**1)What is data normalization? How is it different from database normalization (1st/2nd/3rd)?**

The process of reducing data redundancy in a table and improving data integrity is called data normalization. It is a way of organizing data in a database.

Normalization, also known as database normalization or data normalization, is an important part of relational database design because it helps to improve the speed, accuracy, and efficiency of the database.

1NF: A relation will be 1NF if it contains an atomic value. It states that an attribute of a table cannot

hold multiple values. It must hold only single-valued attribute.

2NF: A relation will be in 2NF if it is in 1NF and all non-key attributes are fully functional

dependent on the primary key.

3NF: A relation will be in 3NF if it is in 2NF and not contain any transitive partial dependency.3NF is

used to reduce the data duplication. It is also used to achieve the data integrity.

BCNF: BCNF is the advance version of 3NF. It is stricter than 3NF. A table is in BCNF if every functional

dependency X → Y, X is the super key of the table.

**2) What is a distribution? What are the uses for frequency and probability distribution?**

A distribution of data represents the number of times each value occurs in a dataset. A statistical distribution can be a frequency distribution or a probability distribution. The data distribution helps in knowing the mean, median, mode and helps in knowing more about the data.

Frequency distribution: A frequency distribution shows the frequency of repeated items in a graphical form or tabular form. It gives a visual display of the frequency of items or shows the number of times they occurred. Let's learn about frequency distribution in this article in detail.

Probability distribution: The **probability distribution** gives the possibility of each outcome of a random experiment or event. It provides the probabilities of different possible occurrences.

3) What is a decision? How's it different from inference?

a decision problem is a computational problem that can be posed as a yes–no question of the input values

Statistical inference refers to the theory, methods, and practice of forming judgements about the parameters of a population and the reliability of statistical relationships, typically based on random samples.

4) Google- what is Gini in probability, and explain in your own terms

Gini coefficient is used to measure income inequality for a given population. It ranges between 0 and 1, where 0 indicates lowest inequality and 1 being the highest level of inequality.

5) What is entropy?

Entropy is a scientific concept as well as a measurable physical property that is mostly associated with a state of disorder, randomness, or uncertainty.

6) What is Euclidean distance?

In mathematics, Euclidean distance between two points in the Euclidean space is the length of a line segment between the two points.

7) What's the difference between correlation and covariance?

**Covariance** and correlation both primarily assess the relationship between variables. Correlation measures the association between two variables while covariance describes the extent to which one variable is related to another.

8) What is mean squared error?

The mean squared error(MSE) is the sum of the squared difference of each value of the dataset and the mean value of the variables in the dataset.

The MSE is used to see the accuracy of the model and the greater the MSE the more unfit a model is.

9) What is the difference between covariance, standard deviation and mean squared error?

Covariance measures the direction of the relationship between two variables. A positive covariance implies that both the variables have high or low values together. A negative covariance implies that if one variable has high value then the other has low values

Mean squared error measures the amount of error in a statistical model.

Standard Deviation is a measure of the dispersion of the actual values in the data from the central tendency. It shows the variance of the data. It is given as the square root of variance.