Week 4: Hypothesis Testing

POP88162 Introduction to Quantitative Research Methods

Tom Paskhalis

Department of Political Science, Trinity College Dublin

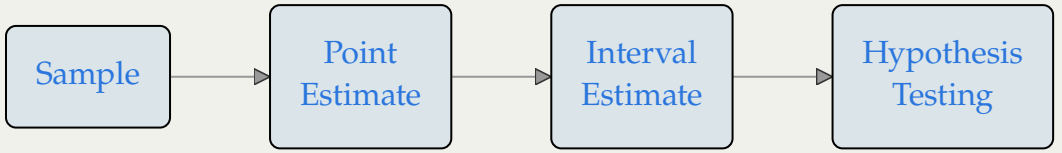
So Far

- The goal of collecting data is usually to calculate *statistics* which can be used to infer *parameters* of a population.
- Variables can be measure on different *scales*, which determine which statistics are applicable.
- *Measures of central tendency* describe a typical case in our data.
- *Measures of variability* show how far away from a typical case are other observations.
- For *discrete random variables* we can calculate point probability of getting specific values.
- For *continuous random variables* we can only estimate the probability of a value falling within some interval.

Topics for Today

- Sampling distributions
- Central Limit Theorem
- Confidence intervals
- Null and alternative hypotheses
- Significance testing

Today's Plan


Sample
Point Estimate
Interval Estimate
Hypothesis Testing

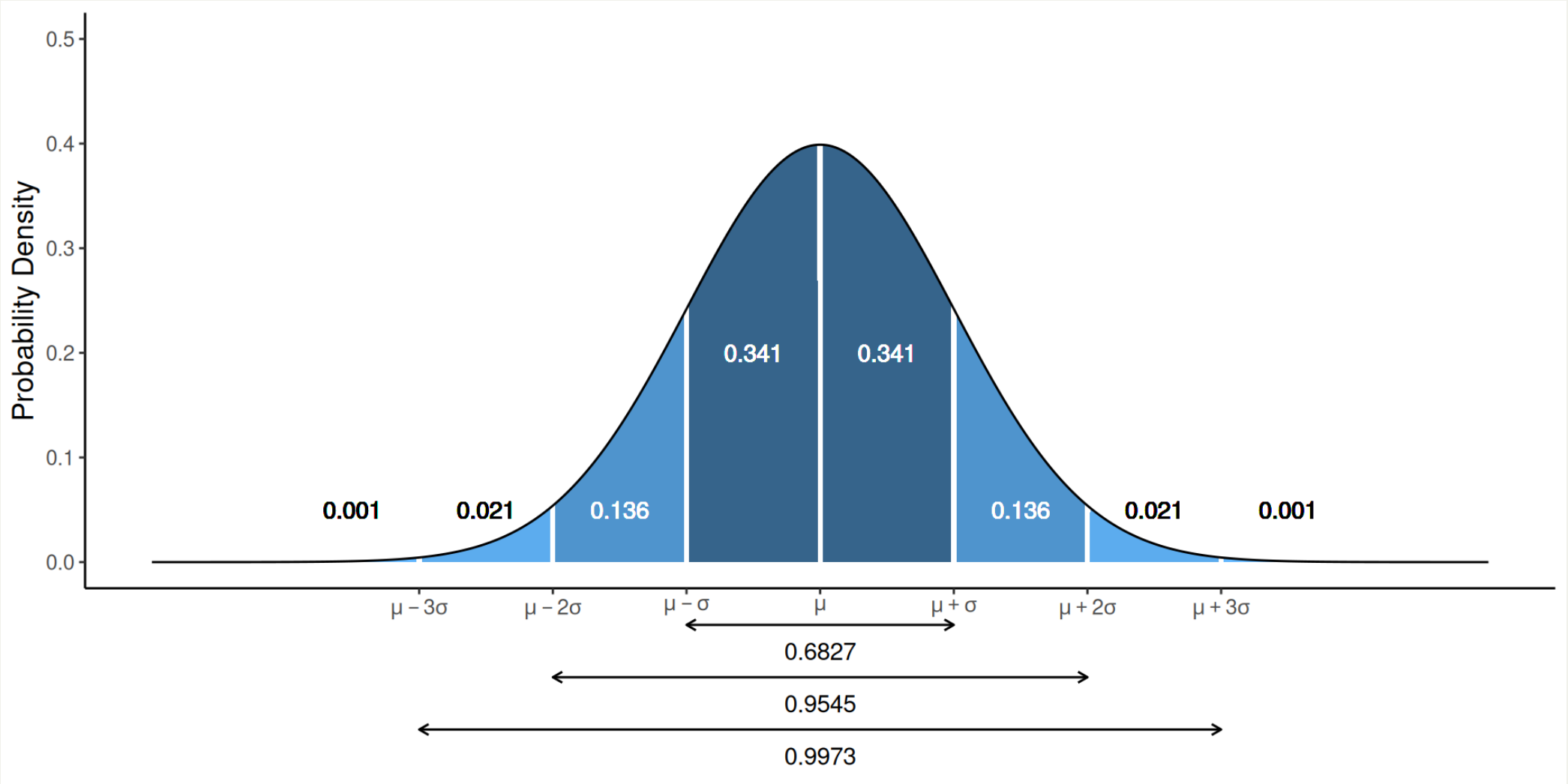
Review: Normal Probability Distribution

- The most important probability distribution.
- As it:
  - Approximates the distribution of many variables in the real world.
  - Is used a lot in inferential statistics.
- It is *symmetric, bell-shaped* and fully described by its **mean** μ and **variance** σ^2.
- Can also be called **normal distribution** for short.
- We can denote it as:

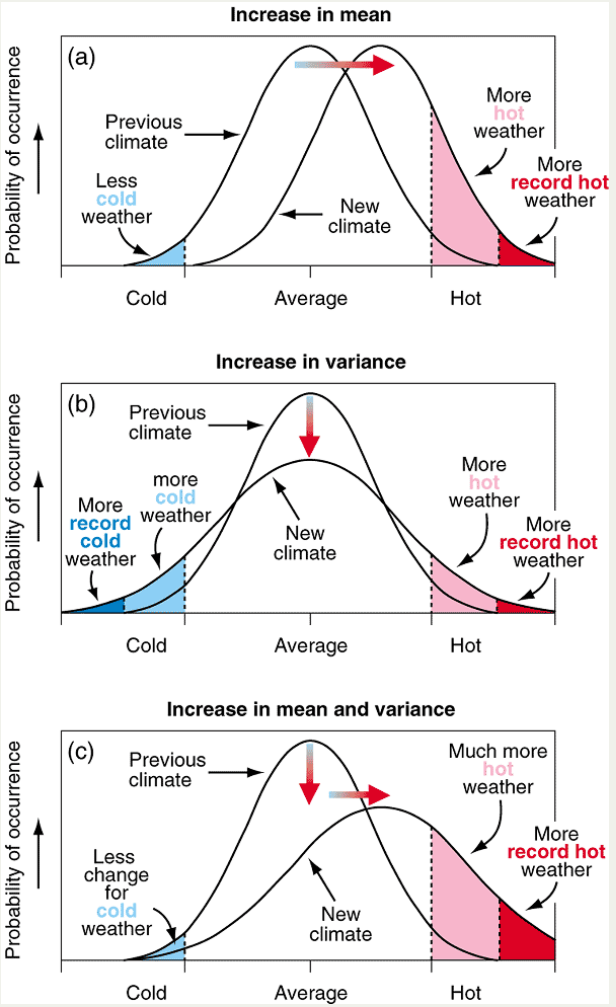
$$Y \sim N(\mu, \sigma^2)$$

"Y is distributed according to a normal distribution with mean μ and variance σ^2."

