intro to data science with probability & statistics

CSCI 3022

October 24, 2018

Introduction to Hypothesis Testing

Department of Computer Science
UNIVERSITY OF COLORADO BOULDER

Dan Larremore

Stuff & Things

- **HW4** due on Friday.
- Check in on App pls.

A thought experiment 🤔

- **Example**: After the introduction of the Euro, Polish mathematicians claimed that the new Belgian 1 Euro coin is not a fair coin. Suppose I hand you a Belgian 1 Euro coin. How could you decide whether or not it is fair?

Statistical Hypotheses

- **Definition**: A *statistical hypothesis* is a claim about the value of a parameter of a population characteristic.
- **Examples**:
  - Suppose the recovery time of a person suffering from disease D be normally distributed with mean μ_1 and standard deviation σ_1. **Hypothesis**: $\mu_1 > 10$ days.
  - Suppose μ_2 is the recovery time of a person suffering from disease D and given treatment for D. **Hypothesis**: $\mu_2 < \mu_1$
  - Suppose μ_1 is the mean internet speed for Comcast and μ_2 is the mean internet speed for Century Link. **Hypothesis**: $\mu_1 \neq \mu_2$

Null vs Alternative Hypotheses

- In any hypothesis testing problem, there are always two competing hypotheses that we consider:

    1.

    2.

- The **objective** of hypothesis testing is to decide, based on the data that we've sampled, *whether the alternative hypothesis is actually supported by the data.*

The classic jury analogy

- Think about a jury in a criminal trial.

- When a defendant is accused of a crime, the jury is supposed to presume that the defendant is not guilty. **“Not guilty” is the null hypothesis**.

- The jury is then presented with **evidence** (data). If the evidence seems implausible under the assumption of not-guilty, they may **reject** the “not guilty” status, and claim that the defendant is likely guilty.

Null vs Alternative Hypotheses

- **Is there strong evidence for the alternative**?
- The burden of proof is placed on those that believe the alternative claim, just like in a jury.
- The initially favored claim, written as H_0, will not be rejected in favor of the alternative claim, written as H_1, unless the sample evidence provides a lot of support for the alternative.
- Two possible conclusions:

1.

2.

Null vs Alternative Hypotheses

- **Why assume the Null Hypothesis**?
  - Sometimes we don't want to accept a particular assertion unless/until data can be shown to strongly support it.
  - Reluctance (measured in cost or time) to change.
- **Example**: A company is considering hiring a new advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200K hits per day. With μ denoting the true average number of hits they'd get per day under the new company's advertising, they would not want to switch companies (because it would be costly) unless evidence strongly suggested that μ exceeds 200K.

Null vs Alternative Hypotheses

- **Example**: A company is considering hiring a new advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200K hits per day. With μ denoting the true average number of hits they'd get per day under the new company's advertising, they would not want to switch companies (because it would be costly) unless evidence strongly suggested that μ exceeds 200K.
- An appropriate problem formulation would involve testing:
- The conclusion that change is justified is identified with the alternative hypothesis and it would take conclusive evidence to justify rejecting H_0 and switching to the new company

Null vs Alternative Hypotheses

- The alternative to the Null Hypothesis $H_0 : \theta = \theta_0$ will look like one of the following assertions (or hypotheses):

- The equals sign is **always** the Null Hypothesis
- The alternative hypothesis is the one for which we are seeking statistical evidence.

Test statistics and evidence

- **Def**: A test statistic is a quantity derived from the sample data and calculated assuming that the Null hypothesis is true. It is used in the decision about whether or not to reject the Null hypothesis.
- **Intuition**:
  - We can think of the test statistics as our evidence about the competing hypotheses.
  - We consider the test statistic under the assumption that H_0 is true by asking: **How likely would we obtain this evidence if the Null were true?**
- **Example**: To determine if the Belgian 1 Euro coin is fair you flip it 100 times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?

