Hypothesis Testing

Computational Social Intelligence - Lecture 03

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

http://www.dcs.gla.ac.uk/vincia
alessandro.Vinciarelli@glasgow.ac.uk

This lecture is based on the following text (available on Moodle):

- D.C.Howell, "Statistical Methods for Psychology", Chapter 4, pp. 92-109, Cengage Learning, 2009.

Outline

- Sampling Distributions
- Hypothesis Testing
- The Null Hypothesis
- Error Types
- Directionality

Outline

- Sampling Distributions
- Hypothesis Testing
- The Null Hypothesis
- Error Types
- Directionality

The height of individual i in the population

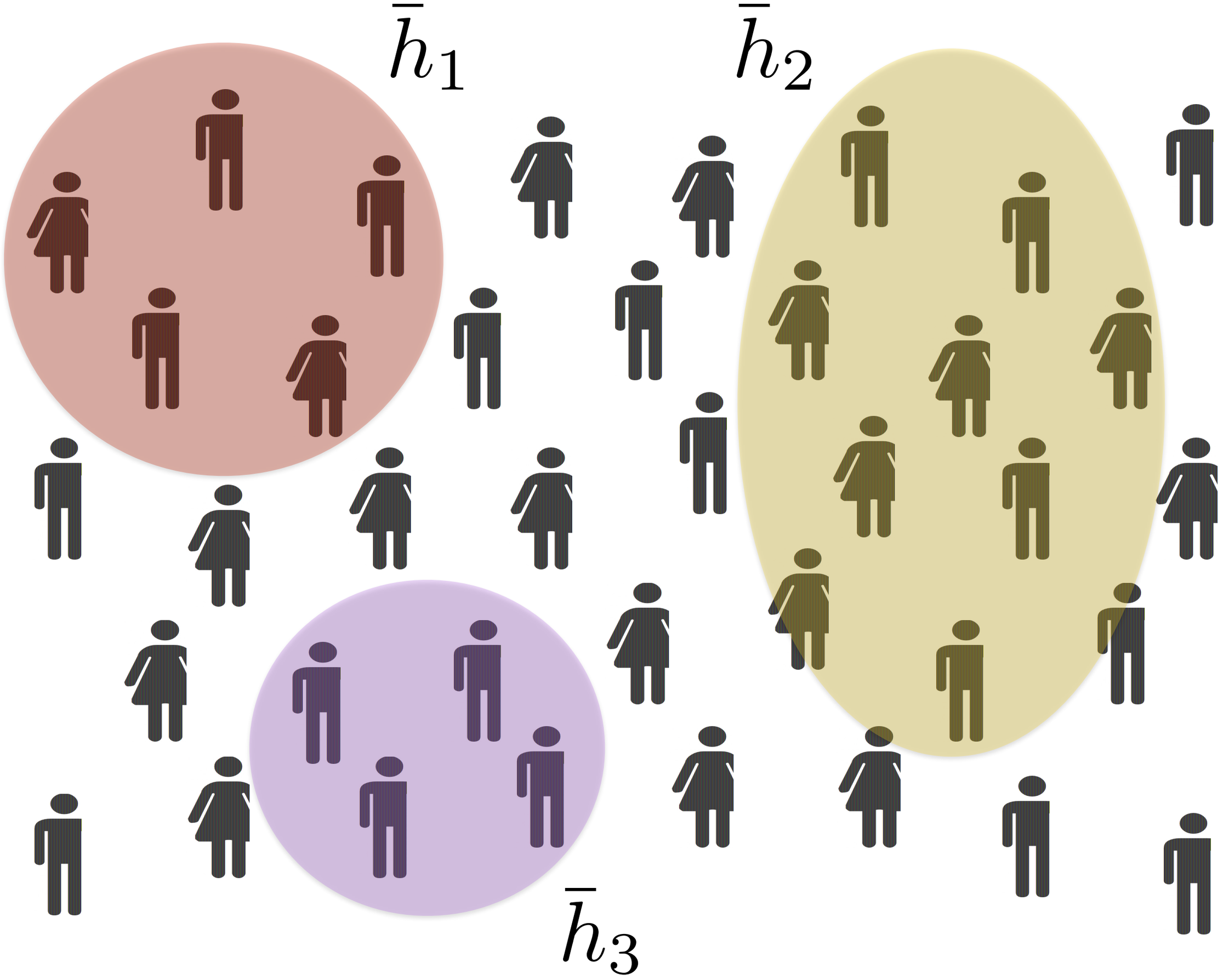
$$\bar{h} = \frac{1}{N} \sum_{i=1}^{N} h_i$$

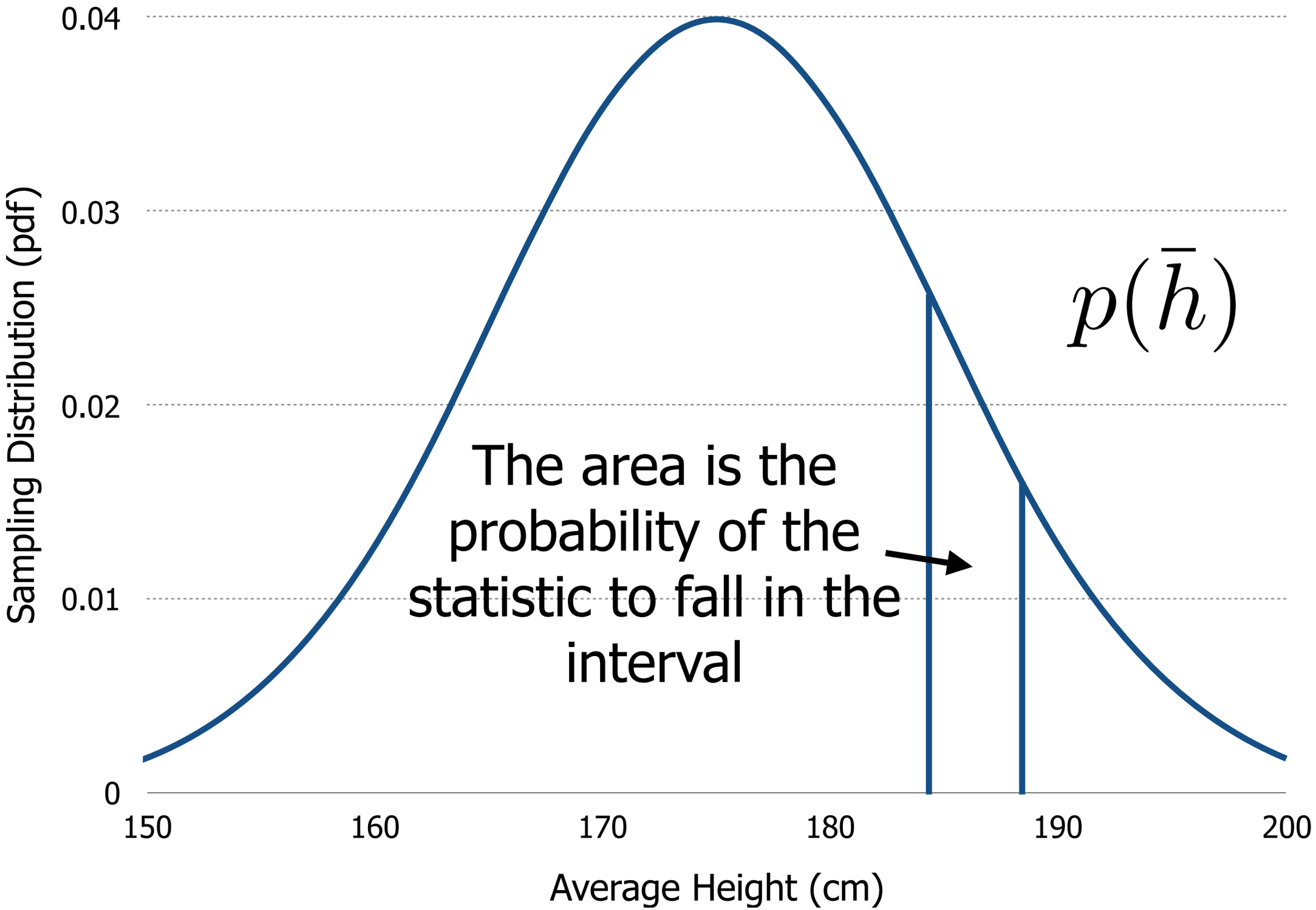
The average height of the individuals in the population

Sampling Distributions

"[…] the distribution of values obtained for [a] statistic over repeated sampling (i.e. running the experiment, or drawing samples, an unlimited number of times)."

D.C.Howell, "Statistical Methods for Psychology", Chapter 4, Cengage Learning, 2009.