Review: Normal Probability Distribution



Example: Climate Change

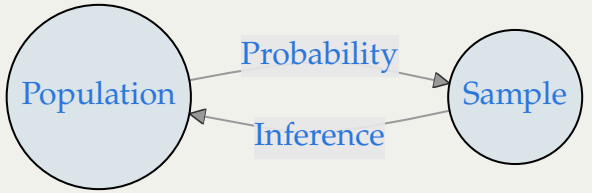


IPCC - Intergovernmental Panel on Climate Change

Sampling

From Sample to Population

- We know how to describe some variable in a sample.
- But we are not typically interested in learning about a sample.
- What we really want to know is:
  - Whether these characteristics of a variable or an association between variables are true *in the population*.



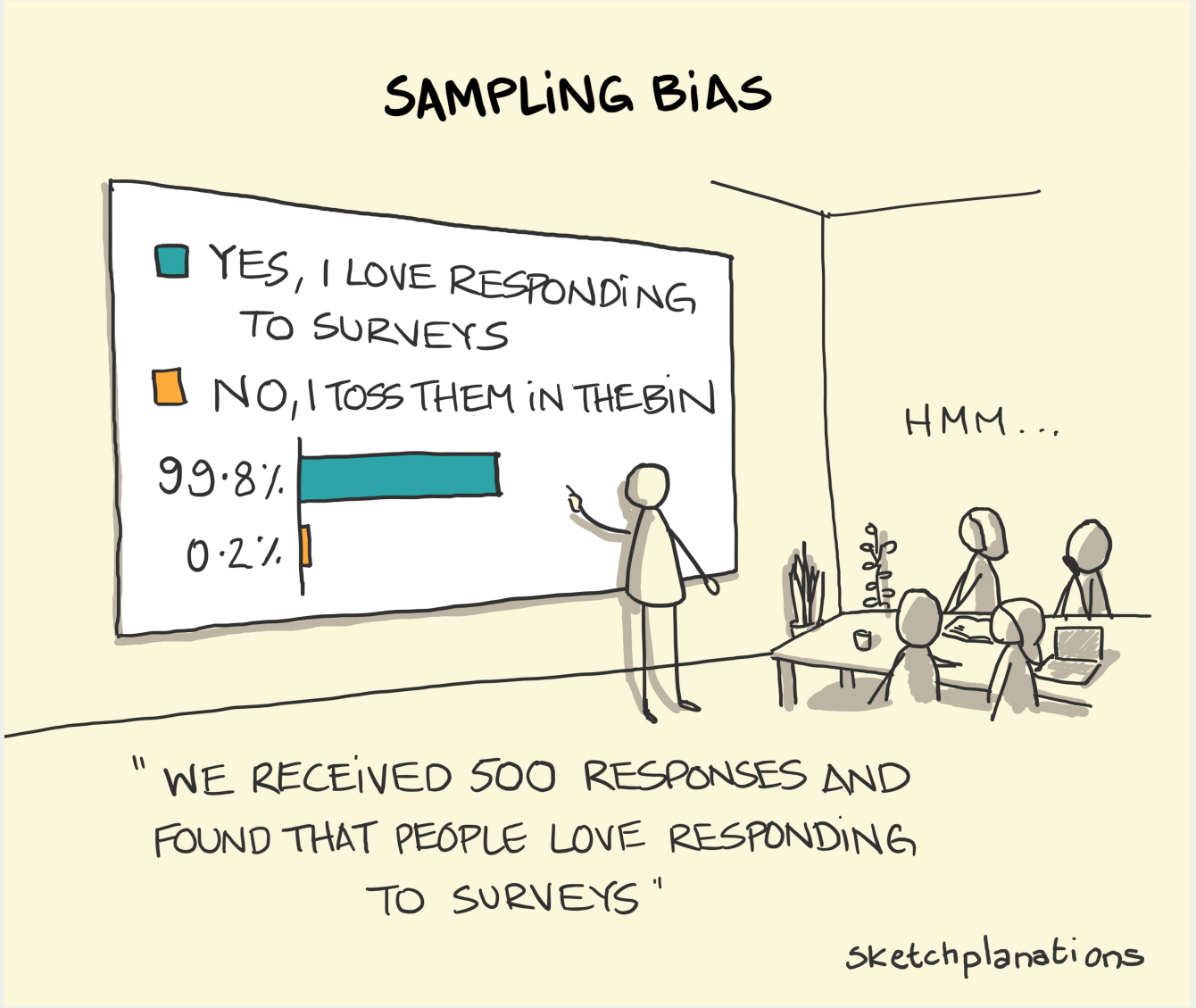
Survey Research

Wikipedia

Not Survey Research

X (Twitter)

Sampling Bias



Sketchplanations

Example: 2016 UK EU Referendum

- On 23 June 2016 51.89% of 🇬🇧 voters cast their ballots in favour of leaving the 🇪🇺 and 48.11% voted to remain.
- Between 14 April and 4 May 2016 British Election Study (BES) conducted Wave 7 of its internet panel.
- This study included 30, 895 respondents, of whom 28, 044 provided answer to the question: *"If there was a referendum on Britain's membership of the European Union, how do you think you would vote?"*
  - 14, 352 - Stay in the EU
  - 13, 692 - Leave the EU
- In other words, 51.2% of the respondents favoured remaining and 48.8% leaving.
- How certain can one be that in the population less than 50% of voters favour Leave?

Source

Fieldhouse et al. (2016), Hobolt (2016)

Point Estimation

- **Point estimate** provides a single 'best guess' about the population parameter.
- Examples of quantities of interest:
  - $\mu = E(Y)$, the population mean
  - $\sigma^2 = Var(Y)$, the population variance
  - $\mu_1 - \mu_0 = E[Y(1)] - E[Y(0)]$, the difference between two groups
- In 🇬🇧 🇪🇺 referendum example we are interested in the proportion of voters who support Leave.

