

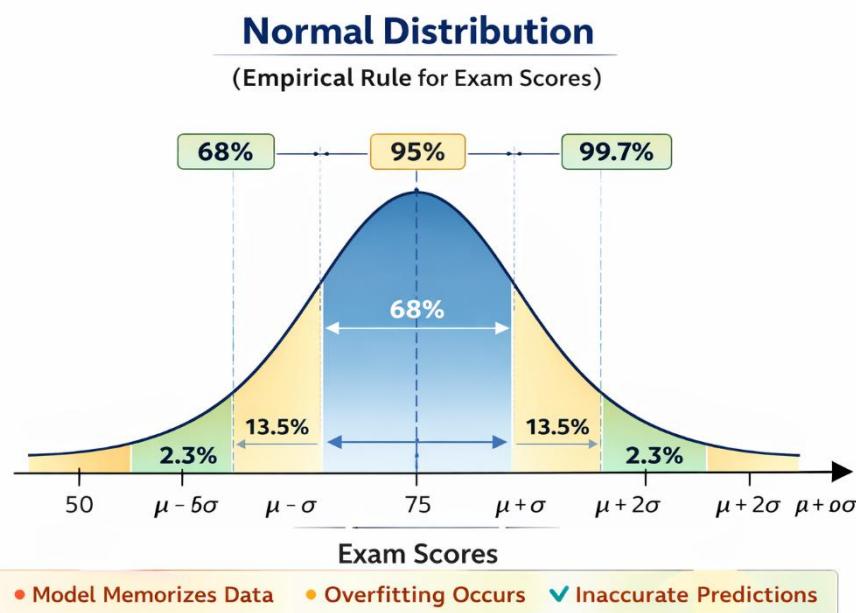
Assignment-2

Take one Domain and draw the graph (Normal distribution) (Empirical rule)

Solution :

➤ Introduction

In statistics and data science, the **normal distribution**, also known as the **Gaussian distribution**, is one of the most important probability distributions. It is widely used to model real-world data because many natural and human-related phenomena follow this pattern. The graph of a normal distribution forms a **bell-shaped curve**, which is symmetric around the mean. The majority of data values lie near the center, while fewer values appear at the extreme ends.



The **empirical rule**, also called the **68–95–99.7 rule**, describes how data is distributed around the mean in a normal distribution. According to this rule, approximately **68% of data lies within one standard deviation ($\pm 1\sigma$)**, **95% within two standard deviations ($\pm 2\sigma$)**, and **99.7% within three standard deviations ($\pm 3\sigma$)**. This rule helps in understanding data spread, variability, and probability.

In this assignment, we select **the Education domain (Students' Exam Marks)** to explain normal distribution and the empirical rule. We also include the graphical representation and detailed interpretation.

1. Selected Domain: Education – Students' Examination Marks

The **Education domain** is one of the best real-world examples of normal distribution. When a large number of students appear for an examination, their **marks usually follow a normal distribution pattern**.

Most students score around the **average marks**, fewer students score **very high marks**, and fewer students score **very low marks**. This creates a **bell-shaped curve** when the marks are plotted on a graph.

❖ Example Dataset

Let us consider a class of **1000 students** where:

- **Mean (μ) = 75 marks**
- **Standard Deviation (σ) = 10 marks**

This means:

- Most students scored near **75 marks**.
- The marks range approximately between **45 to 105**.

2. Concept of Normal Distribution

A **normal distribution** is a continuous probability distribution that is symmetric about its mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

❖ Key Properties of Normal Distribution

- The curve is **bell-shaped and symmetric**.
- The **mean = median = mode**.
- The total area under the curve is **1 (100%)**.
- The curve approaches the horizontal axis but never touches it.
- Most values cluster around the center.

❖ Mathematical Representation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where:

