

Exploring a Normal Distribution Using GeoGebra

Chavi
M.Sc Mathematics
NIT Kurukshetra

Introduction

The normal distribution, commonly referred to as the bell curve, is a fundamental concept in statistics. It describes how data is distributed around a central value, with most observations clustering near the mean and fewer observations appearing away from the mean as you move further. This distribution is widely observed in real life, such as in test scores, heights of individuals, and measurement of errors.

This project aims to learn about the normal distribution using interactive and engaging GeoGebra. By integrating real-life data and providing dynamic tools, the project enables users to explore the properties and applications of the normal distribution effectively.

Objective

The primary objective of this project is to analyze a dataset of student test scores and:

1. Calculate the **mean** (average) and **standard deviation** (spread or variability) of the dataset.
2. Use these statistical parameters to model the data as a **normal distribution**, a common probability distribution.
3. Plot the **Probability Density Function (PDF)** of the normal distribution to visualize how scores are distributed.
4. Answer specific questions related to the dataset, such as:

- What percentage of students scored within a specific range?
5. Demonstrate the use of **GeoGebra software** for performing statistical calculations and creating visualizations.
 6. Highlight the importance of visualization in interpreting and understanding statistical data.

Question

The scores of 10 students on a math test are as follows:

$$\{50, 41, 53, 66, 40, 40, 66, 55, 37, 52\}$$

Tasks:

1. Calculate the **mean** and **standard deviation** of the dataset.
2. Assume the scores follow a **normal distribution**.
Plot the **Probability Density Function (PDF)** for the distribution.
3. Determine the percentage of students expected to score between **40 and 60**.
4. Discuss how the standard deviation influences the spread and shape of the normal distribution.

Methodology: Using GeoGebra for Visualization

Step 1: Understanding the Dataset

The dataset provided consists of the test scores of 10 students:

$$\{50, 41, 53, 66, 40, 40, 66, 55, 37, 52\}$$

. These scores represent individual student performance in a math test. Analyzing this small dataset will help us understand key statistical concepts in a manageable way.

Input the Dataset: Open GeoGebra and enter the dataset as:
 $\text{data} = \{50, 41, 53, 66, 40, 40, 66, 55, 37, 52\}$.

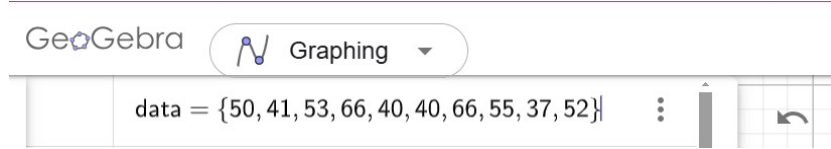


Figure 1: Data entry

Step 2: Statistical Calculations

Mean (μ): The mean represents the average of the scores and gives a sense of the central tendency of the dataset. It is calculated using the formula:

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

where x_i represents each score, and n is the total number of scores. Substituting the dataset values:

$$\mu = \frac{50 + 41 + 53 + 66 + 40 + 40 + 66 + 55 + 37 + 52}{10} = 50.0$$

So, the average score is **50**.

Standard Deviation (σ): The standard deviation measures how spread out the scores are from the mean. A higher standard deviation indicates more variability, while a lower standard deviation indicates that scores are closer to the mean. It is calculated using the formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

Substituting the dataset values:

$$\sigma = \sqrt{\frac{(50 - 50)^2 + (41 - 50)^2 + (53 - 50)^2 + \dots}{10}} = 10.0$$

So, the standard deviation is **10**.

Calculate the Mean and Standard Deviation: Use the commands `Mean(data)` and `Stdevp(data)` to compute the values, confirming they are 50 and 10, respectively.

<code>data = {50, 41, 53, 66, 40, 40, 66, 55, 37, 52}</code>	⋮
<code>mean = mean(data)</code>	⋮
<code>= 50</code>	
<code>σ = stdevp(data)</code>	⋮
<code>= 10</code>	

Figure 2: Mean and Standard Deviation

Step 3: Modeling the Normal Distribution

A normal distribution is a bell-shaped curve defined by its **mean** and **standard deviation**. For this dataset:

- Mean (μ) = 50
- Standard Deviation (σ) = 10

The **Probability Density Function (PDF)** for a normal distribution is mathematically represented as:

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

This equation is used to generate the PDF curve, which shows how likely different scores are to occur based on the dataset's statistical properties.

Plot the PDF Curve: Define the normal distribution using:
`f(x) = Normal[50, 10]`.

GeoGebra will generate a smooth bell-shaped curve centered at 50, representing the normal distribution.