Estimating Parameters

- How do we estimate the population parameter of interest?
- Using **estimators**.
- There are many different possible esimators. E.g.:
  - $\bar{Y}$, the sample proportion of voters supporting Leave
  - Y_1, just use the first observation
  - 0.5, always guess 50% support
- Note that sample proportion is the same as the sample mean:

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{\text{\# of respondents supporting Leave}}{n}$$

- Sample proportion is on average equal to the population proportion (unbiased estimator).

Inference for UK EU Referendum

- The total number of eligible voters in the 🇬🇧 was 46.5 million.
- The study collects responses from 28, 044 people.
- How sure can one be in the accuracy of the results in this sample?
- Let's outline the statistical way of thinking about it:
  - Suppose in the population the true percentage of pro-Leave voters is 50%.
  - How likely we are to get a point estimate of 48.8% in a sample?
  - Note that we can formulate this question either relative to pro-Leave or pro-Remain voters.
  - Here we will denote pro-Leave voters with 1 and pro-Remain with 0.

Simulating Sample

- Let's start by creating our population and then drawing one sample from it.
- We will simulate the entire process in R.

```r
voters <- c(rep(1, 46500000/2), rep(0, 46500000/2)) # Create population
```

```r
voters <- sample(voters, length(voters)) # Shuffle voters
```

```r
sample_voters <- sample(voters, 28044) # Draw a random sample of 28,044
```

```r
prop.table(table(sample_voters)) # Calculate the proportions of Leave/Remain
```

```
sample_voters
        0         1
0.5072386 0.4927614
```

```r
prop.table(table(sample_voters))[2] # Select the proportion of pro-Leave voters
```

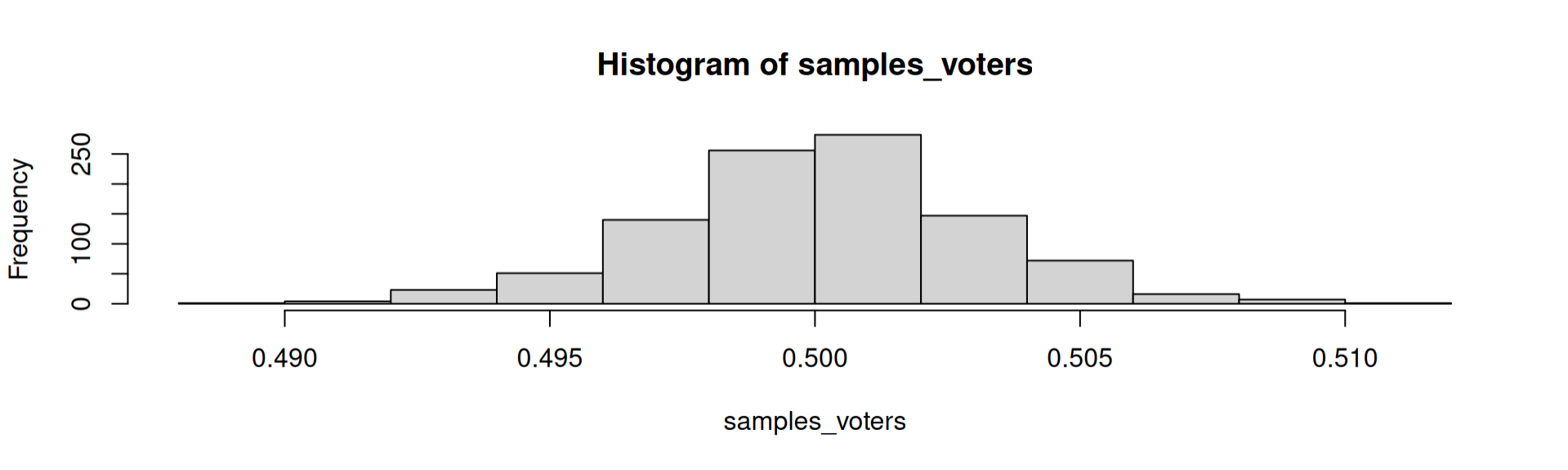
```
        1
0.4927614
```

Simulating Sample Continued

- Now let's repeat drawing a sample 1,000 times.

```
# Draw a sample of 28,044 from the population of UK voters 1000 times
samples_voters <- sapply(1:1000, function(x) prop.table(table(sample(voters, 28044)))[2])
```

```
hist(samples_voters)
```



- Nearly all of the simulated sample proportions fall between 0.49 and 0.51.
- How unusual would it be to observe that 48.8% of voters in a sample favour Leave?

Sampling Distribution

- **Sample statistic** (e.g. sample mean) is itself a *random variable*.
- Each sample gives its own sample mean.
- **Sampling distribution** is a probability distribution that assigns probabilities to values that a statistic can take.
- E.g. sampling distribution of a sample mean, sampling distribution of a sample proportion, etc.
- Sampling distribution helps us predict how close our sample statistic is to the population parameter we estimate.
- Each sample statistic has a sampling distribution.

Sampling Distribution of Sample Mean

- Sample mean $\bar{Y}$ is a random variable.
- For *random samples* it varies around the population mean μ_Y.
- The mean of the sampling distribution of $\bar{Y}$ equals μ_Y.
- **Standard error** is the standard deviation of the sampling distribution of $\bar{Y}$

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

- Sampling distribution of a sample mean is a normal distribution with mean μ and standard error $\sigma_{\bar{Y}}$.
- Note that the larger the sample size n, the smaller is the standard error.
- In other words, our estimate of population mean gets more precise.

Central Limit Theorem

- For random sampling with large sample sizes, the sampling distribution of the sample mean is approximately a normal distribution.
- **Central Limit Theorem** (CLT) applies no matter what the shape of the population distribution is.
- How large the sample size n must be depends on how skewed or irregular the population distribution is.
- If the population distribution is bell-shaped than the sampling distribution is bell-shaped for all sample sizes.
- CLT can be proved mathematically, but we will verify it by looking at an illustration of a simulation.