Test statistics and evidence

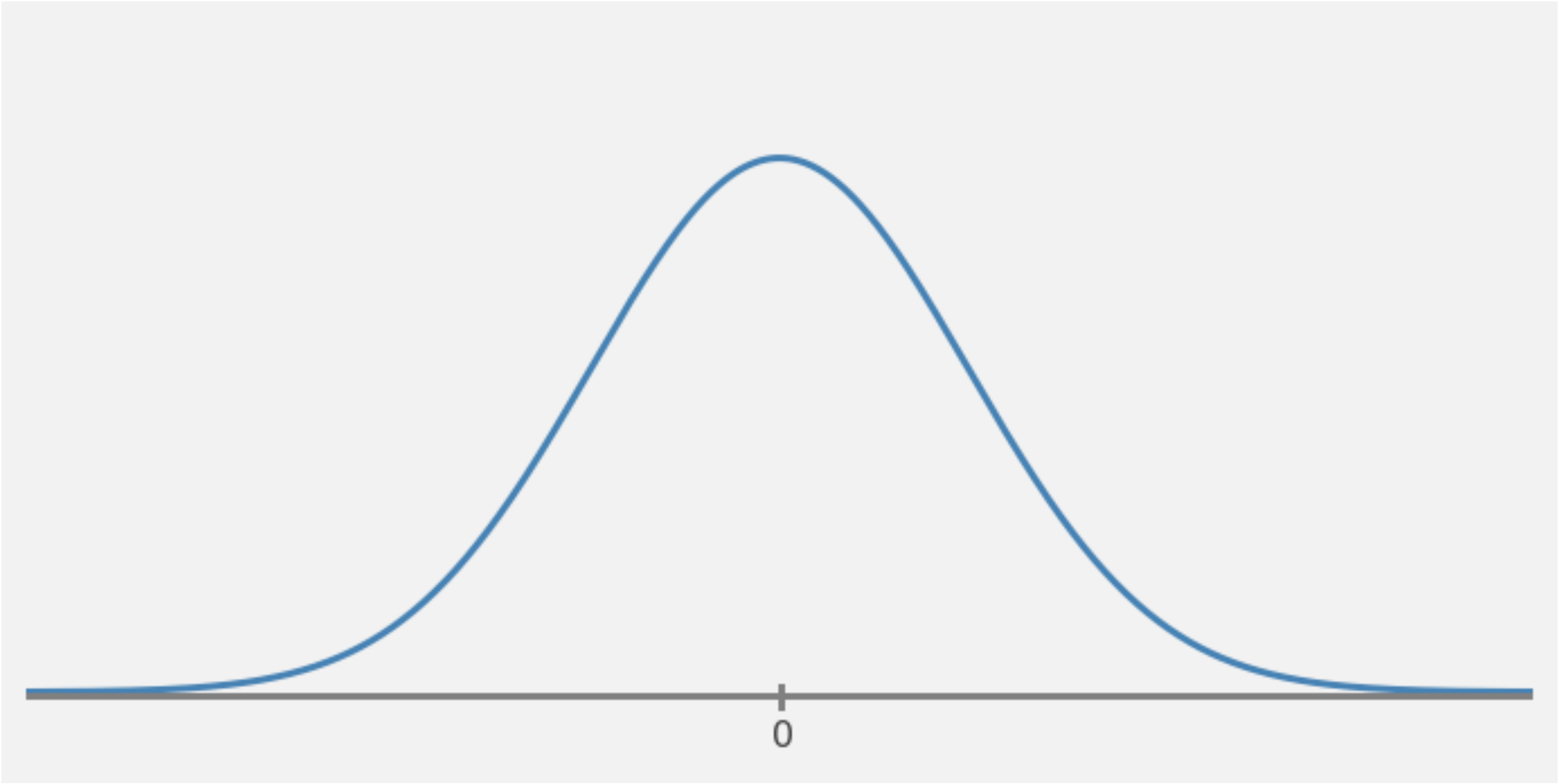
- **Example**: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?

Test statistics and evidence

- **Example**: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?
- **Question**: What would it take to convince you that the coin is not fair?

