American University of Beirut

STAT 210: Elementary Statistics for Sciences

2022–2023 Fall

Siamak Taati

Chapter 9

Testing hypotheses about mean and proportion

Part 1

Statistical hypothesis testing

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Courtroom

Is the suspect guilty?

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Courtroom

Is the suspect guilty?

Medical research

Is a certain treatment effective?

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Courtroom

Is the suspect guilty?

Medical research

Is a certain treatment effective?

Quality control

Does a certain product have the claimed specification?

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Courtroom

Is the suspect guilty?

Medical research

Is a certain treatment effective?

Quality control

Does a certain product have the claimed specification?

Policy making

Is a certain policy effective?

Testing hypotheses

Very often, we are in a situation that we would like to judge between competing claims in absence of complete information.

Objective

Make a rational judgment based on the available evidence.

In statistical hypothesis testing, the judgment is made based on statistical evidence.

Innocent until Proven Guilty

Is the suspect guilty?

We do not accept the prosecutor's claim of guilt unless it is supported by sufficient evidence.

Innocent until Proven Guilty

Is the suspect guilty?

We do not accept the prosecutor's claim of guilt unless it is supported by sufficient evidence.

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Innocent until Proven Guilty

Is the suspect guilty?

We do not accept the prosecutor's claim of guilt unless it is supported by sufficient evidence.

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Is a certain treatment effective?

We do not accept the effectiveness claim made by a drug company unless it is supported by sufficient evidence.

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Warning

Being skeptical about a claim $\neq$ accepting the opposite claim!

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Warning

Being skeptical about a claim $\neq$ accepting the opposite claim!

- Is a certain treatment effective?

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Warning

Being skeptical about a claim $\neq$ accepting the opposite claim!

- Is a certain treatment effective?
- Is a certain treatment ineffective?

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Warning

Being skeptical about a claim $\neq$ accepting the opposite claim!

- Is a certain treatment effective?
- Is a certain treatment ineffective?

We remain skeptical of both claims unless provided with evidence one way or the other.

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Warning

Being skeptical about a claim $\neq$ accepting the opposite claim!

- Is a certain treatment effective?
- Is a certain treatment ineffective?

We remain skeptical of both claims unless provided with evidence one way or the other.

However, there is also a counter principle:

Skeptical until Provided with Evidence

Principle of Scientific Skepticism

Do not accept a claim until it is supported by evidence!

Warning

Being skeptical about a claim $\neq$ accepting the opposite claim!

- Is a certain treatment effective?
- Is a certain treatment ineffective?

We remain skeptical of both claims unless provided with evidence one way or the other.

However, there is also a counter principle:

Principle of Parsimony (Occam's razor)

Simpler explanations are favourable!

Statistical hypothesis testing

Objective

Judge the validity of a claim based on statistical evidence.

Statistical hypothesis testing

Objective

Judge the validity of a claim based on statistical evidence.

We focus on scenarios in which the statistical evidence comes from a random sample.

Statistical hypothesis testing

Example (Is my coin fair?)

If we flip a fair coin, the two possible outcomes `H` and `T` are equally likely: each occurs with probability $1/2$.

Statistical hypothesis testing

Example (Is my coin fair?)

If we flip a fair coin, the two possible outcomes `H` and `T` are equally likely: each occurs with probability $1/2$.

So, if we flip the same fair coin a large number of times, we expect to see heads about 50% of the times.

Statistical hypothesis testing

Example (Is my coin fair?)

If we flip a fair coin, the two possible outcomes `H` and `T` are equally likely: each occurs with probability $1/2$.

So, if we flip the same fair coin a large number of times, we expect to see heads about 50% of the times.

Suppose we have a coin, and we do not know whether it is fair or unfair.

Statistical hypothesis testing

Example (Is my coin fair?)

If we flip a fair coin, the two possible outcomes `H` and `T` are equally likely: each occurs with probability $1/2$.

So, if we flip the same fair coin a large number of times, we expect to see heads about 50% of the times.

Suppose we have a coin, and we do not know whether it is fair or unfair.

We flip the coin 1000 times and we get 550 heads.

Statistical hypothesis testing

Example (Is my coin fair?)

If we flip a fair coin, the two possible outcomes `H` and `T` are equally likely: each occurs with probability $1/2$.

So, if we flip the same fair coin a large number of times, we expect to see heads about 50% of the times.

Suppose we have a coin, and we do not know whether it is fair or unfair.

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not?

A Based on the available evidence we cannot be 100% sure one way or the other. Indeed, 550 heads out of 1000 flips is a possible event, whether the coin is fair or not.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not?

A Based on the available evidence we cannot be 100% sure one way or the other. Indeed, 550 heads out of 1000 flips is a possible event, whether the coin is fair or not.

Nevertheless, the result of the experiment contains some information about the coin.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not?

A Based on the available evidence we cannot be 100% sure one way or the other. Indeed, 550 heads out of 1000 flips is a possible event, whether the coin is fair or not.

