universität wien

BAK3: Introduction to Quantitative Methods

Week 7: Probability

Leonardo Carella

universität wien

Big Picture

- So far: “point estimation”. What’s the mean? What’s the proportion?
  - Also in the last class: What’s the correlation between X and Y ?
- **Next four weeks: how confident can we be of our estimates?**
- Final four weeks of the course: putting it all together.
  - Linear regression: what’s the relationship between X and Y and how confident are we that there is one? (Very informally)

universität wien

The Plan for today

- Statistics: talking about **uncertainty** and **quantifying** it.
  - Probability distributions
  - Sampling and the Central Limit Theorem
  - The Standard Error
- R: no coding, just reinforcing concepts in the lecture.

universität wien

Probability

- A way of talking about “randomness” or “uncertainty” or “chance”:
  - If I flip a coin, I have a 50% chances that it lands on “tails”.
  - There’s a 40% probability of rain tomorrow.
  - The Social Democrats have a 60% probability of being the largest party in the next election.

universität wien

Probability

- Formally, two ways of understanding probability in statistics:
  - **Frequentist**: probabilities reflect the relative frequency of events over very large number of identical trials.
  - **Bayesian**: subjective belief about the relative likelihood of events, combining preexisting knowledge with new observation.
- **Not our fight here.** The math is the same anyway.
- In this course, we're (informally) frequentist, although it's not very intuitive to imagine "repeating the same election many times over".

universität wien

Axioms of Probability

1. **Non-negativity.** Probability is always positive. $\mathrm{P}(A) \geq 0$
2. **Normalisation.** The sum of the probability of all possible events is 1.
   - Formally, we write that $\mathrm{P}(\Omega) = 1$.
   - Where Ω is known as the sample space. For instance, if you're flipping a coin, $\Omega = \{\text{heads}, \text{tails}\}$. If you're rolling a die, then $\Omega = \{1,2,3,4,5,6\}$. In the second round of the 2016 Austrian presidential election, $\Omega = \{\text{VdB}, \text{Höfer}, \text{abstain}\}$.

universität wien

Axioms of Probability

3. **Addition Rule.** If two events A and B are mutually exclusive, then the probability of either happening is equal to the sum of the probability of A and the probability of B.

   - Formally, $\mathrm{P}(A \text{ or } B) = \mathrm{P}(A) + \mathrm{P}(B)$
   - The probability of rolling a die and obtaining a number less than 3 is $\mathrm{P}(1 \text{ or } 2) = \mathrm{P}(1) + \mathrm{P}(2) = 1/6 + 1/6 = 1/3.$

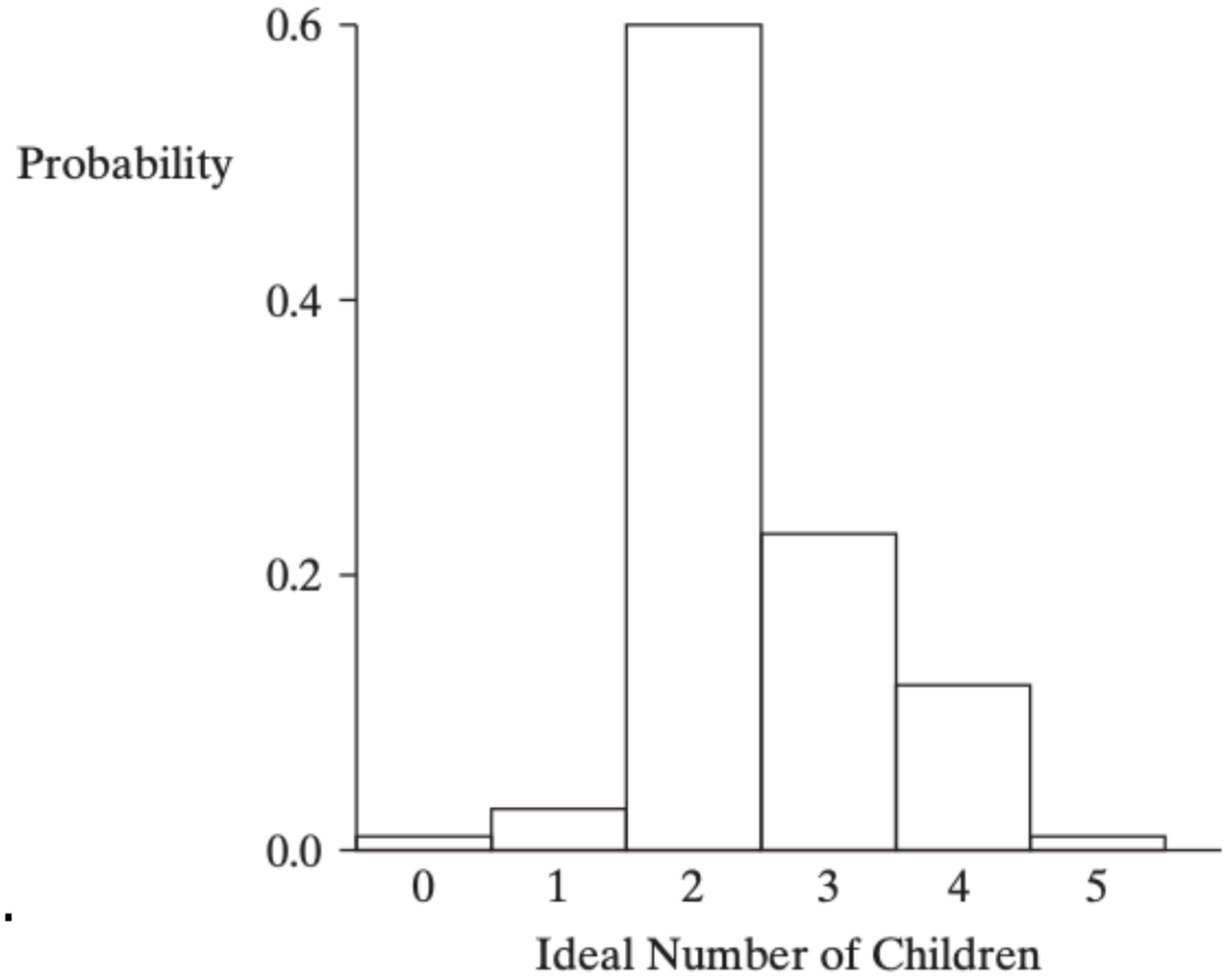
universität wien

Probability Distributions

- A **trial** is an action that generates outcomes of interest: for instance, flipping a coin a single time, sampling from a population.
- An **event** is an outcome, or set of outcomes that occurs with a certain probability. The set of all possible outcomes is the sample space Ω.
- A **probability distribution** lists the possible **outcomes** and their probabilities. By definition, all probabilities must add up to 1.