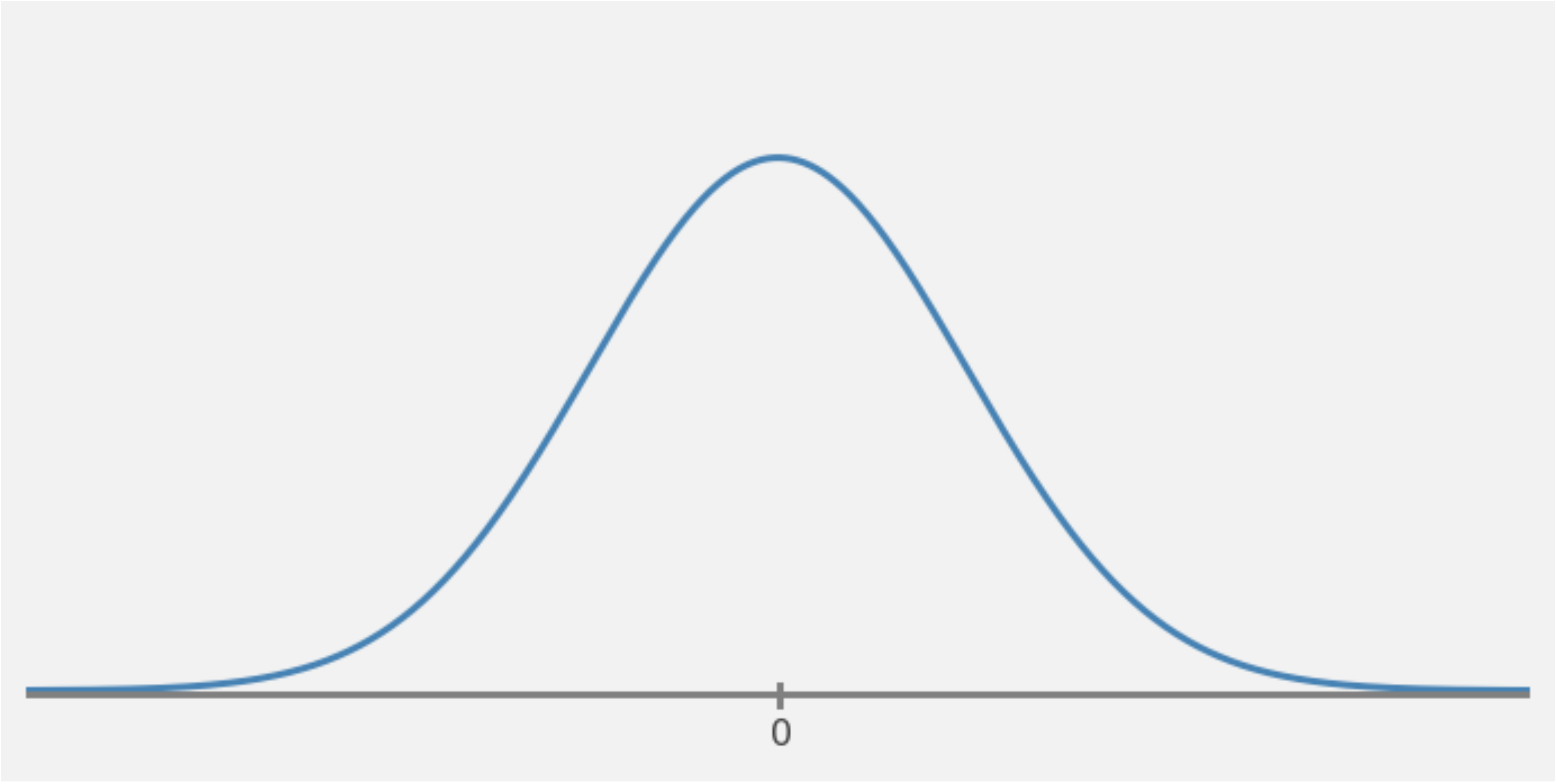
Test statistics and evidence

- **Example**: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads. What is the test statistic? What are the Null and alternative hypotheses?
- **Question**: What would it take to convince you that the coin is not fair?



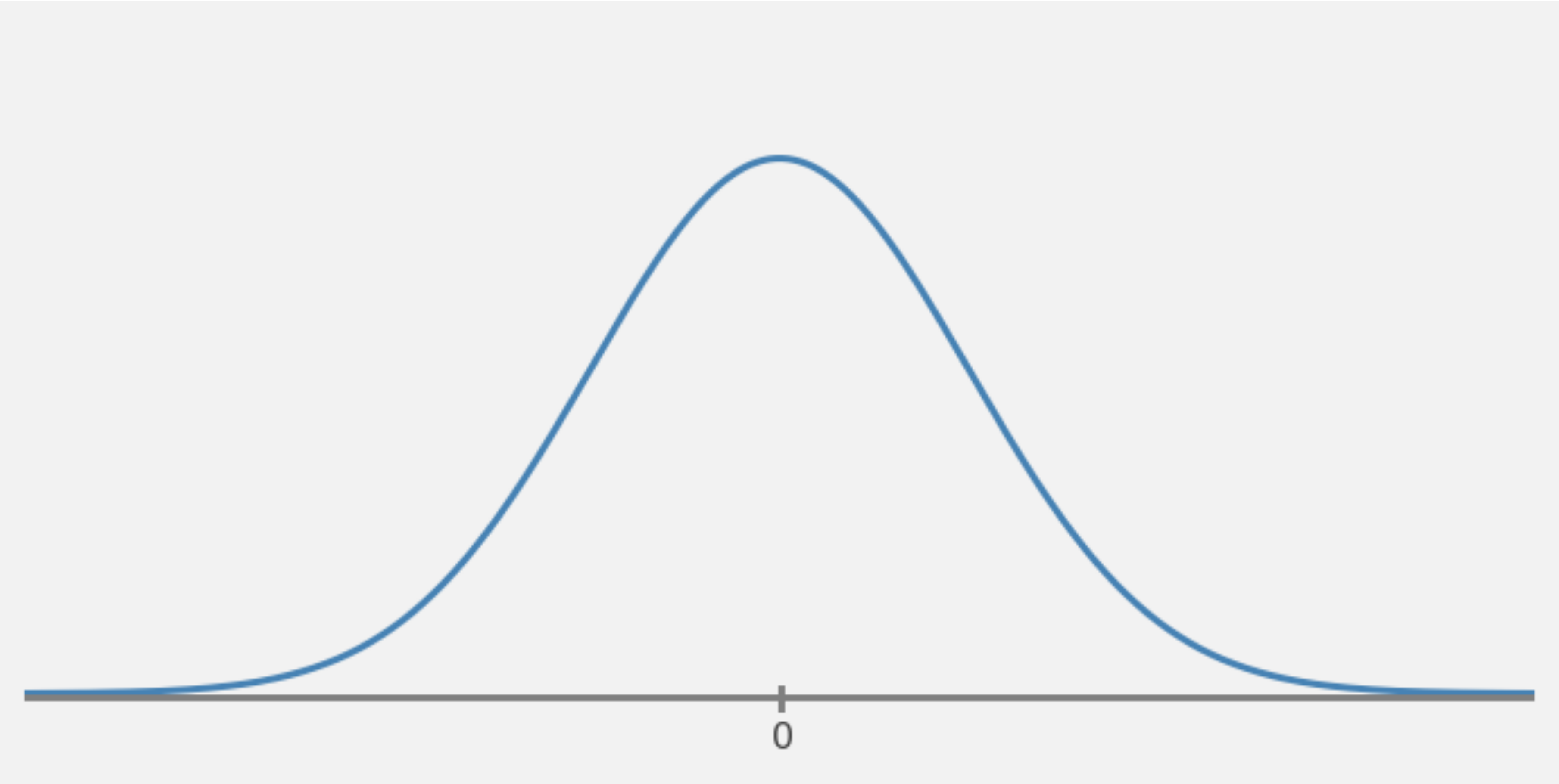
Rejection regions and significance level

- **Example**: To determine if the Belgian 1 Euro coin is fair you flip it n times and record the number of Heads.
- **Def**: The **rejection region** is a range of values of the test statistic that would lead you to **reject** the Null hypothesis.
- **Def**: The **significance level** α indicates the largest probability of the test statistic occurring under the Null hypothesis that would lead you to reject the Null hypothesis.