Nevertheless, the result of the experiment contains some information about the coin. Is this information enough to judge the coin unfair with a reasonable degree of confidence?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.
- We flip the coin 1000 times and we get 550 heads.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.
- We flip the coin 1000 times and we get 550 heads.
- We flip the coin 1000000 times and we get 550000 heads.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.
- We flip the coin 1000 times and we get 550 heads.
- We flip the coin 1000000 times and we get 550000 heads.

The first scenario hardly contains any indication of the unfairness of the coin.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.
- We flip the coin 1000 times and we get 550 heads.
- We flip the coin 1000000 times and we get 550000 heads.

The first scenario hardly contains any indication of the unfairness of the coin. The last scenario is (for all practical purposes) a conclusive evidence that the coin is biased.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.
- We flip the coin 1000 times and we get 550 heads.
- We flip the coin 1000000 times and we get 550000 heads.

The first scenario hardly contains any indication of the unfairness of the coin. The last scenario is (for all practical purposes) a conclusive evidence that the coin is biased.

Where does 550-out-of-1000 stand?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Ⓠ Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

To appreciate the difficulty, consider the following hypothetical scenarios:

- We flip the coin 10 times and we get 6 heads.
- We flip the coin 100 times and we get 55 heads.
- We flip the coin 1000 times and we get 550 heads.
- We flip the coin 1000000 times and we get 550000 heads.

The first scenario hardly contains any indication of the unfairness of the coin. The last scenario is (for all practical purposes) a conclusive evidence that the coin is biased.

Where does 550-out-of-1000 stand?

We need a quantitative assessment of the strength of the evidence!

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A1 Consider the random experiment of flipping a fair coin 1000 times.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A1 Consider the random experiment of flipping a fair coin 1000 times. Let X denote the number of heads.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A1 Consider the random experiment of flipping a fair coin 1000 times. Let X denote the number of heads.

Q What is the distribution of X?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A1 Consider the random experiment of flipping a fair coin 1000 times. Let X denote the number of heads.

Q What is the distribution of X?

A Binomial with parameters $n = 1000$ and $p = 0.5$

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A1 Consider the random experiment of flipping a fair coin 1000 times. Let X denote the number of heads.

Compare the distribution of X with the observed value 550.

Statistical hypothesis testing

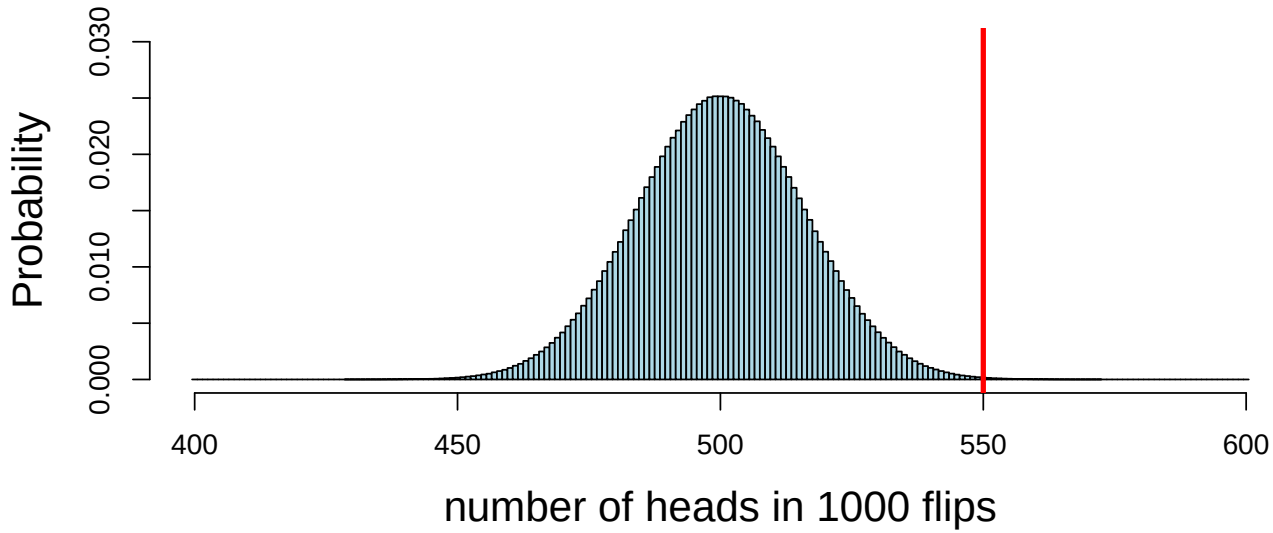
Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A1 Consider the random experiment of flipping a fair coin 1000 times. Let X denote the number of heads.

Compare the distribution of X with the observed value 550.