$\bar{h}_1$
$\bar{h}_2$
$\bar{h}_3$

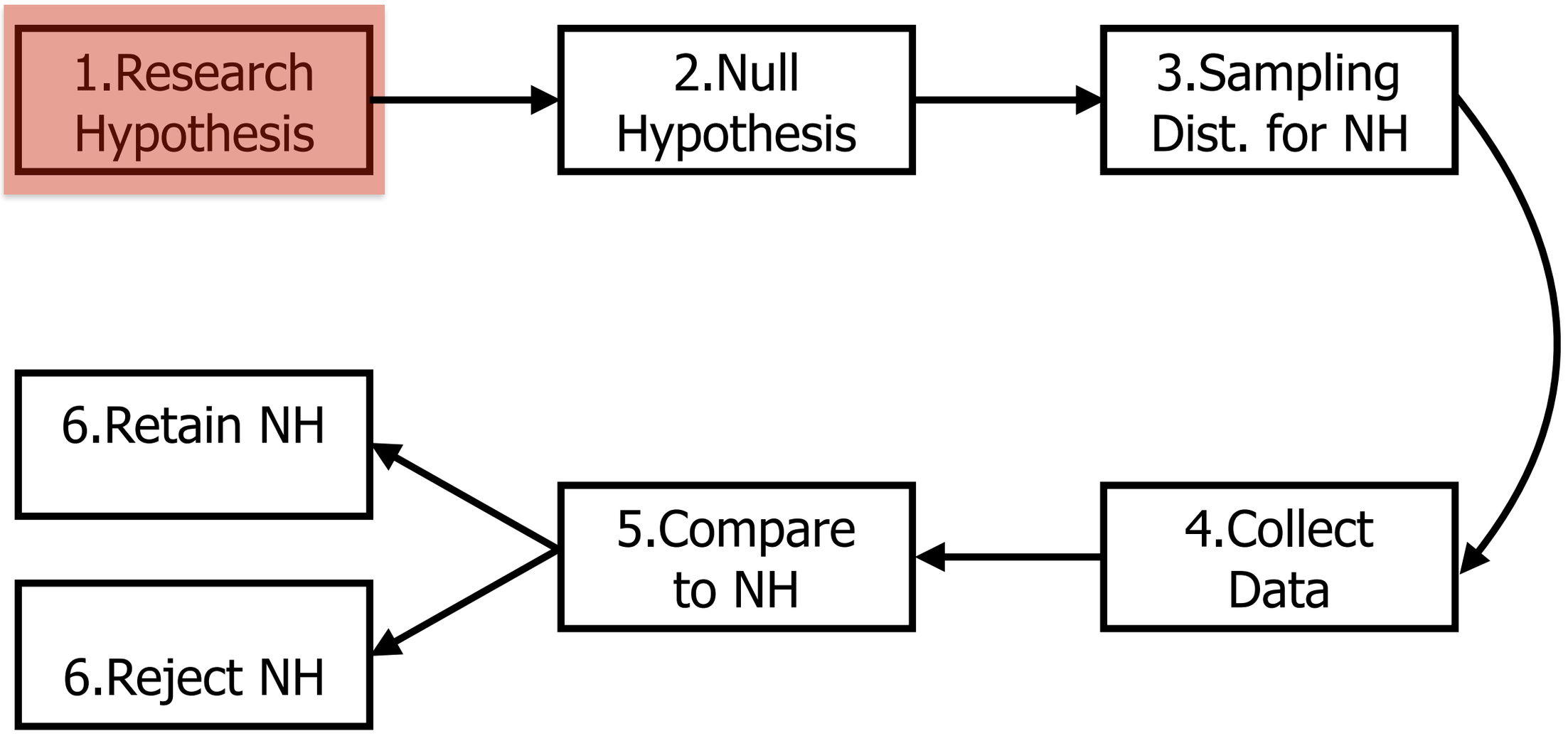

0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
$p(\bar{h})$
The area is the probability of the statistic to fall in the interval
150
160
170
180
190
200
Average Height (cm)

Outline

- Sampling Distributions
- Hypothesis Testing
  - The Null Hypothesis
  - Error Types
  - Directionality

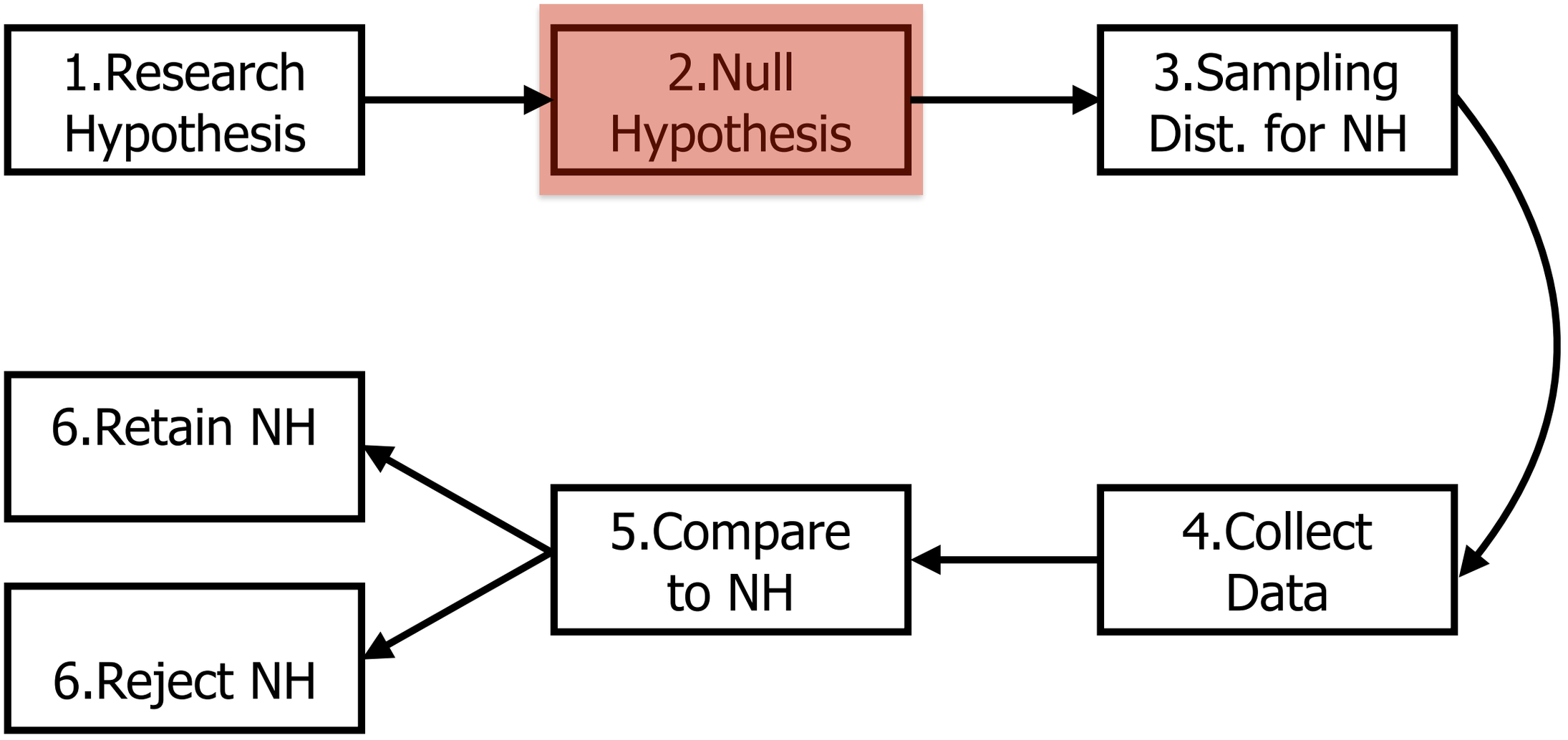
Hypotesis Testing

- State a research hypothesis;
- Setup the null hypothesis (NH);
- Construct the sampling distribution of the statistic when the null hypothesis is true;
- Collect data;
- Compare sample statistic to its distribution when the null hypothesis is true;
- Retain or reject the null hypothesis.


