Muhammad Nadeem

2019800 50

Introduction to Data Science.

Section - A

Task # 2

Data Distribution:-

Data Distribution refers to way data is spread or distributed across different values or ranges in a dataset. It provides insights into the patterns, trends and characteristics of the data. Understanding the data distribution is important in various fields including statistics, data analytics and machine learning.

There are different types of data distribution:

i) Normal Distribution

ii) Uniform Distribution

iii) Skewed Distribution

iv) Bimodal Distribution

v) Exponential Distribution.

vi) Log-Normal Distribution.

i) **Normal Distribution:-** (Gaussian Distribution)

It is a symmetrical probability distribution that follows the bell-shaped curve.

It is used in statistics and data analytics. Because many natural phenomena, such as human height or errors in measurements, tend to follow this pattern. It allows for the application of various statistical techniques and simplifies calculations.

Distribution is defined by its mean and standard deviation. It can described using mathematical formulas and statistical parameters, allowing us to calculate probabilities and perform statistical inference.

ii) Uniform Distribution:- (Rectangler Distribution)

It is a probability distribution where all values within a specified range are equally likely.

It is used when all outcomes or values are equally likely to occur. They are common in simmulations, random number generation, and scenarious where there is no particular bias towards any specific value.

The distribution is defined by specifying the minimum and maximum values of the range. Random numbers can be generated within the range with equal probabilities.

iii) **Skewed Distribution:-**

It is an asymetric probabilities distribution where the data cont concentrated more on one side of the distribution than the other

It is useful for representing real-world phenomenas that exhibit a natural bias or asymmetry, such as income distribution, exam scores or stock market returns.

Skewness is measured using statistical techniques, and there are different types of skewed distribution such as positive skewed (tail to the right) and negatively skewed (tail to the left). The choice of analysis and interpretation. depends on the specific skewness present in the data.

iv) Bimodal Distribution :-

It is a probability distribution with two distinct peaks or modes. It is used to represent situations where data can be classified into two separate groups or categories, each with its own characteristics or properties.

It can be identified through data visualization techniques such as histogram or density plots, which show the presence of two prominent peaks. Analysis and Interpretation of Bimodal distributions may involve examining each mode separately or considering the interaction between the two modes.

(v) Exponential Distribution:-

It describe the time between events occuring at a constant rate.

It commonly used in reliability analysis, queuing theory and survival analysis, where the occurence of events over time follows a constant or exponential decay pattern.

It is defined by the rate parameters, which determines the average time between events. It can be used to calculate probabilities of waiting times, survival probabilities or failure rates.

(vi) Log-Normal Distribution:-

It describe where the logarithm of the data follows a normal distribution.

It is used to model data that is skewed to the right, such as the distribution of income, stock prices or natural phenomena that are constrained to positive values

The distribution is defined by its mean and standard deviation after taking the logarithm of the data. It can be used to analyze and make inference about the original non-logarithmic data.