

# STATISTICS WORKSHEET-2

Q1 to Q15 have only one correct answer.

Choose the correct option to answer your question.

1. What represent a population parameter?

ANS C) both

2. What will be median of following set of scores (18,6,12,10,15)?

ANS C) 12

3. What is standard deviation?

ANS D) All of the above

4. The intervals should be \_\_\_\_\_ in a grouped frequency distribution

ANS C) Both of these

5. What is the goal of descriptive statistics?

ANS B) Summarizing and explaining a specific set of data

6. A set of data organized in a participant by variables format is called

ANS B) Data set

7. In multiple regression, \_\_\_\_\_ independent variables are used

ANS B) 2

8. Which of the following is used when you want to visually examine the relationship between 2 quantitative variables?

ANS B) Scatterplot

9. Two or more groups means are compared by using

ANS D) Analysis of variance

10. \_\_\_\_\_ is a raw score which has been transformed into standard deviation units?

ANS A) Z-score

11. \_\_\_\_\_ is the value calculated when you want the arithmetic average?

ANS C) mean

12. Find the mean of these set of number (4,6,7,9,2000000)?

ANS D) 400005.2

13. \_\_\_\_\_ is a measure of central tendency that takes into account the magnitude of scores?

ANS D) Mean

14. \_\_\_\_\_ focuses on describing or explaining data whereas \_\_\_\_\_ involves going beyond immediate data and making inferences

ANSA) Descriptive and inferences

15. What is the formula for range?

ANS D) H-L

## MACHINE LEARNING

1. Movie Recommendation systems are an example of:

i) Classification ii) Clustering iii) Regression

ANS Options: b) 1 and 2

2. Sentiment Analysis is an example of:

i) Regression ii) Classification iii) Clustering iv) Reinforcement

ANS Options: d) 1, 2 and 4

3. Can decision trees be used for performing clustering?

ANS a) True

4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:

i) Capping and flooring of variables ii) Removal of outliers

ANS Options: a) 1 only

5. What is the minimum no. of variables/ features required to perform clustering?

ANS b) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?

ANS b) No

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?

ANS a) Yes

8. Which of the following can act as possible termination conditions in K-Means?

- i) For a fixed number of iterations.
- ii) Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
- iii) Centroids do not change between successive iterations.
- iv) Terminate when RSS falls below a threshold.

ANS Options: d) All of the above

9. Which of the following algorithms is most sensitive to outliers?

ANS a) K-means clustering algorithm

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

- i) Creating different models for different cluster groups.
- ii) Creating an input feature for cluster ids as an ordinal variable.
- iii) Creating an input feature for cluster centroids as a continuous variable.
- iv) Creating an input feature for cluster size as a continuous variable.

ANS Options: d) All of the above

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

ANS d) All of the above

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?

ANS One important drawback of K means is the lack of robustness to

13. Why is K means better?

ANS Advantages of k-means

- Relatively simple to implement.
- Scales to large data sets.
- Guarantees convergence.
- Can warm-start the positions of centroids.
- Easily adapts to new examples.
- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm

ANS The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.

## WORKSHEET 2 SQL

Q1 to Q13 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following constraint requires that there should not be duplicate entries?

ANS D) Unique

2. Which of the following constraint allows null values in a column?

ANS C) Null

3. Which of the following statements are true regarding Primary Key?

ANS A) Each entry in the primary key uniquely identifies each entry or row in the table

4. Which of the following statements are true regarding Unique Key?

ANS A) There should not be any duplicate entries

5. Which of the following is/are example of referential constraint?

ANS B) Foreign Key

For Questions 6-13 refer to the below diagram and answer the questions:

6. How many foreign keys are there in the Supplier table?

ANS C) 2

7. The type of relationship between Supplier table and Product table is:

A) one to many

8. The type of relationship between Order table and Headquarter table is:

C) one to one

9. Which of the following is a foreign key in Delivery table?

A) delivery id

10. The number of foreign keys in order details is:

A) 1 C)

11. The type of relationship between Order Detail table and Product table is:

ANS. many to one

12. DDL statements perform operation on which of the following database objects?

ANS C) Table

13. Which of the following statement is used to enter rows in a table?

ANS A) Insert in to

14. Which of the following is/are entity constraints in SQL?

ANS B) Unique

C) Primary Key

D) Null

15. Which of the following statements is an example of semantic Constraint?

ANS A) A blood group can contain one of the following values - A, B, AB and O.

B) A blood group can only contain characters