universität wien

BAK3: Introduction to Quantitative Methods

Week 9: Hypothesis Testing I

Leonardo Carella

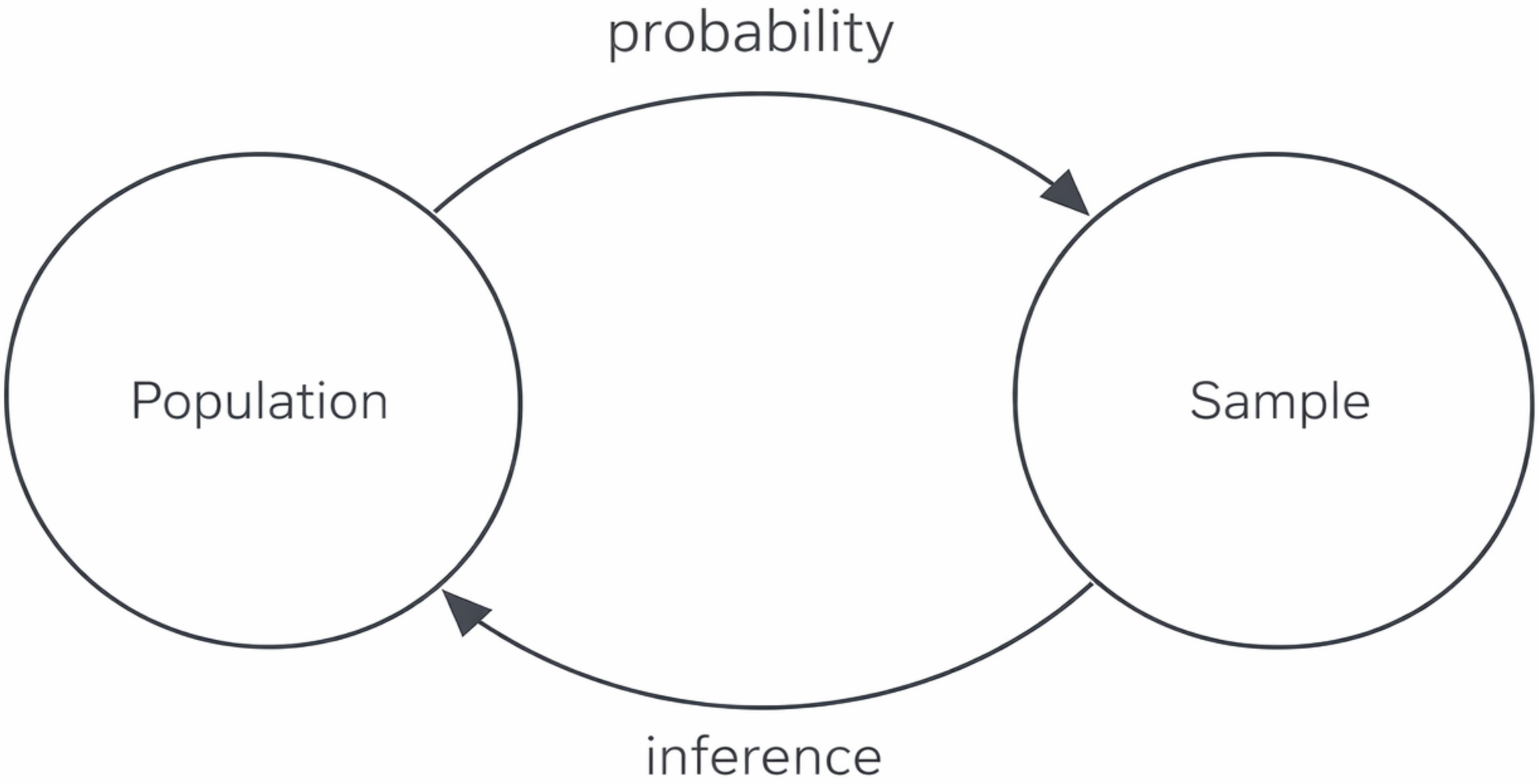
universität wien

The Plan for today

- Statistics:
  - **Hypothesis testing**: mechanics and interpretation.
  - One-sample **t-tests** and tests for proportions.
  - Interpreting test statistics and **p-values**.
- Coding in R:
  - Working with Autnes data as publicly available.
  - One-sample **t-tests** and tests for proportions in R.

