

Malaria Molecular Surveillance Study Design Workshop

Module 3: Hypothesis testing

Sometimes we are simply trying to estimate something, e.g. prevalence. **We have seen how to perform sample size calculation based on precision arguments.**

In other cases, we have a specific question that we want to answer. This questions may be phrased as a **null hypothesis test.**

Sometimes we are simply trying to estimate something, e.g. prevalence. **We have seen how to perform sample size calculation based on precision arguments.**

In other cases, we have a specific question that we want to answer. This questions may be phrased as a **null hypothesis test**.

A null hypothesis (H_0) is a statement of **no effect or difference** between groups. This is often a statement that nothing interesting is happening*

*Rather than trying to prove there is an effect, in null hypothesis testing we try to **disprove** that there is **no effect**.*

* Sometimes it can be very interesting if the null hypothesis is true

- Q: Has prevalence increased over the last 5 years?
- Q: Are certain genetic variants associated with gender, or occupation?
- Q: Does vaccine efficacy vary based on genetic markers?

- Q: Has prevalence increased over the last 5 years?
 H_0 : Prevalence has remained the same over the last 5 years.
- Q: Are certain genetic variants associated with gender, or occupation?
- Q: Does vaccine efficacy vary based on genetic markers?

- Q: Has prevalence increased over the last 5 years?

H_0 : Prevalence has remained the same over the last 5 years.

- Q: Are certain genetic variants associated with gender, or occupation?

H_0 : There is no association between genetic variant and gender or occupation.

- Q: Does vaccine efficacy vary based on genetic markers?

- Q: Has prevalence increased over the last 5 years?

H_0 : Prevalence has remained the same over the last 5 years.

- Q: Are certain genetic variants associated with gender, or occupation?

H_0 : There is no association between genetic variant and gender or occupation.

- Q: Does vaccine efficacy vary based on genetic markers?

H_0 : Vaccine efficacy is the same irrespective of genetic markers.

