Normal Distribution and Empirical Rule

Problem Statement : Take one domain and draw the Normal Distribution graph.

Introduction

Data analysis plays an important role in understanding patterns and behavior in real-world problems. Many natural and human-related phenomena follow a specific statistical pattern known as the Normal Distribution. This project explains the concept of Normal Distribution and the Empirical Rule using a selected real-world domain.

Selected Domain: Student Exam Score Analysis

Student marks in large examinations usually follow a predictable pattern:

- Most students score average marks.
- Fewer students score very high marks.
- Few students score very low marks.

This pattern forms a bell-shaped curve, which represents a Normal Distribution.

What is Normal Distribution?

Normal Distribution is a continuous probability distribution where data values are symmetrically distributed around the mean.

Characteristics

- The graph is bell-shaped.

- Mean, Median, and Mode are equal.

- Data spreads equally on both sides of the mean.

- Extreme values occur less frequently.

- The total area under the curve equals 1.

- Normal Distribution is also called Gaussian Distribution.

Importance of Normal Distribution

Normal distribution helps in:

- Understanding data behavior

- Predicting future outcomes

- Statistical analysis

- Machine Learning model assumptions

- Quality control and decision making

- Many real-life examples follow normal distribution such as:

- Exam scores

- Human height and weight

Mean and Standard Deviation

Mean (μ)

The mean represents the average value of the dataset.

It is the center of the distribution.

Example: If average student marks = 70, then mean = 70.

Standard Deviation (σ)

Standard deviation measures how far data spreads from the mean.

Small standard deviation → data close to mean

Large standard deviation → data widely spread

Example: Standard deviation = 10 means most scores vary around 10 marks from the average.

Empirical Rule (68–95–99.7 Rule)

The Empirical Rule explains how data is distributed in a normal distribution.

Rule Explanation

68% Rule:

About 68% of data lies within one standard deviation from the mean.

Range = μ ± 1σ

95% Rule

About 95% of data lies within two standard deviations.

Range = μ ± 2σ

99.7% Rule

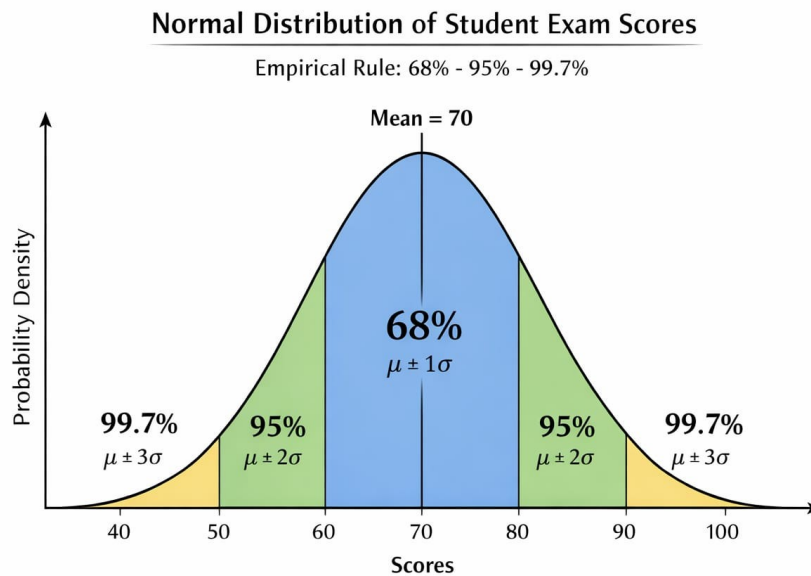About 99.7% of data lies within three standard deviations.

Range = μ ± 3σ

Example using student scores

Example Using Student Scores

Mean (μ) = 70 Standard Deviation (σ) = 10

Range	Student Score Interval	Percentage
μ ± 1σ	60 – 80	68% students
μ ± 2σ	50 – 90	95% students
μ ± 3σ	40 – 100	99.7% students

Graph

Normal Distribution of Student Exam Scores
Empirical Rule: 68% - 95% - 99.7%

Explanation

- The Normal Distribution graph forms a bell-shaped curve.

- The center peak represents the mean score.

- The curve decreases symmetrically on both sides.

- Majority of students lie near the average score.

- Extreme high or low scores are rare.

- The shaded regions in the graph represent the Empirical Rule percentages (68%, 95%, and 99.7%).

Applications of Empirical Rule

- The Empirical Rule is used in:

- Educational performance analysis

- Business sales prediction

- Healthcare data analysis

- Financial risk assessment

- Machine learning data preprocessing

Advantages

- Easy to understand data distribution

- Helps identify outliers

- Useful for prediction and decision making

- Widely applicable in statistics and AI

Limitations

- Works only for normally distributed data

- Not suitable for skewed datasets

- Real-world data may not perfectly follow normal

Conclusion

Normal Distribution is one of the most important concepts in statistics and data science. Using the Empirical Rule, we can understand how data spreads around the mean and analyze real-world situations effectively. In the student exam score domain, the majority of students perform around the average, while extreme performances occur less frequently, confirming the characteristics of a normal distribution.