- μ = Mean
- σ = Standard Deviation

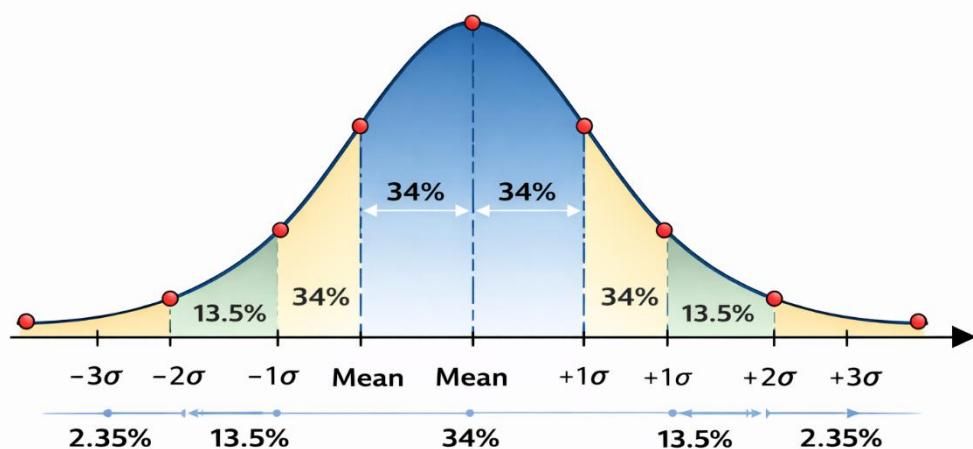
3. Empirical Rule (68–95–99.7 Rule)

The **empirical rule** describes how data is distributed in a normal distribution. It gives the approximate percentage of data lying within one, two, and three standard deviations from the mean.

❖ Empirical Rule Breakdown

Range	Percentage of Data
$\mu \pm 1\sigma$	68%
$\mu \pm 2\sigma$	95%
$\mu \pm 3\sigma$	99.7%

➤ Graphical Representation (Standard Format)



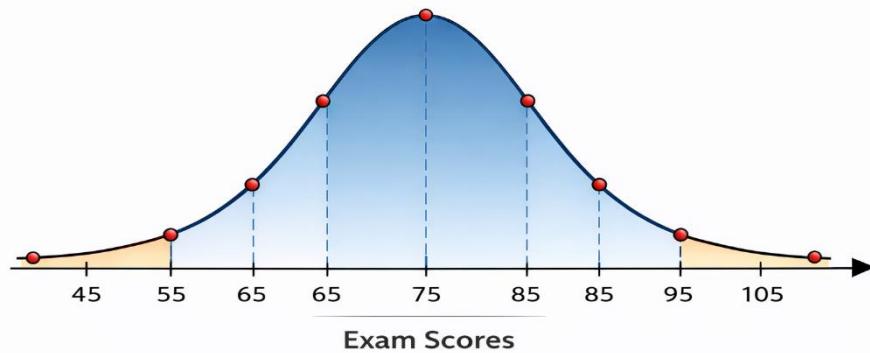
❖ Interpretation

- $34\% + 34\% = 68\% \rightarrow$ within $\pm 1\sigma$

- $13.5\% + 13.5\% = 27\%$ → between $\pm 1\sigma$ and $\pm 2\sigma$
- $2.35\% + 2.35\% = 4.7\%$ → between $\pm 2\sigma$ and $\pm 3\sigma$

4. Graph of Normal Distribution (Students' Marks Example)

Graph Representation



❖ Axis Explanation

- **X-axis:** Marks scored by students
- **Y-axis:** Frequency (Number of students)

❖ Graph Interpretation

- Peak of the curve is at **75 marks (mean)**.
- Most students scored between **65 and 85**.
- Very few students scored below **55** or above **95**.

5. Application of Empirical Rule to Education Domain

Let:

- Mean = 75
- Standard Deviation = 10

❖ Using Empirical Rule

Range Marks Interval % of Students

$\mu \pm 1\sigma$	65 to 85	68%
$\mu \pm 2\sigma$	55 to 95	95%
$\mu \pm 3\sigma$	45 to 105	99.7%

❖ Interpretation

- **68% students scored between 65 and 85 marks.**
- **95% students scored between 55 and 95 marks.**
- **Almost all students scored between 45 and 105 marks.**
- This helps teachers and institutions **analyze performance, grading patterns, and student capability.**

6. Importance of Normal Distribution and Empirical Rule

❖ Advantages Helps in performance evaluation.

- Supports **grading and ranking systems.**
- Useful for **quality control.**
- Helps in **prediction and forecasting.**
- Widely used in **statistics, ML, and AI.**

7. Real-Life Applications

❖ Domains Using Normal Distribution

- **Education** – Exam marks, IQ levels
- **Medical** – Blood pressure, height, weight
- **Finance** – Stock price fluctuations
- **Manufacturing** – Product quality measurement
- **Weather** – Temperature variations

➤ Conclusion

The **normal distribution and empirical rule** are essential statistical concepts used to understand and analyze real-world data. By selecting the **Education domain**, we demonstrated how students' exam marks naturally form a bell-shaped curve. The **empirical rule (68–95–99.7)** helps in measuring variability, identifying trends, and predicting performance. This knowledge is extremely valuable in **data analysis, machine learning, education systems, and research fields.**