



STATISTICS

Adam Klepáč

September 14, 2023

WHAT EVEN IS STATISTICS?

STATISTICS

Statistics is a mathematical discipline concerned with predicting future state of a system based *solely* on its past behaviour.

WHAT EVEN IS STATISTICS?

STATISTICS

Statistics is a mathematical discipline concerned with predicting future state of a system based *solely* on its past behaviour.

The collective information about a system's past state is called **data**.

WHAT EVEN IS STATISTICS?

STATISTICS

Statistics is a mathematical discipline concerned with predicting future state of a system based *solely* on its past behaviour.

The collective information about a system's past state is called **data**.
It assigns **probabilities** to each possible future state of system based on data.

WHAT EVEN IS STATISTICS?

STATISTICS

Statistics is a mathematical discipline concerned with predicting future state of a system based *solely* on its past behaviour.

The collective information about a system's past state is called **data**.
It assigns **probabilities** to each possible future state of system based on data.
It also assigns probabilities to the **possibility of wrong prediction**.

EXAMPLE – BIASED COIN?

We throw a coin 10 times with the following outcome:

$$\{H, H, H, T, H, T, H, H, H, T\},$$

H for ‘heads’, *T* for ‘tails’.

EXAMPLE – BIASED COIN?

We throw a coin 10 times with the following outcome:

$$\{H, H, H, T, H, T, H, H, H, T\},$$

H for 'heads', T for 'tails'. We can ask two questions:

EXAMPLE – BIASED COIN?

We throw a coin 10 times with the following outcome:

$$\{H, H, H, T, H, T, H, H, H, T\},$$

H for ‘heads’, T for ‘tails’. We can ask two questions:

- What is the probability that the **next toss** will come out ‘heads’/‘tails’?

EXAMPLE – BIASED COIN?

We throw a coin 10 times with the following outcome:

$$\{H, H, H, T, H, T, H, H, H, T\},$$

H for ‘heads’, T for ‘tails’. We can ask two questions:

- What is the probability that the **next toss** will come out ‘heads’/‘tails’?
- Is this coin is **biased towards** ‘heads’/‘tails’ with *allowed probability of error* α ?

EXAMPLE – BIASED COIN?

We throw a coin 10 times with the following outcome:

$$\{H, H, H, T, H, T, H, H, H, T\},$$

H for ‘heads’, T for ‘tails’. We can ask two questions:

- What is the probability that the **next toss** will come out ‘heads’/‘tails’?
 - We got 7 heads out of 10 tosses, so the probability for the next toss being heads is $7/10$.
- Is this coin is **biased towards** ‘heads’/‘tails’ with *allowed probability of error* α ?

EXAMPLE – BIASED COIN?

We throw a coin 10 times with the following outcome:

$$\{H, H, H, T, H, T, H, H, H, T\},$$

H for ‘heads’, T for ‘tails’. We can ask two questions:

- What is the probability that the **next toss** will come out ‘heads’/‘tails’?
 - We got 7 heads out of 10 tosses, so the probability for the next toss being heads is $7/10$.
- Is this coin is **biased towards** ‘heads’/‘tails’ with *allowed probability of error* α ?
 - **No**, for $\alpha = 0.05$.
 - **Yes**, for $\alpha = 0.2$.

CONTENTS



Data

Types of Data



DATA

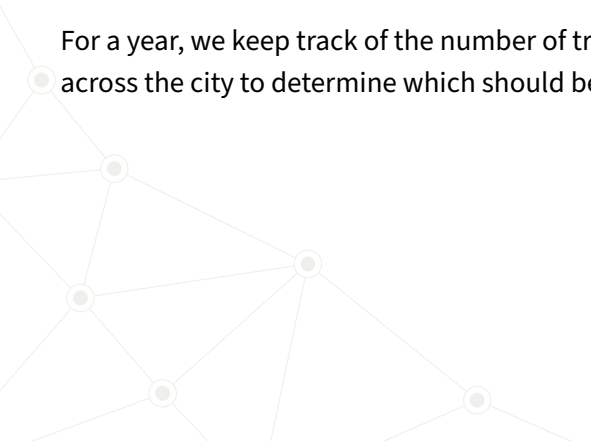
WHAT DO WE MEAN BY DATA?

DATA

Two sets (called *inputs* and *outputs*) describing the studied system.