Each test has a **test statistic**

One-sample z-test for proportions: tests prevalence against a known value

Each test has a **test statistic**

One-sample z-test for proportions: tests prevalence against a known value

H_0 : The population prevalence equals p_0

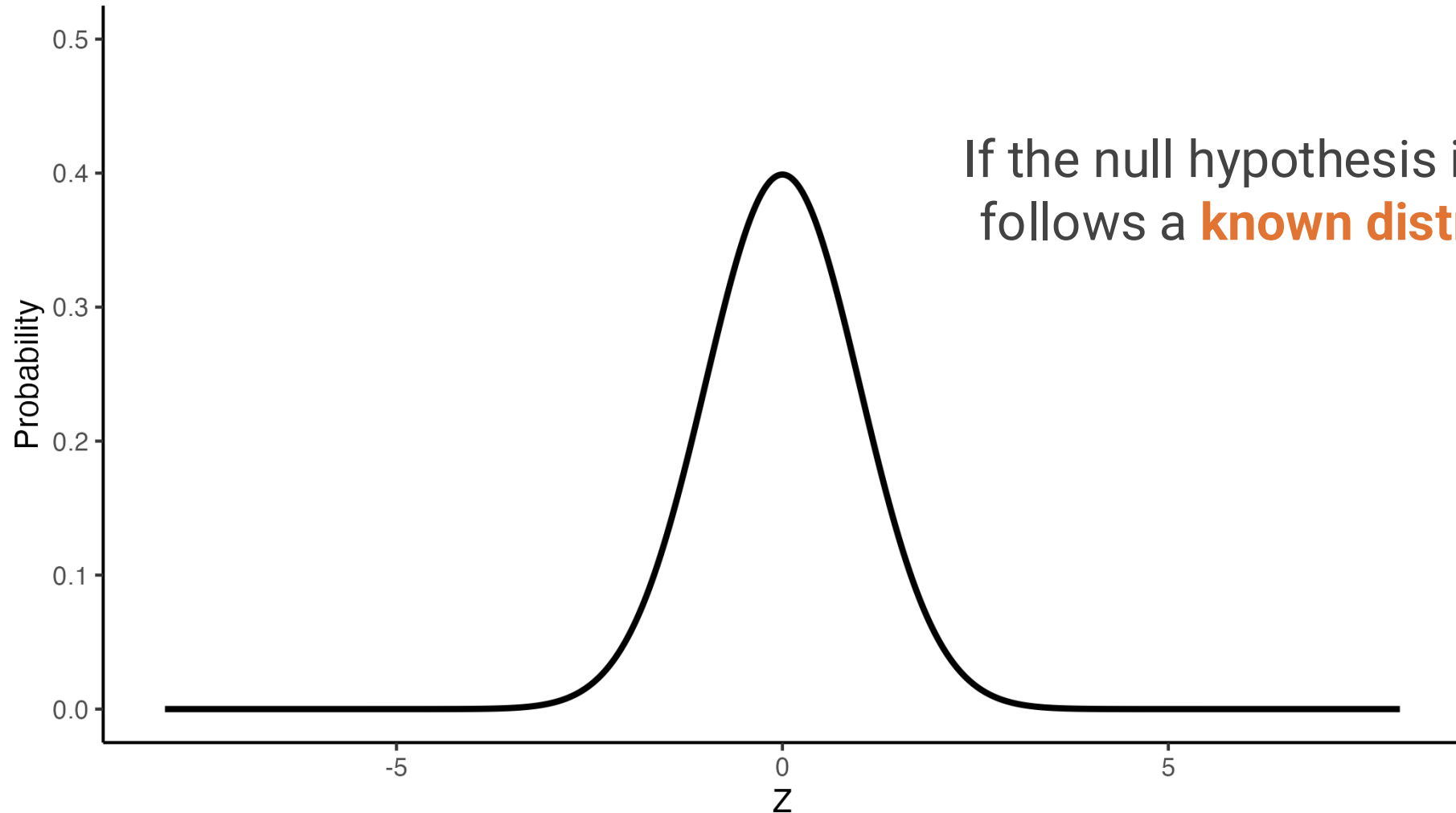
Each test has a **test statistic**

One-sample z-test for proportions: tests prevalence against a known value

H_0 : The population prevalence equals p_0

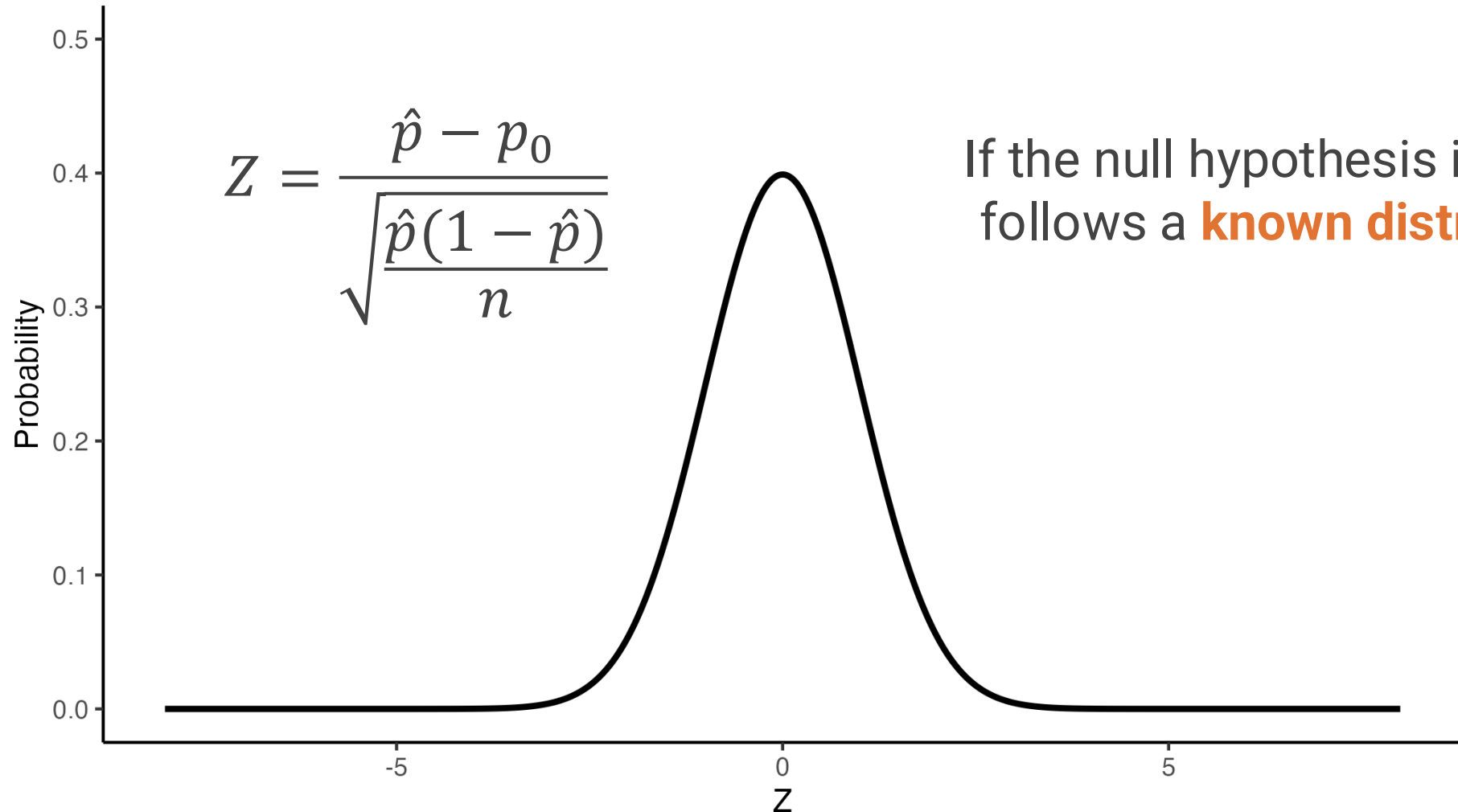
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}}$$

