Probability Distributions

What is a Probability Distribution?

Imagine you're analyzing real-world data — sales, weather, clicks, heights, etc. You'll notice:

- Some values are more **likely** than others.
- There's a **pattern** to how data is spread.

That pattern is described by a **probability distribution**.

In short:

> A probability distribution tells you how likely different outcomes are.

What is a Probability Distribution?

A **probability distribution** is a **mathematical function** or **table** that:

- Assigns **probabilities** to all possible outcomes of a random process.

There are two types:

H T

1. Discrete Probability Distribution

- For outcomes you can **count** (e.g. number of heads in 3 coin tosses).
- Example: **Binomial Distribution**

> Think: "What's the probability I get exactly 2 heads in 3 coin tosses?"

---

height 175 cm
178.2
173.881

2. Continuous Probability Distribution

- For outcomes that can be **any number in a range** (e.g. height, weight).
- Example: **Normal Distribution**

180.7721

> Think: "What's the probability someone's height is between 165 cm and 170 cm?"

weight → 63.7721
75.2
78.9

A Simple Analogy

Let's say you roll a die:

- Outcomes = {1, 2, 3, 4, 5, 6}
- Each has a probability of `1/6`

$P(1) = 1/6$
$P(2) = \frac{1}{6}$
⋮
$P(6) = \frac{1}{6}$

This is a **uniform distribution** (discrete).

Now imagine measuring people's **heights**:

- You don't get fixed values.
- Instead, you get a **curve** — most people around average height, fewer very short or very tall.

That's a **normal distribution** (continuous).

Why Are Distributions Useful in Data Science?

1. **Model real-world randomness** (user behavior, errors, arrivals, etc.)
2. **Make predictions** (how likely is a customer to buy?)
3. **Run simulations** (A/B testing, Monte Carlo)
4. **Assume underlying patterns** (e.g., linear regression assumes errors are normally distributed)

$P(171.2) =$

$P(180.7) =$

$P(150.8) =$

Summary:

Type	Example	Used For
Discrete	Binomial, Poisson	Count of events
Continuous	Normal, Uniform	Measuring quantities

76 ~~kg~~