Central Limit Theorem

Confidence Intervals

Interval estimates

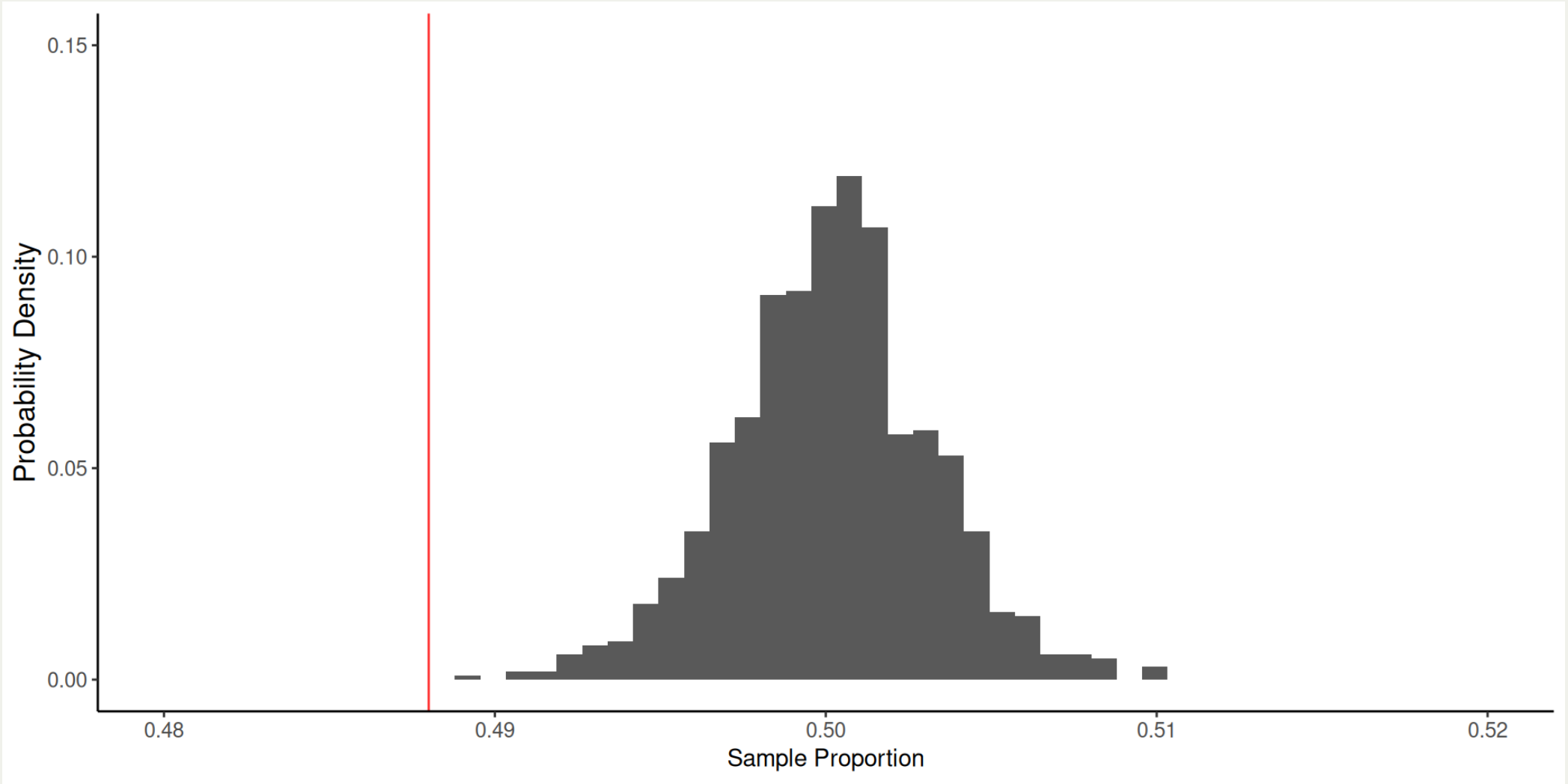
- A **point estimate** is our best guess about the population parameter.
- However, we want to be able to assess how good (accurate) our point estimate is.
- We know that a sampling distribution for a large enough sample size is approximately normal.
- And we know where the probability mass of a normal distribution lies.
- Hence, we can calculate an **interval estimate** around our point estimate.

Confidence Interval

- **Confidence interval** is an interval around the point estimate where the population parameter is believed to fall.
- Probability that our estimator produces an interval that contains the parameter is called **confidence level**.
- This number is chosen to be close to 1: 0.95, 0.99, 0.999
- The form of the confidence interval is:

$$\text{CI} = \text{Point estimate} \pm \text{Margin of error}$$

Example: Sampling Distribution for UK EU Refendum Poll



Confidence Interval for Proportions

- Numerically, *confidence interval* is point estimate ± margin of error.
- Margin of error is some multiple (a z-score) of a standard error.
- For 95% confidence level margin of error is $\pm 1.96\sigma$
- For proportions:

$$\sigma_{\hat{\pi}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- Where π is a population proportion and $\hat{\pi}$ (pi-hat) is our estimate of it.
- For variables where we code category of interest as 1 and 0 otherwise, $\hat{\pi}$ is just sample mean.

Example: Confidence Interval for UK EU Referendum

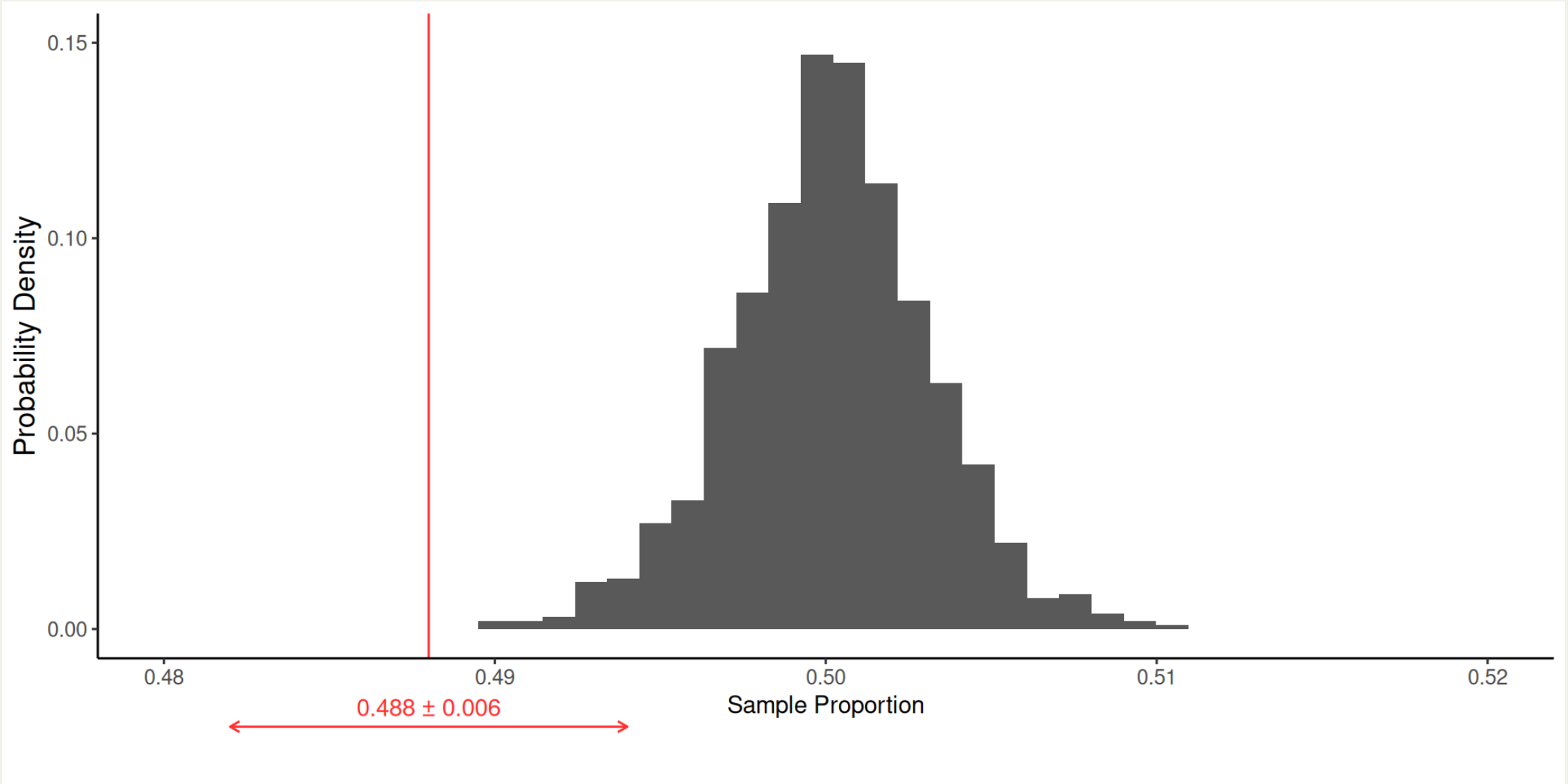
- In the BES survey the sample proportion of Leave voters was 0.488 ($\hat{\pi} = 0.488$).
- 95% confidence interval would, thus, be: $\hat{\pi} \pm 1.96\sigma_{\hat{\pi}}$
- And $\sigma_{\hat{\pi}} = \sqrt{\pi(1-\pi)/n}$
- But as we don't actually know population parameter π, we substitute it with our sample estimate $\hat{\pi}$:

$$se_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = \sqrt{\frac{0.488(0.512)}{28044}} = 0.003$$

- Then, a 95% confidence interval is:

$$\hat{\pi} \pm 1.96se_{\hat{\pi}} = 0.488 \pm 1.96(0.003) = 0.488 \pm 0.006 \text{ or } [0.482, 0.494]$$

Example: Confidence Interval for UK EU Referendum



Hypothesis Testing

Hypothesis Testing

- This part is critical to this module - you will be expected to perform at least one hypothesis test for your research paper.
- Using theory, you will be formulating a hypothesis of the form, $X \rightarrow Y$, or the variation in Y can be explained by the variation in X.
- But how can we test that using data?
- We use what is know in statistics as **classical hypothesis testing**.
- This approach is based on **proof by contradiction**.
- We are formulating two hypothesis (**null hypothesis** and **alternative hypothesis**) and using the collected data to calculate statistics and reject the null hypothesis.

The Lady Tasting Tea

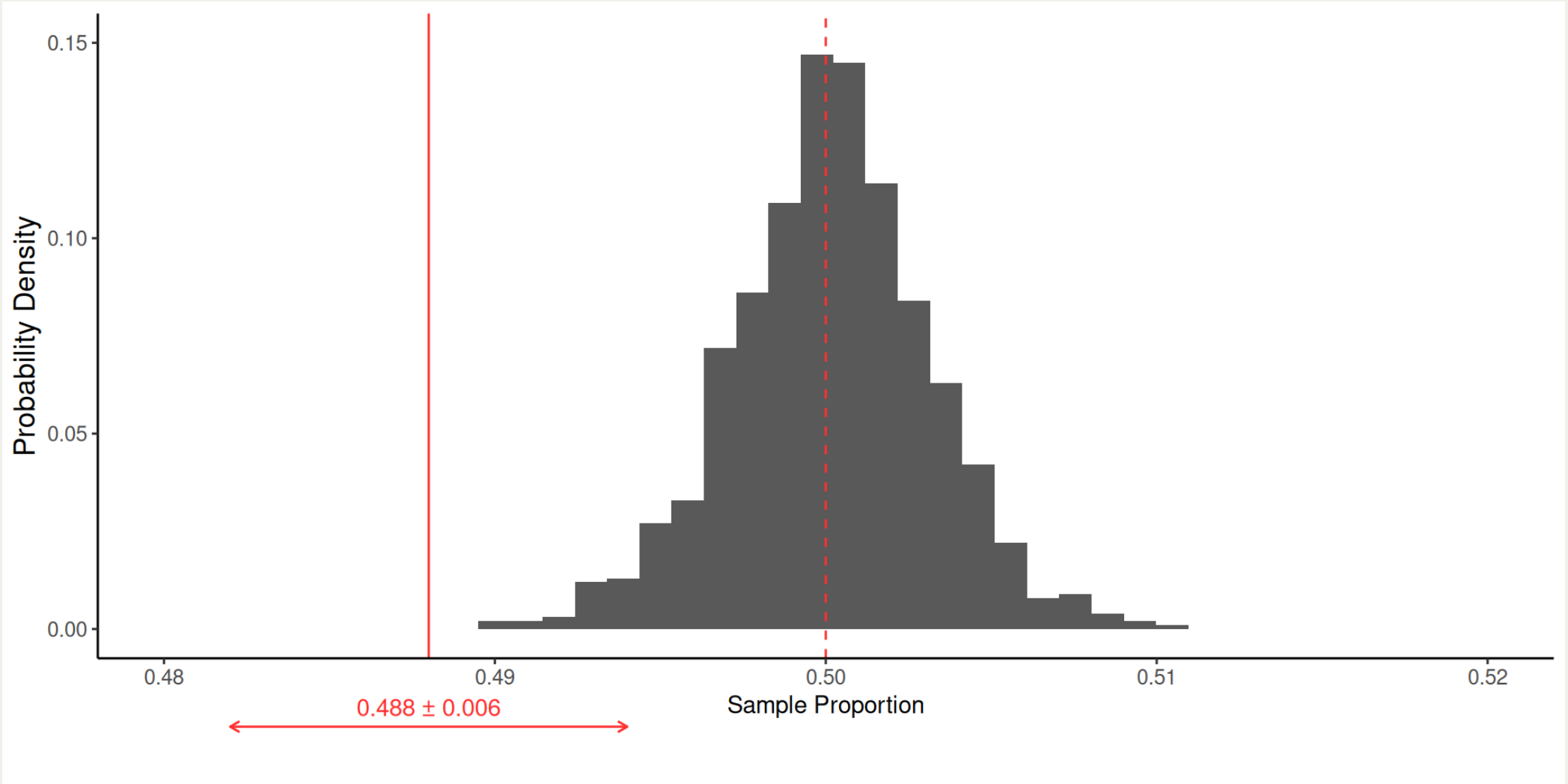
- One of the scientists at an early 1920s agricultural station in England, Dr Muriel Bristol, claims to be able to distinguish whether the milk or the tea had been poured into the cup first.
- To test this claim (*hypothesis*) dubious colleagues set up an experiment:
  - They arrange 8 ☕ (4 of each type) in *random* order
  - Dr Bristol correctly identified all 4 ☕ into which the milk was poured first
- How much evidence is this for Dr Bristol's claim?
- Chances of guessing all 4 correctly are $\frac{1}{70} \approx 0.014$ or 1.4%
- This looks implausible…