EXAMPLE – JUNCTIONS

For a year, we keep track of the number of traffic accidents per day on road junctions across the city to determine which should be first replaced by roundabouts.



EXAMPLE – JUNCTIONS

For a year, we keep track of the number of traffic accidents per day on road junctions across the city to determine which should be first replaced by roundabouts. An **input** is a day in a year.

EXAMPLE – JUNCTIONS

For a year, we keep track of the number of traffic accidents per day on road junctions across the city to determine which should be first replaced by roundabouts.

An **input** is a day in a year.

An **output** is the number of traffic accidents in a given day.

EXAMPLE – FIRST BABY

We study the age that women bear children for the first time across Europe.

EXAMPLE – FIRST BABY

We study the age that women bear children for the first time across Europe.

An **input** would be a name of a European country.

An **output** is the average age of a first-time mother in that country.

1

TYPES OF DATA



DISCRETE DATA VS. CONTINUOUS DATA

DISCRETE DATA

We call a data **discrete** if the set of *inputs* (and therefore also that of *outputs*) is **countable**.

DISCRETE DATA VS. CONTINUOUS DATA

DISCRETE DATA

We call a data **discrete** if the set of *inputs* (and therefore also that of *outputs*) is **countable**.

Both previous examples feature **discrete** data.

- There are only *finitely many* junctions in a city.

DISCRETE DATA VS. CONTINUOUS DATA

DISCRETE DATA

We call a data **discrete** if the set of *inputs* (and therefore also that of *outputs*) is **countable**.

Both previous examples feature **discrete** data.

- There are only *finitely many* junctions in a city.
- There are only *finitely many* countries on a continent.

DISCRETE DATA VS. CONTINUOUS DATA

CONTINUOUS DATA

We call a data **continuous** if the set of inputs is **uncountable**. In this case, the data is actually a **function**: set of inputs \rightarrow set of outputs.

DISCRETE DATA VS. CONTINUOUS DATA

CONTINUOUS DATA

We call a data **continuous** if the set of inputs is **uncountable**. In this case, the data is actually a **function**: set of inputs \rightarrow set of outputs.

More often than not, the inputs in a continuous data are **moments in time** or **coordinates in space**.

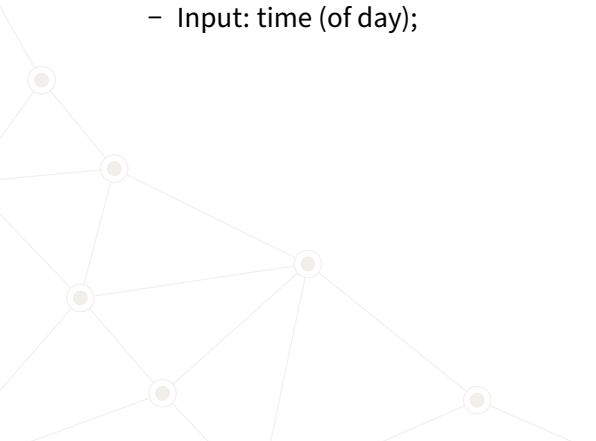
CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.



CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);



CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);
 - Output: number of trains in the station.

CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);
 - Output: number of trains in the station.
 - The data is a function $f : [0, 24] \rightarrow \mathbb{N}$.

CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);
 - Output: number of trains in the station.
 - The data is a function $f : [0, 24] \rightarrow \mathbb{N}$.
- Another example is the density of air per cubic meter.

CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);
 - Output: number of trains in the station.
 - The data is a function $f : [0, 24] \rightarrow \mathbb{N}$.
- Another example is the density of air per cubic meter.
 - Input: Coordinates of a unit cube in space.

CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);
 - Output: number of trains in the station.
 - The data is a function $f : [0, 24] \rightarrow \mathbb{N}$.
- Another example is the density of air per cubic meter.
 - Input: Coordinates of a unit cube in space.
 - Output: The combined weight of air molecules.

CONTINUOUS DATA – EXAMPLES

- We study the number of trains in a railway station at any given time.
 - Input: time (of day);
 - Output: number of trains in the station.
 - The data is a function $f : [0, 24] \rightarrow \mathbb{N}$.
- Another example is the density of air per cubic meter.
 - Input: Coordinates of a unit cube in space.
 - Output: The combined weight of air molecules.
 - The data is a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$.