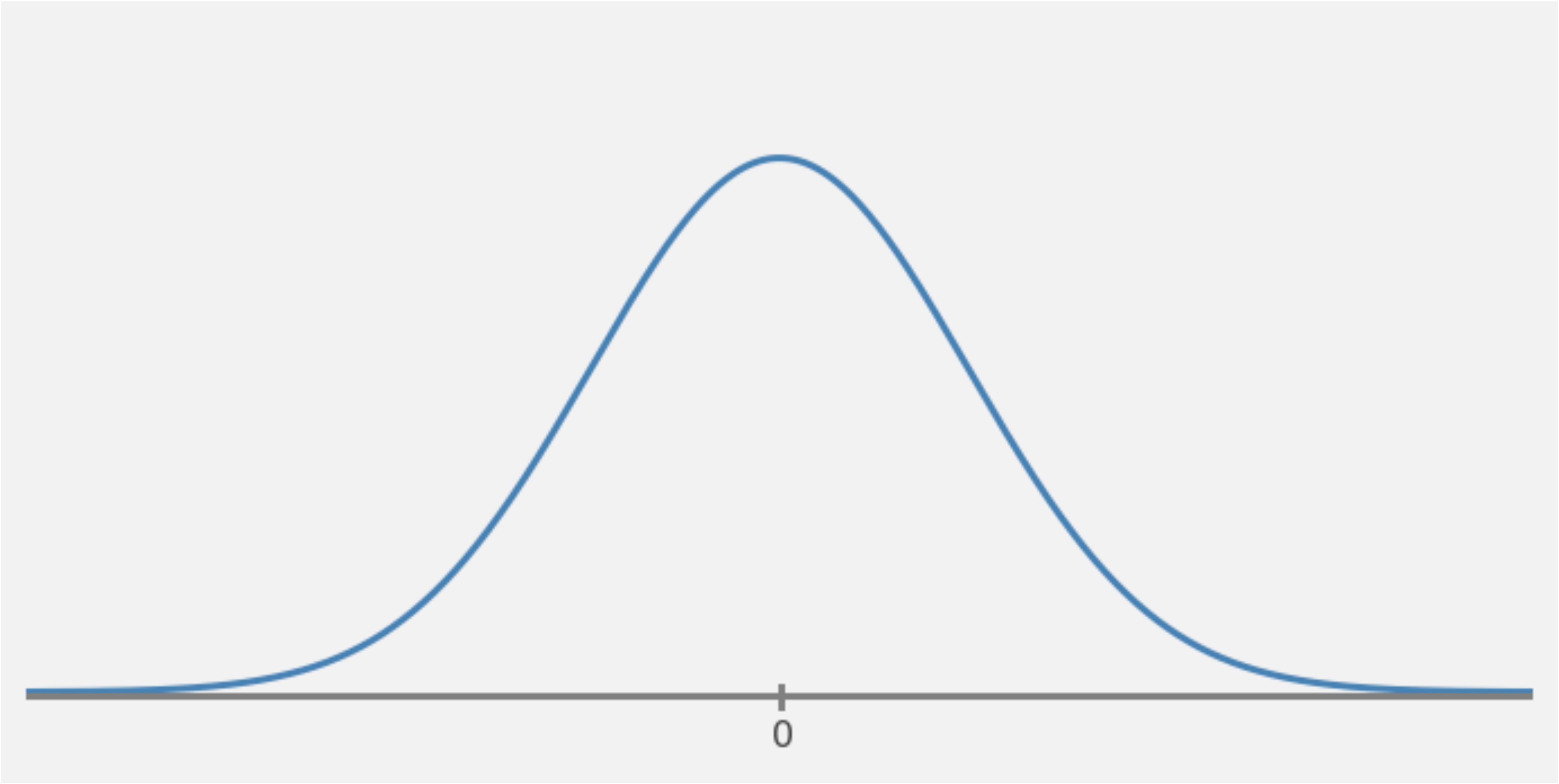
Detecting Biased Coins

- **Example**: To test if the Belgian 1 Euro coin is fair you flip it 100 times and get 38 Heads. Do you reject the Null at the .05 significance level or not?