universität wien

The Big Picture



universität wien

Inference

- So, far we have learnt two things that we can infer from a sample **estimate** about some population **parameter**:
- **Standard Error:** quantifies how much the estimate of that parameter would vary across hypothetical repeated sample, and thus how precise or uncertain is our estimate of the population parameter.
- **Confidence Interval:** it is a range of plausible values for the population parameter. If we sampled from the population many times, e.g. 95% of the confidence intervals constructed this way would contain the population parameter.

universität wien

Hypothesis Testing

- Framework to evaluate the plausibility of **hypotheses about the population** employing **proof of contradiction**:
- Suppose we have a random survey sample, and our hypothesis is that “women are more supportive of gay marriage than men”. We find that, on average, this is the case in our sample. But what about the population?
- We test how plausible it is that — if “men and women are **equally** supportive of gay marriage” were true ***in the population*** — we would obtain a result ‘like’ the one we did ***in the sample***.
- If this probability is **low**, then we can be **more** confident of our inference from the sample to the population.

universität wien

Hypothesis Testing

- Many types of hypothesis tests. But common procedure:
  - State the “null hypothesis”: the thing we want to disprove, or H_0.
  - Compare our sample estimate to the parameter value that we are hypothesising under H_0.
  - *p*-value: summary of the evidence against the “null hypothesis”.

universität wien

Hypothesis Testing

- Many types of hypothesis tests:
- **Depending on the hypothesis**: directional (the parameter is greater than…) or non-directional (the parameter is different from…). In this course, we focus on **non-directional hypotheses** (two-tailed tests).
- **Depending on the parameter to test**: today and next week, we focus on means and proportions; from week 11 onwards, regression coefficients.
- **Depending on the number of samples compared**: today, one-sample tests (comparing a sample estimate to a known or hypothesised value). Next week: two-sample tests (comparing estimates in two groups to each other).

universität wien

One-sample t-test

- Example: We have a sample of size 500, with information on respondents' approval rating for Trump (x) on a scale from 0 to 100. We find $\bar{x} = 44$.
- Our hypothesis: on average, *in the population*, American adults either approve ($\mu > 50$) or disapprove ($\mu < 50$) to Trump.
  - The null hypothesis: $\mu = 50$.
  - The alternative hypothesis: $\mu \neq 50$.
- The point of the the test: if μ were 50, how likely is it that we get an $\bar{x}$ as far from 50 as the one we actually observe (44)?