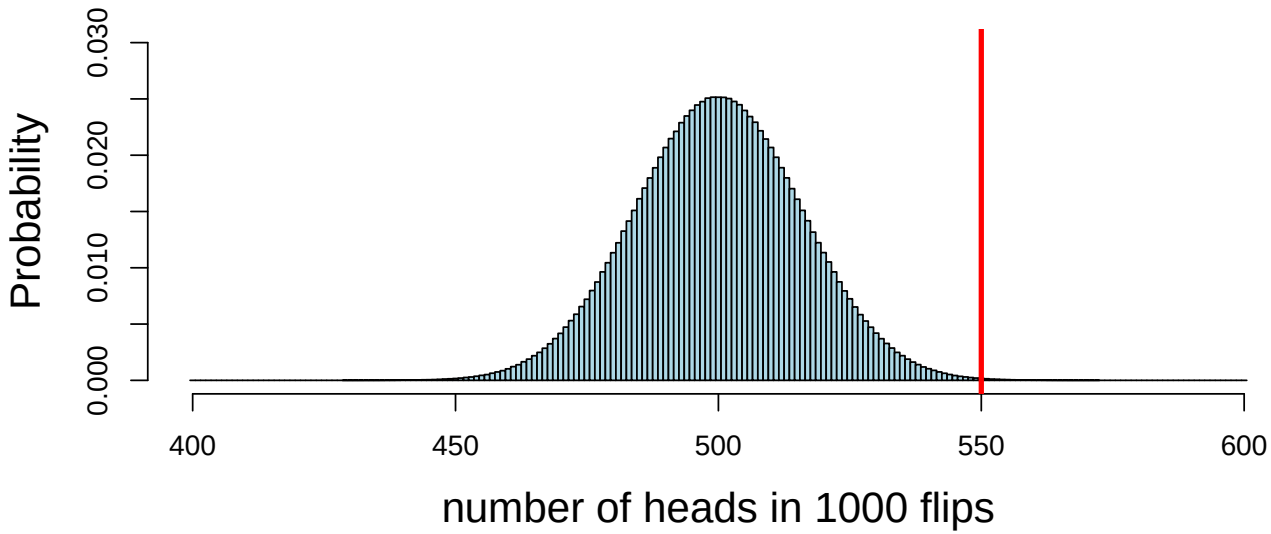


Distribution of X

Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.



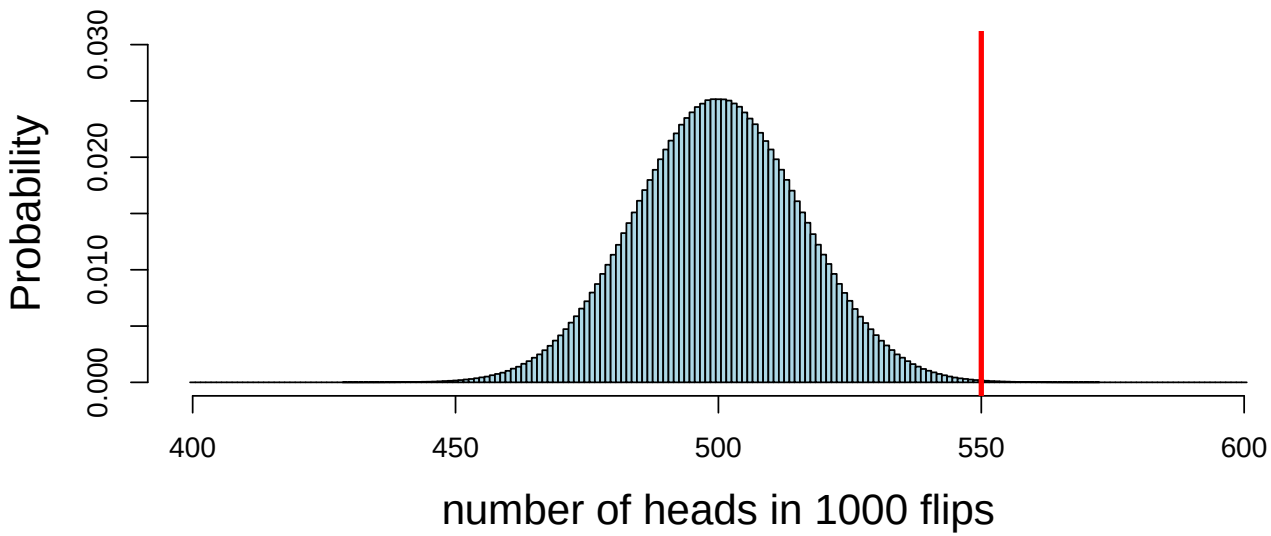
Distribution of X

Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?