Null hypothesis testing



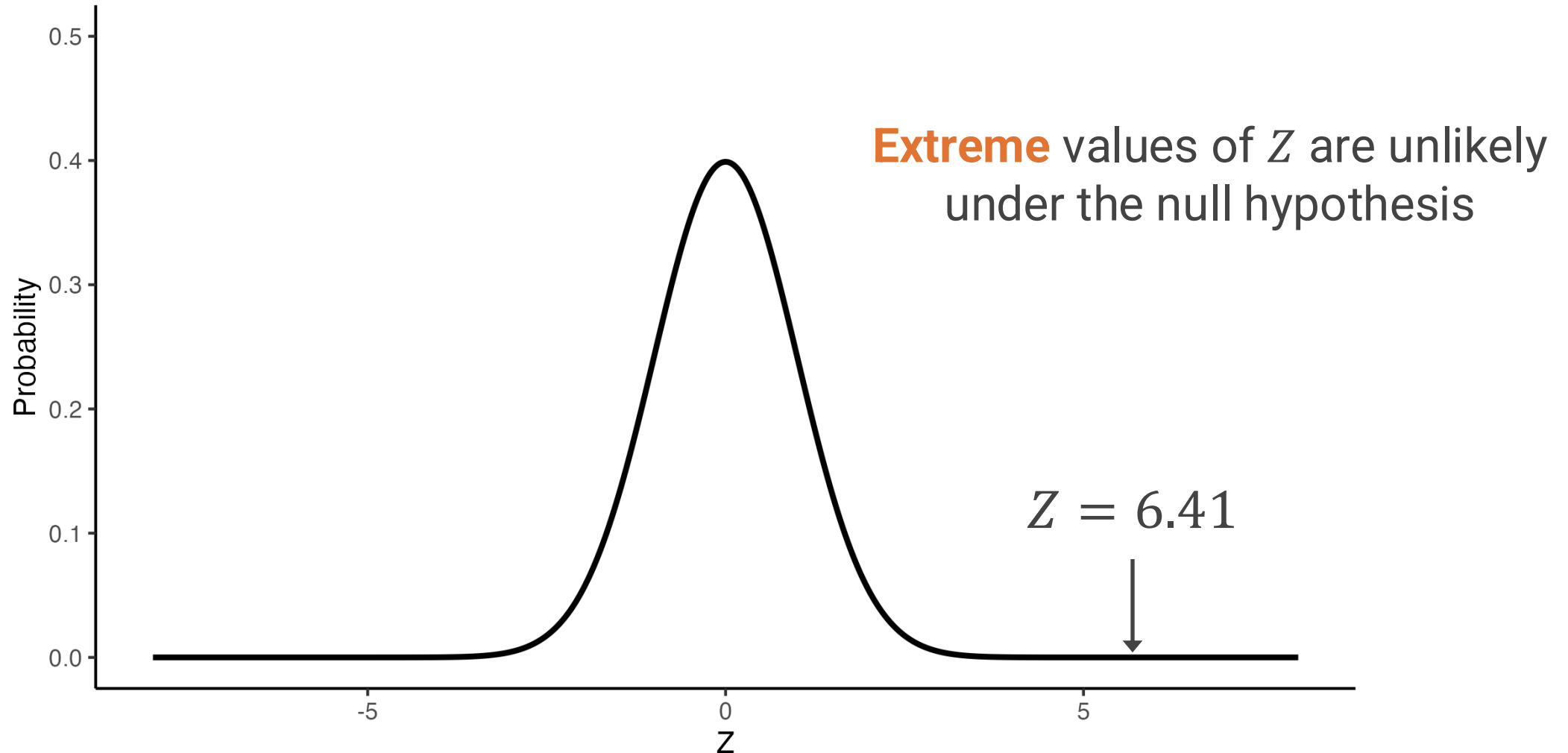
If the null hypothesis is true, Z follows a **known distribution**