Detecting Biased Coins

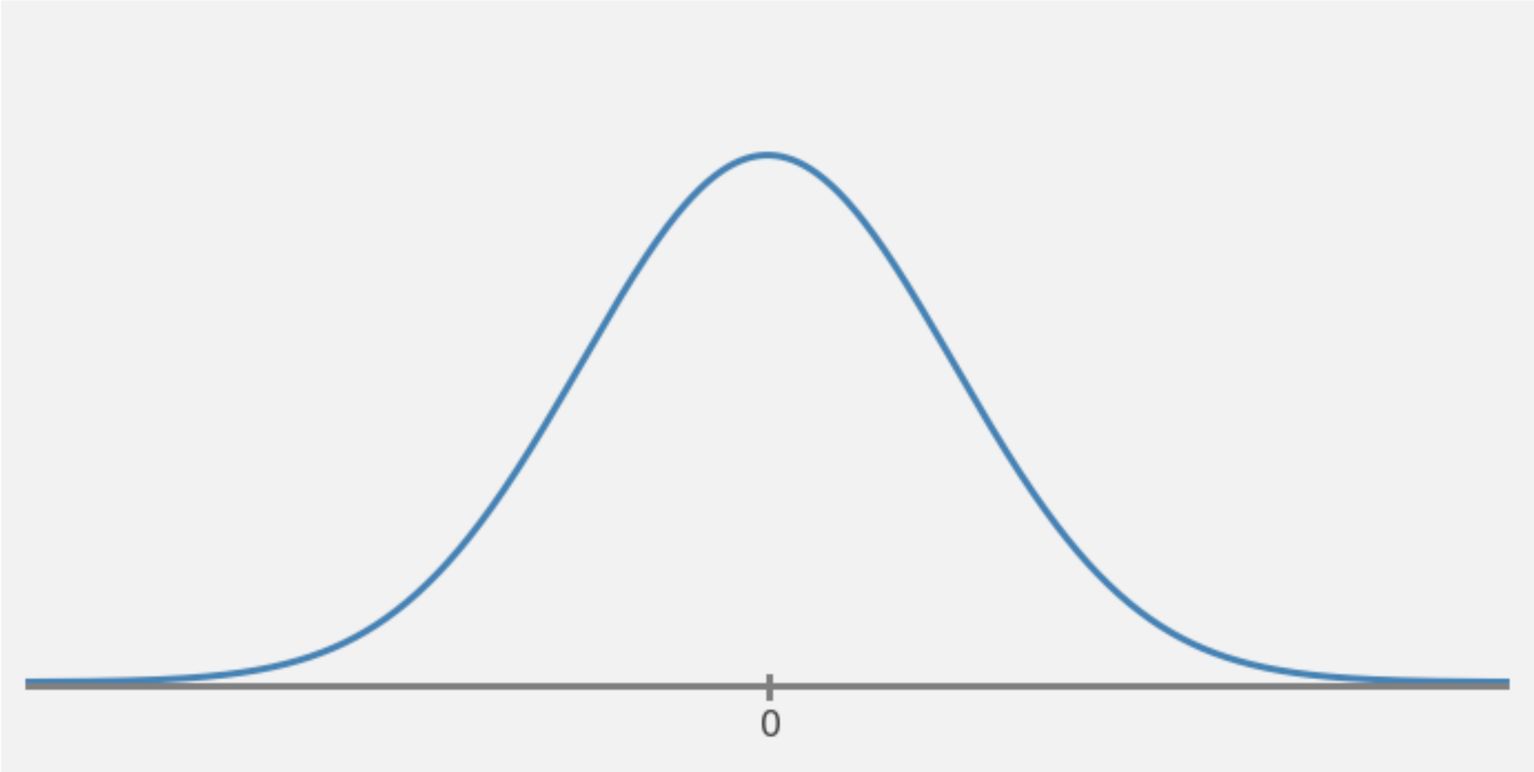
- **Example**: To test if the Belgian 1 Euro coin is fair you flip it 100 times and get 38 Heads. Do you reject the Null at the .01 significance level or not?