Distribution of X

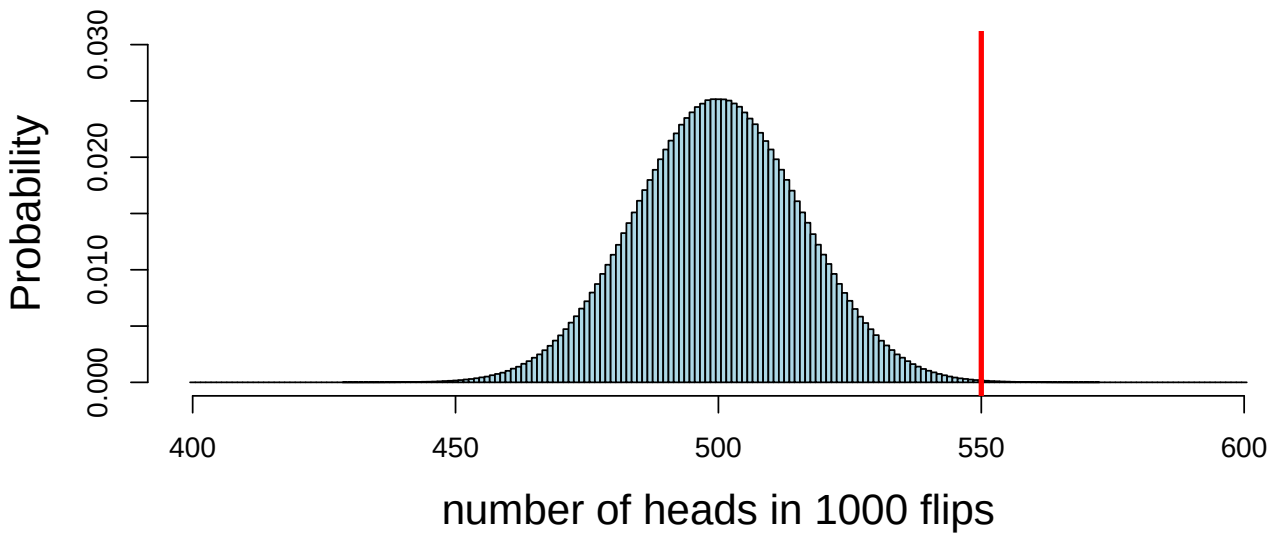
Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?

A $\mathbb{P}(X \text{ is at least as extreme as } 550)$



Distribution of X

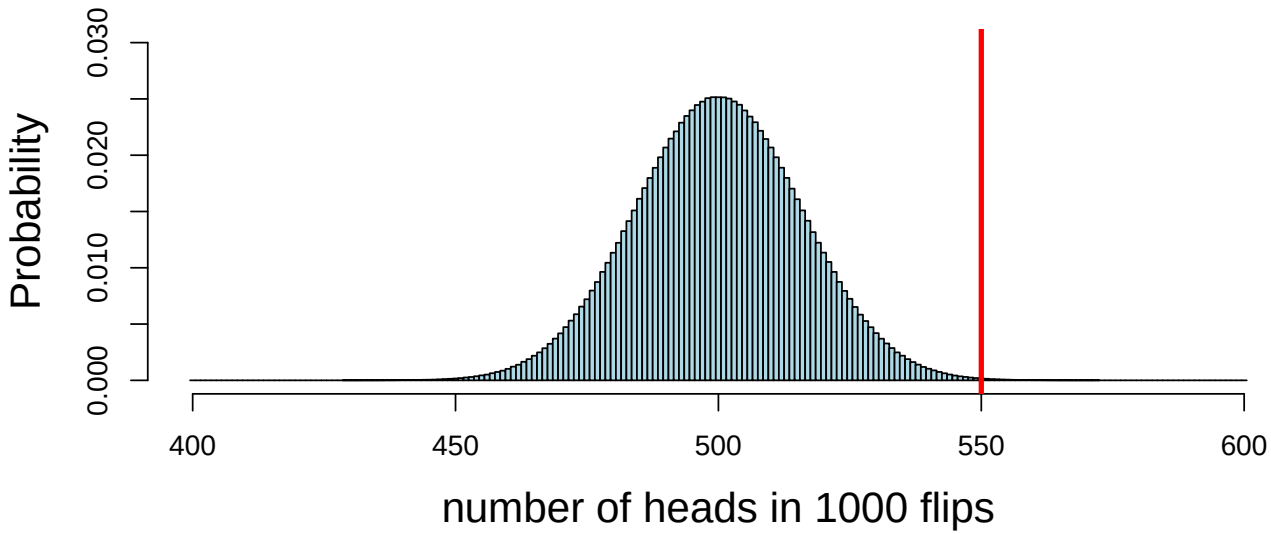
Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?

A
$$\begin{aligned}&\mathbb{P}(X \text{ is at least as extreme as } 550)\\ &= \mathbb{P}(X \geq 550) + \mathbb{P}(X \leq 450)\\ &\approx\end{aligned}$$



