network analysis b. Market trend prediction c. Topic modeling d. All of the above

2. On which data type, we cannot perform cluster analysis?

Ans. None

a. Time series data b. Text data c. Multimedia data d. None

3. Netflix's movie recommendation system uses

Ans. c. Reinforcement learning and Unsupervised learning

a. Supervised learning b. Unsupervised learning c. Reinforcement learning and Unsupervised learning d. All of the above

4. The final output of Hierarchical clustering is

ANS. b. The tree representing how close the data points are to each other

a. The number of cluster centroids b. The tree representing how close the data points are to each other c. A map defining the similar data points into individual groups d. All of the above

5. Which of the step is not required for K-means clustering?

Ans. d. None

a. A distance metric b. Initial number of clusters c. Initial guess as to cluster centroids d. None

6. Which of the following is wrong?

Ans. c. k-nearest neighbour is same as k-means

a. k-means clustering is a vector quantization method b. k-means clustering tries to group n observations into k clusters c. k-nearest neighbour is same as k-means d. None

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

Ans. d. 1, 2 and 3

i. Single-link ii. Complete-link iii. Average-link Options:

a. 1 and 2 b. 1 and 3 c. 2 and 3 d. 1, 2 and 3

8. Which of the following are true?

Ans. a. 1 only

i. Clustering analysis is negatively affected by multicollinearity of features ii. Clustering analysis is negatively affected by heteroscedasticity Options:

a. 1 only b. 2 only c. 1 and 2 d. None of them

9. In the figure above, if you draw a horizontal line on y-axis for $y=2$. What will be the number of clusters formed?

Ans. a. 2

10. For which of the following tasks might clustering be a suitable approach?

Ans. B. Given a database of information about your users, automatically group them into different market segments.

11. Given, six points with the following attributes:

Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

Ans. a

12. Given, six points with the following attributes:

Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

Ans. b

Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly

13. What is the importance of clustering?

Ans. The goal of cluster analysis is to discover natural grouping of a set of patterns, points or objects. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other

14. How can I improve my clustering performance

Ans.

WORKSHEET 3 SQL

Refer the following ERD and answer all the questions in this worksheet. You have to write the queries using MySQL for the required Operation.

1. Write SQL query to create table Customers.

ANS. CREATE TABLE customers (customerNumber INT, customerName VARCHAR(50), LastName VARCHAR(50), FirstName VARCHAR(50), phone INT, address VARCHAR(255), city VARCHAR(50), state VARCHAR (50), postalCode INT, country VARCHAR(50), salesRepEmployeeNumber INT, creditLimit INT);

2. Write SQL query to create table Orders.

ANS. CREATE TABLE orders (orderNumber VARCHAR(25), orderdate DATE, requireDate DATE, shippedDate date, Status VARCHAR(50), comments VARCHAR(255),customerNumber INT);

3. Write SQL query to show all the columns data from the Orders Table.

ANS. SELECT * FROM orders;

4. Write SQL query to show all the comments from the OrdersTable.

ANS. SELECT comments FROM orders;

5. Write a SQL query to show orderDate and Total number of orders placed on that date, from Orders table.

ANS. SELECT orderDate, orderNumber FROM orders WHERE orderDate=shippedDate;

6. Write a SQL query to show employeeNumber, lastName, firstName of all the employees from employees table.

ANS. SELECT employeeNumber,lastName,firstName FROM employees;

7. Write a SQL query to show all orderNumber, customerName of the person who placed the respective order.

ANS. SELECT orders.orderNumber,customerName FROM orders INNER JOIN customers ON orders.customerNumber = customers.customerNumber WHERE orders.shippedDate <= orders.requiredDate;

8. Write a SQL query to show name of all the customers in one column and salerepemployee name in another column.

ANS. SELECT customerName, salesRepEmployeeNumber FROM customers;

9. Write a SQL query to show Date in one column and total payment amount of the payments made on that date from the payments table.

ANS. SELECT paymentDate, sum(amount) FROM payments GROUP BY paymentDate;

10. Write a SQL query to show all the products productName, MSRP, productDescription from the products table.

ANS. `SELECT productName,MSRP,productDescription FROM products;`

11. Write a SQL query to print the productName, productDescription of the most ordered product.

ANS. `SELECT productName, productDescription FROM products WHERE productCode=`
`(SELECT productCode FROM orderdetails WHERE quantityOrdered = (SELECT max(quantityOrdered) FROM orderdetails));`

12. Write a SQL query to print the city name where maximum number of orders were placed.

ANS. `SELECT customers.city FROM customers WHERE customerNumber = (SELECT customerNumber FROM orders where`
`orderNumber = (SELECT orderNumber FROM orderdetails ORDER BY quantityOrdered desc LIMIT 1));`

13. Write a SQL query to get the name of the state having maximum number of customers.

ANS.

14. Write a SQL query to print the employee number in one column and Full name of the employee in the second column for all the employees.

ANS. `SELECT employeeNumber, CONCAT(firstName,' ',lastName) AS fullname FROM employees;`

15. Write a SQL query to print the orderNumber, customer Name and total amount paid by the customer for that order
(quantityOrdered × priceEach)

ANS.