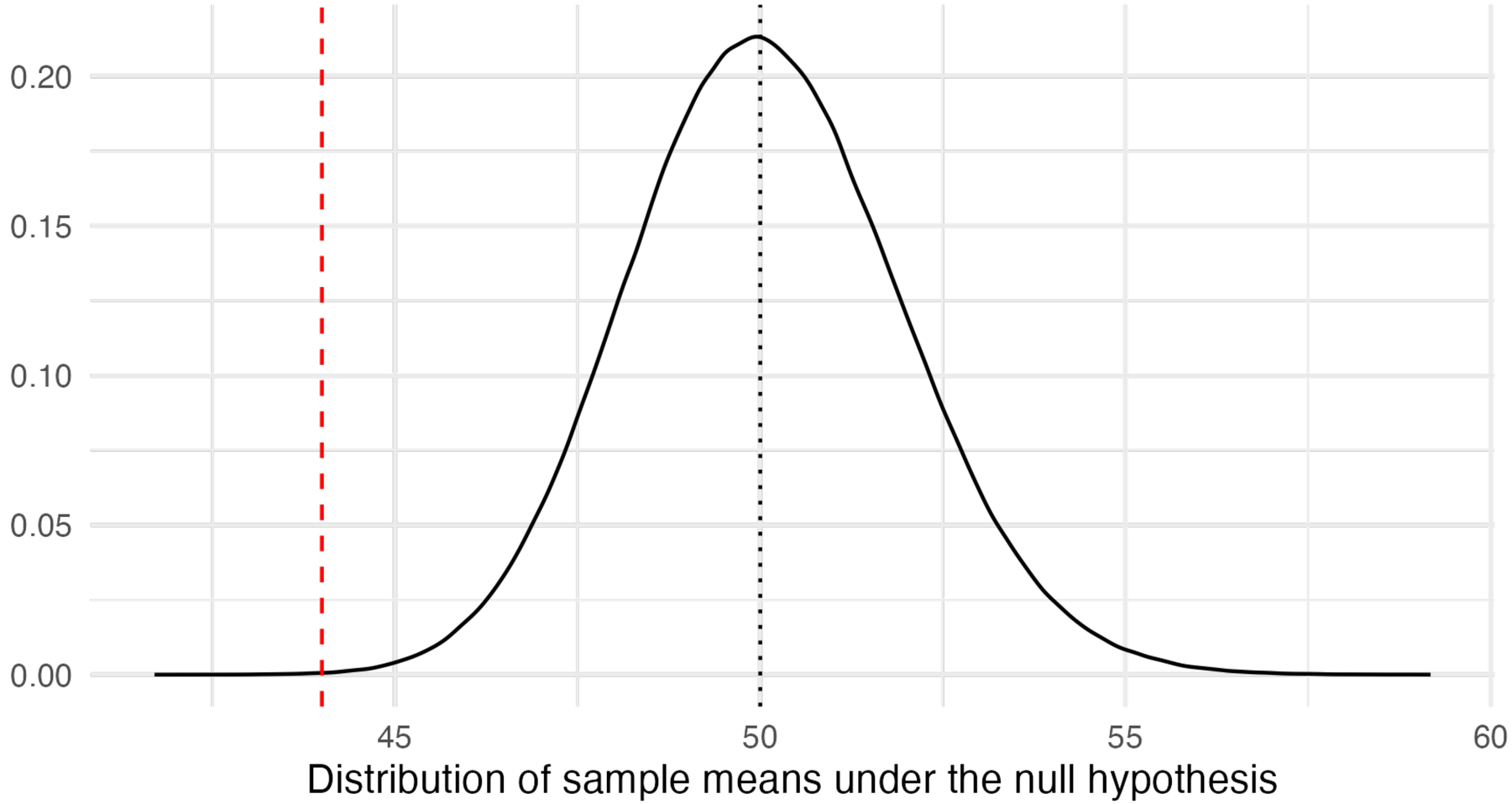
universität wien

One-sample t-test

- The distribution of sample means if $\mu = 50$ would be (approximately) a normal distribution centred at 50, and with a standard deviation equal to the standard error of the sample mean — which we can get from our sample!
- In fact, we commonly assume that the sampling distribution is *t*-distributed: the *t*-distribution is like a normal distribution, but slightly flatter when the sample size is small. For all intents and purposes, think of the *t*-distribution as a normal distribution.

universität wien

One-sample t-test



universität wien

The t-statistic

- How unlikely is it to get a sample mean of 44 in a random sample of size 500, under the null hypothesis ($\mu = 50$)? One indicator: the **t-statistic**.

- t-statistic: the distance between $\bar{x}$ and 50, expressed in number of standard deviations of the sampling distribution (i.e. standard errors):