Distribution of X

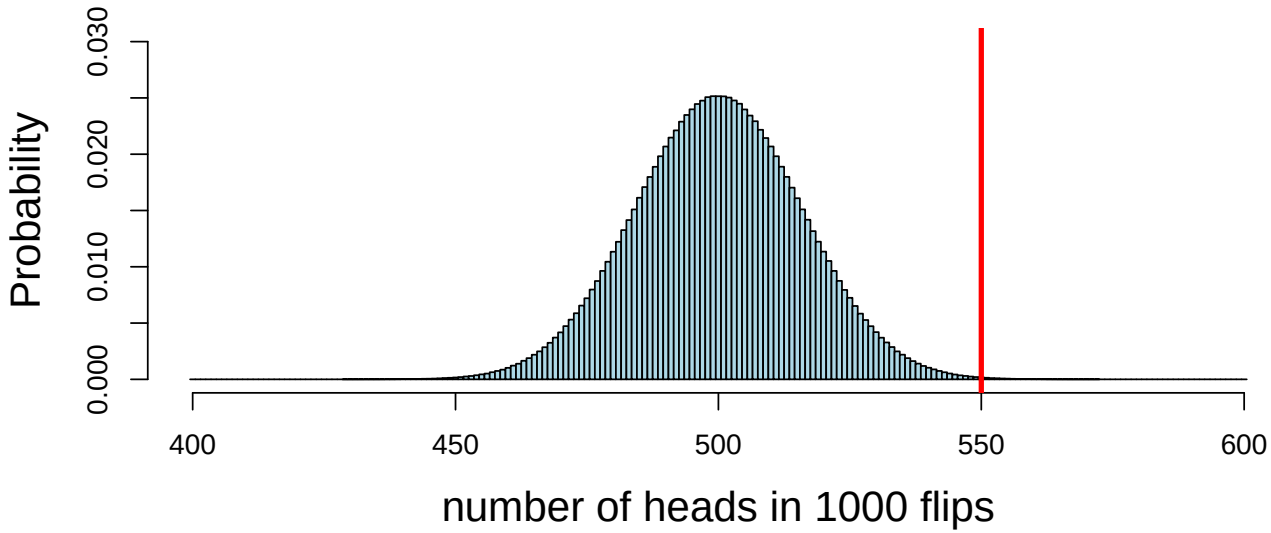
Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?

A
$$\begin{aligned}&\mathbb{P}(X \text{ is at least as extreme as } 550)\\ &= \mathbb{P}(X \geq 550) + \mathbb{P}(X \leq 450)\\ &\approx 0.00087 + 0.00087 = \boxed{0.00174} \approx 0.2\%\end{aligned}$$



Distribution of X

Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?

A
$$\begin{aligned}\mathbb{P}(X \text{ is at least as extreme as } 550) &= \mathbb{P}(X \geq 550) + \mathbb{P}(X \leq 450) \\ &\approx 0.00087 + 0.00087 = \boxed{0.00174} \approx 0.2\%\end{aligned}$$

Q Which is more plausible?

Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?

A
$$\begin{aligned}\mathbb{P}(X \text{ is at least as extreme as } 550) &= \mathbb{P}(X \geq 550) + \mathbb{P}(X \leq 450) \\ &\approx 0.00087 + 0.00087 = \boxed{0.00174} \approx 0.2\%\end{aligned}$$

Q Which is more plausible?

- H_0 The coin is fair. The observed deviation is due to chance.
- H_1 The coin is biased.

Statistical hypothesis testing

Example (Is my coin fair?)

A1 Compare the distribution of X with the observed value 550.

Q What is the chance that the value of X is at least as extreme as 550?

A
$$\begin{aligned}\mathbb{P}(X \text{ is at least as extreme as } 550) &= \mathbb{P}(X \geq 550) + \mathbb{P}(X \leq 450) \\ &\approx 0.00087 + 0.00087 = \boxed{0.00174} \approx 0.2\%\end{aligned}$$

Q Which is more plausible?

- H_0 The coin is fair. The observed deviation is due to chance.
- H_1 The coin is biased.

Here, H_1 appears more plausible,
but in order to make a fair and rigorous judgment, we need to decide on a criterion before seeing the evidence!

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A2 Consider the random experiment of flipping a fair coin 1000 times.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A2 Consider the random experiment of flipping a fair coin 1000 times. Let $\widehat{p}$ denote the proportion of heads.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A2 Consider the random experiment of flipping a fair coin 1000 times. Let $\widehat{p}$ denote the proportion of heads.

Q What is the distribution of $\widehat{p}$?

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A2 Consider the random experiment of flipping a fair coin 1000 times. Let $\widehat{p}$ denote the proportion of heads.

Q What is the distribution of $\widehat{p}$?

A Approximately normal with $\mu = 0.5$ and $\sigma = \sqrt{\frac{0.5(1-0.5)}{1000}} \approx 0.01581139$.

Statistical hypothesis testing

Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A2 Consider the random experiment of flipping a fair coin 1000 times. Let $\widehat{p}$ denote the proportion of heads.

Compare the distribution of $\widehat{p}$ with the observed value $\frac{550}{1000} = 0.55$.