Null hypothesis testing

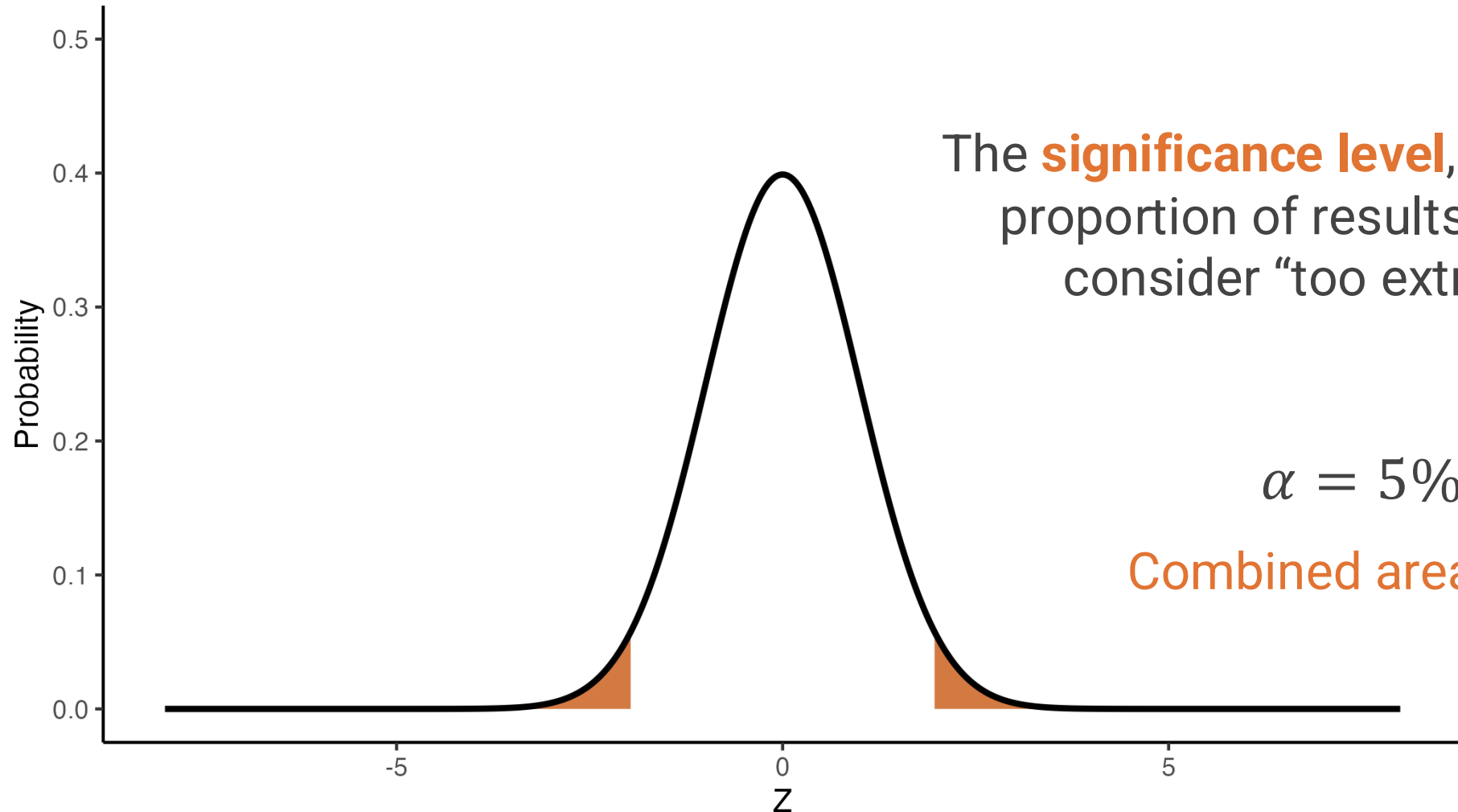


If the null hypothesis is true, Z follows a **known distribution**

Null hypothesis testing