$$t\text{-statistic} = \frac{\bar{x} - 50}{\text{SE}_{\bar{x}}}$$

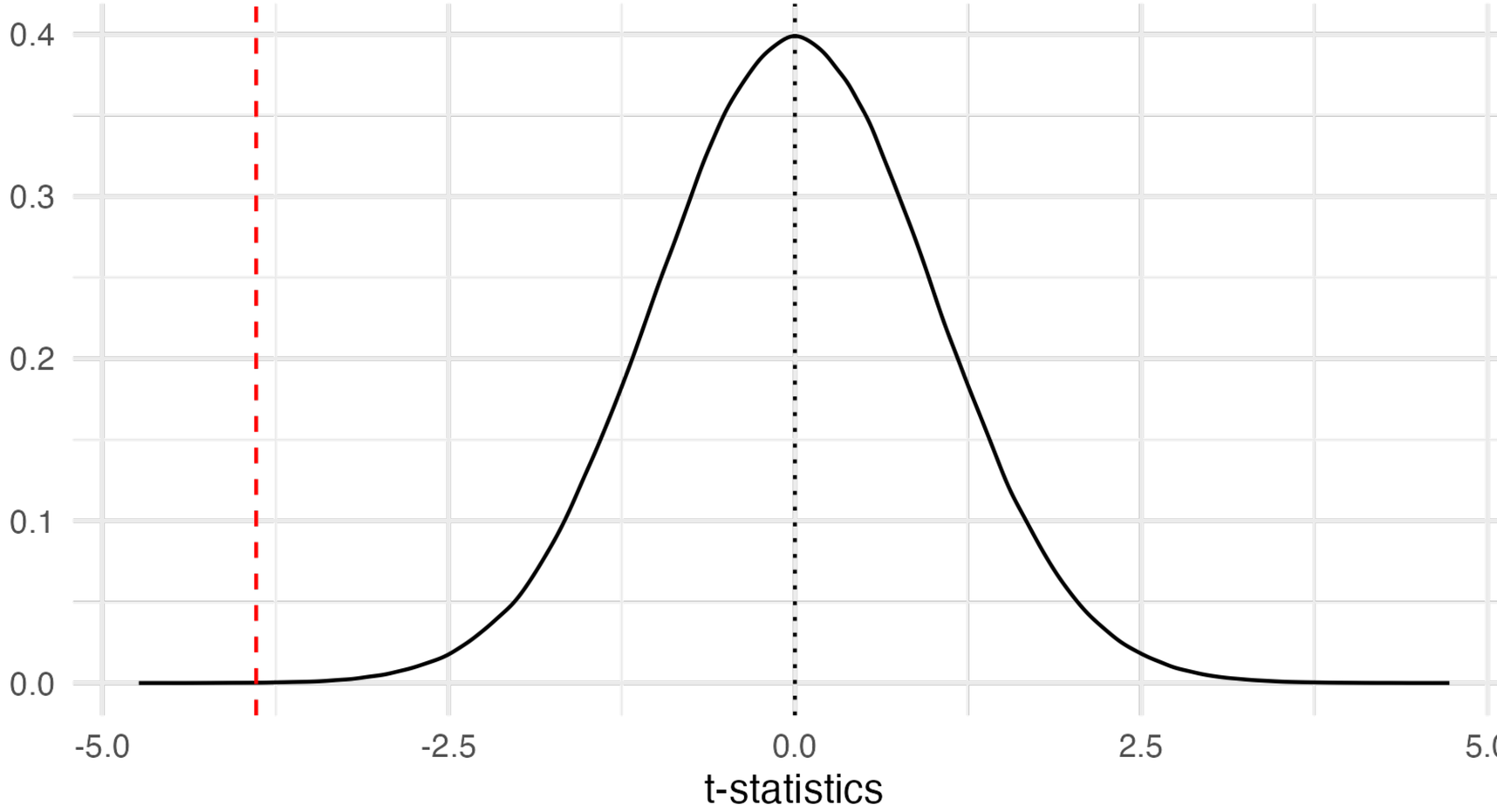
- More generally, substituting any arbitrary μ_0 for “50” and the formula for the SE:

$$t\text{-statistic} = \frac{\bar{x} - \mu_0}{\text{sd}(x)/\sqrt{n}}$$

universität wien

t-statistic

- We do our calculations, and we get a t-statistic of -3.89.
- This already tells us that our sample mean is quite unlikely if the null were true, as we know that 95% of sample means are between -1.96 and 1.96 SEs from the mean.