Statistical hypothesis testing

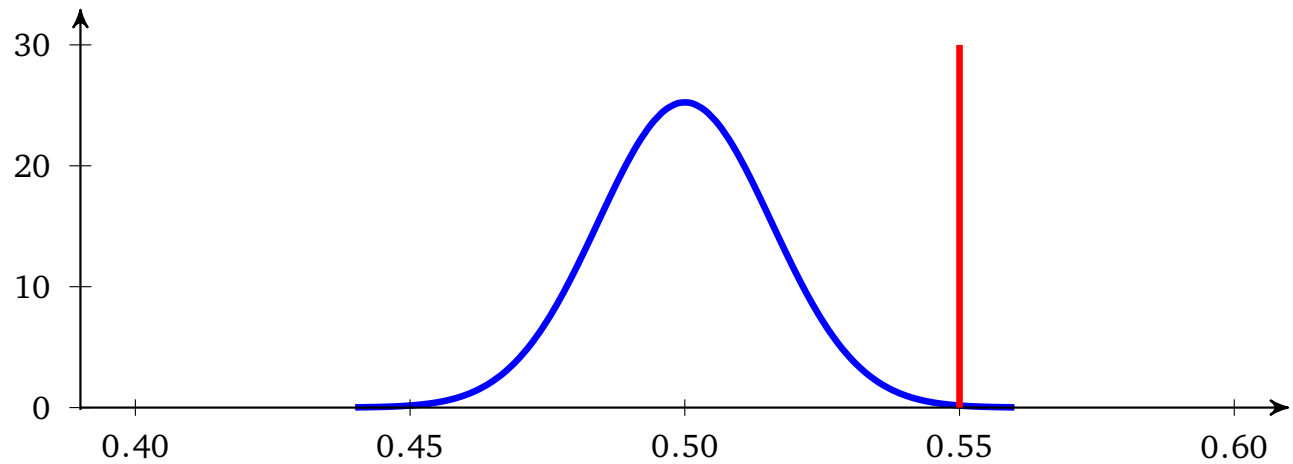
Example (Is my coin fair?)

We flip the coin 1000 times and we get 550 heads.

Q Based on this evidence, can we judge whether the coin is fair or not with a reasonable degree of confidence?

A2 Consider the random experiment of flipping a fair coin 1000 times. Let $\widehat{p}$ denote the proportion of heads.

Compare the distribution of $\widehat{p}$ with the observed value $\frac{550}{1000} = 0.55$.

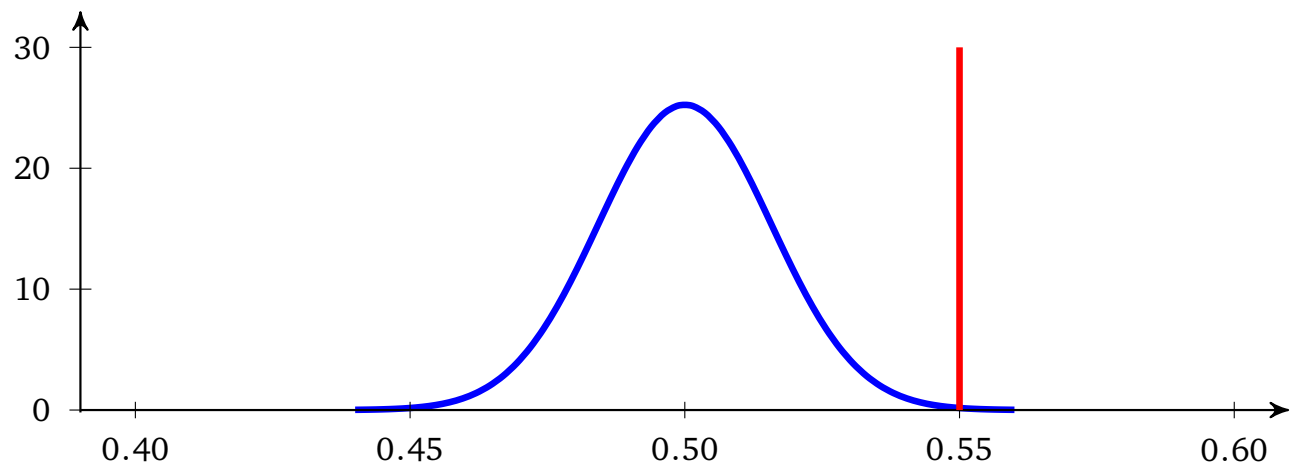


Distribution of $\widehat{p}$

Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.