Probability Distributions

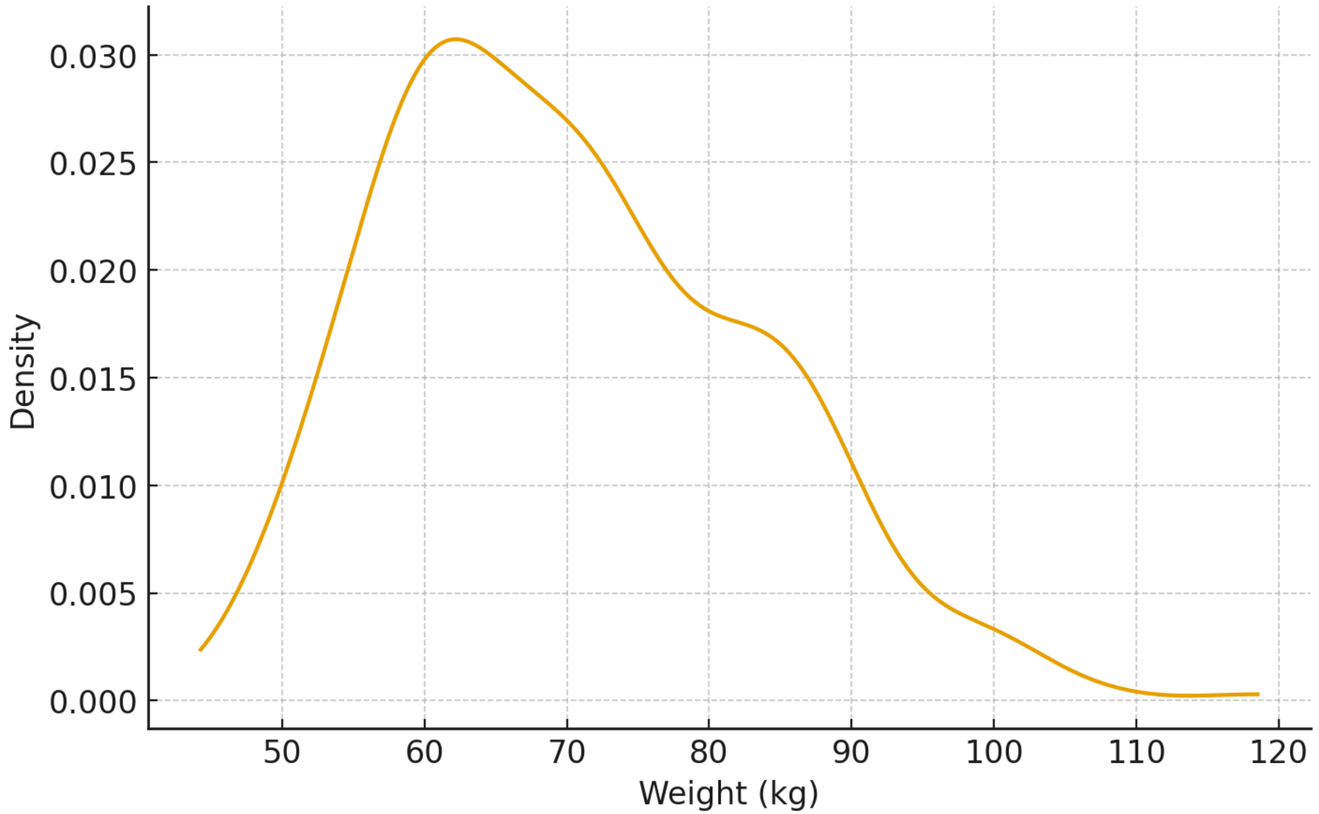
- **Discrete Probability Distributions**
  - Each event has non-zero chances of happening.
  - Represented by a **probability mass function**.
  - Special case: Bernoulli distribution, for binary variables.



universität wien

Probability Distributions

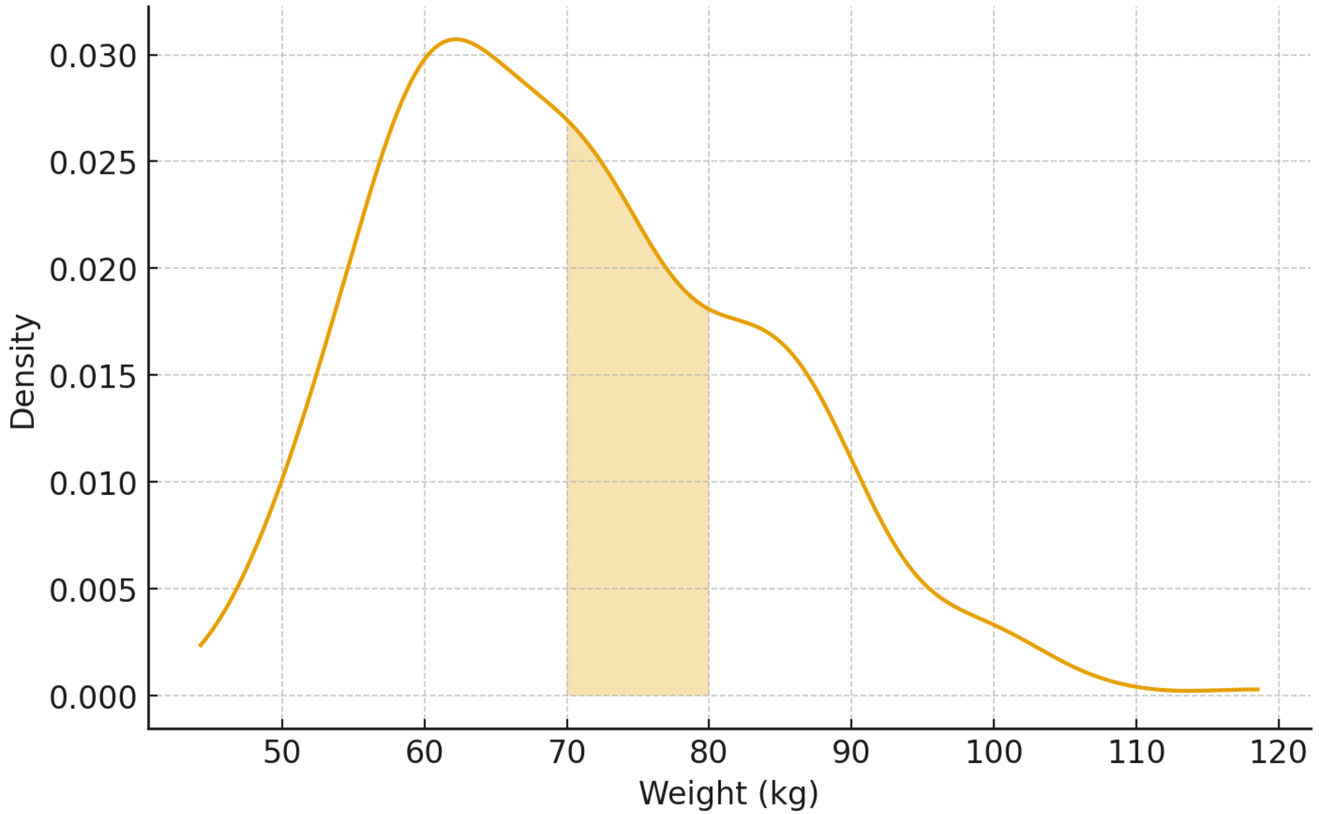
- **Continuous Probability Distributions**
  - Represented by a curve, known as **probability density function.** You can think of it as a histogram with infinitely small bins.
  - Each individual event may have (essentially) zero chances of happening. E.g. the probability that a patient's weight is exactly 75kg, as opposed to 75.00001kg or 74.99999kg is basically zero. But it's still "more likely" than 7.5kg.