> ⓘ **Source**
>
> Fisher (1935)

Example: Hypotheses in UK EU Referendum

- We are interested in finding out whether the true population proportion of pro-Leave voters is different from 0.5 (null hypothesis).
- We can formulate several different alternative hypothesis:
  - population proportion **is not** 0.5: $H_a \neq 0.5$ (two-sided)
  - population proportion **is smaller** than 0.5: $H_a < 0.5$ (one-sided)
  - population proportion **is larger** than 0.5: $H_a > 0.5$ (one-sided)
- Null hypothesis and alternative hypothesis are formulated at the same time.
- This determines the type of a statistical test.
- How likely are to be observe the value in our sample data given the null hypothesis is true?

Example: Hypothesis Testing in UK EU Referendum



Hypothesis

- **Hypothesis** is a statement about the population.
- E.g. UK voters are equally likely to support Leave or Remain.
- We formulate two hypothesis:
  - H_0 - null hypothesis (e.g. no difference, no association)
  - H_a - alternative hypothesis (there is a difference/association)
- Alternative hypotheses can be *one-sided* or *two-sided*:
  - **one-sided**: the direction of difference/association is included
  - **two-sided**: no direction of difference/association is hypothesised

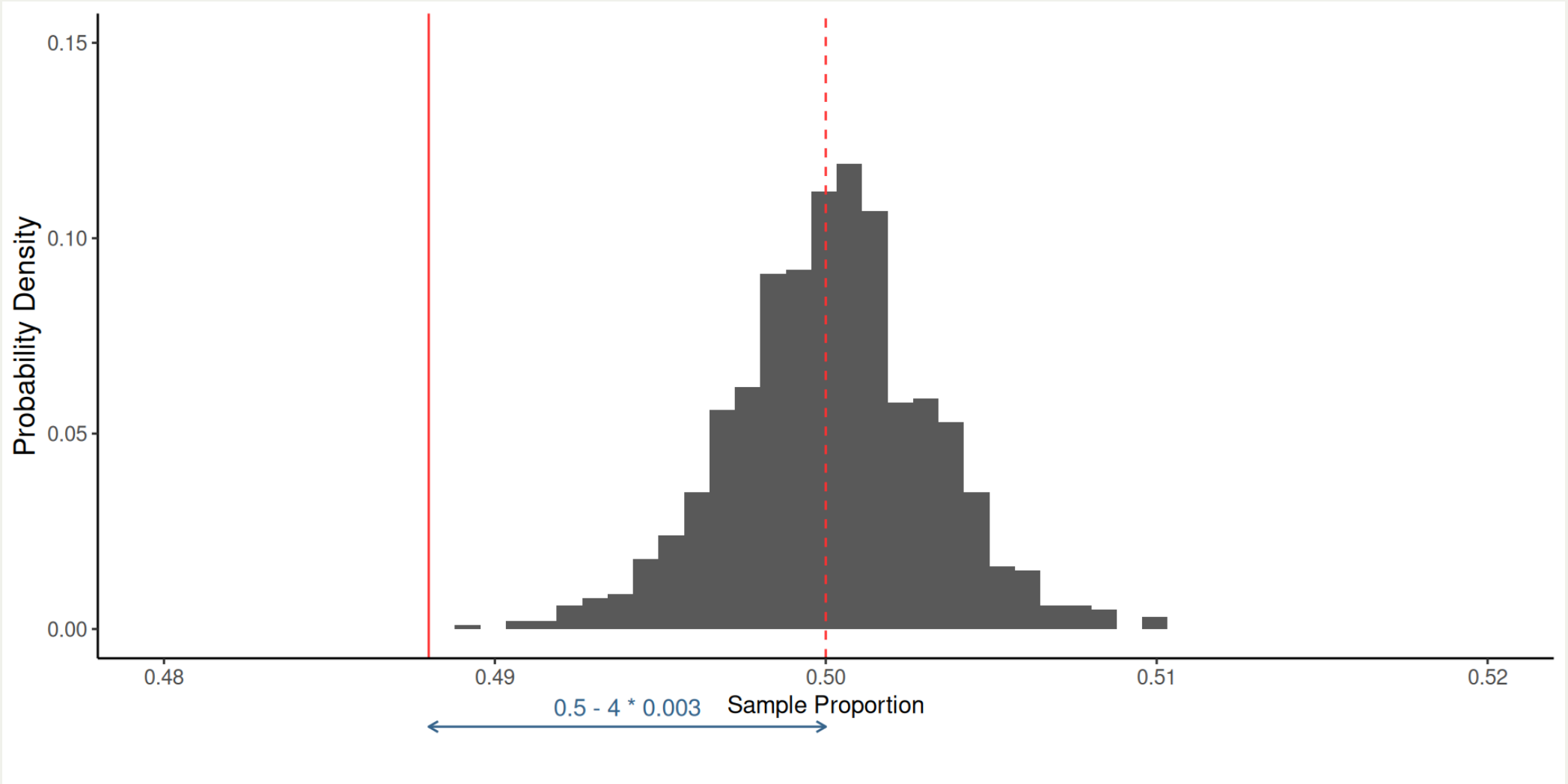
Significance Test

- **Significance test** is used to summarize the evidence about a hypothesis.
- It does so by comparing the our estimates with those predicted by a null hypothesis.
- 5 components of a significance test:
  - **Assumptions**: scale of measurement, randomization, population distribution, sample size
  - **Hypotheses**: null H_0 and alternative H_a hypothesis
  - **Test statistic**: compares estimate to those under H_0
  - **P-value**: weight of evidence against H_0, smaller P indicate stronger evidence
  - **Conclusion**: decision to *reject* or *fail to reject* H_0.