Null hypothesis testing

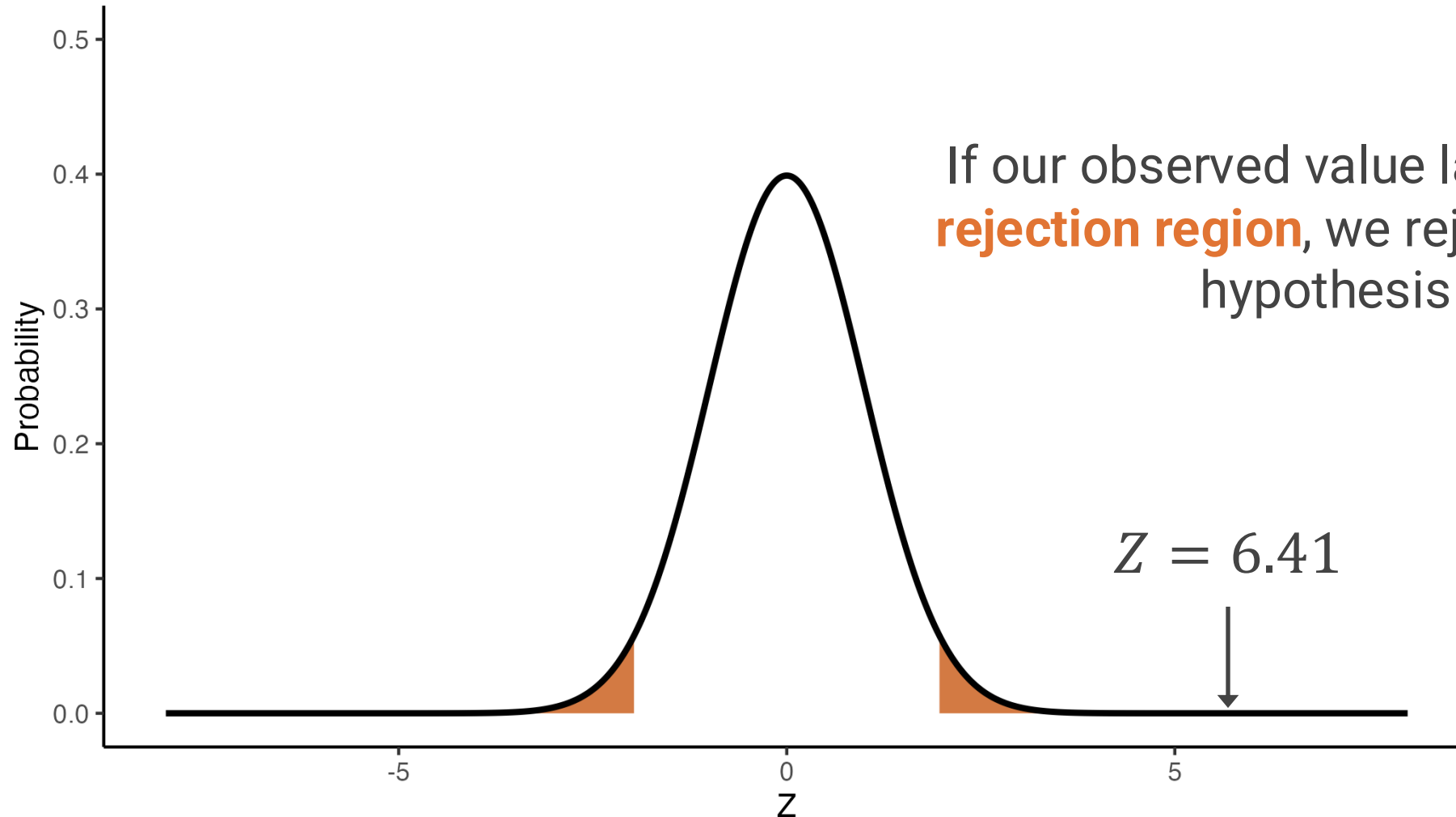


The **significance level**, α , sets the proportion of results that we consider “too extreme”