universität wien

Probability Distributions

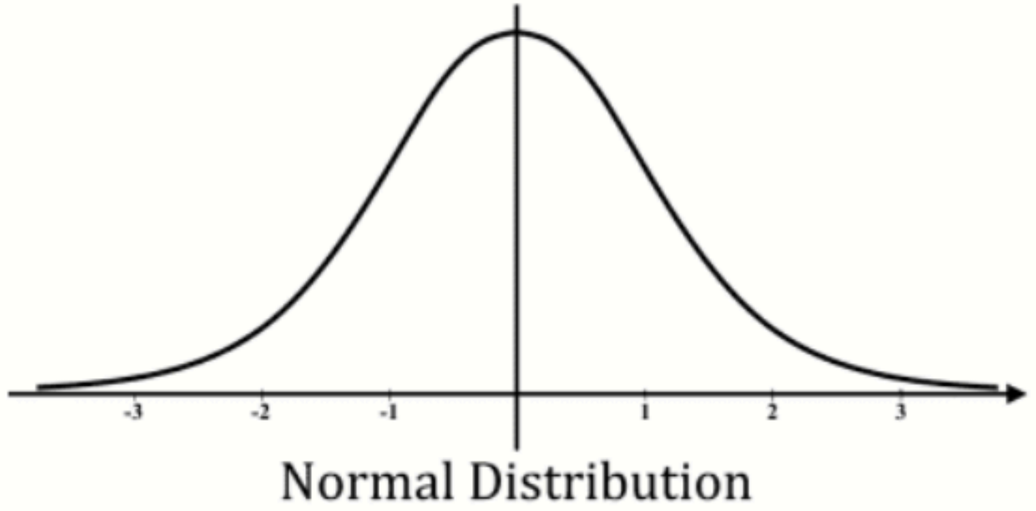
- **Continuous Probability Distributions**
  - One way of thinking about this: each individual value may have probability equal approximately to zero
  - But the **area under the curve** gives you the probability of obtaining outcomes in an interval, as a fraction of the total area (which sums up to 1).



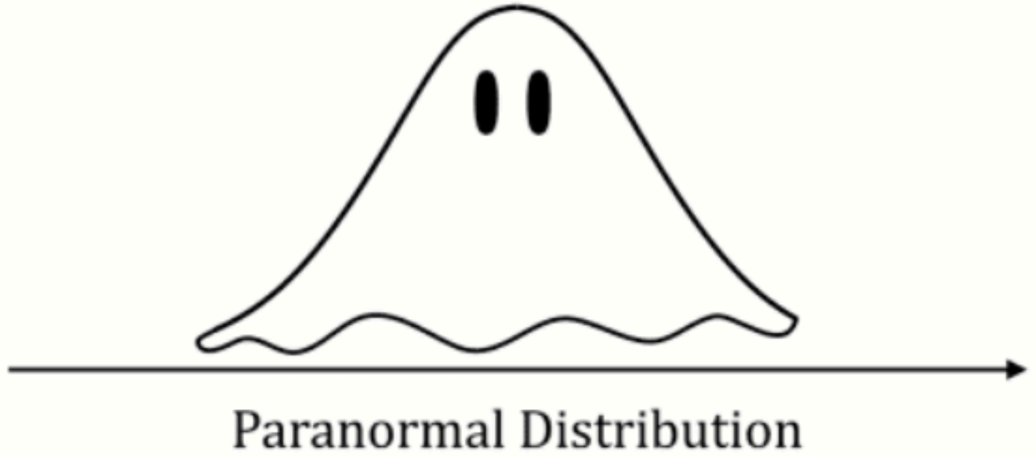
universität wien

Normal Distribution

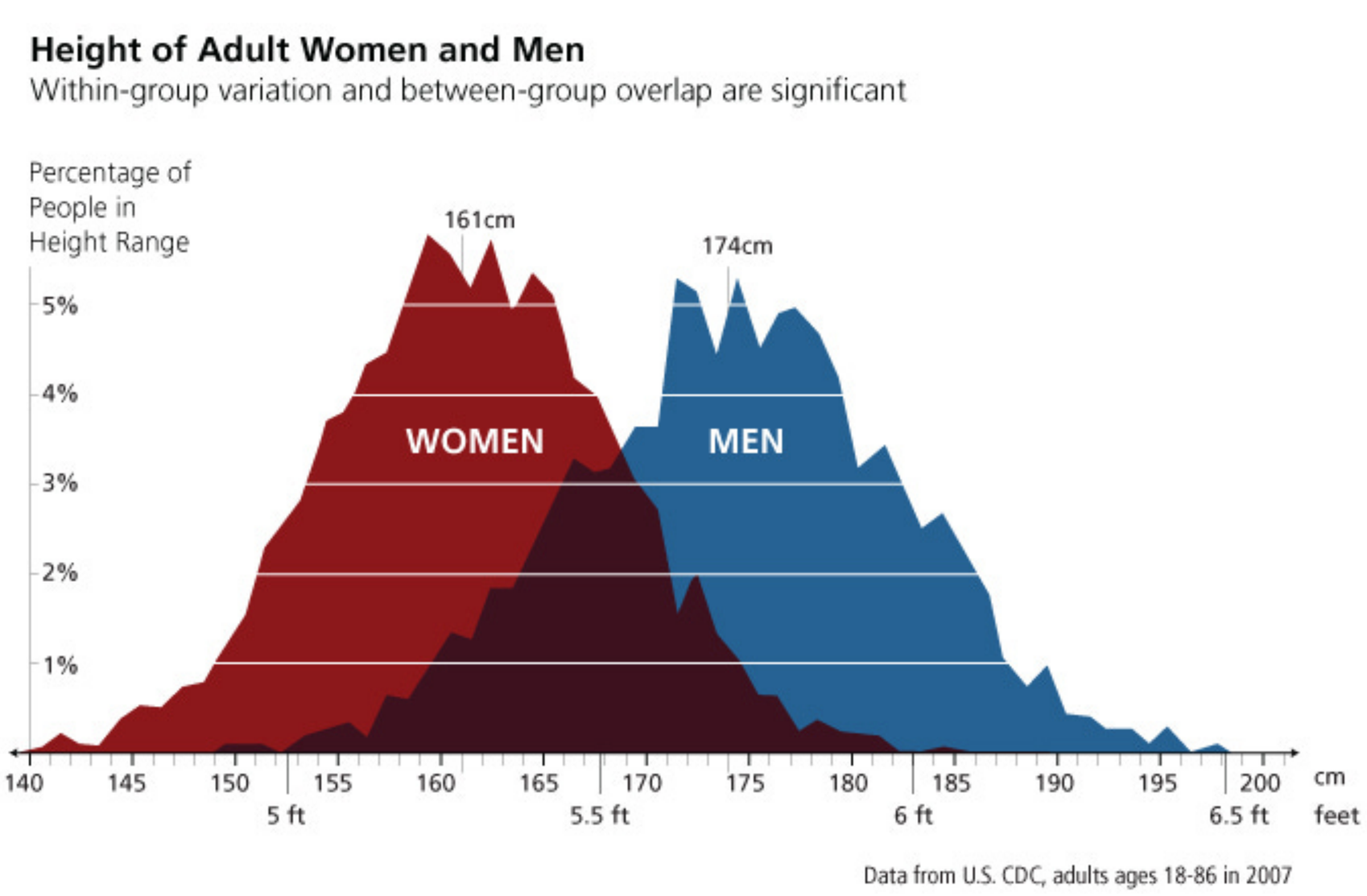
- The normal distribution represents a particularly common (and useful) type of probability density function.
- It has a **single peak**, equal to the mean and the median.
- The distribution is **symmetric**