1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

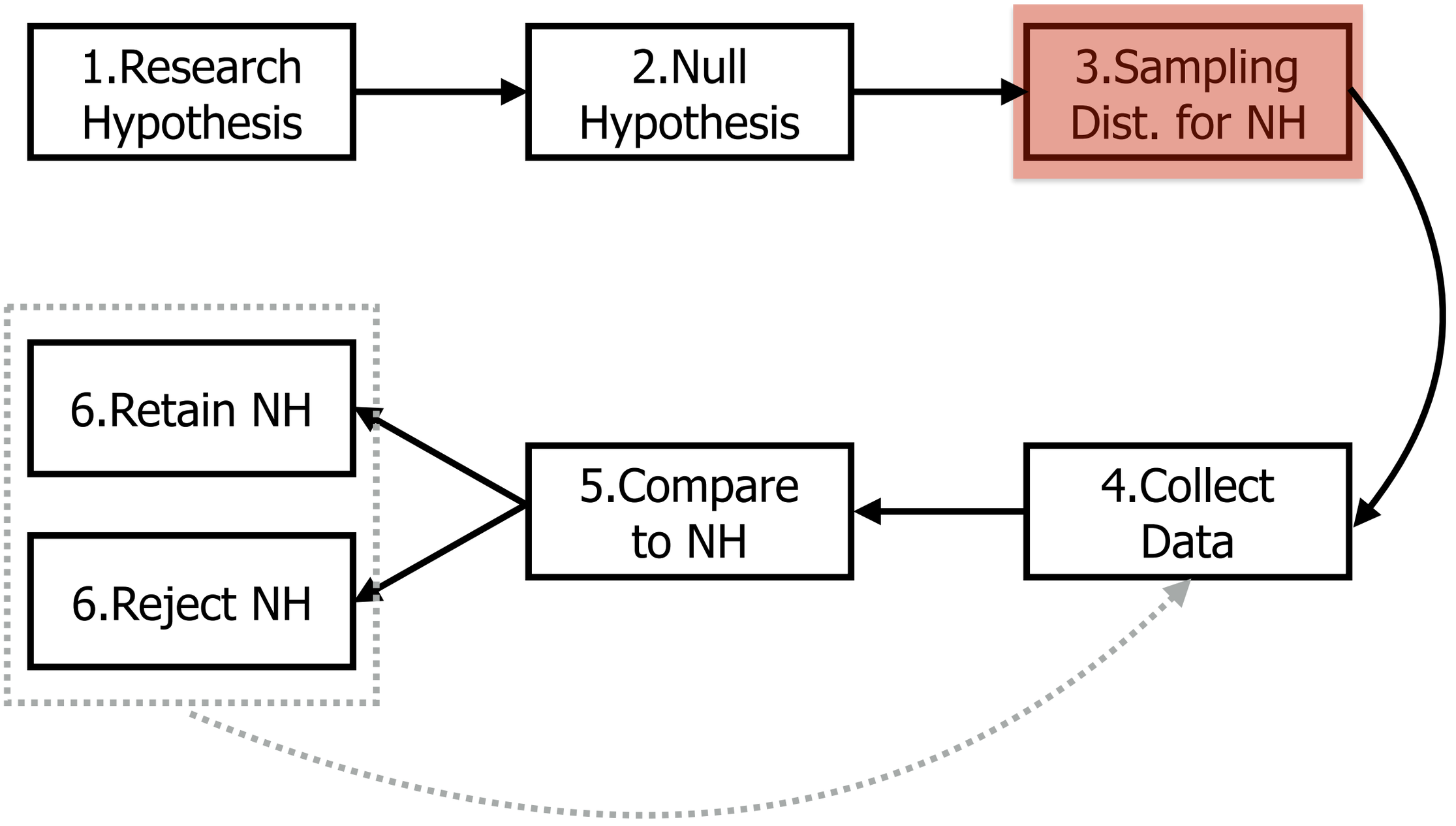
Research Hypothesis

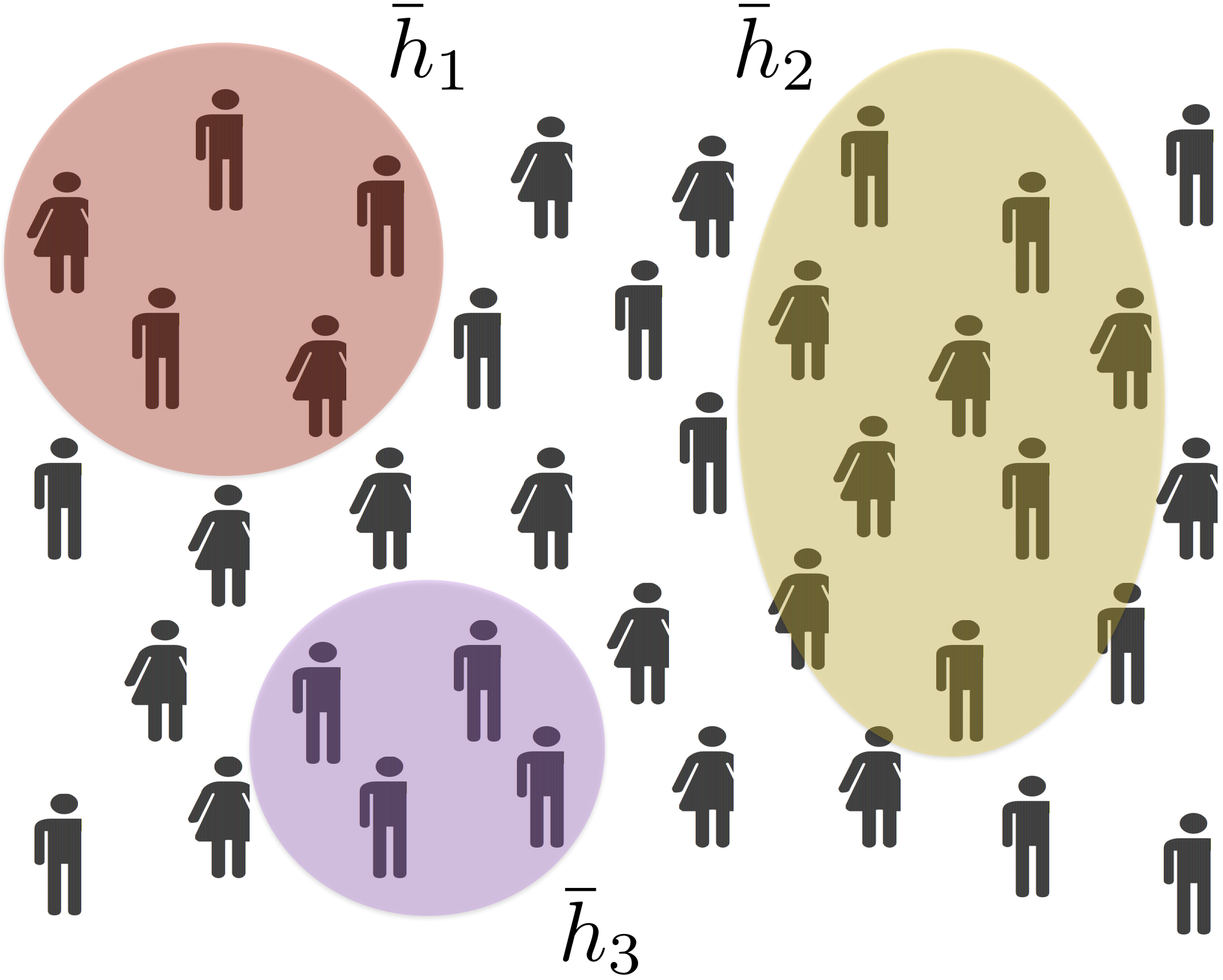
- Basketball players tend to be taller than the rest of the population;
- Primary school children tend to be smaller than the rest of the population.