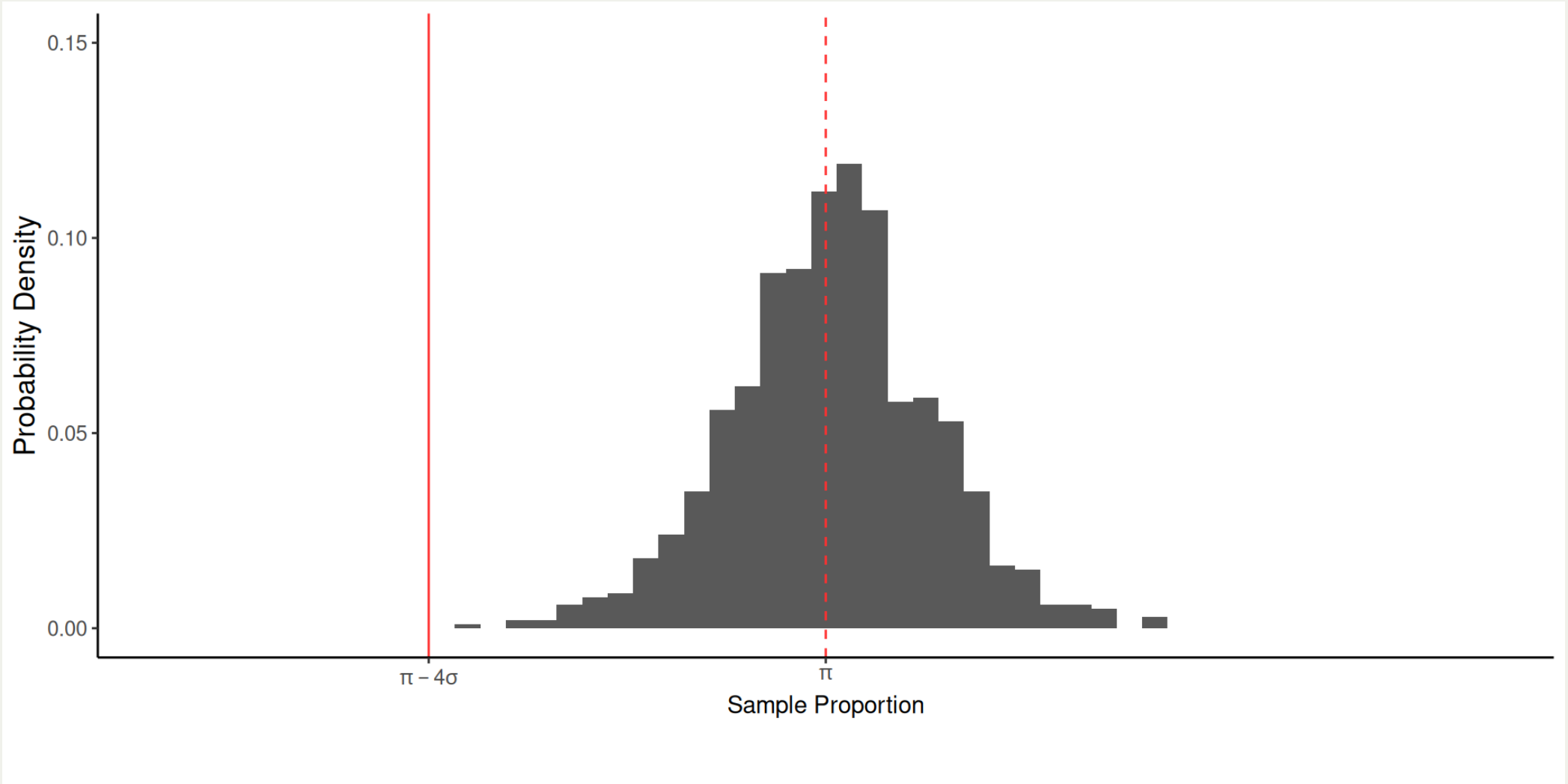
Example: Hypothesis Testing in UK EU Referendum

- **Assumptions**: categorical scale, random sample
- **Hypotheses** (two-sided):
  - $H_0 : \pi_{Leave} = 0.5$
  - $H_a : \pi_{Leave} \neq 0.5$
- **Conclusion**:
  - Calculated 95% confidence interval [0.482, 0.494] does not include 0.5
  - Thus, we can reject the H_0.
  - In the population the proportion of pro-Leave voters does not equal 0.5 at $p < 0.05$ level.

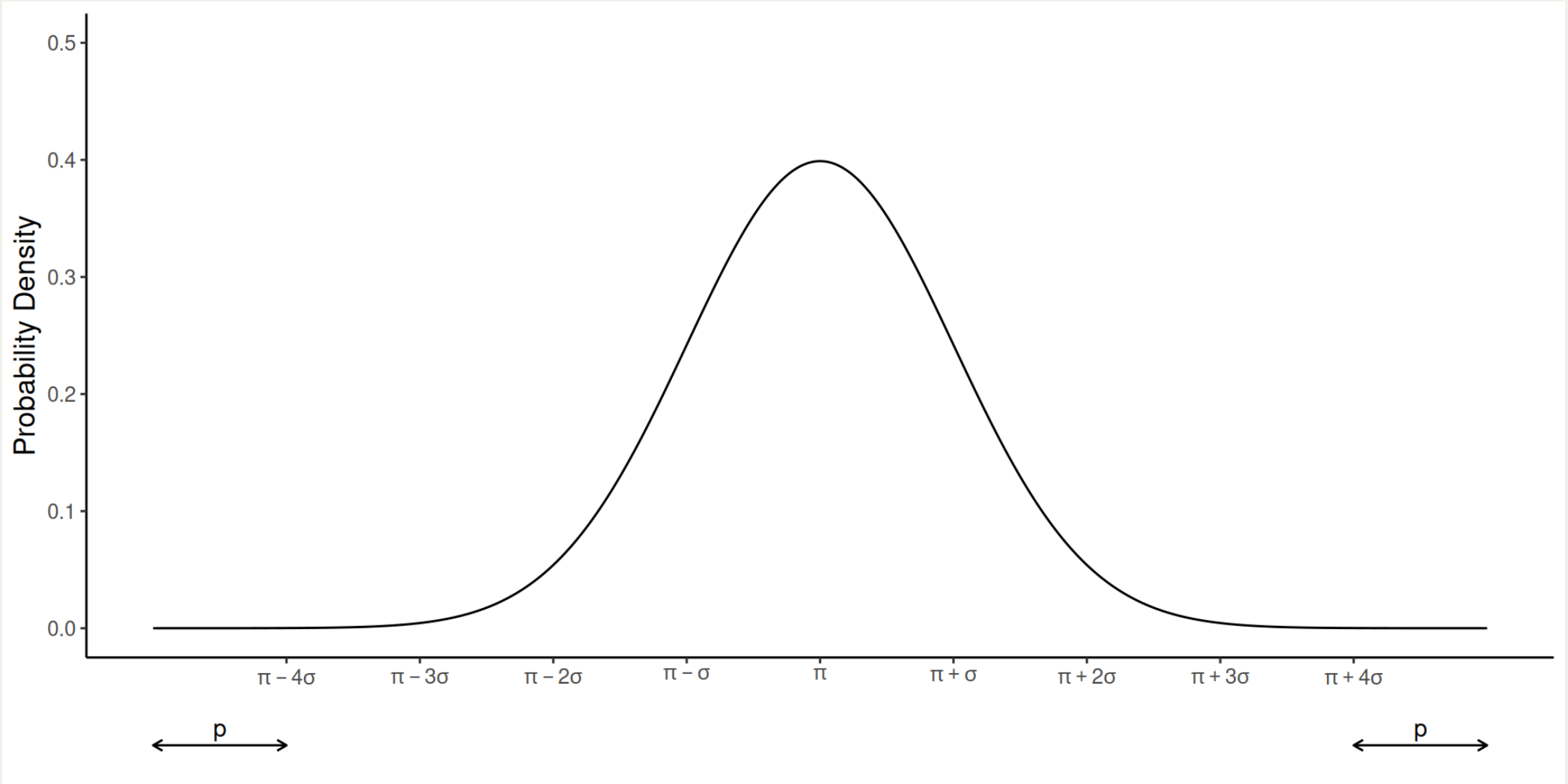
Example: What is the P-value in UK EU Referendum?