$$\alpha = 5\%$$

Combined area = 5%

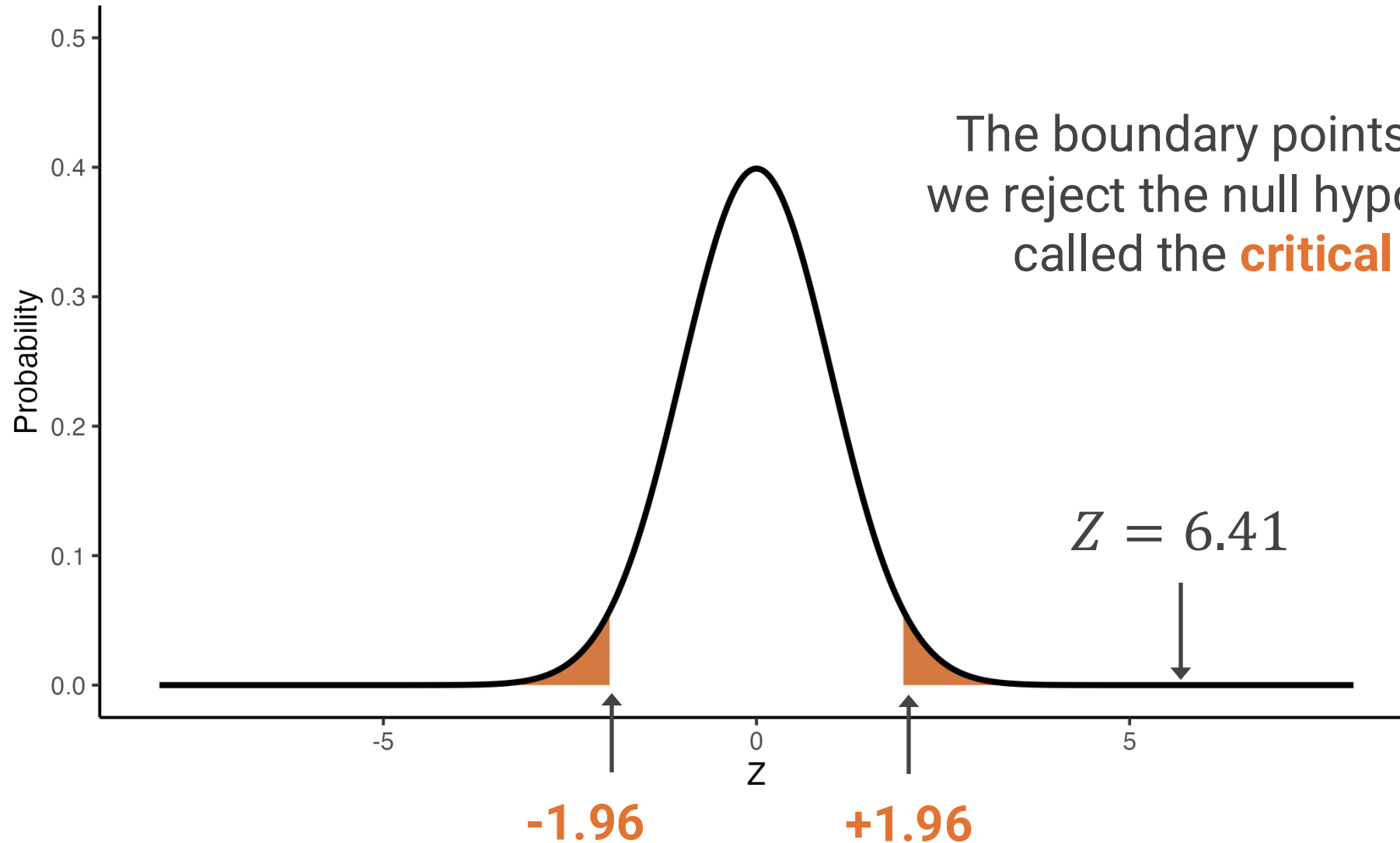
Null hypothesis testing



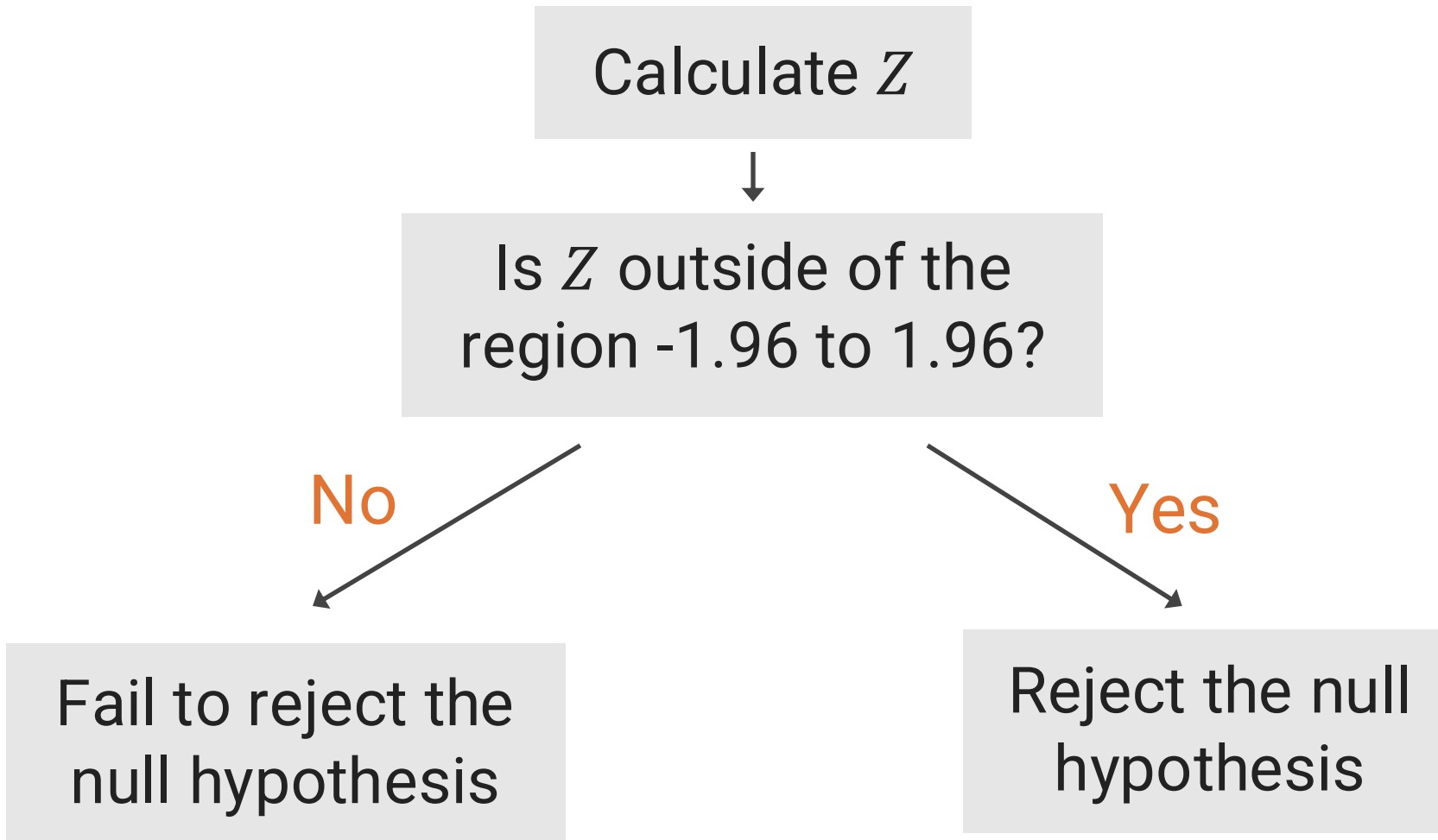
If our observed value lands in the **rejection region**, we reject the null hypothesis

$$Z = 6.41$$

Null hypothesis testing



Null hypothesis testing



Null hypothesis testing

		Conclusion about H_0	
		Fail to reject	Reject
Truth about H_0	True	True negative	False positive
	False		

Null hypothesis testing

		Conclusion about H_0	
		Fail to reject	Reject
Truth about H_0	True	True negative $1 - \alpha$	False positive α
	False		

Null hypothesis testing

		Conclusion about H_0	
		Fail to reject	Reject
Truth about H_0	True	True negative $1 - \alpha$	False positive α
	False		

α sets the **false positive rate** of a test. Using α we can control how often we incorrectly conclude that there is a real effect when there is none.

Null hypothesis testing

		Conclusion about H_0	
		Fail to reject	Reject
Truth about H_0	True	True negative $1 - \alpha$	False positive α
	False	What about this!? (next lecture!)	

α sets the **false positive rate** of a test. Using α we can control how often we incorrectly conclude that there is a real effect when there is none.

- We can ask questions using **null hypothesis tests**
- A null hypothesis is a statement of **no effect/difference** between groups
- The significance level α controls the **false-positive rate**

Format: Interactive R code, accessed through the web

- Short quiz on hypothesis testing
- Test prevalence against a threshold
- Calculating the test statistic



Workshop materials

https://mrc-ide.github.io/MMS-SD_workshop/