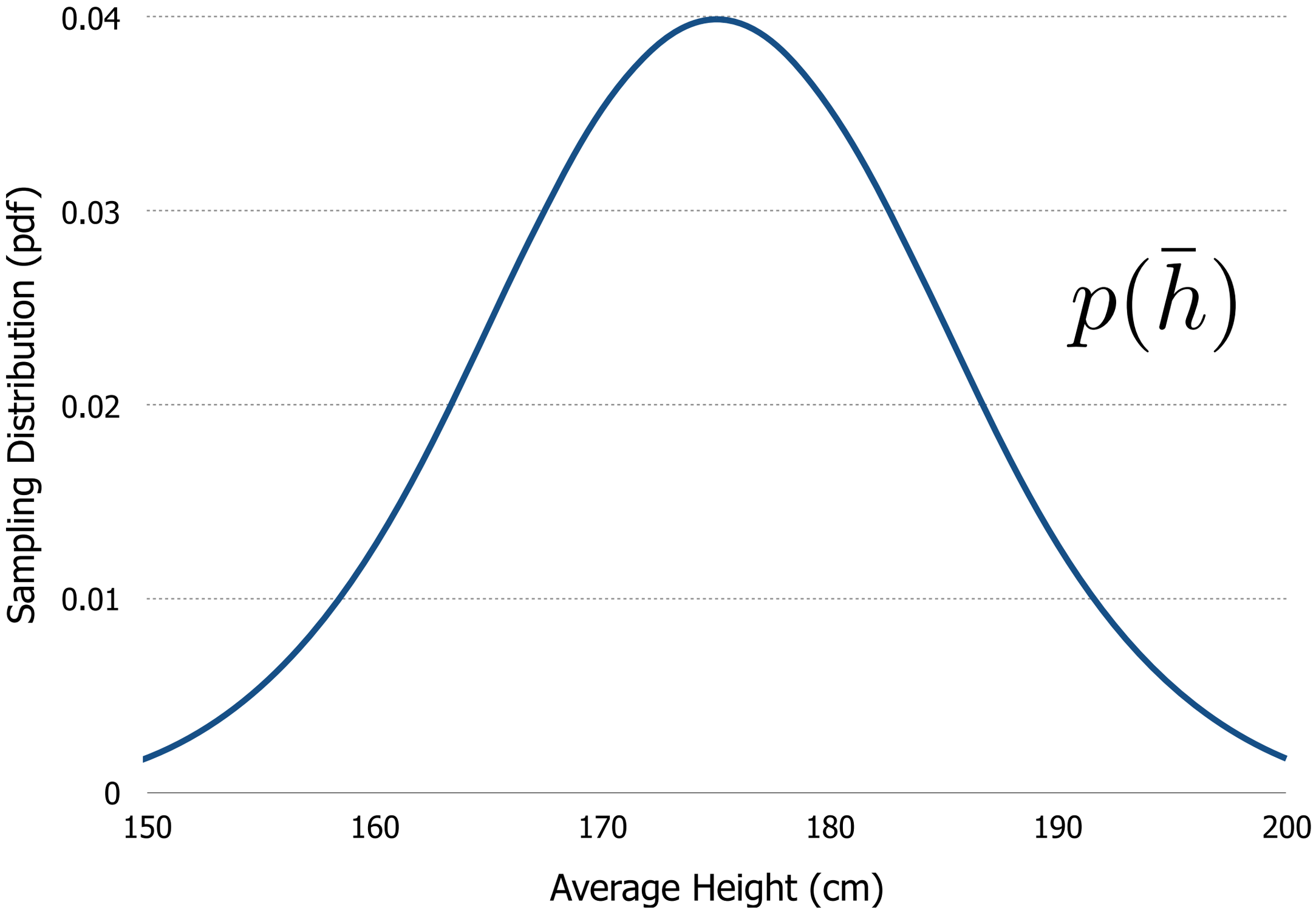

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

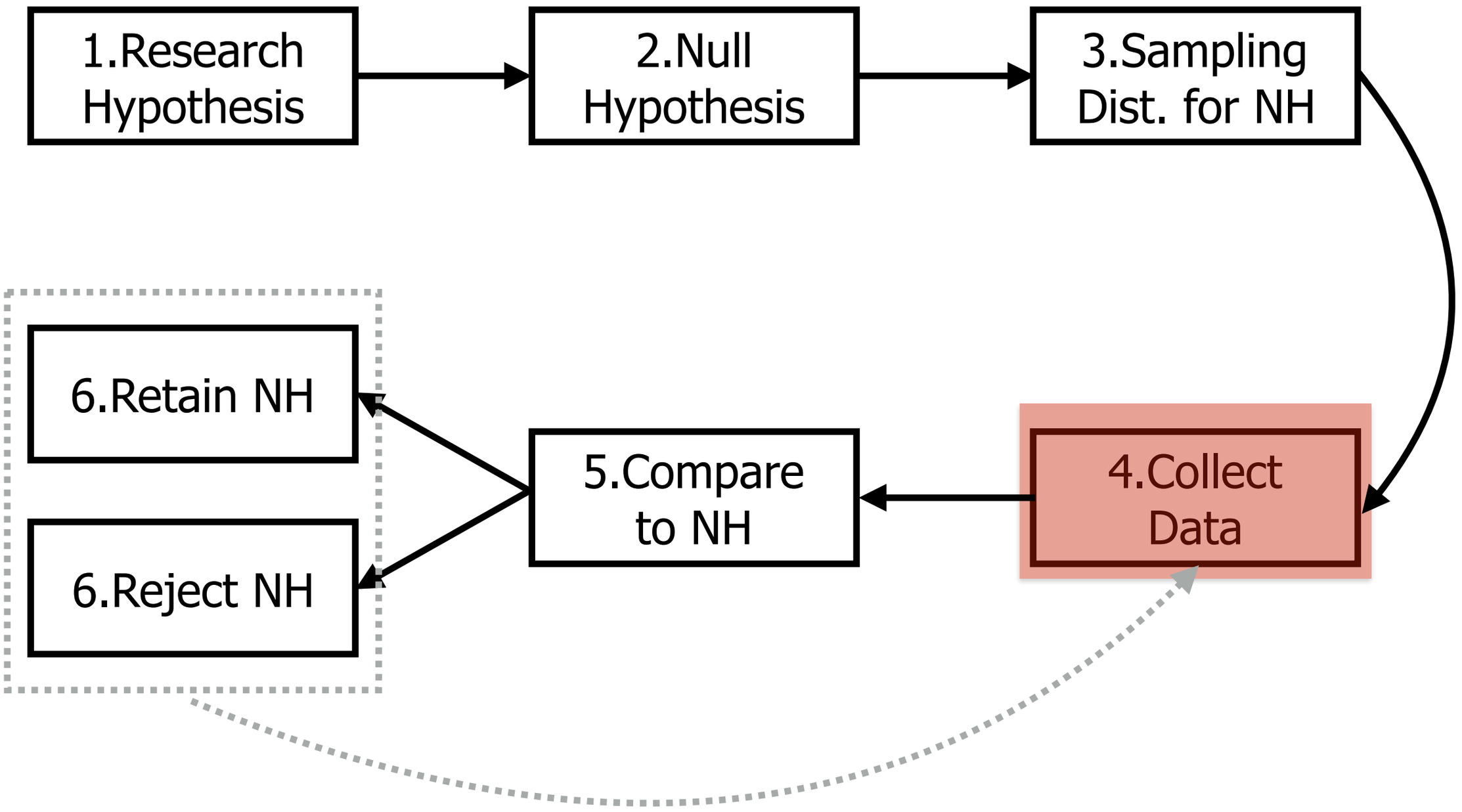
Null Hypothesis

- Basketball players do not tend to be taller than the rest of the population;
- Primary school children do not tend to be smaller than the rest of the population.