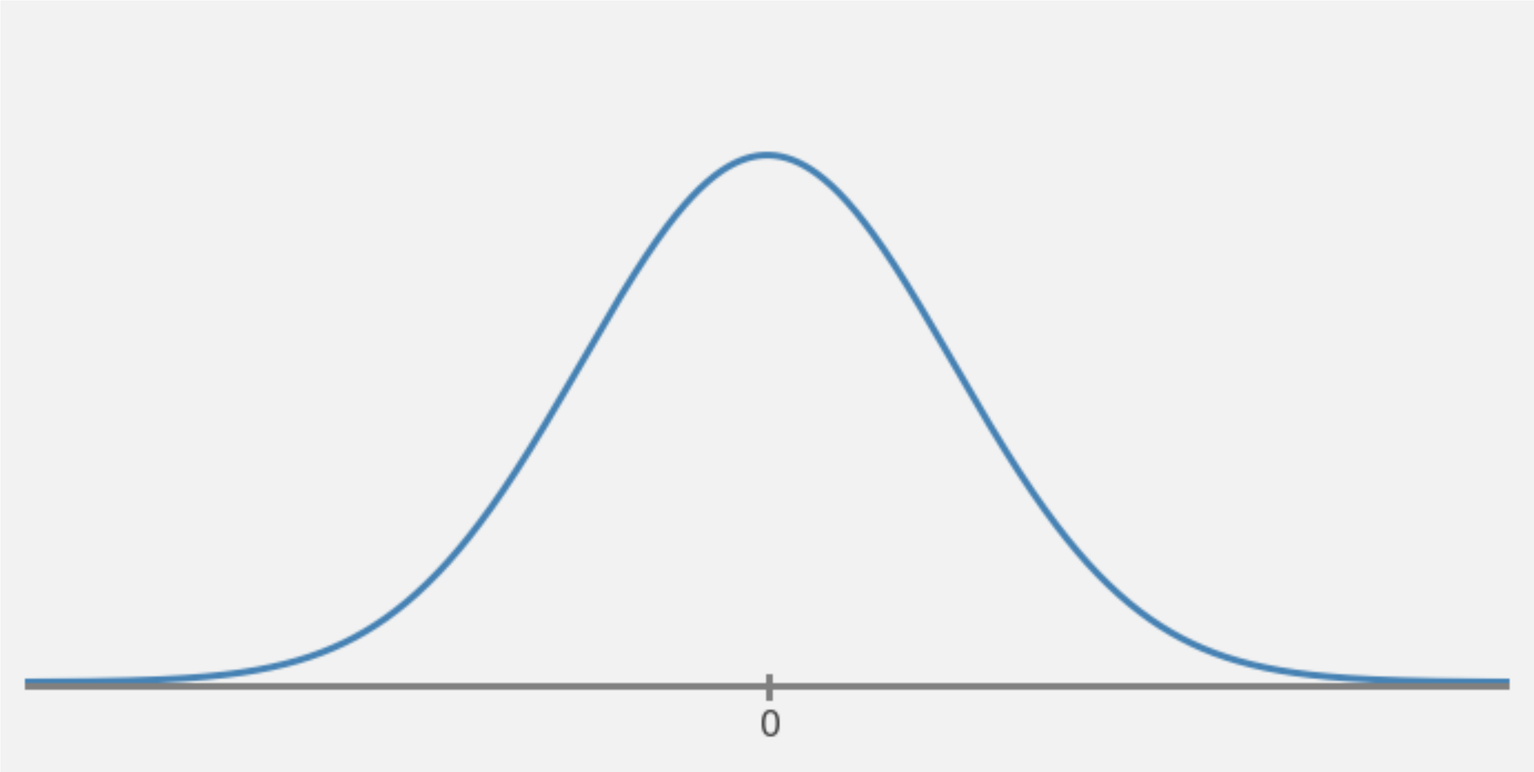


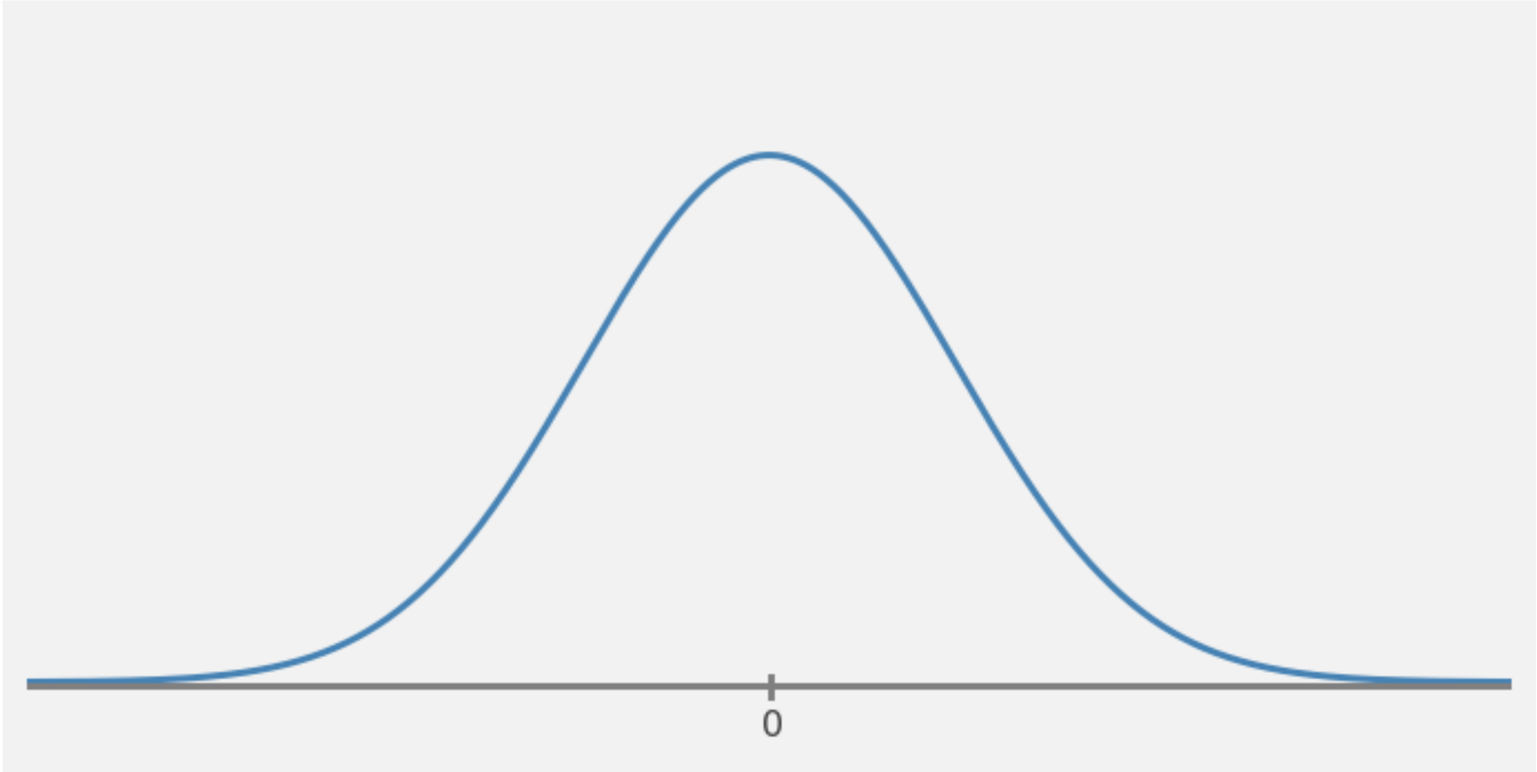
Different tests for different hypotheses

- The coin example was an exampled of a **two-tailed hypothesis test**, because we would have rejected the Null hypothesis had the coin been been biased towards heads OR tails.