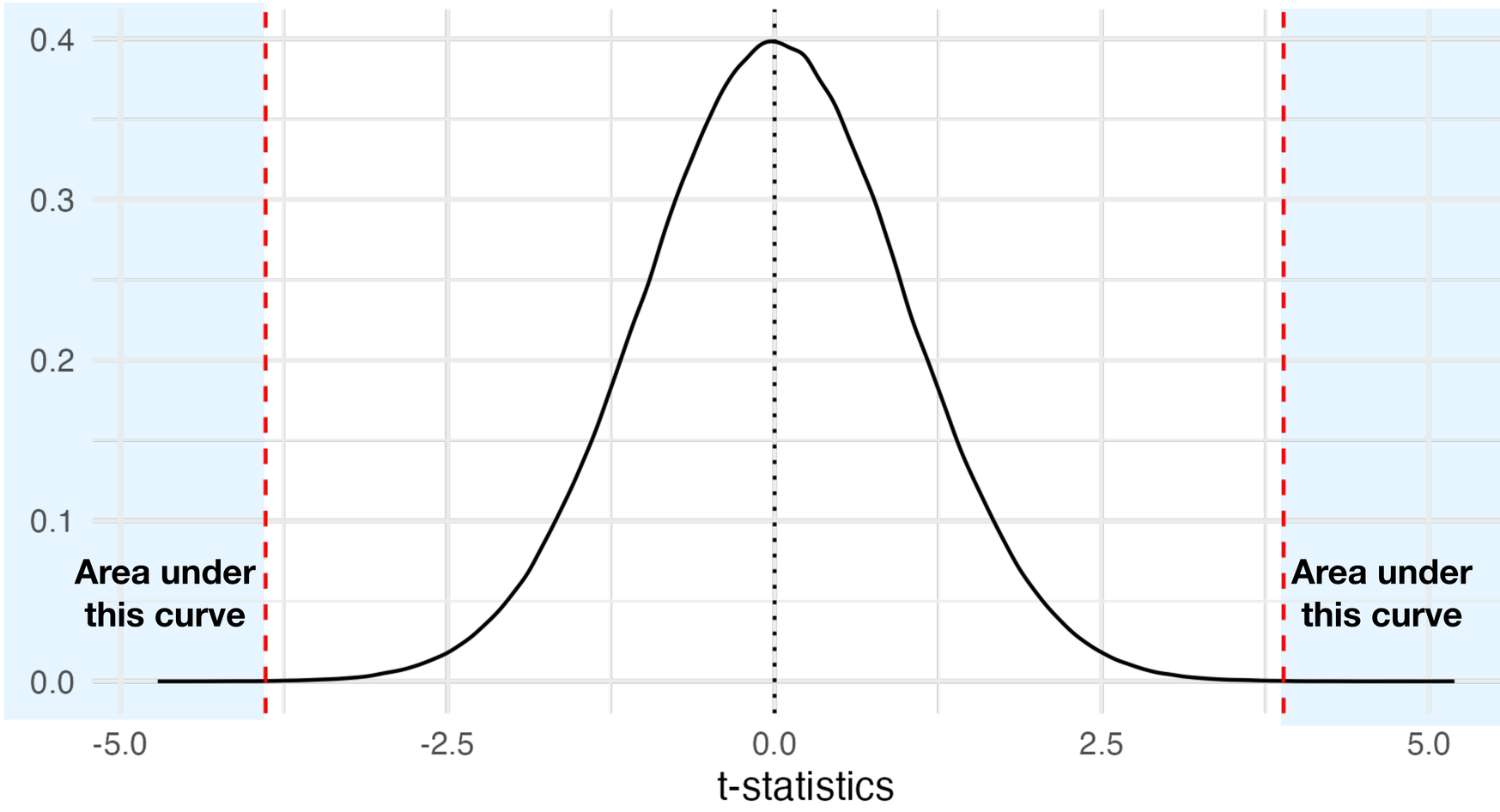
universität wien

p-value

- **How unlikely *exactly*?**
- The (two-tailed) *p*-value expresses the same idea as the *t*-statistic, but in terms of probability.
- In general, the *p*-value is **the probability of obtaining a test statistic at least as extreme than the one we get, under the null hypothesis.**
- Because the test statistic is associated with a sample mean here, the *p*-value of a one-sample t-test is probability of obtaining a **sample mean** at least as extreme as the one observed, under the null hypothesis.

p-value

- Therefore, the p-value is the probability of obtaining a test statistic below -3.89 or above 3.89.
- In our example, this is very, very small: ***p* = 0.0002**



universität wien

p-value

- Smaller *p*-values give us stronger evidence against the null.
- Conventionally, we say that:
  - if $p < 0.05$, we can **reject the null** at the 95% confidence level.
  - if $p \geq 0.05$, we **cannot** reject the null at the 95% confidence level.
- We **never** "accept the null", because the probability that μ is exactly 50 is vanishingly small (it could be 50.0001, 49.9999 etc.).

universität wien

p-value

- Other (conventional) significance cutoffs exist:
  - if $p < 0.10$, we can **reject the null** at the 90% confidence level.
  - if $p < 0.01$, we can **reject the null** at the 99% confidence level.
- Substantive interpretation of a *p*-value: it is the probability of obtaining a sample mean **at least as extreme** as the one observed (44), under the null hypothesis that the population mean is 50.
- It is ***not*** “the probability that the population mean is 50”.

universität wien

p-value

- Other acceptable ways of talking about a p-value:
- “The sample mean of 44 is far enough from 50 that such a difference is unlikely to be due to chance alone ($p < 0.01$).”
- “The sample mean of 44 is ***significantly different*** from 50 ($p < 0.01$).”

universität wien

Test for Proportions

- Imagine now we have an exit poll of 400 voters (randomly sampled!) in a referendum, and we find that “Yes” voters make up 52% of the sample ($p = 0.52$) . We want to know whether this is significantly different from 50%.
- Same story as before…
  - The null hypothesis: $\pi = 0.5$
  - The alternative hypothesis: $\pi \neq 0.5$