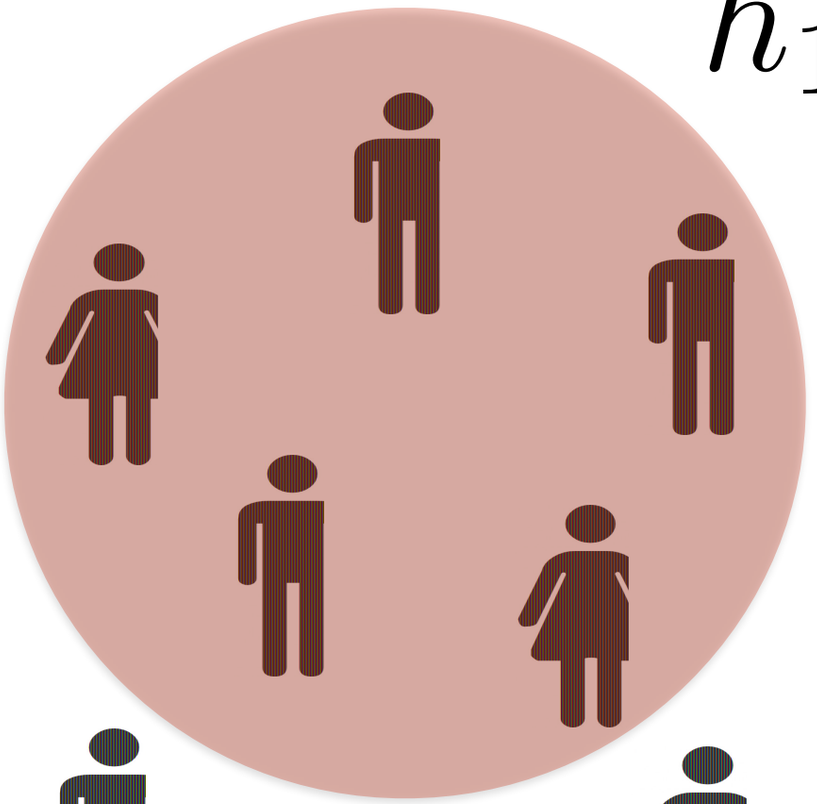

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

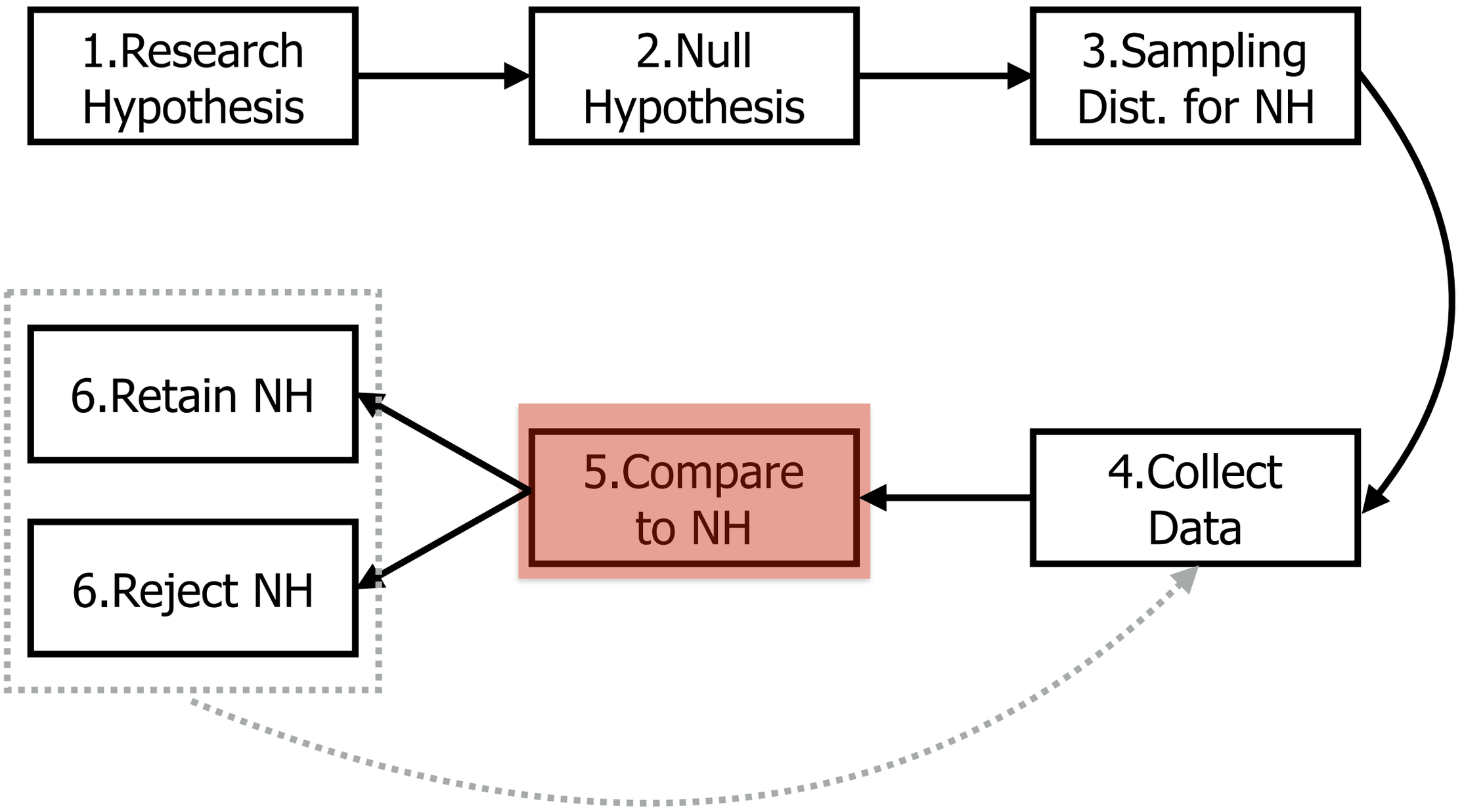

$\bar{h}_1$
$\bar{h}_2$
$\bar{h}_3$

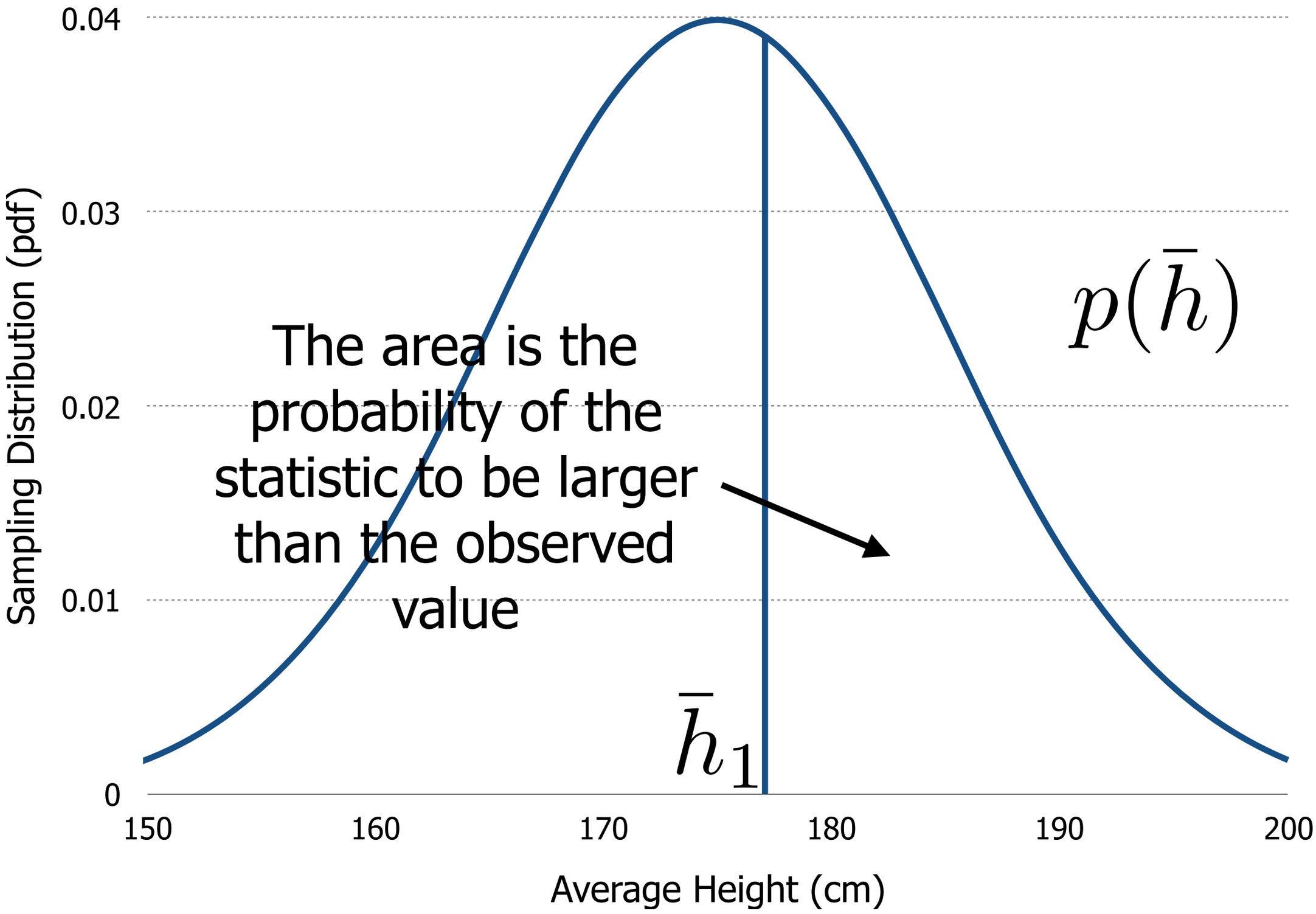

0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
$p(\bar{h})$
150
160
170
180
190
200
Average Height (cm)