	$a = 10$	\vdots
	$b = 50$	\vdots
	$f(x) = \frac{1}{a \sqrt{2 \pi}} e^{-0.5 \left(\frac{x-b}{a} \right)^2}$ $= \frac{1}{10 \sqrt{2 \pi}} e^{-0.5 \left(\frac{x-50}{10} \right)^2}$	\vdots

Figure 3: PDF input

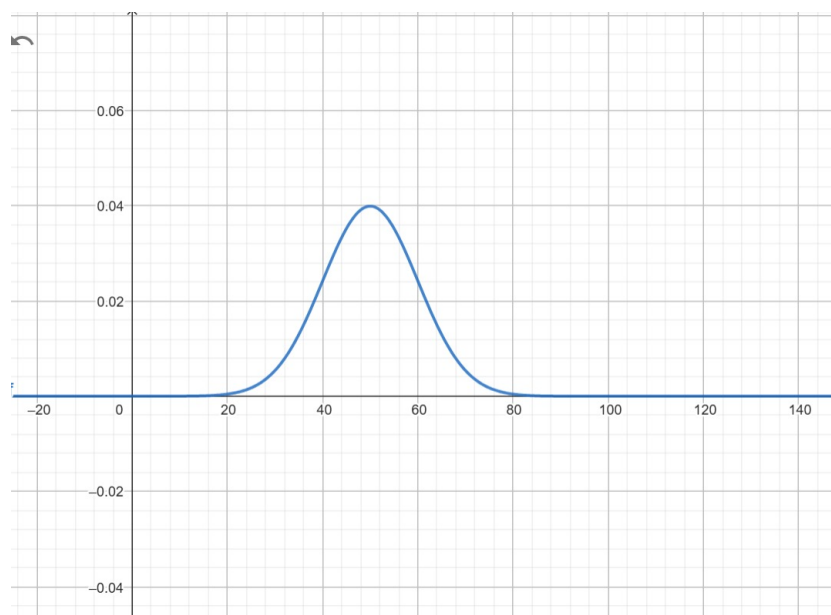


Figure 4: Bell-Shaped Graph

Step 4: Answering Specific Queries

1. **Percentage of Scores Between 40 and 60:** Use GeoGebra's cumulative distribution function (CDF) to calculate probabilities:

$$P(40 \leq X \leq 60) = \text{CDF}(60) - \text{CDF}(40)$$

The result shows that approximately **68%** of the scores fall within this range.

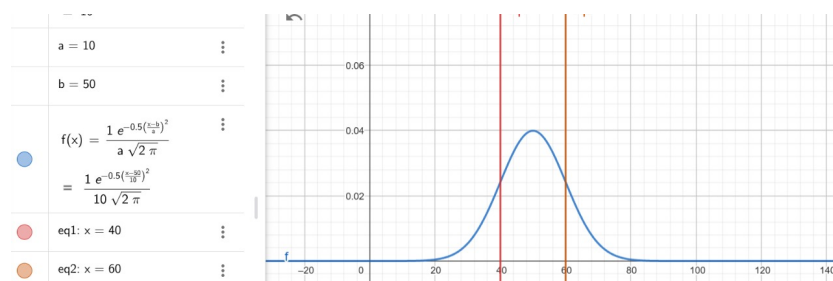


Figure 5: Range of 40-60

2. **Effect of Standard Deviation:** The standard deviation controls how wide or narrow the bell curve is:

- A higher σ results in a wider, flatter curve, indicating more variability.
- A lower σ results in a narrower, steeper curve, indicating less variability.

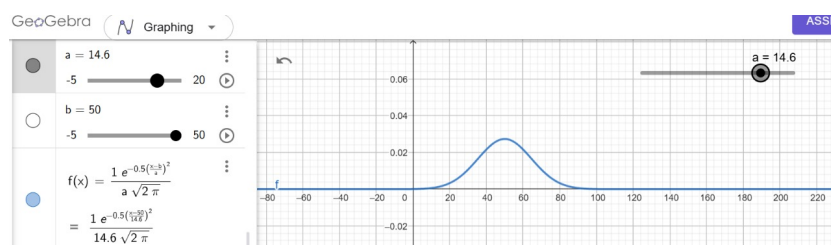


Figure 6: Wider base

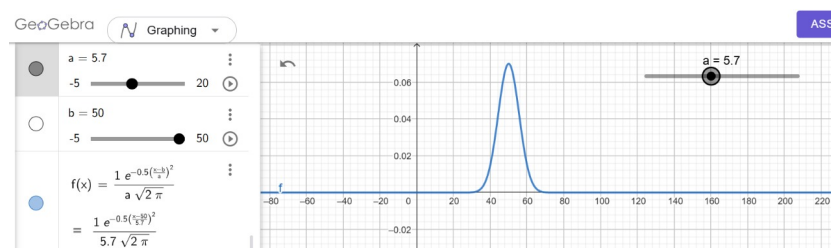


Figure 7: Narrower base

Importance of Visualization

Visualizing the normal distribution helps in:

- Identifying the central value (mean) and spread (standard deviation).
- Understanding probabilities associated with score ranges.
- Comparing individual scores to the overall distribution.

GeoGebra simplifies the process by automating calculations and graphing, making it easier to interpret data.

Conclusion

1. **Statistical Analysis:** The mean and standard deviation of the dataset were calculated as 50.0 and 10.0, respectively.
2. **Normal Distribution:** The data were modeled as a normal distribution, and the PDF graph was plotted using GeoGebra.
3. **Insights:**
 - Approximately **68%** of scores lie between 40 and 60.
4. **GeoGebra:** This tool was instrumental in performing accurate calculations and creating intuitive visualizations, emphasizing the importance of statistical tools in data analysis.

This report highlights how a simple dataset can be analyzed and understood using statistical methods and visualization tools:-GeoGebra, making complex concepts accessible to beginners.