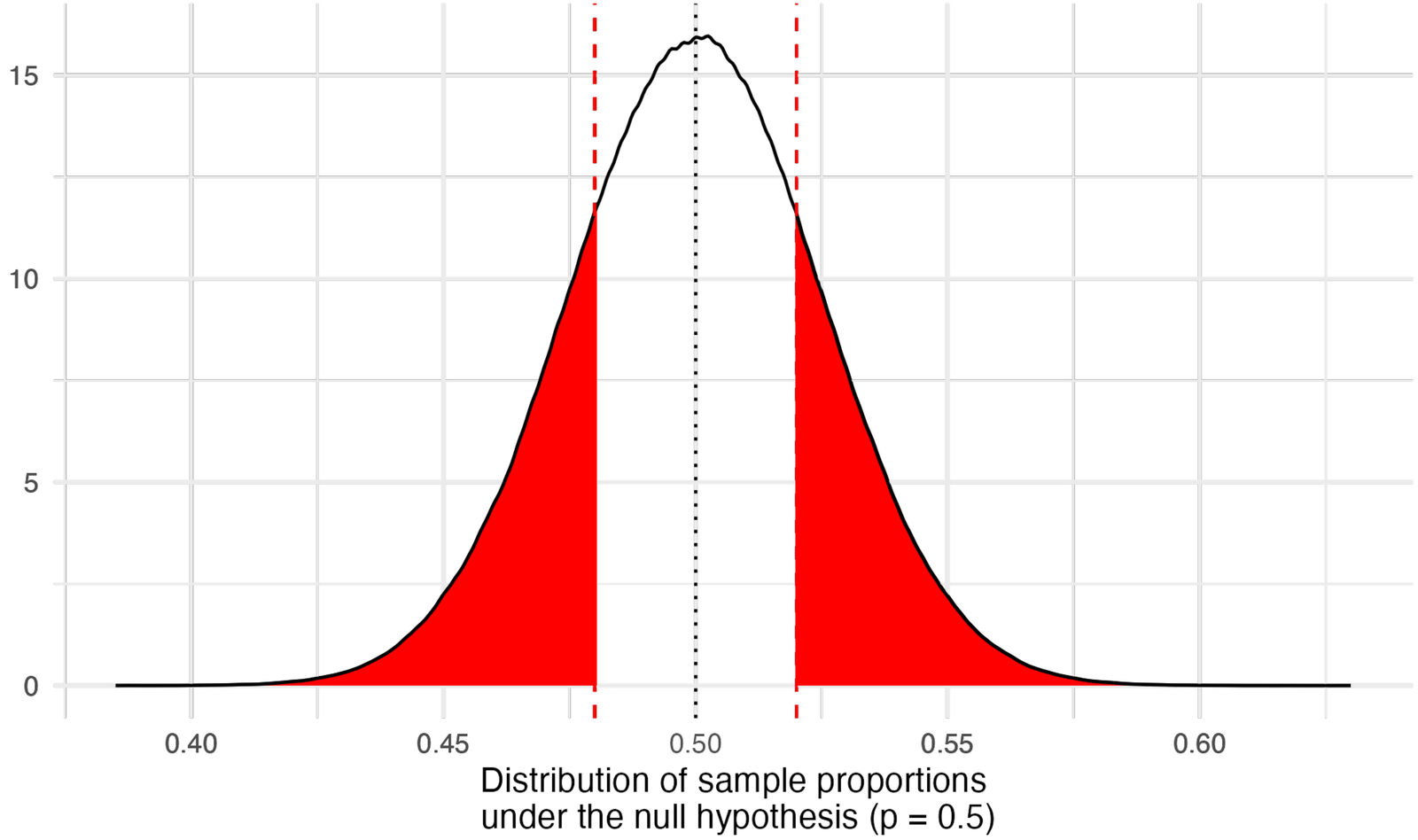
universität wien

Test for Proportions

- Normally, we can treat a proportion as a mean.
- But ***technically***, we cannot assume that the distribution of sample proportions is a *t*-distribution (which, like the normal, assigns some small non-zero probability to any value).
- This is because a sample proportion cannot exceed 0 and 1.
- So we use a different set of assumptions about the distribution of the test statistic. For this reason, test for proportions is known as **'chi-squared' test**, rather than 't'-test. Not a big deal now.

Test for Proportions

- p-value = **0.45.**
- We **cannot reject** the null that the ‘true’ population proportion is 50%.
- Informally: we can’t know from this sample if “Yes” won in the population.



universität wien

Summing Up

- Hypothesis testing for sample means and proportions:
  - State a “null” and an “alternative” hypothesis.
  - The test returns a **p-value**, which summarises the evidence against the null in terms of a probability.
  - In a t-test, you can also interpret the **t-statistic**: difference between the sample mean and the population mean under the null, in terms of SEs.
- Reject/not reject the null at some confidence level, depending on p-value.
- And now, let’s implement this in R…