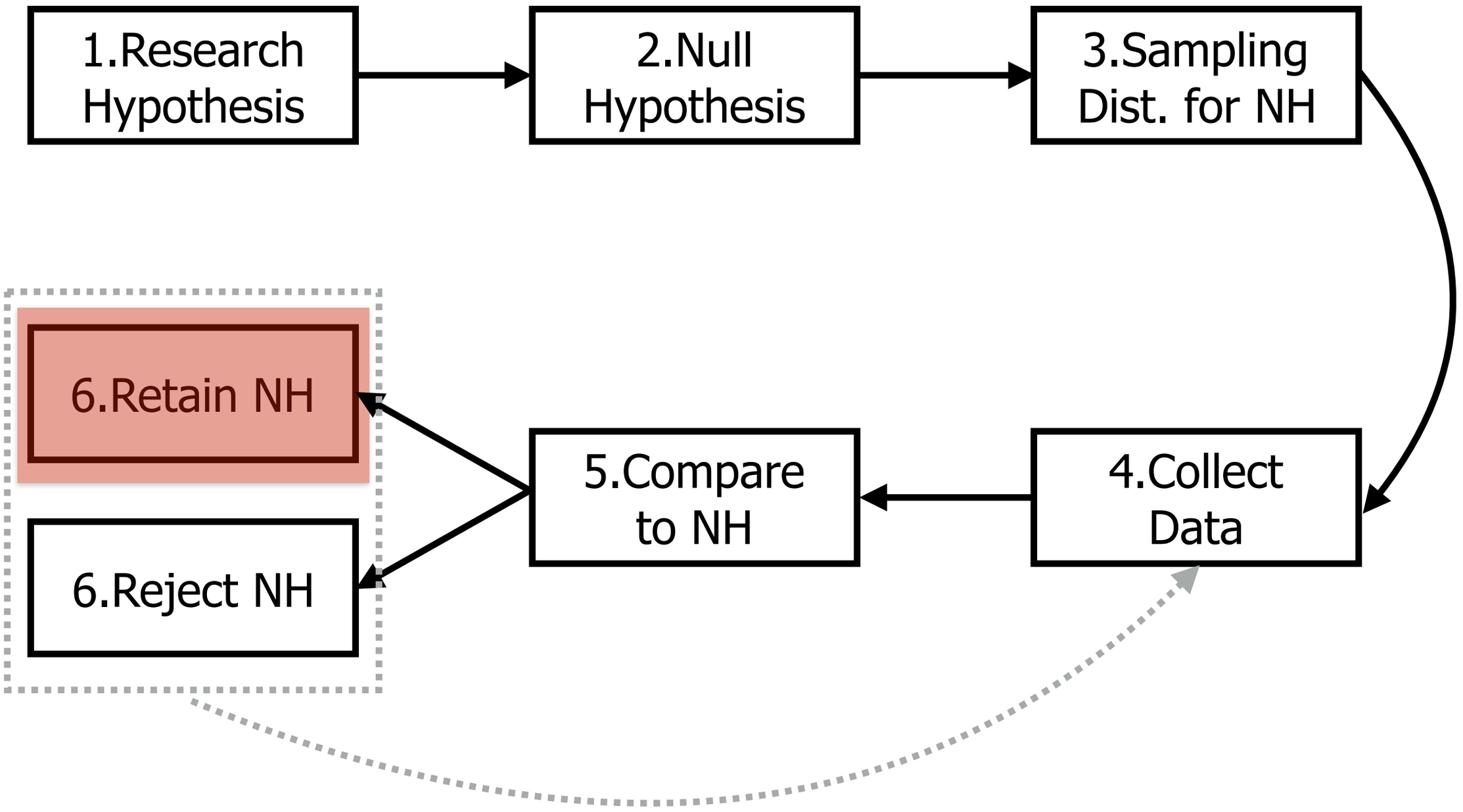

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

$\bar{h}_1$



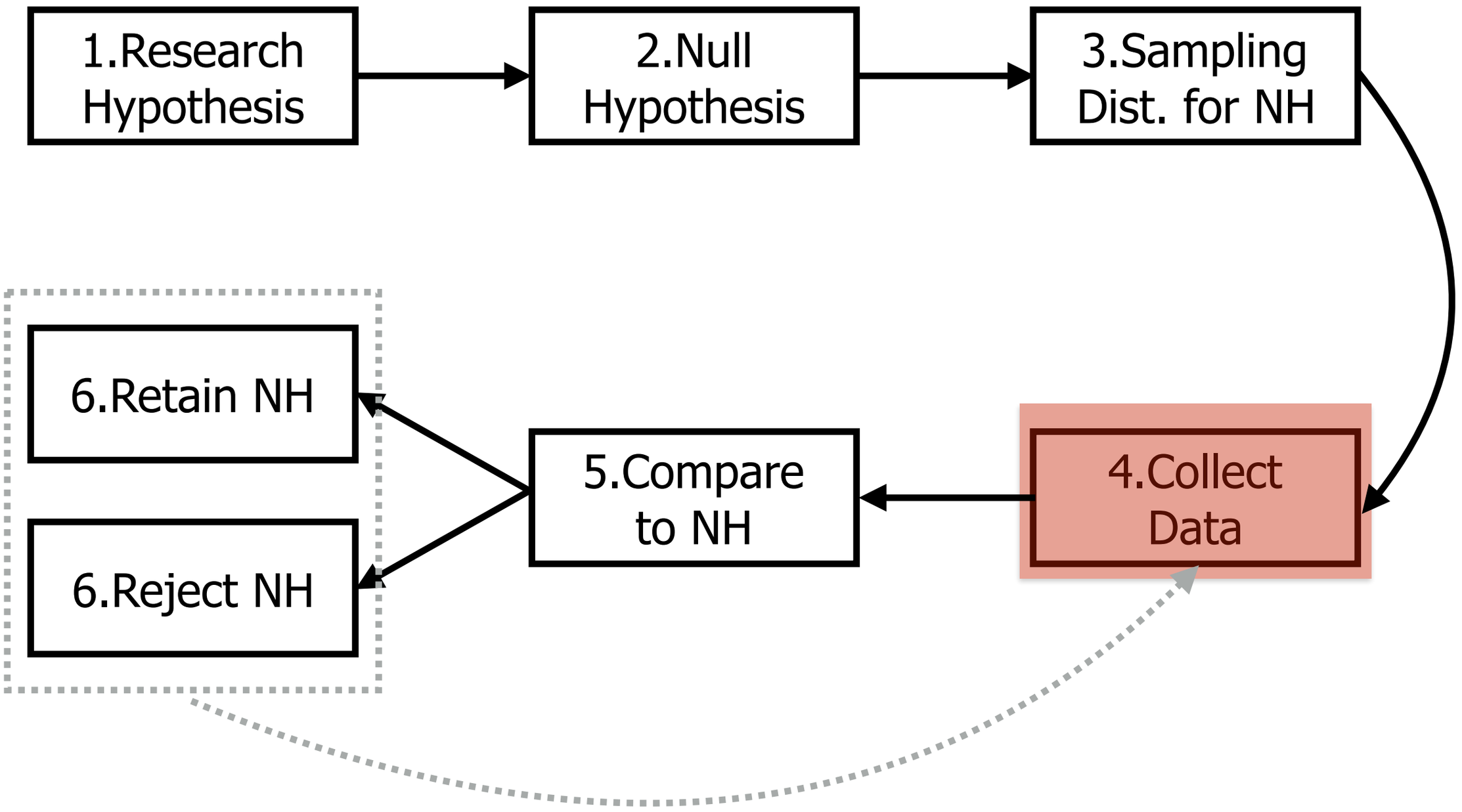

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH


0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
The area is the probability of the statistic to be larger than the observed value
$p(\bar{h})$
$\bar{h}_1$
150
160
170
180
190
200
Average Height (cm)