Example: What is the P-value in UK EU Referendum?



Example: What is the P-value in UK EU Referendum? (two-sided)



Example: What is the P-value in UK EU Referendum? (two-sided)

- We can use R to calculate the area under the normal curve.

```
pnorm(-4)
```

```
[1] 3.167124e-05
```

- As we need the area under both ends:

```
pnorm(-4) * 2
```

```
[1] 6.334248e-05
```

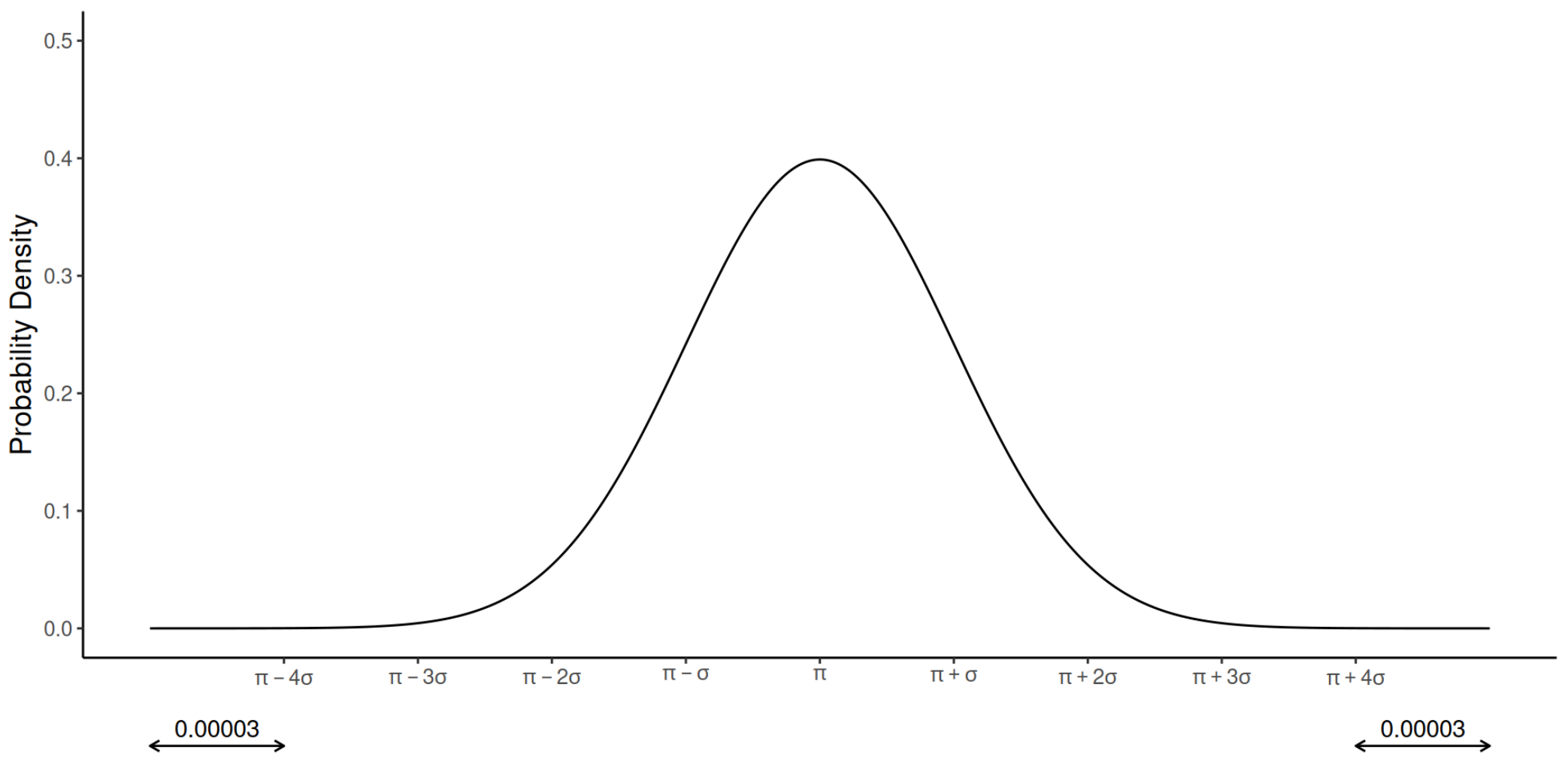
- Recall our discussion of scientific notation:

```
6 * 10 ^ -5
```

```
[1] 6e-05
```

Which is equivalent to $6 \times 10^{-5} = \frac{6}{10^5} = \frac{6}{100000} = 0.00006$

Example: What is the P-value in UK EU Referendum? (two-sided)



Example: Significance Test in UK EU Referendum

- **Assumptions**: categorical scale, random sample
- **Hypotheses** (two-sided):
  - $H_0 : \pi_{Leave} = 0.5$
  - $H_a : \pi_{Leave} \neq 0.5$
- **Test statistic**:

$$z = \frac{\hat{\pi} - \pi_{H_0}}{\sigma_{\hat{\pi}}} = \frac{0.488 - 0.5}{0.003} = -4$$

- **P-value**: $p = 0.00006$
- **Conclusion**:
  - Reject the H_0.
  - In the population the proportion of pro-Leave voters does not equal 0.5 at $p < 0.001$ level.

Error Types

		Null hypothesis (H_0) is	
		True	**False**
Decision about Null hypothesis (H_0)	**Don't reject**	Correct inference (true negative) (probability = 1−α)	Type II Error (false negative) (probability = β)
	Reject	Type I Error (false positive) (probability = α)	Correct inference (true positive) (probability = 1−β)

Next

- Workshop:
  - Data Frames
- Next week:
  - Analysis of Proportions and Means