## Histogram

* X-axis (horizontal): Ranges like **(1, 5), (5, 9), (9, 13)** etc. These are **intervals** of values
* Y-axis (vertical): Tells you **how many values** (or frequency) fall into each interval
* The taller the bar → the more values are in that range

**🔍 Analysis of Histogram**

* The bar for **(13, 17)** is the tallest — around 24 values fall into that range
* So most of your data is centered around that interval

## Bernaulli Distribution

Imagine you're flipping a coin.

* **Heads = 1** (Success)
* **Tails = 0** (Failure)

This is a **Bernoulli Trial** — **only 2 possible outcomes**: 1 or 0

**💡 Bernoulli Distribution:**

It is the **probability distribution** of a random variable which takes only two values:

* **1** with probability **p**
* **0** with probability **1 - p**

**📘 Example:**

If the probability of getting heads is 0.7 (p = 0.7), then:

P(X = 1) = 0.7

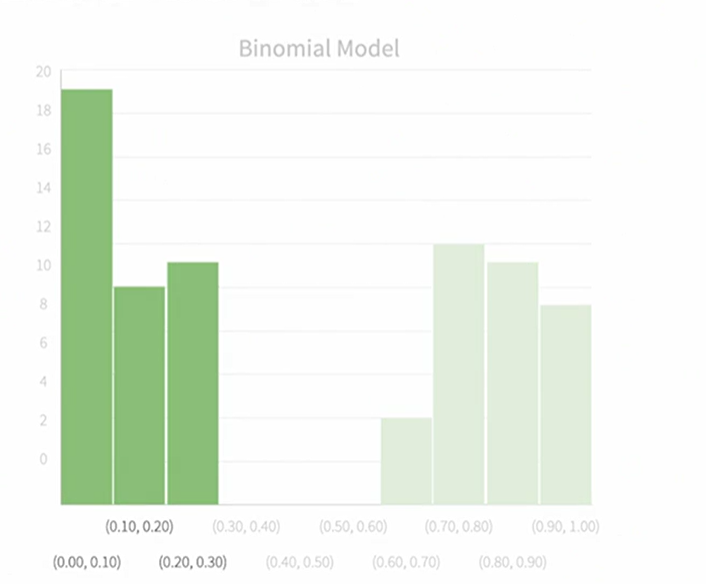
P(X = 0) = 0.3

That’s a Bernoulli distribution! It’s like the simplest distribution in probability

## Line : Mean is not evenly distributed

5, 6, 8, 10, 90, 91, 95

Here the mean may be 50, but almost no data is around 50. So even though the mean exists mathematically, it's not representative of the data. This often happens in skewed or multi-modal distributions



## Line : The mean itself is normally distributed, but outcomes are not

**Suppose:**

You run a test on 10 students and record how many hours they study:

[1, 2, 2, 3, 10, 12, 13, 15, 18, 20]

🔸 These numbers are **all over the place**.  
🔸 Some study very little (1 hour), some a lot (20 hours).  
🔸 The data is **not a nice bell curve** → so **outcomes are not normal**

**🎯 Now do this:**

You repeat this test with **many different groups** of 10 students each.

Each time, you calculate the **average (mean)** of that group.

You get results like:

Group 1 → mean = 10

Group 2 → mean = 11

Group 3 → mean = 9.5

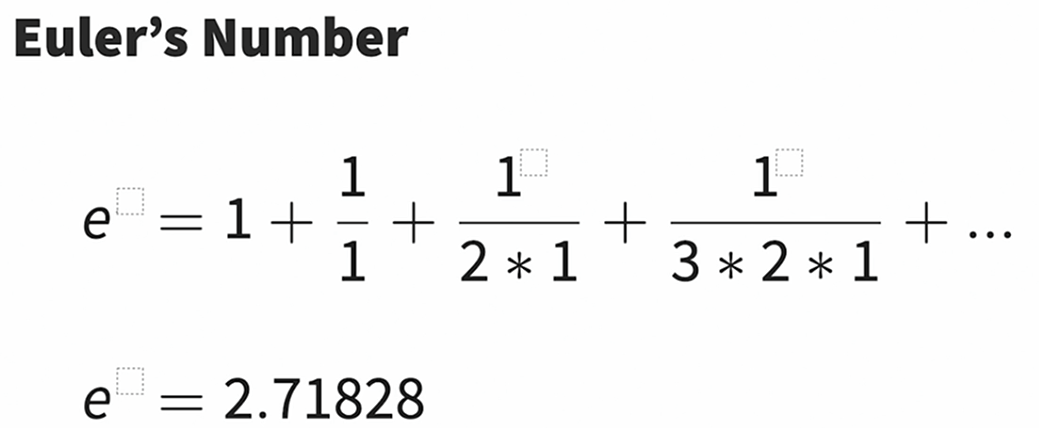
Group 4 → mean = 10.2

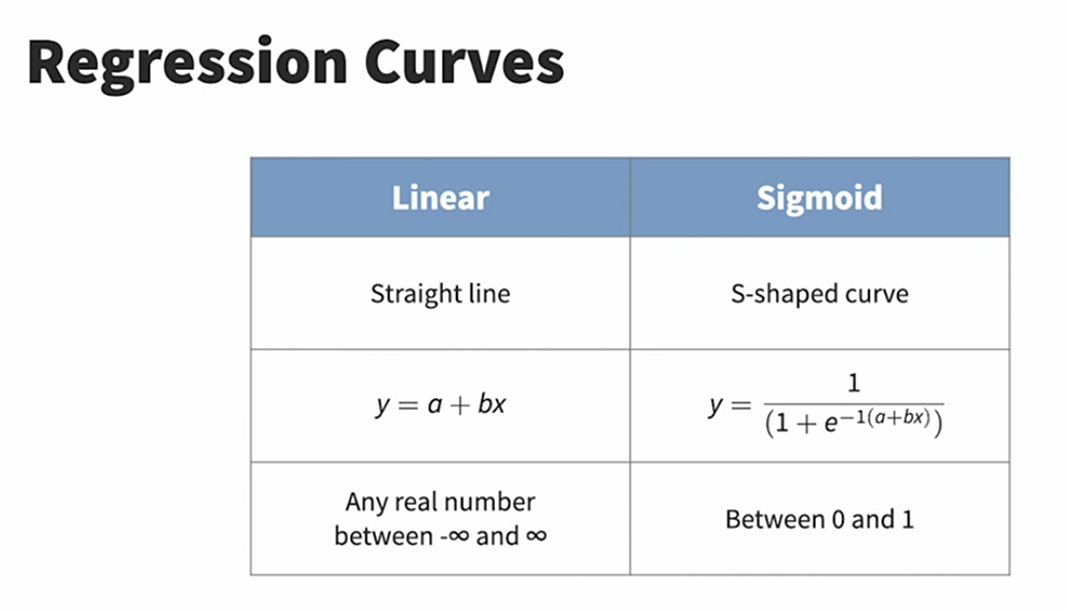
Group 5 → mean = 10.1

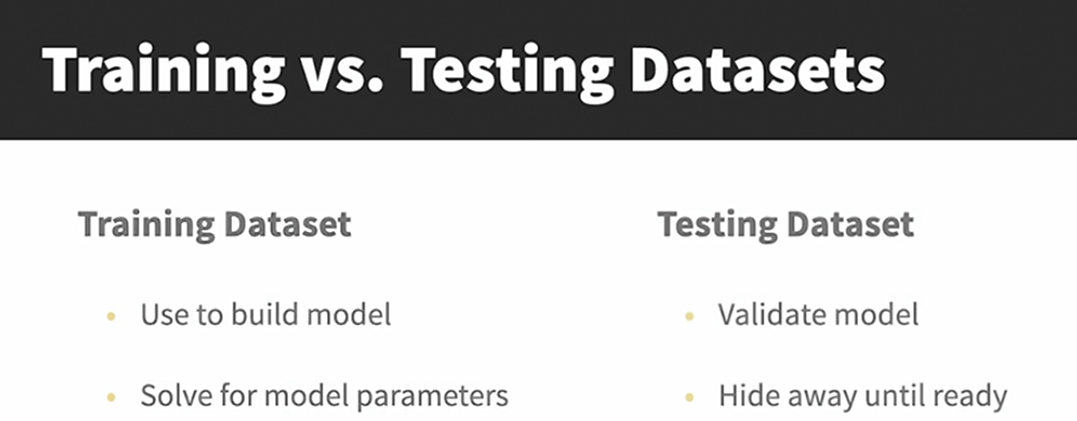
These **means** (averages) start forming a **bell-shaped curve** when plotted — that’s **Gaussian distribution**

**Super Simple Analogy:**

* Imagine a classroom full of students shouting random numbers — chaos! 🎤🔊
* But if you take the **average** of each small group — those group averages are calm and centred
* So the **crowd (outcomes)** is wild, but the **average (mean)** is peaceful and predictable







**Algorithms are designed to make the best possible prediction even when the data isn’t perfect**

They’re trying to **learn patterns and generalize**, even when there’s uncertainty or confusion in the input

A graph of a training error

AI-generated content may be incorrect.

The model should learn the **underlying pattern**, not every tiny detail

This way, it performs **well on both training and testing data**. That’s called **generalization**, and it's the goal of any good machine learning model

**We don’t want to peek at the testing dataset until we’ve fit our model using the training dataset**

We should **not look at the testing data** while training our model — only use it **after** the model is fully trained