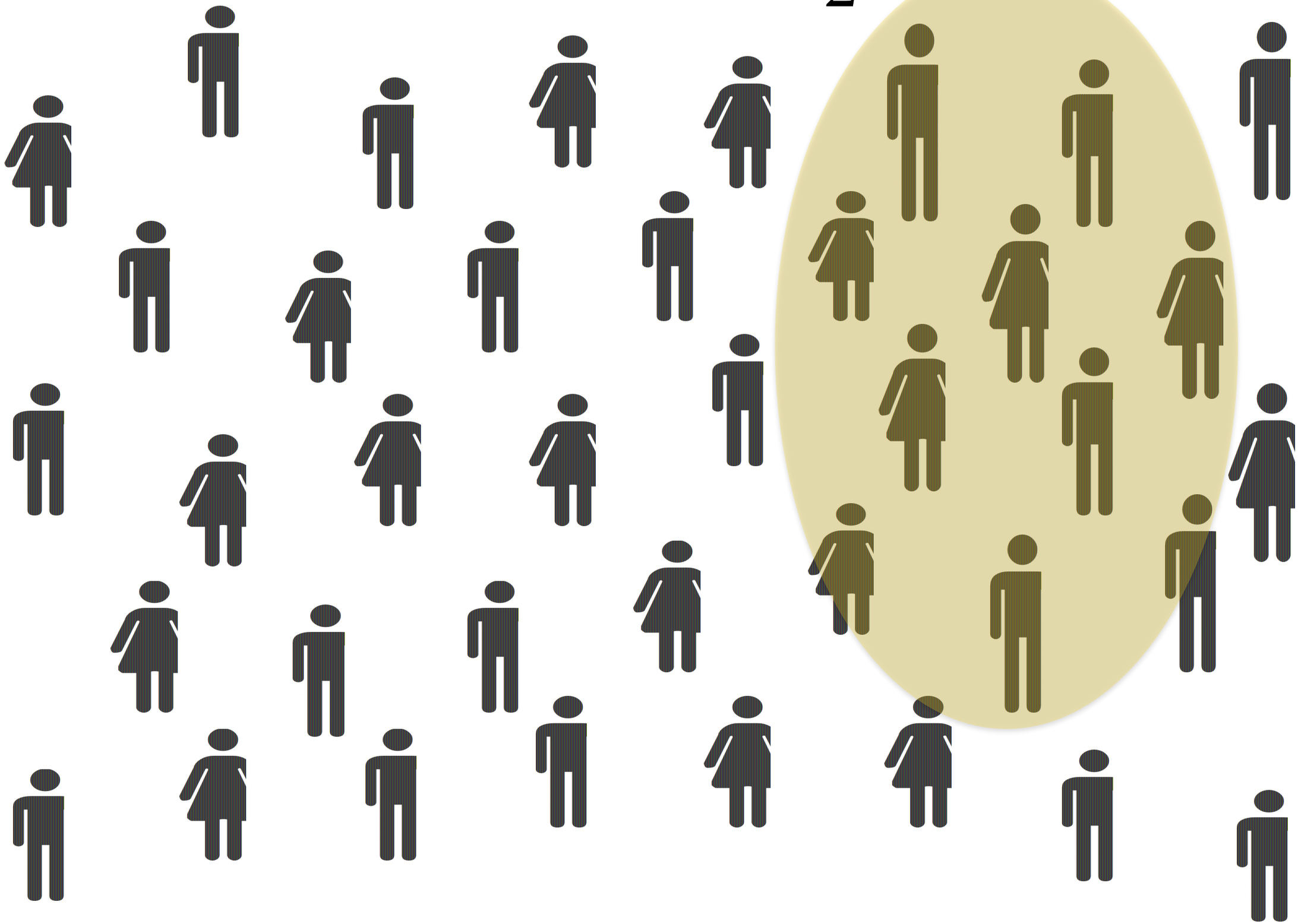

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

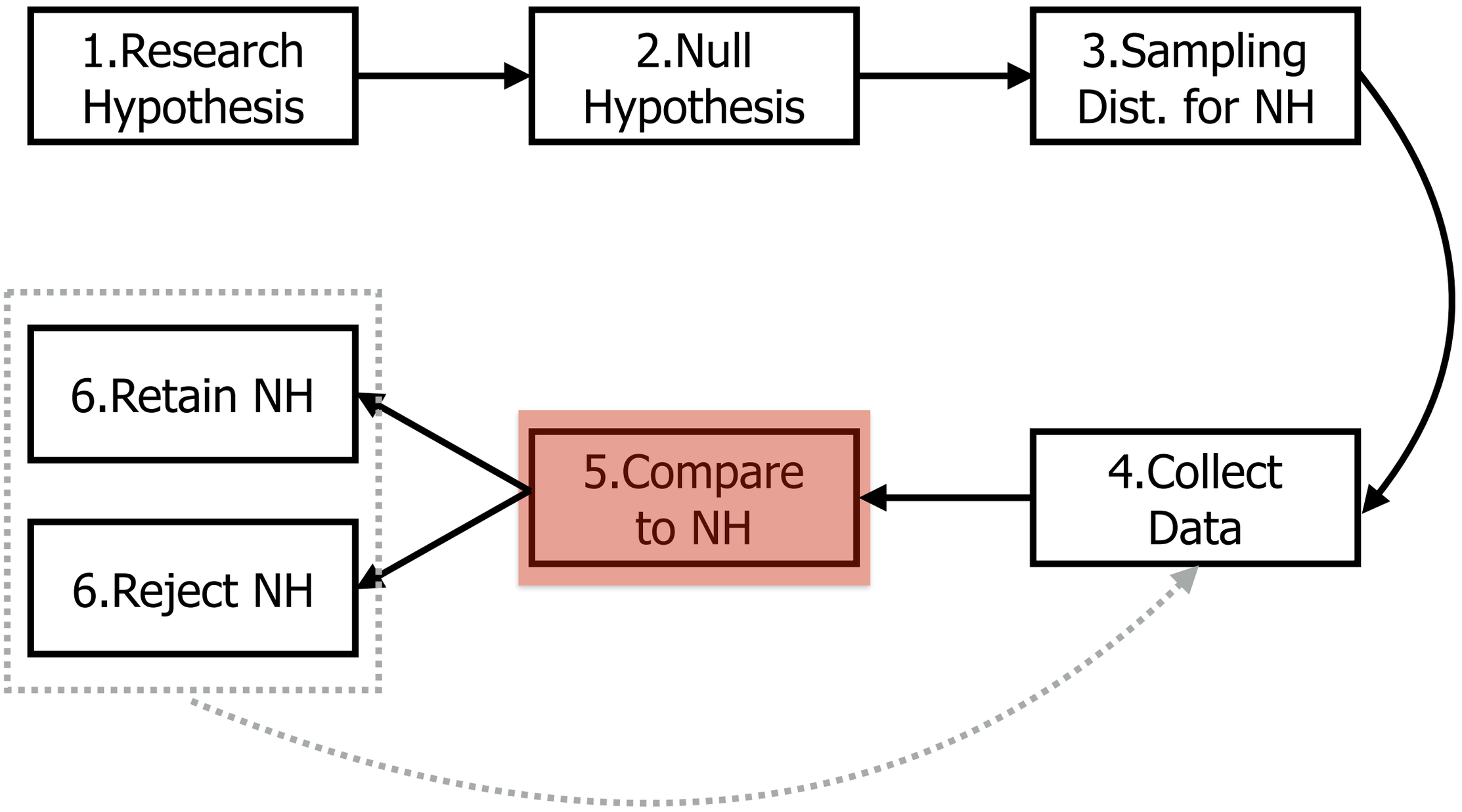
Fail to Reject the Null Hypothesis

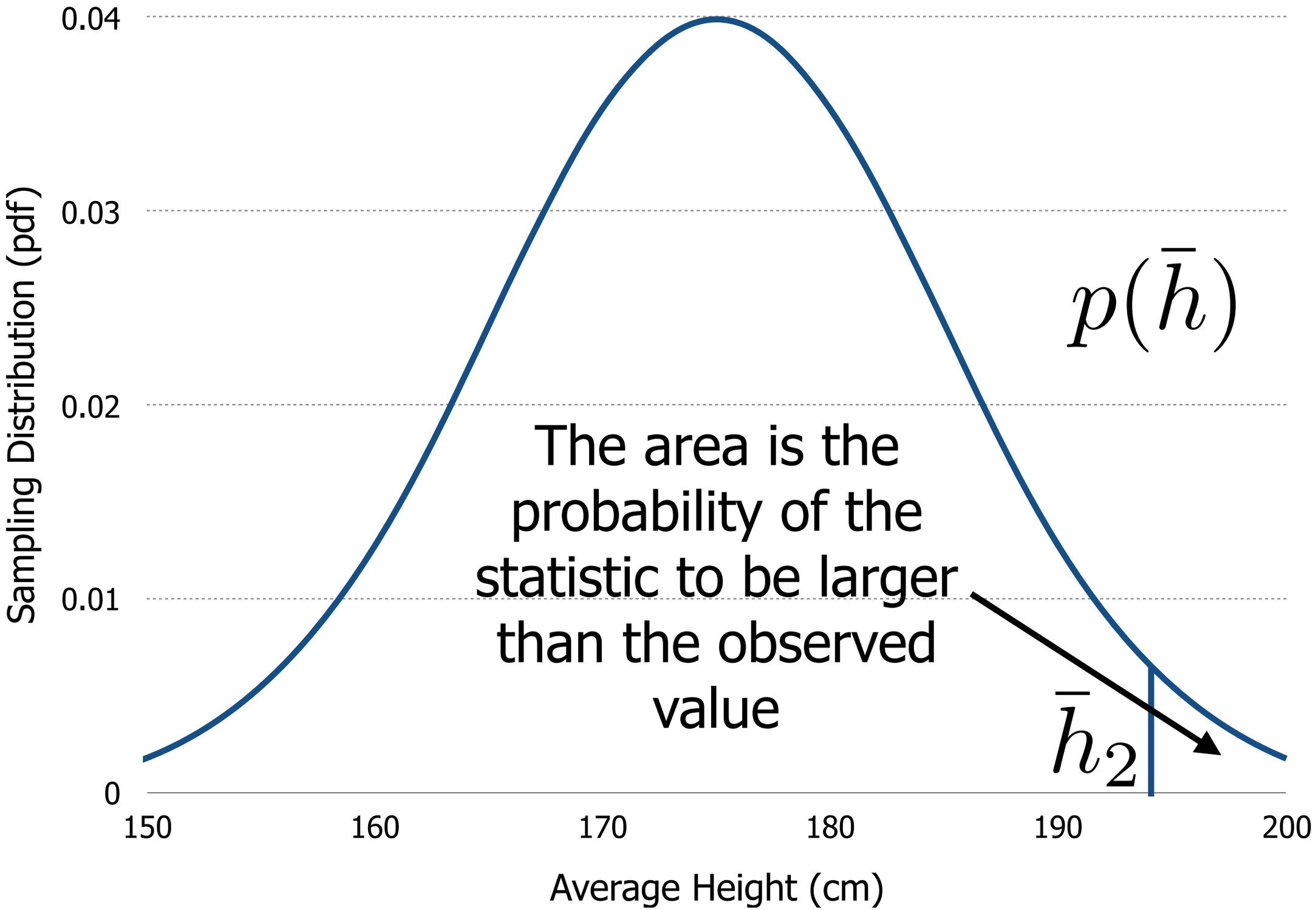
- When the probability to observe a average height higher or equal to the one I observe is large enough, the null hypothesis fails to be rejected;
- What "large enough" means exactly will be explained later.

