# Math 1501 - Elements of Statistics with Applications

1.1 - Definitions of Statistics, Probability, and Key Terms

# Chapter 1: Sampling and Data



By the end of 1.1 you should be able to answer:

- Who are we studying, and who did we actually measure?
- What is the difference between describing data and concluding from it?
- Which numbers can we compute, and which would be nonsense?

## Let's collect some data — and then let Python draw it

On average, to the nearest half hour, how many hours do you sleep per night?

We will type every answer in the room into a list and get a picture back.

```
sleep = [5, 5.5, 6, 6, 6, 6.5, 6.5, 6.5, 7, 7, 8, 8, 9]
```

Google Colab Link

## Let's collect some data — and then let Python draw it

On average, to the nearest half hour, how many hours do you sleep per night?

We will type every answer in the room into a list and get a picture back.

```
sleep = [5, 5.5, 6, 6, 6, 6.5, 6.5, 6.5, 7, 7, 8, 8, 9]
```

That is a *list* in Python. It is how data gets typed in and stored.

This is not a coding course. Python is a tool we point at data, the same way we point a calculator at arithmetic. Google Colab Link

# The two halves of the subject

**Statistics** — the science of collecting, analyzing, interpreting, and presenting data.

# The two halves of the subject

**Statistics** — the science of collecting, analyzing, interpreting, and presenting data.

**Descriptive statistics** — organizing and summarizing data – with a graph, or with a number such as an average.

# The two halves of the subject

**Statistics** — the science of collecting, analyzing, interpreting, and presenting data.

**Descriptive statistics** — organizing and summarizing data – with a graph, or with a number such as an average.

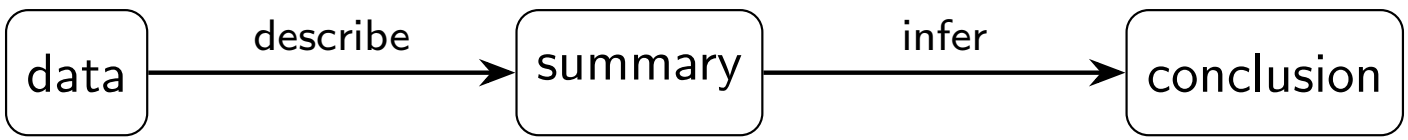
**Inferential statistics** — using data from a small group to draw a conclusion about a much larger group.

# The two halves of the subject

**Statistics** — the science of collecting, analyzing, interpreting, and presenting data.

**Descriptive statistics** — organizing and summarizing data – with a graph, or with a number such as an average.

**Inferential statistics** — using data from a small group to draw a conclusion about a much larger group.



# Why probability shows up in a statistics course

Toss a fair coin 4 times. You may well get 3 heads. Nothing is wrong.

Toss it 4,000 times and the fraction of heads sits very close to one half.

Karl Pearson (Mathematician) actually did this. He tossed a coin 24,000 times and got 12,012 heads.

# What that fraction is actually saying

$$\frac{12{,}012}{24{,}000} = 0.5005$$

# What that fraction is actually saying

$$\frac{12{,}012}{24{,}000} = 0.5005$$

$$\frac{\text{the part we counted}}{\text{the whole group}}$$

# What that fraction is actually saying

$$\frac{12{,}012}{24{,}000} = 0.5005$$

0.50 → %

× 100 →

50%

$$\frac{\text{the part we counted}}{\text{the whole group}}$$

The two extremes tell you everything:

$$\frac{0}{24{,}000} = 0 \quad \text{never happened}$$

0.0 × 100 = 0%

$$\frac{24{,}000}{24{,}000} = 1 \quad \text{happened every time}$$

1.00 × 100 = 100%

# What that fraction is actually saying

$$\frac{12{,}012}{24{,}000} = 0.5005$$

$$\frac{\text{the part we counted}}{\text{the whole group}}$$ ← Heads

The two extremes tell you everything:

$$\frac{0}{24{,}000} = 0 \quad \text{never happened}$$

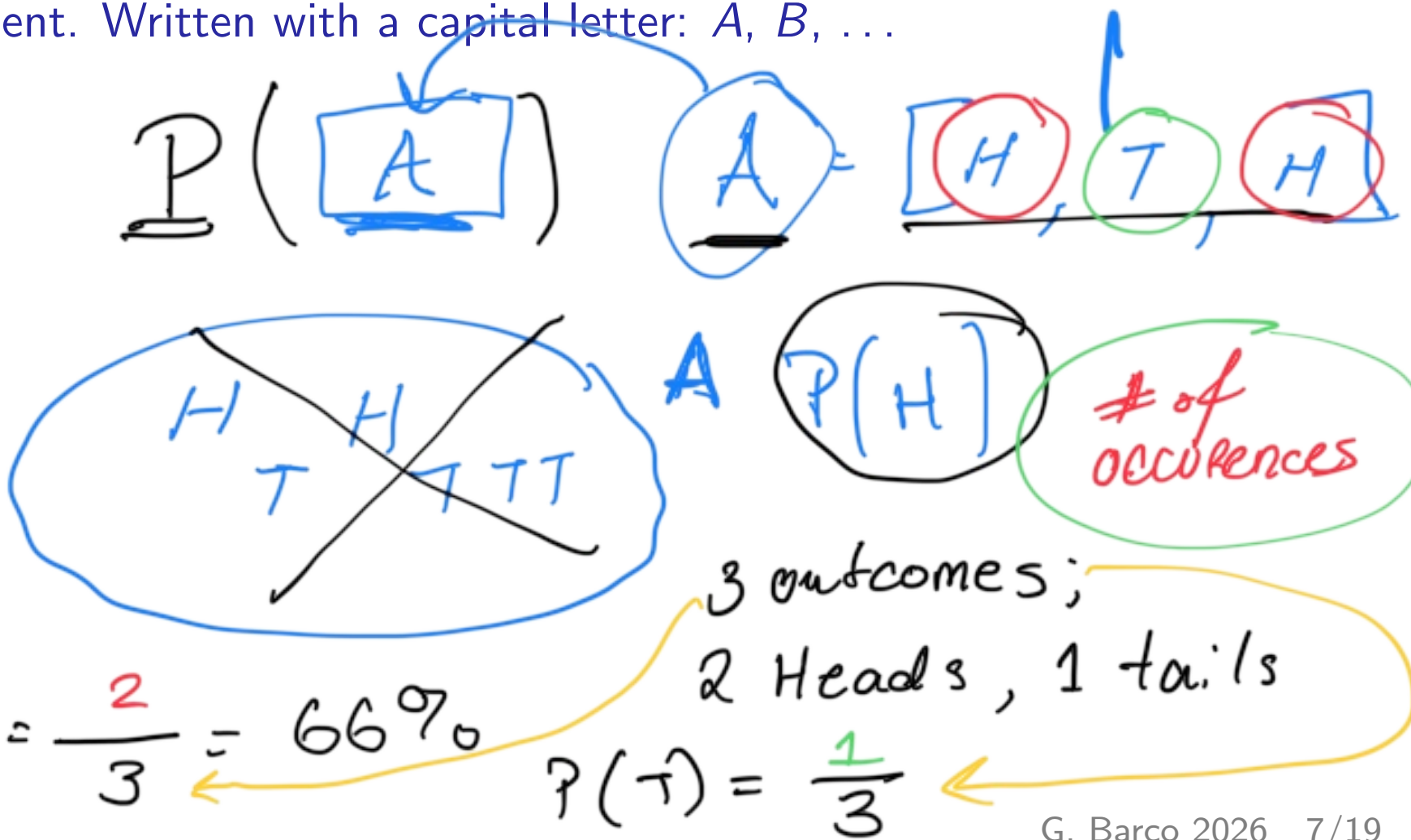
$$\frac{24{,}000}{24{,}000} = 1 \quad \text{happened every time}$$

The part can never exceed the whole. So this number always lands between 0 and 1.