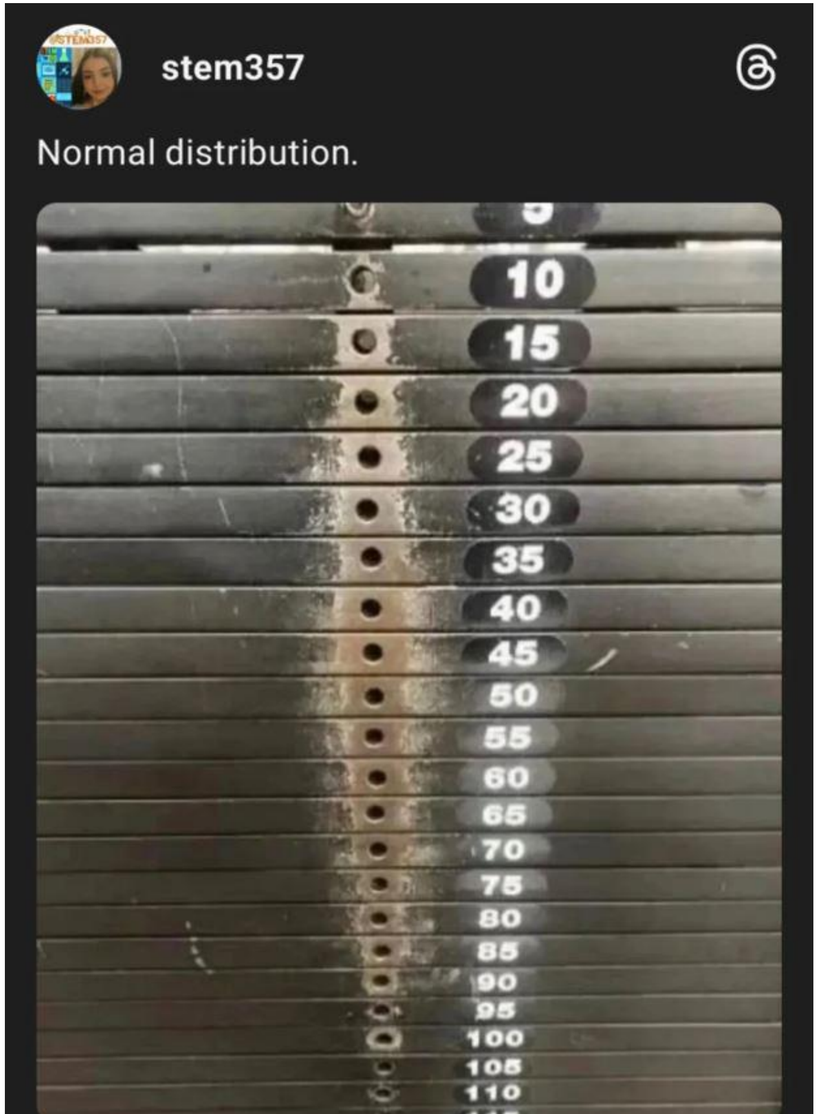


Normal Distribution


Paranormal Distribution

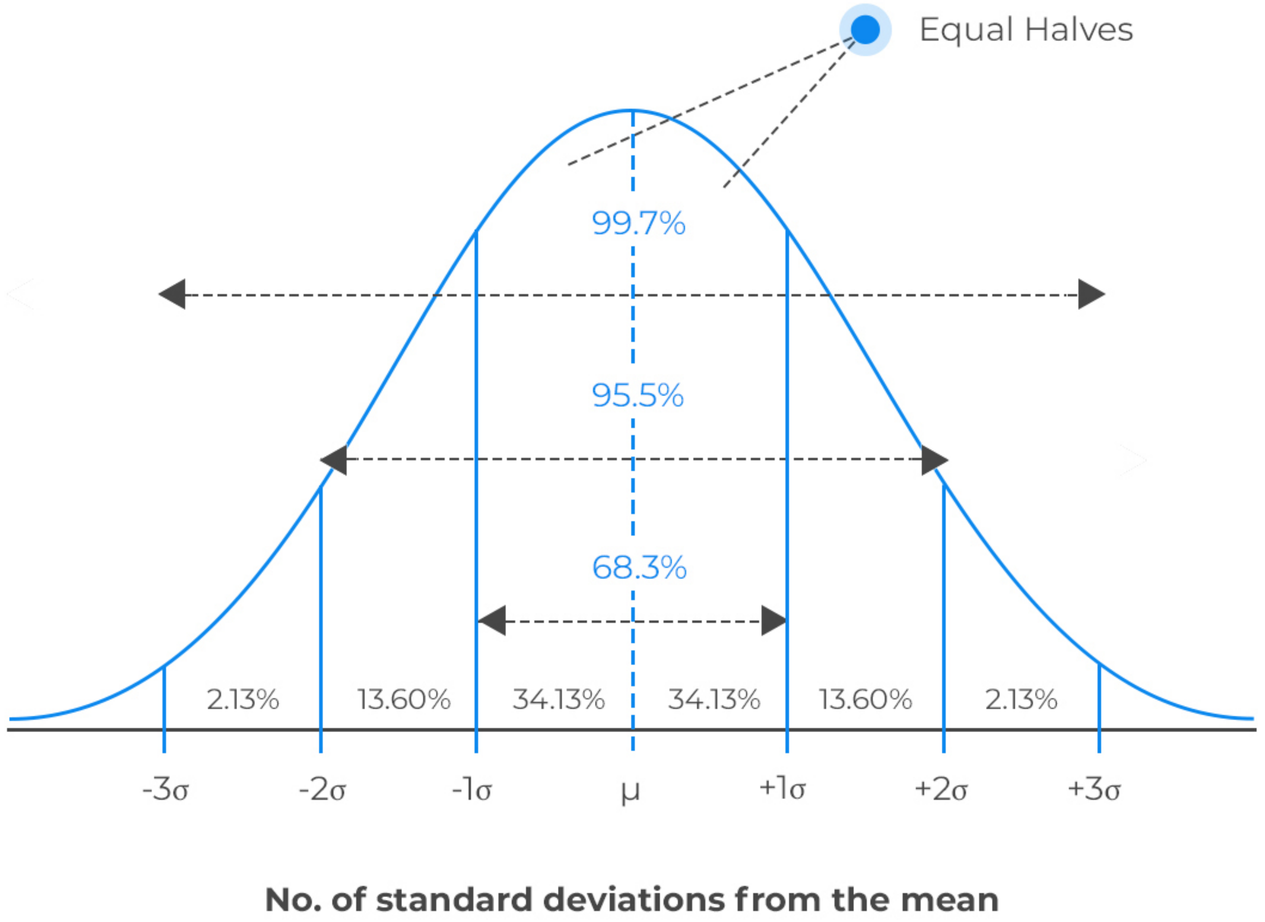
Normal Distribution





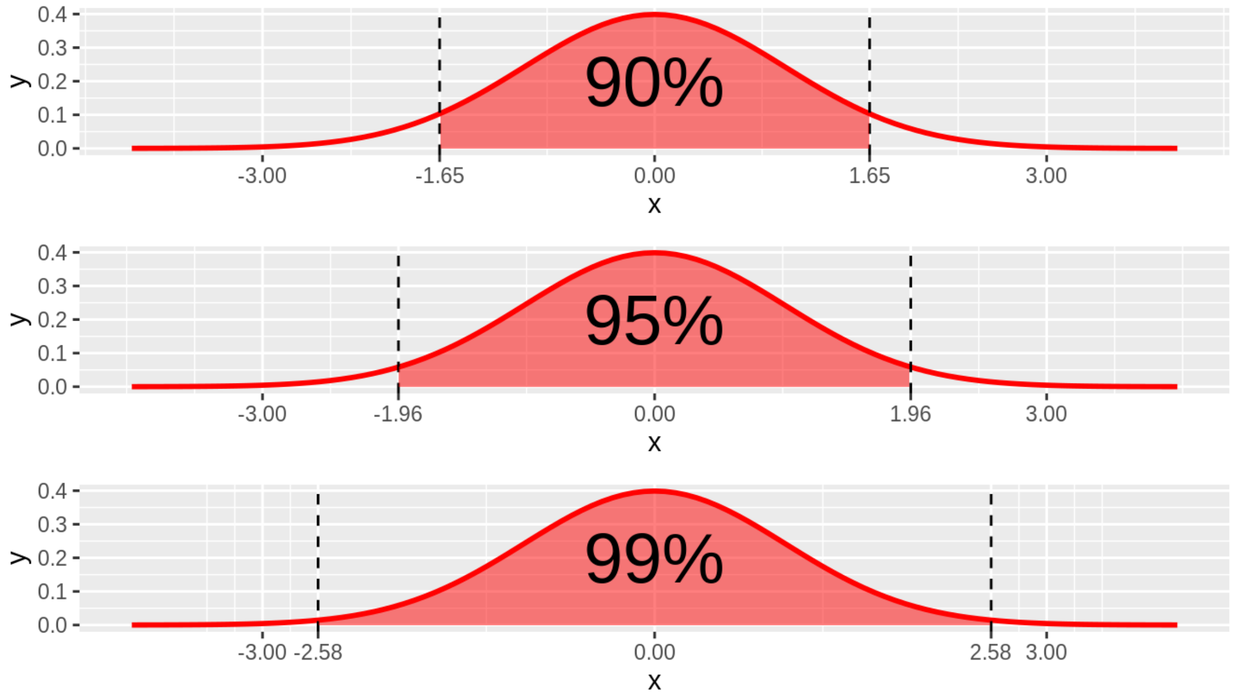
Normal Distribution

- In a normal distribution, standard deviations (σ or *s*) describe what percentage of the values fall within a certain distance of the mean (area under the curve).



Normal Distribution

- Stick a pin on this:
- 90% of the outcomes fall within 1.64σ of the mean.
- 95% of the outcomes fall within 1.96σ of the mean.
- 99% of the outcomes fall within 2.58σ of the mean.
- Values known as *z*-scores.



Sampling

- Selecting a smaller group of individuals or observations from a larger population so that by studying the smaller group, we can learn something about the whole population.
- Here we’re mostly interested in **random sampling**.

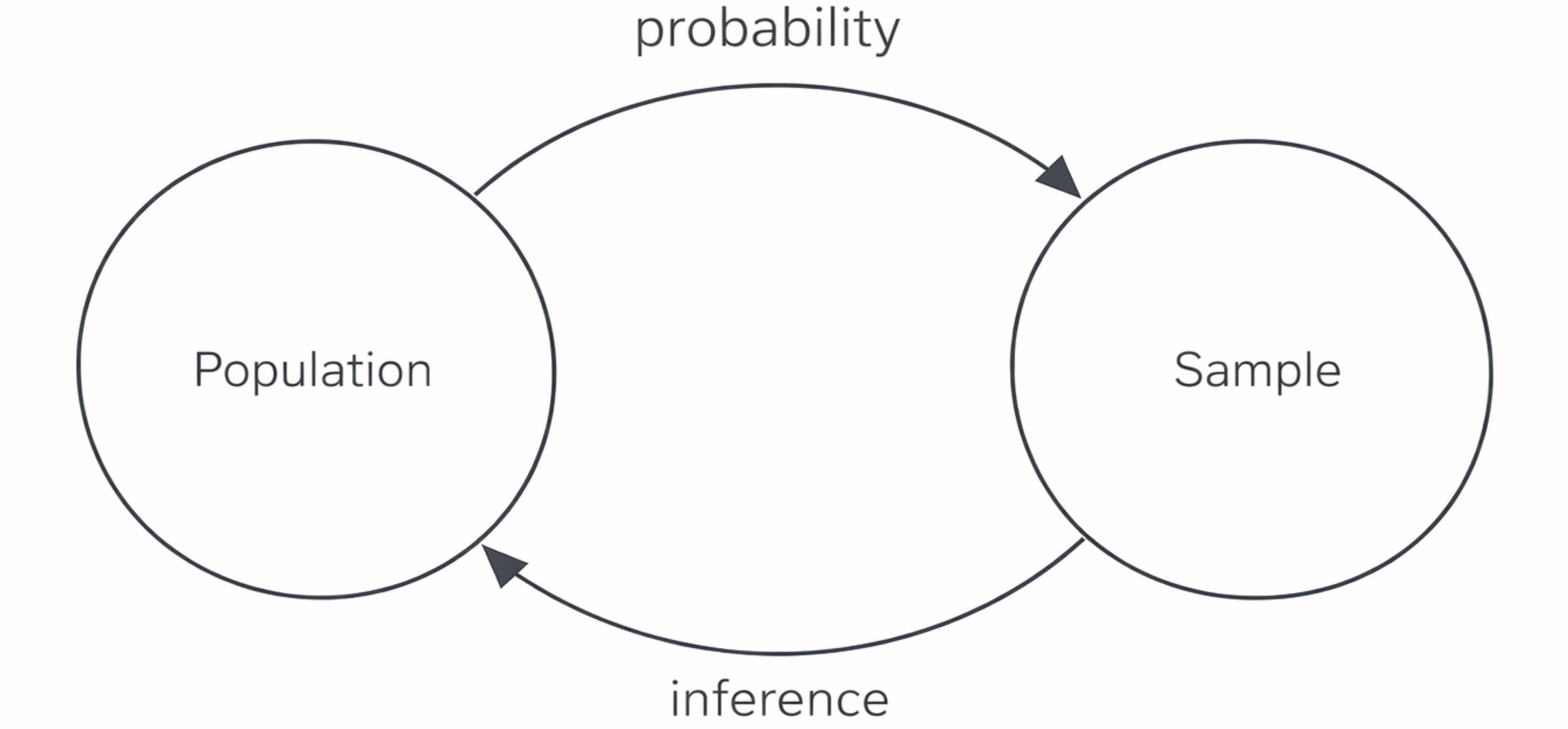
universität wien

Sampling

- Key terms:
  - **Population**: the universe of cases to which we want to generalise.
  - **Sample**: a subset of units from the population we actually observe.
  - **Parameter**: statistic (mean, standard deviation, median etc.) that describes a variable in the population (which we don't observe).
  - **Estimate**: statistic (mean, standard deviation, median etc.) that describes a variable in a sample (which we observe).

universität wien

Sampling and Probability



universität wien

Sampling and 'Errors'

- Suppose we knew (magically) that 36.45% of voters support the FPÖ.
- If we **randomly** sample 100 people who intend to vote in the next election, how many would be FPÖ voters?
- What if we **randomly** sampled a different 100 voters?
- What if we **randomly** sampled 1000 voters instead?
- What if we sampled 2000 people passing by the UniVie courtyard?
- ***Even if we take random samples***, there will be a discrepancy between the "estimate" (in this case, the sample proportion) and the "parameter" (in this case, the population proportion that we magically get to know).