Alternative Hypothesis	Rejection Region
$H_1 : \theta > \theta_0$	$z \geq z_\alpha$
$H_1 : \theta < \theta_0$	$z \leq -z_\alpha$
$H_1 : \theta \neq \theta_0$	$z \leq -z_\alpha$ or $z \geq z_\alpha$

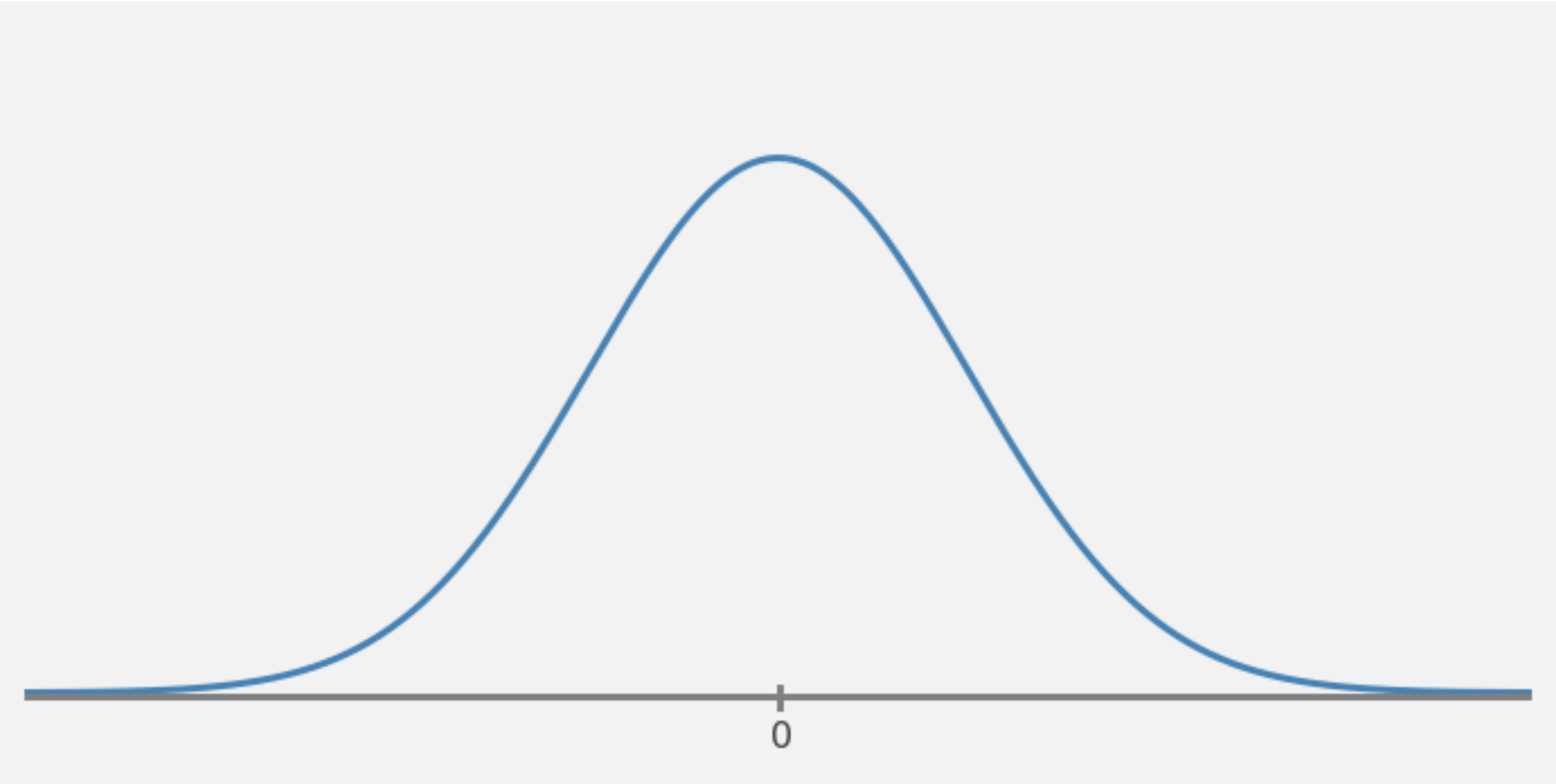






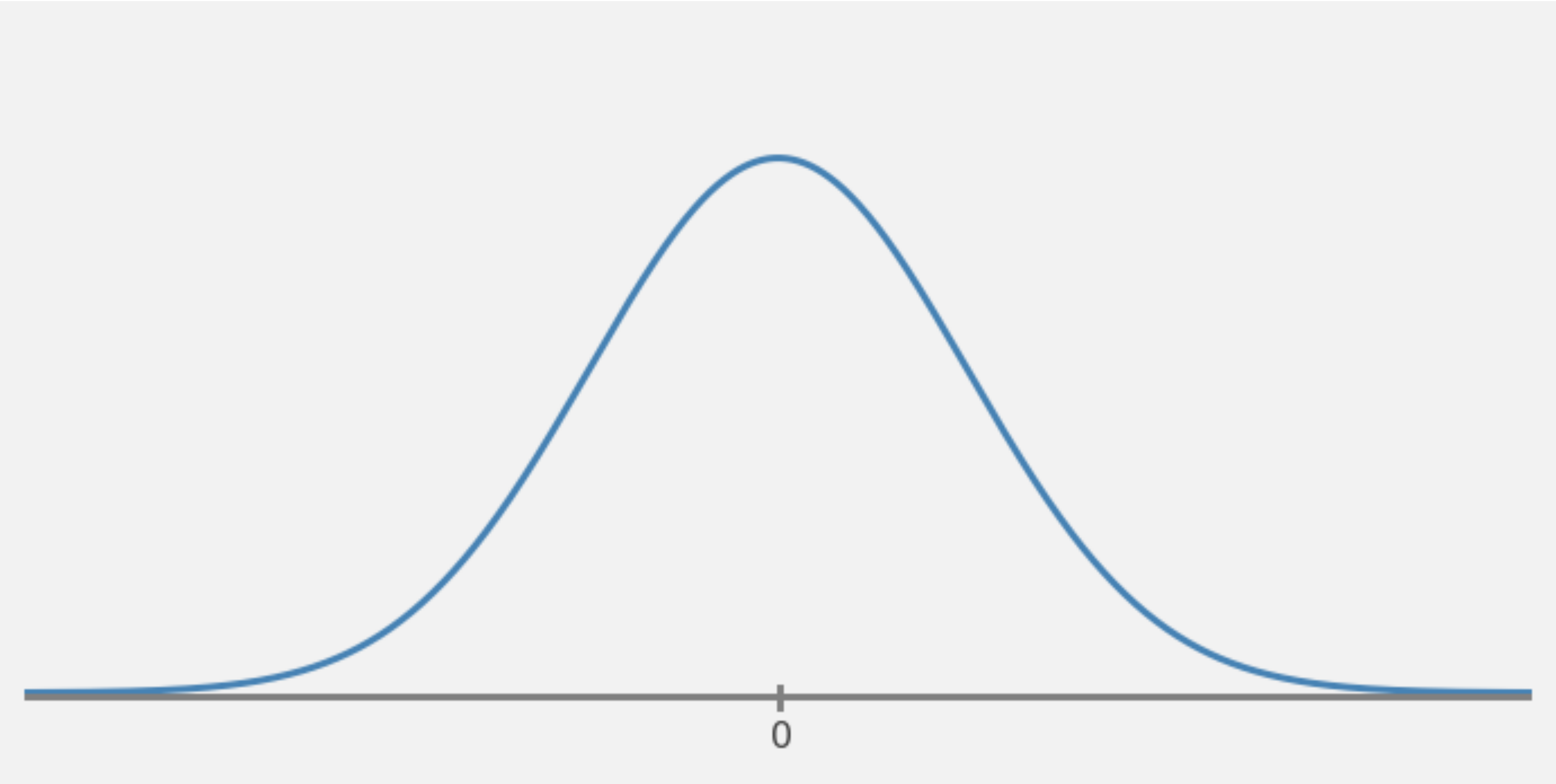
Switching advertising strategies

- **Example**: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200 thousand hits per day with a standard deviation of 50 thousand hits per day. You decide to hire the new ad company for a 30 day trial. During those 30 days, your website gets 210 thousand hits per day. Perform a hypothesis test to determine if the new ad campaign outperforms the old one at the .05 significance level.



Switching advertising strategies

- **Example**: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising they get, on average, 200 thousand hits per day with a standard deviation of 50 thousand hits per day. You decide to hire the new ad company for a 30 day trial. During those 30 days, your website gets 210 thousand hits per day. Perform a hypothesis test to determine if the new ad campaign outperforms the old one at the .05 significance level.



Important assumptions

- **Question**: What assumptions did we make in the previous example?

Errors in hypothesis testing

- **Definitions**:
- A **Type I Error** occurs when the Null hypothesis is rejected, but the Null hypothesis is in fact true (**False Positive**)
- A **Type II Error** occurs when the Null hypothesis is not rejected, but the Null hypothesis is in fact false (**False Negative**)
- **Question**: What is the probability that we commit a **Type I Error**?

Errors in hypothesis testing

- **Definitions**:
- A **Type I Error** occurs when the Null hypothesis is rejected, but the Null hypothesis is in fact true (**False Positive**)
- A **Type II Error** occurs when the Null hypothesis is not rejected, but the Null hypothesis is in fact false (**False Negative**)
- **Question**: What is the probability that we commit a **Type I Error**?
- **Answer**: this is exactly the significance level α
- **Consequence**: choose α by considering willingness to risk a Type I error.

Jetta Rejection

- **Example**: The 1999 Volkswagen Jetta was one of the first VW models produced both in Germany and Mexico. The life expectancy of Jettas produced in Germany was found to follow a normal distribution with mean 300K miles and standard deviation 150K miles. Life expectancy of models made in Mexico were recorded for a sample of size 100. The sample mean of these Jettas was found to be 250K miles.
- **Question 1**: What are the Null hypothesis and alternative hypothesis to test the claim that there is statistical evidence that 1999 Jettas made in Mexico have a smaller life expectancy that those made in Germany?

Jetta Rejection

- **Example**: The 1999 Volkswagen Jetta was one of the first VW models produced both in Germany and Mexico. The life expectancy of Jettas produced in Germany was found to follow a normal distribution with mean 300K miles and standard deviation 150K miles. Life expectancy of models made in Mexico were recorded for a sample of size 100. The sample mean of these Jettas was found to be 250K miles.
- **Question 2**: Is there sufficient evidence to conclude that, in fact, 1999 Jettas made in Mexico have a shorter life expectancy that those made in Germany? Carry out a rejection region test at the 0.01 significance level.