# Naming what just happened: $P(A)$

*So far:* statistics, descriptive, inferential

**Event** — one outcome, or a group of outcomes, from a random experiment. Written with a capital letter: $A$, $B$, ...

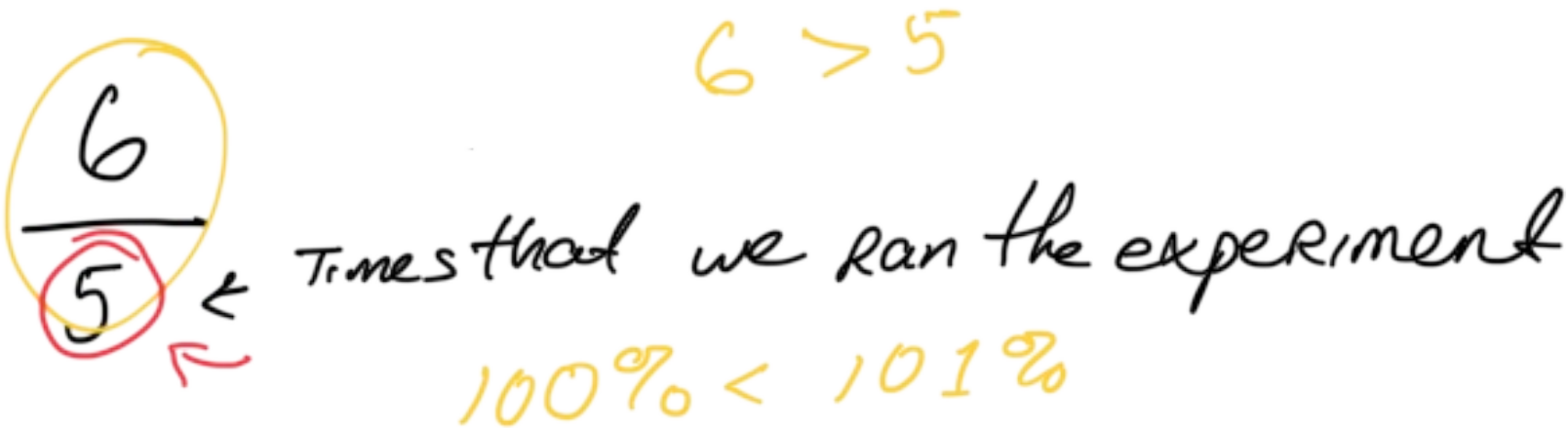


# Naming what just happened: $P(A)$

*So far:* statistics, descriptive, inferential

**Event** — one outcome, or a group of outcomes, from a random experiment. Written with a capital letter: $A$, $B$, ...

**Sample space** — the set of every possible outcome. Written $S$.



# Naming what just happened: $P(A)$

*So far:* statistics, descriptive, inferential

**Event** — one outcome, or a group of outcomes, from a random experiment. Written with a capital letter: $A$, $B$, ...

**Sample space** — the set of every possible outcome. Written $S$.

$P(A)$ — read "the probability of $A$." The same part-over-whole fraction from a moment ago, now with a name for each piece.

# Naming what just happened: $P(A)$

*So far:* statistics, descriptive, inferential

**Event** — one outcome, or a group of outcomes, from a random experiment. Written with a capital letter: $A$, $B$, ...

**Sample space** — the set of every possible outcome. Written $S$.

$P(A)$ — read "the probability of $A$." The same part-over-whole fraction from a moment ago, now with a name for each piece.

$$P(A) = \frac{\text{outcomes that count as } A}{\text{outcomes in the sample space } S}$$

# Naming what just happened: $P(A)$

*So far:* statistics, descriptive, inferential

**Event** — one outcome, or a group of outcomes, from a random experiment. Written with a capital letter: $A$, $B$, ...

**Sample space** — the set of every possible outcome. Written $S$.

$P(A)$ — read "the probability of $A$." The same part-over-whole fraction from a moment ago, now with a name for each piece.

$$P(A) = \frac{\text{outcomes that count as } A}{\text{outcomes in the sample space } S}$$

Roll a die. Let $A =$ "you roll a 2." $S = \{1, 2, 3, 4, 5, 6\}$. $P(A) = \frac{1}{6}$.

$S = \{1, 2, 3, 4, 5, 6\}$

$\frac{2}{6} = \frac{1}{3}$

# Three rules, and that is all

## Definition

**Probability** is a number between 0 and 1, inclusive, that gives the likelihood that a specific event will occur.

$0 \times 100 = 0\%$

$1.00 \times 100 = 100\%$