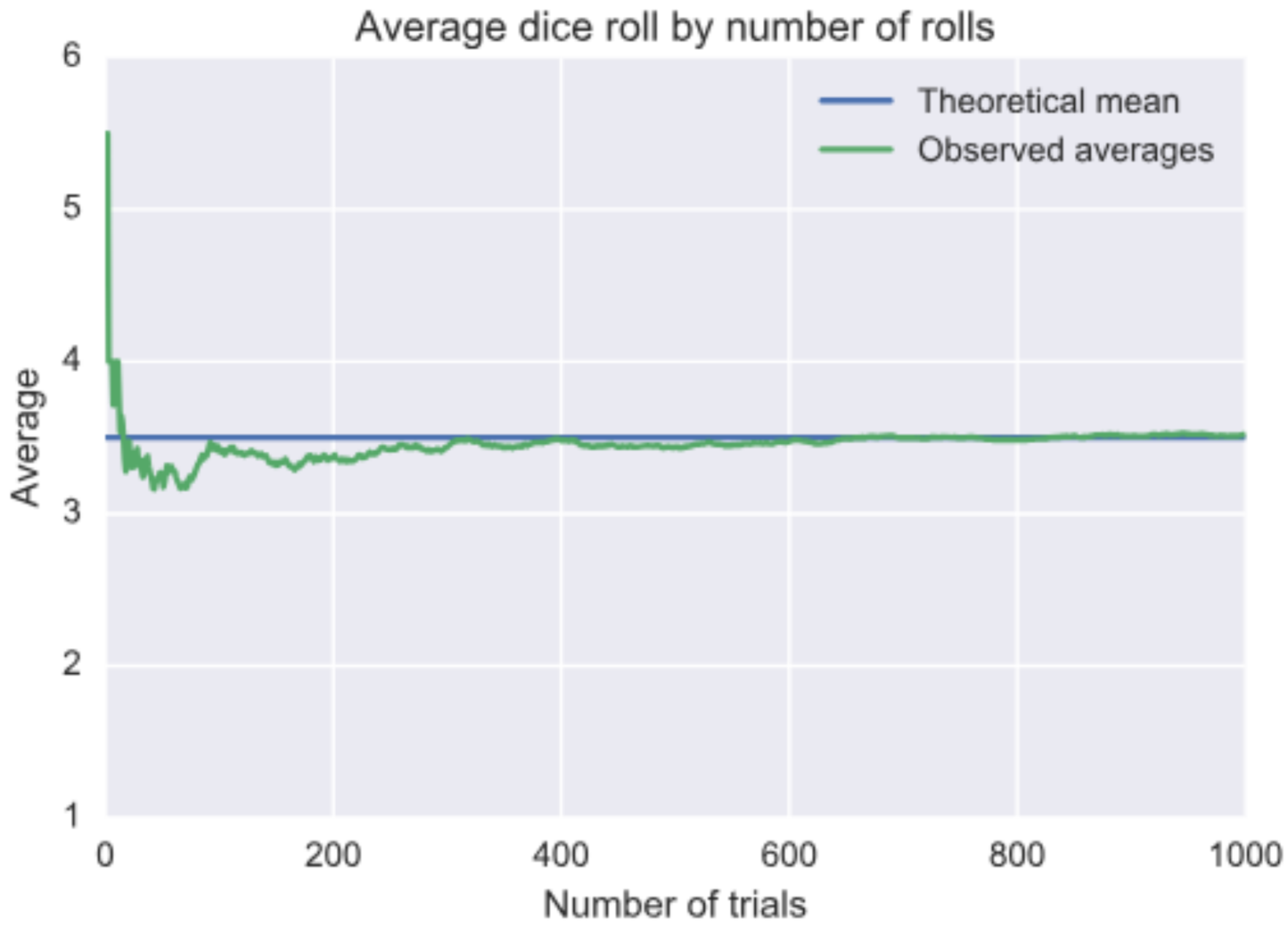
universität wien

The Law of Large Numbers

- We can derive from probability two very important facts about the relationship between **sample estimates** and **population parameters**:
- The **central limit theorem (CLT)**, which we will discuss shortly.
- The **law of large numbers**, which is an implication of the CLT. It states that **as the sample size increases**, the sample estimate can be expected to get closer and closer to the population parameter.
  - Only the sample size matters, independently of the population size.