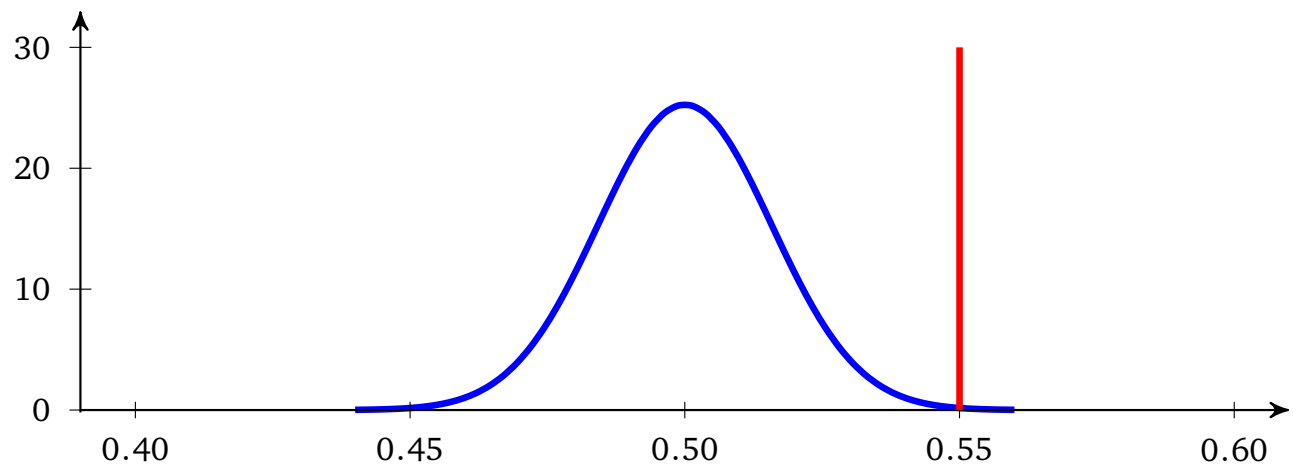
Distribution of $\widehat{p}$

Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?



Distribution of $\widehat{p}$

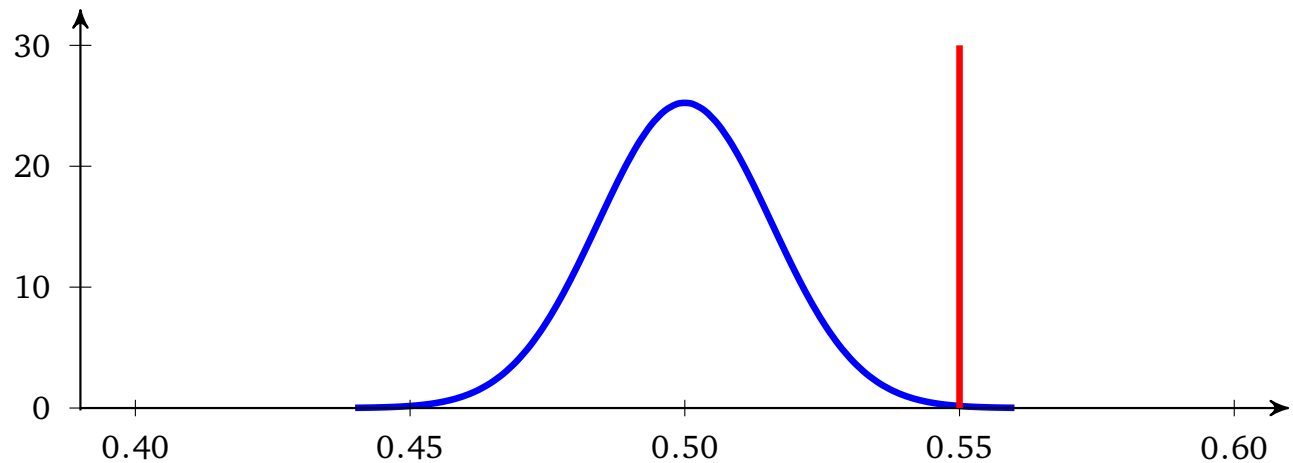
Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?

A $\mathbb{P}(\widehat{p}$ is at least as extreme as $0.55)$



Distribution of $\widehat{p}$

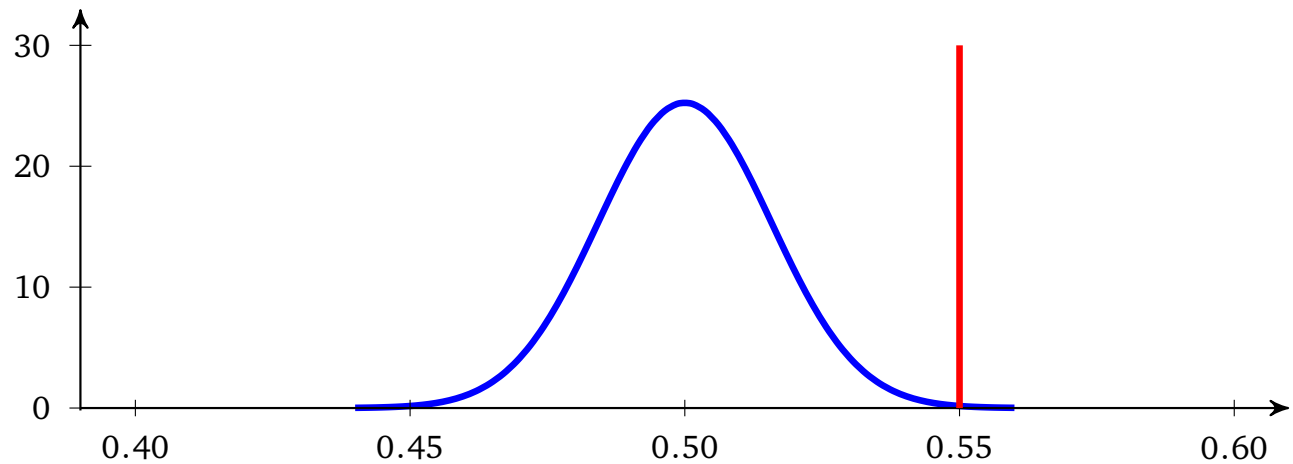
Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?

