

SQL

- 1 Creat , Alter
- 2 Update , Delete, Select
- 3 Structured Query Language
- 4 Data Definition Language
- 5 Data Management Language
- 6 Create Table A (B int,C float)
- 7 Alter Table A ADD COLUMN D float
- 8 Alter Table A Drop Column
- 9 Alter Table A Alter column D int
- 10 Alter Table A Add Primary key B
- 11 Data warehouse is a Central repository of information that can be analyzed to make more informed decision.
- 12 OLTP Is Operational data that we get from our day to day transaction thus the oltps are the original source of data.
OLAP data is larger amount of operative and historical data that comes from the various oltp database such as BW(Business Warehouse) system.
- 13 Characteristics of A Data Warehouse are the four characteristics of data warehouse also called feature ware house of data ware house .
 - 1 subject oriented
 - 2 Time Variant
 - 3 Integrated
 - 4 Non Volatile
- 14 A star Schema is a multi dimensional data model used to organized data in a data base so that it is easy to under stand and analyze. Star schema can be applied to data Warehouse, Database , data marts, and tools. The star schema design is optimized for querying large datasets.
- 15 SETL is a very high level programming language based on the mathematical theory of sets.

STATISTICS WORKSHEET

- 1 True
- 2 Central Limit Theorem
- 3 Modeling bounded count data
- 4 all of the mentioned
- 5 Poisson

6 False

7 Hypothesis

8 a) 0

9 Outliers cannot conform to the regression relationship

10 A normal distribution is a type of continuous probability distribution in which most data points cluster towards the middle range, while the rest taper off symmetrically toward either extreme.

11 Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical program will make the decision.

Imputation techniques

a. mean imputation

b. Regression imputation

c. substitution

12 A/B testing is a way to compare two versions of figure out which performs better, while it's all about data of these days.

13 mean imputation is typically considered a terrible practice, it ignores feature correlation.

14 Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15

Types of statistics

1 Descriptive statistics

2 Inferential statistics

A) Measure of central tendency

(Mean, Mode, Median)

b) Measure of Variability

(Range, Variance, Dispersion)

MACHINE LEARNING

1 4

2 D) 1, 2 and 4

3 formulating the clustering problem

4 Euclidean distance

5 Divisive clustering

- 6 All answers are correct
- 7 All of the above
- 8 Unsupervised learning
- 9 K- Means clustering
- 10 K-means clustering algorithm
- 11 All of the above
- 12 Labeled data
- 13 The cluster analysis follows three basic
 - 1. Calculate the distances.
 - 2. Link the clusters.
 - 3. Right number of cluster .
- 14 To measure the quality of a clustering , we can use the average Silhoutte coefficient value of all object in the data set.
- 15 Cluster analysis is a data analysis technique that explores the naturally occurring groups with in a data set known as cluster.
 - They are in 6 types in cluster analysis
 - 1 partitioning method.
 - 2 Hierarchical method.
 - 3 Density based method
 - 4 Gird based method
 - 5 Model based method
 - 6 Constraint based method