universität wien

The Law of Large Numbers



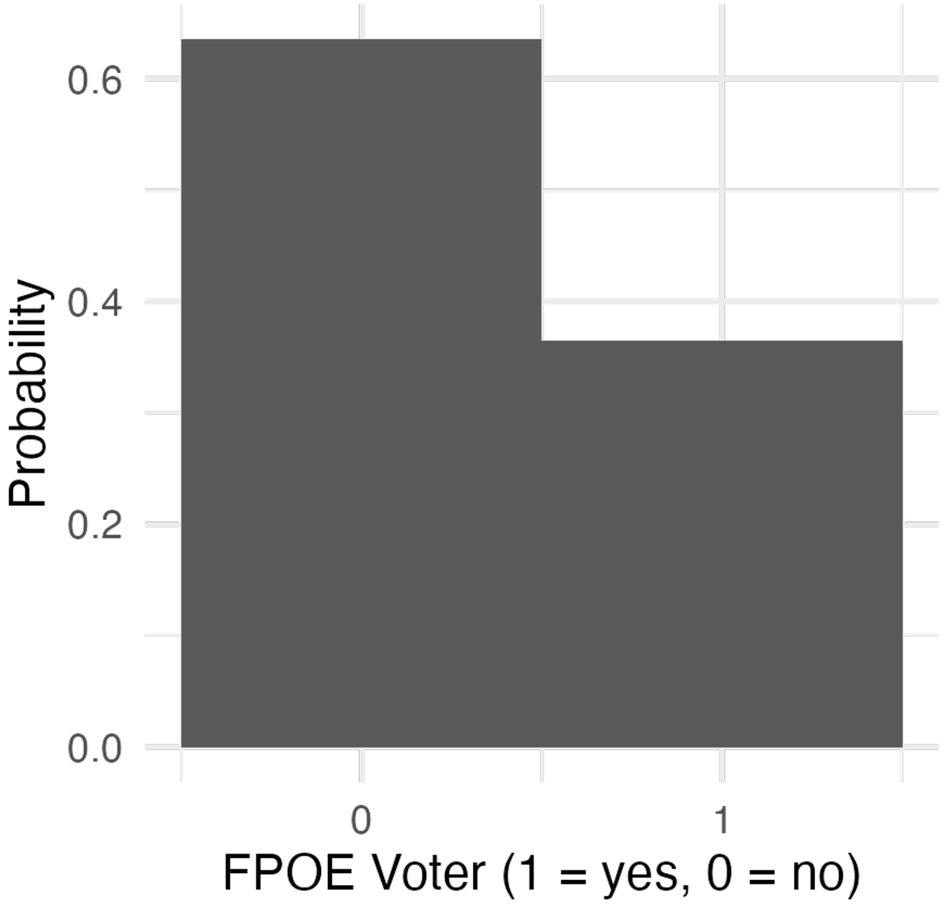
universität wien

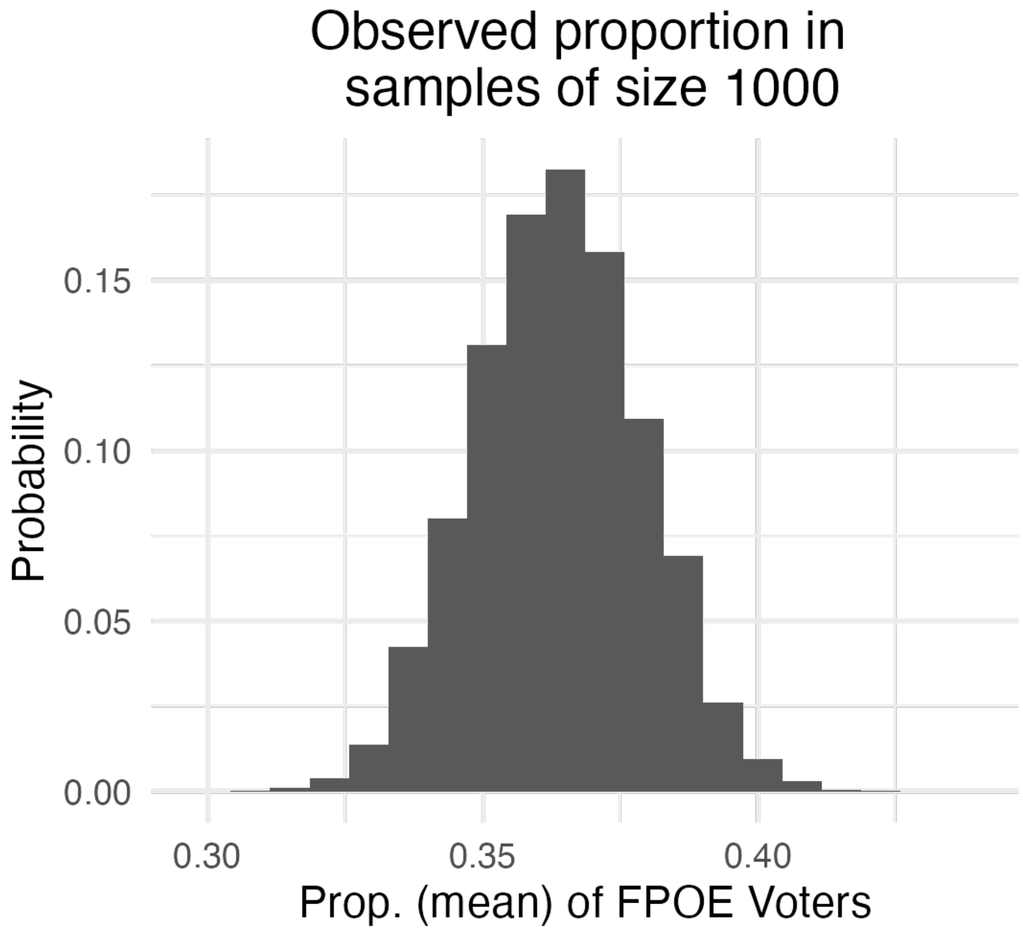
The Central Limit Theorem

- The central limit theorem tells us that:
  - If we take lots and lots of random samples, the distribution of **sample means** approximates a normal distribution centred at the population mean — **no matter what the original population looks like.**
  - Provided sufficiently large sample size (conventionally $N > 30$).
  - Applies also to estimates/parameters other than mean.