A
$$\begin{aligned}\mathbb{P}(\widehat{p} \text{ is at least as extreme as } 0.55) &= \mathbb{P}(\widehat{p} \geq 0.55) + \mathbb{P}(\widehat{p} \leq 0.45) \\ &\approx\end{aligned}$$



Distribution of $\widehat{p}$

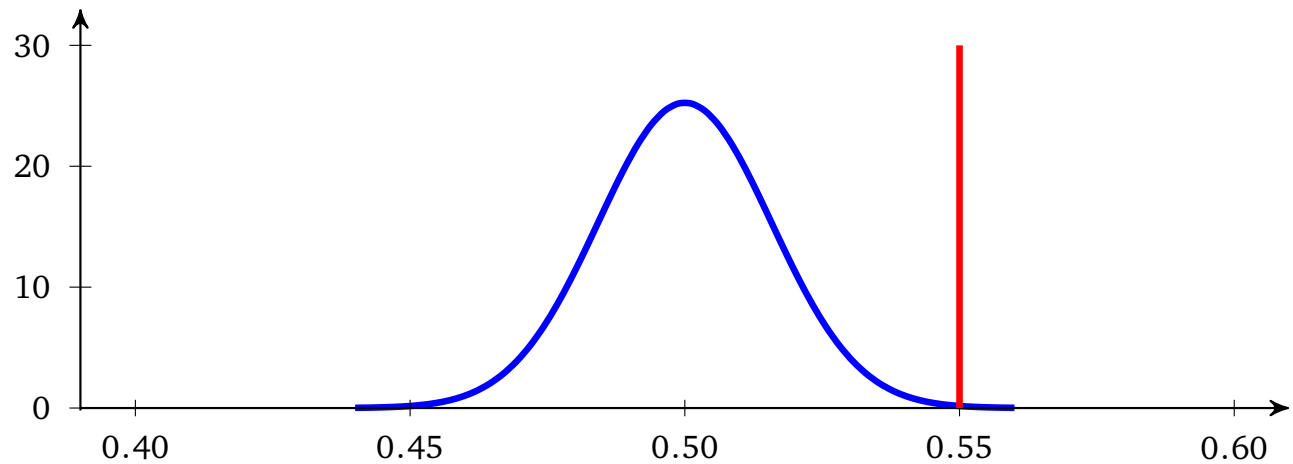
Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?

A
$$\begin{aligned}\mathbb{P}(\widehat{p} \text{ is at least as extreme as } 0.55) &= \mathbb{P}(\widehat{p} \geq 0.55) + \mathbb{P}(\widehat{p} \leq 0.45) \\ &\approx 0.00078 + 0.00078 = \boxed{0.00156} \approx 0.2\%\end{aligned}$$



Distribution of $\widehat{p}$

Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?

A
$$\begin{aligned}\mathbb{P}(\widehat{p} \text{ is at least as extreme as } 0.55) &= \mathbb{P}(\widehat{p} \geq 0.55) + \mathbb{P}(\widehat{p} \leq 0.45) \\ &\approx 0.00078 + 0.00078 = \boxed{0.00156} \approx 0.2\%\end{aligned}$$

Q Which is more plausible?

Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?

A
$$\begin{aligned}\mathbb{P}(\widehat{p} \text{ is at least as extreme as } 0.55) &= \mathbb{P}(\widehat{p} \geq 0.55) + \mathbb{P}(\widehat{p} \leq 0.45) \\ &\approx 0.00078 + 0.00078 = \boxed{0.00156} \approx 0.2\%\end{aligned}$$

Q Which is more plausible?

- H_0 The coin is fair. The observed deviation is due to chance.
- H_1 The coin is biased.

Statistical hypothesis testing

Example (Is my coin fair?)

A2 Compare the distribution of $\widehat{p}$ with the observed value 0.55.

Q What is the chance that the value of $\widehat{p}$ is at least as extreme as 0.55?

A
$$\begin{aligned}\mathbb{P}(\widehat{p} \text{ is at least as extreme as } 0.55) &= \mathbb{P}(\widehat{p} \geq 0.55) + \mathbb{P}(\widehat{p} \leq 0.45) \\ &\approx 0.00078 + 0.00078 = \boxed{0.00156} \approx 0.2\%\end{aligned}$$

Q Which is more plausible?

- H_0 The coin is fair. The observed deviation is due to chance.
- H_1 The coin is biased.

Here, H_1 appears more plausible,
but in order to make a fair and rigorous judgment, we need to decide on a criterion before seeing the evidence!