universität wien

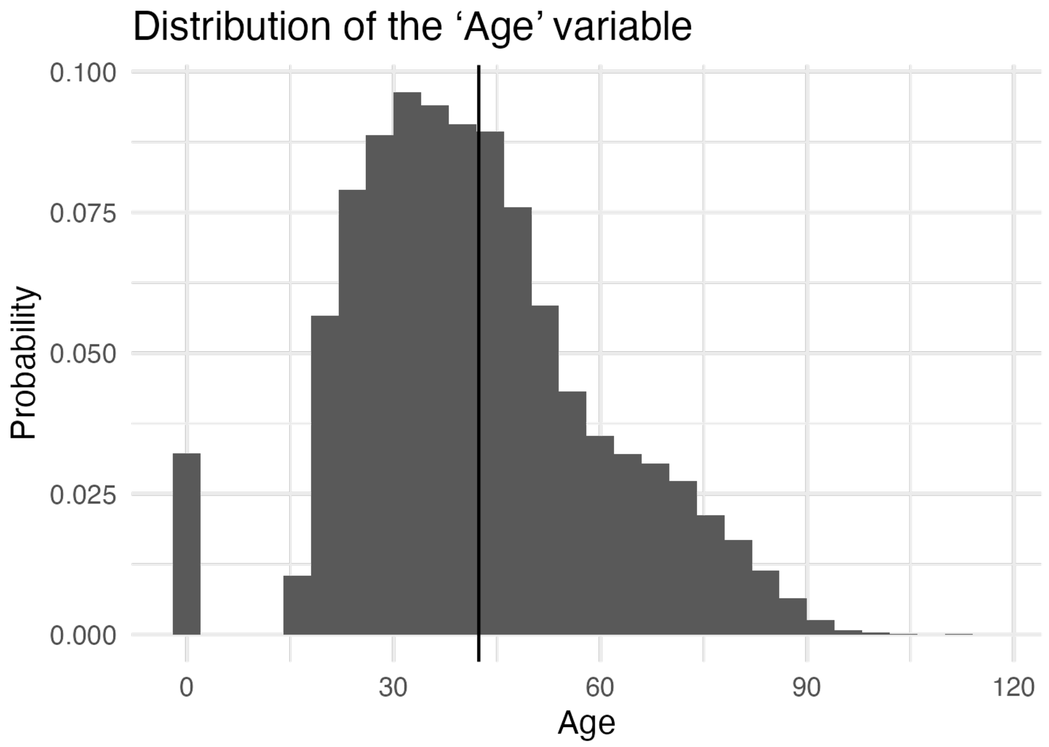
The Central Limit Theorem

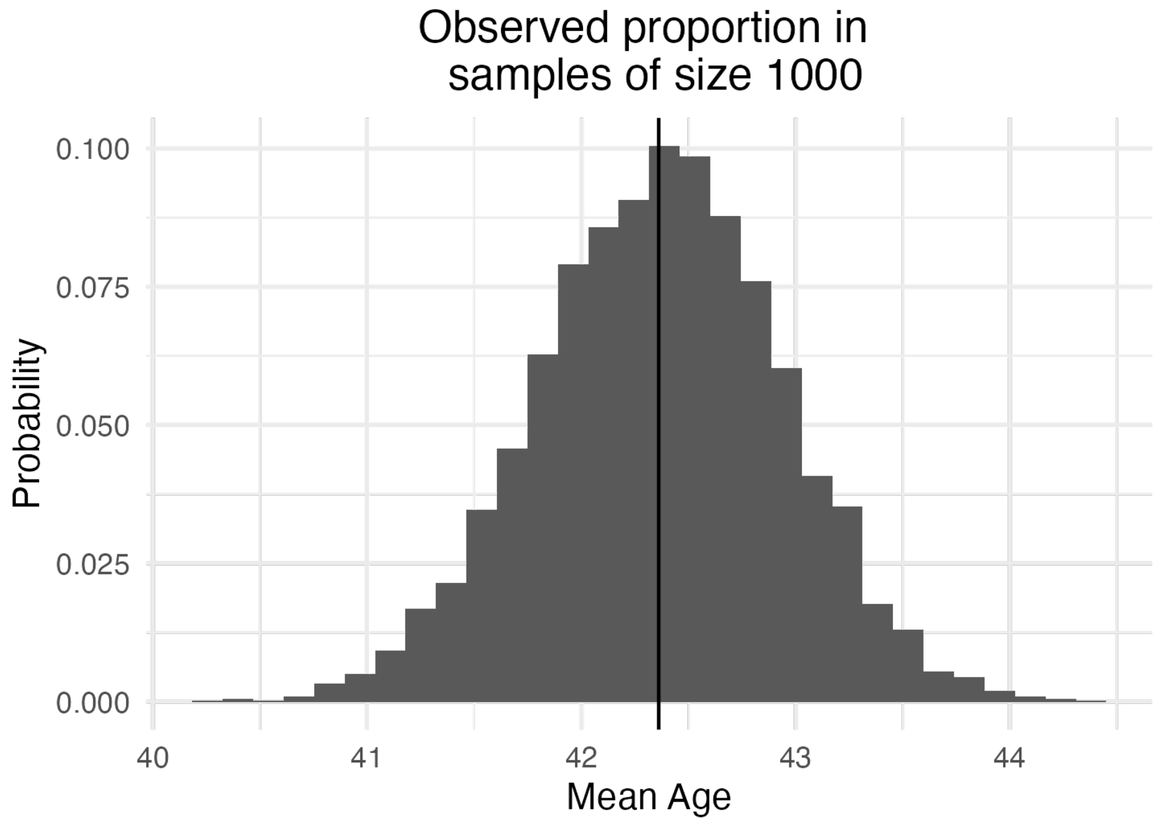




universität wien

The Central Limit Theorem

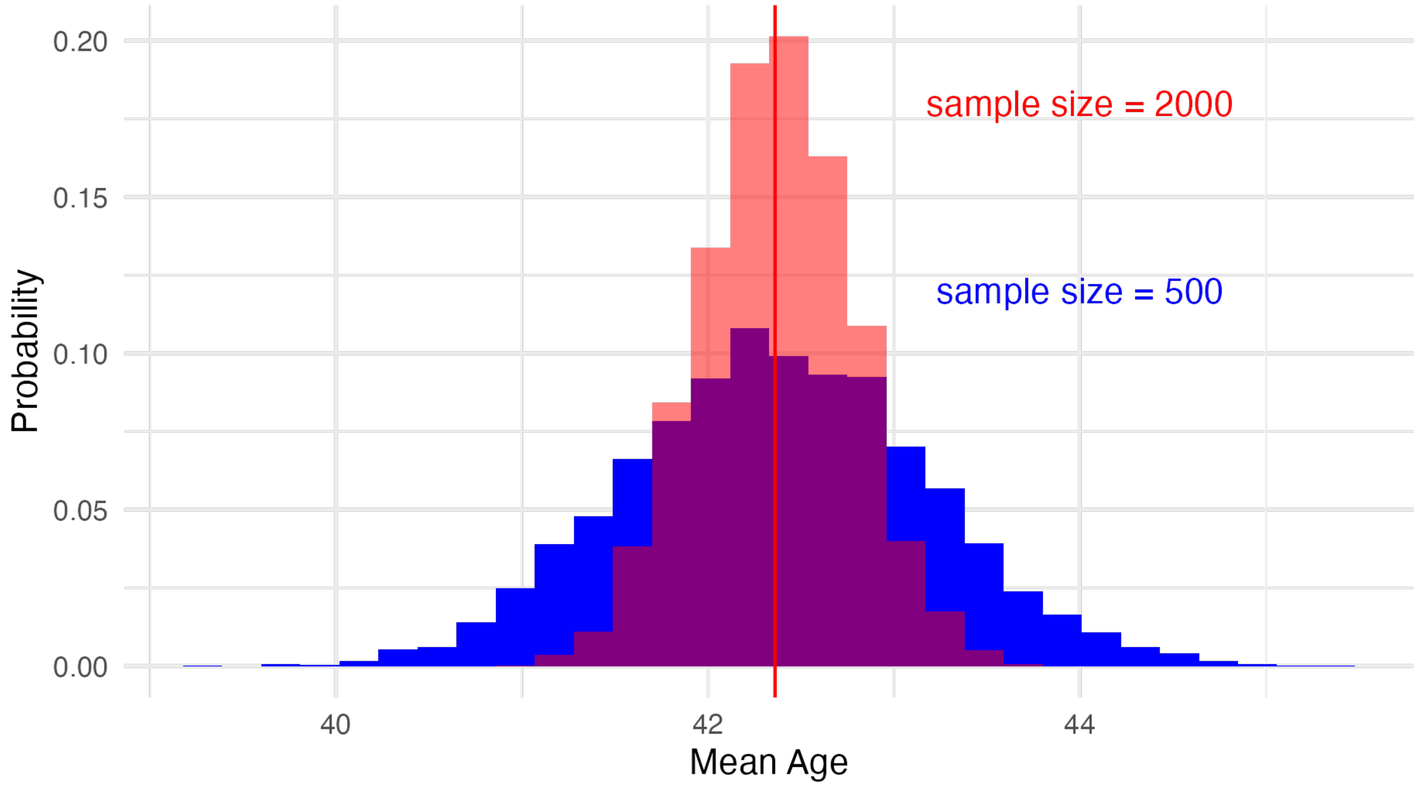




universität wien

The Central Limit Theorem

- The normal distribution will be flatter for smaller sample sizes, and steeper for bigger sample sizes.



universität wien

The Standard Error

- By how much is our sample estimate likely to be ‘wrong’?
- One way of quantifying it: the “variability” of its distribution across repeated sampling trials.
- That is, the standard deviation of the **sampling distribution**.
- This is known as **standard error**, and can be computed as…

$$\mathrm{SE}_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

universität wien

The Standard Error

- This is known as **standard error**, and can be computed as

$$\mathrm{SE}_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Where σ is the standard deviation of the population and n is the sample size.
- The problem: we normally don't observe the population standard deviation σ, because we only observe a sample, not the population.

universität wien

The Standard Error

- The solution: we use the **sample standard deviation** (s) instead.
  Thanks to the LLN, this approximates σ for large enough samples.
- Therefore…

$$\mathrm{SE}_{\bar{x}} \approx \frac{s}{\sqrt{n}}$$

- Which means that we can quantify the uncertainty of our estimate (the mean) $\bar{x}$ from quantities of a single sample (s and n), without going through the trouble of taking many, many samples from the population.

universität wien

The Standard Error

- Important to distinguish:
  - **Standard deviation** of a variable. Describes spread of all values.
  - **Standard error** of a sample estimate (mean, proportion). Describes the uncertainty of that specific value.
- Important fact we will use next time:
  - The standard error is the standard deviation of the distribution of estimates across repeated samples. For large enough sample sizes, that distribution is approximately **normal**.

universität wien

Summing Up…

- Probability is the language of uncertainty and randomness. We can think of variables as having **probability distributions**.
- The CLT theorem tells us that the probability distribution of an estimate (mean, proportion etc.) across many random samples is (1) **normal** and (2) centred at the population parameter.
- The standard error of an estimate is the standard deviation of this sampling distribution. We can approximate it from **a single sample**.
- And now, let's open RStudio…