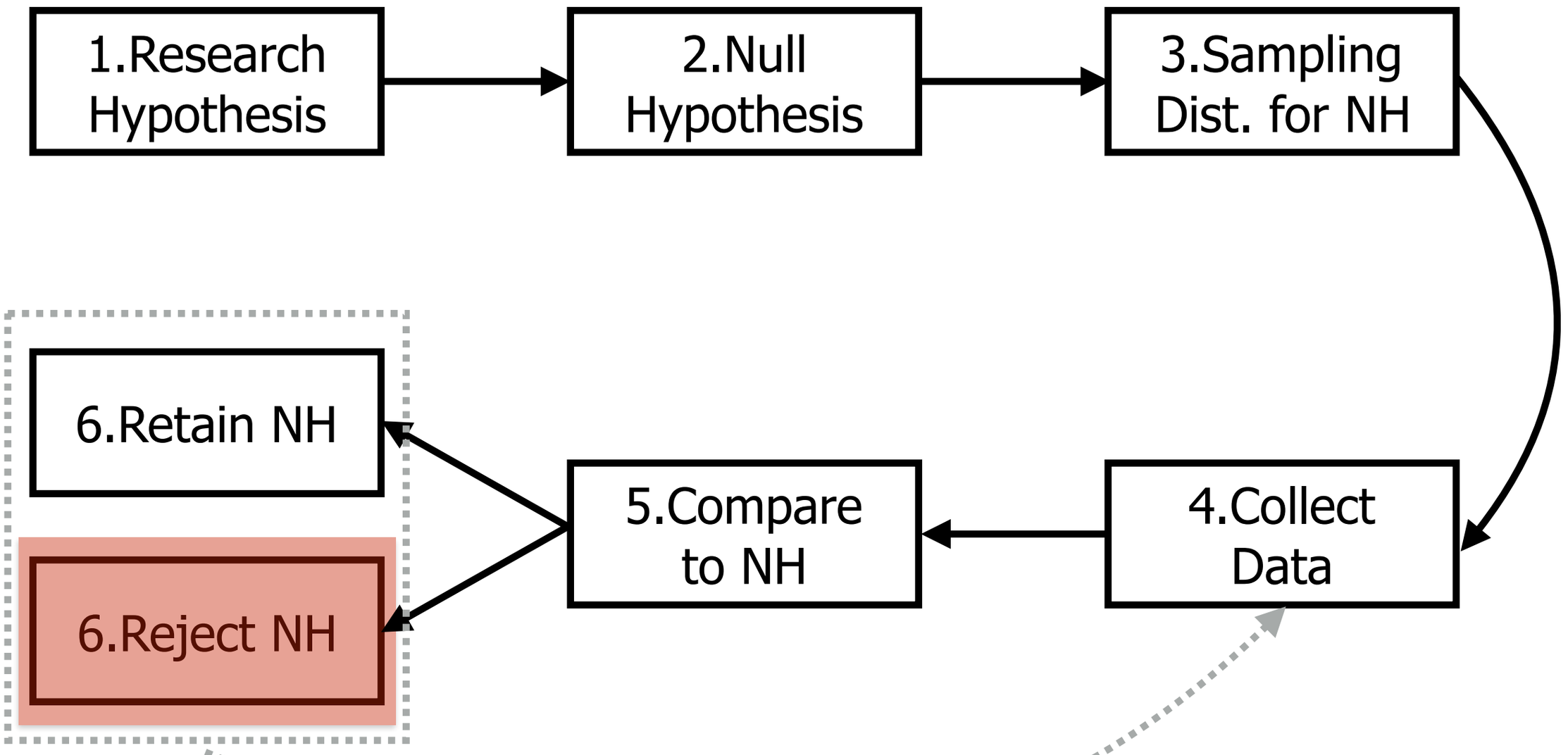

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

$\bar{h}_2$



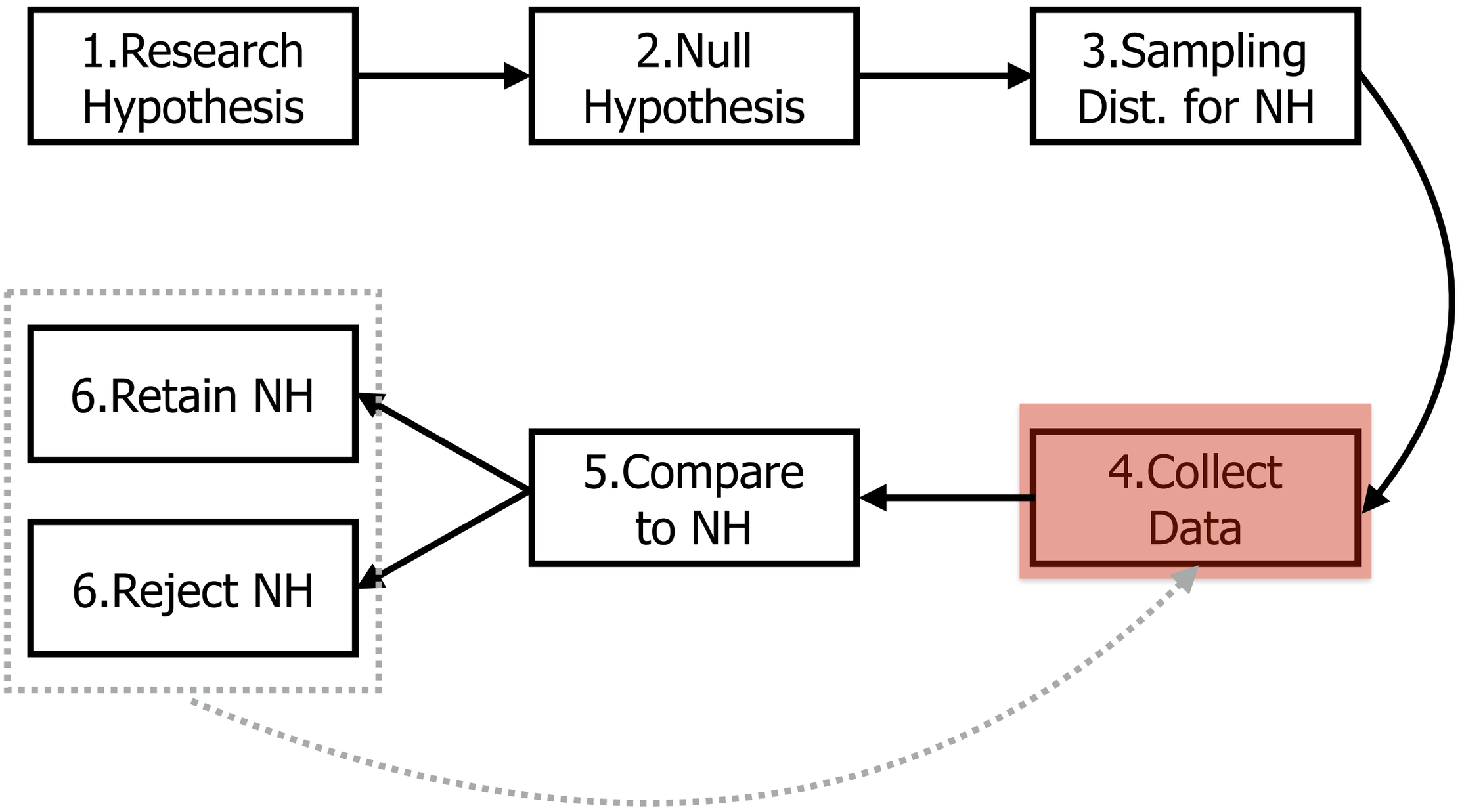

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH


0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
$p(\bar{h})$
The area is the probability of the statistic to be larger than the observed value
$\bar{h}_2$
150
160
170
180
190
200
Average Height (cm)

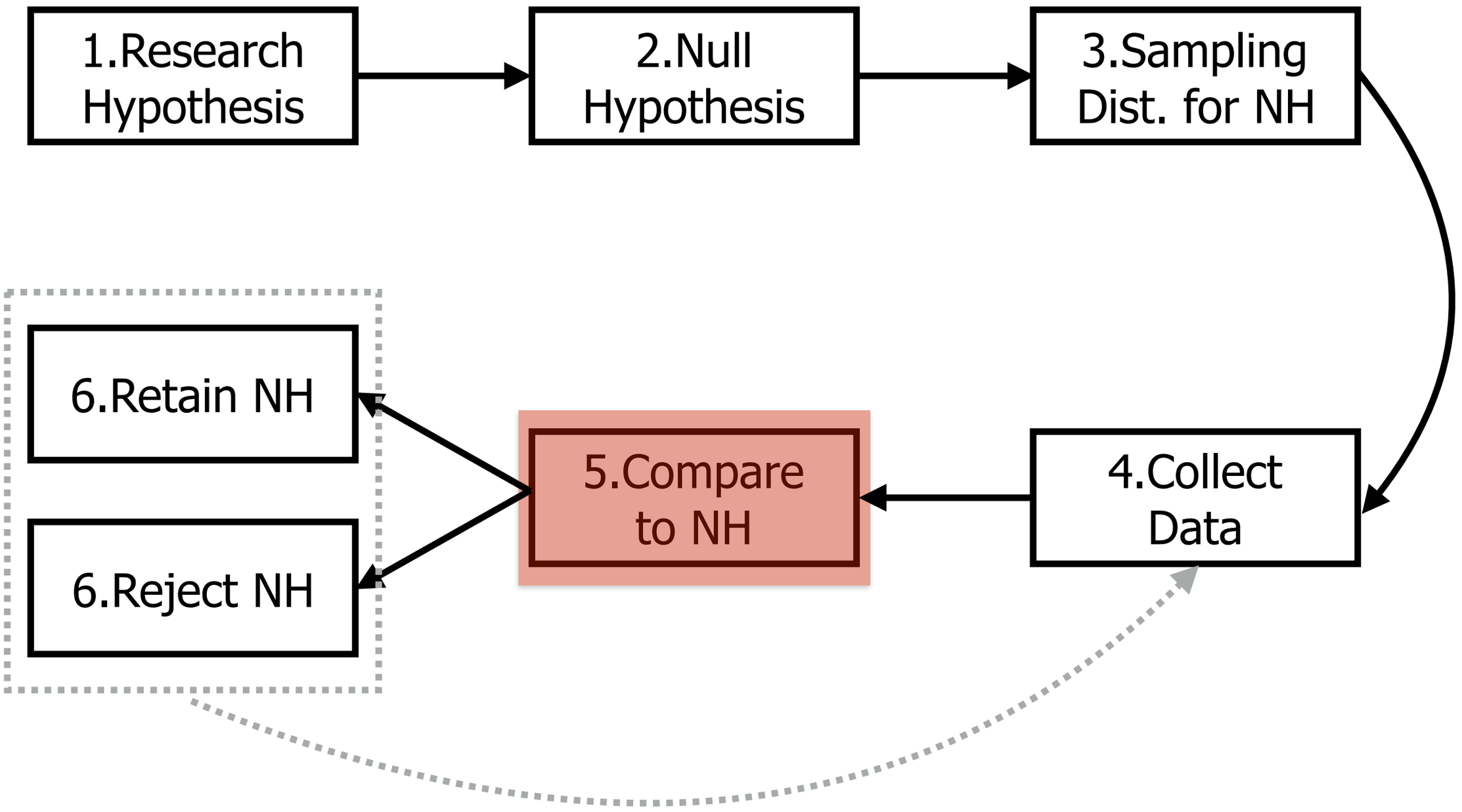

1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

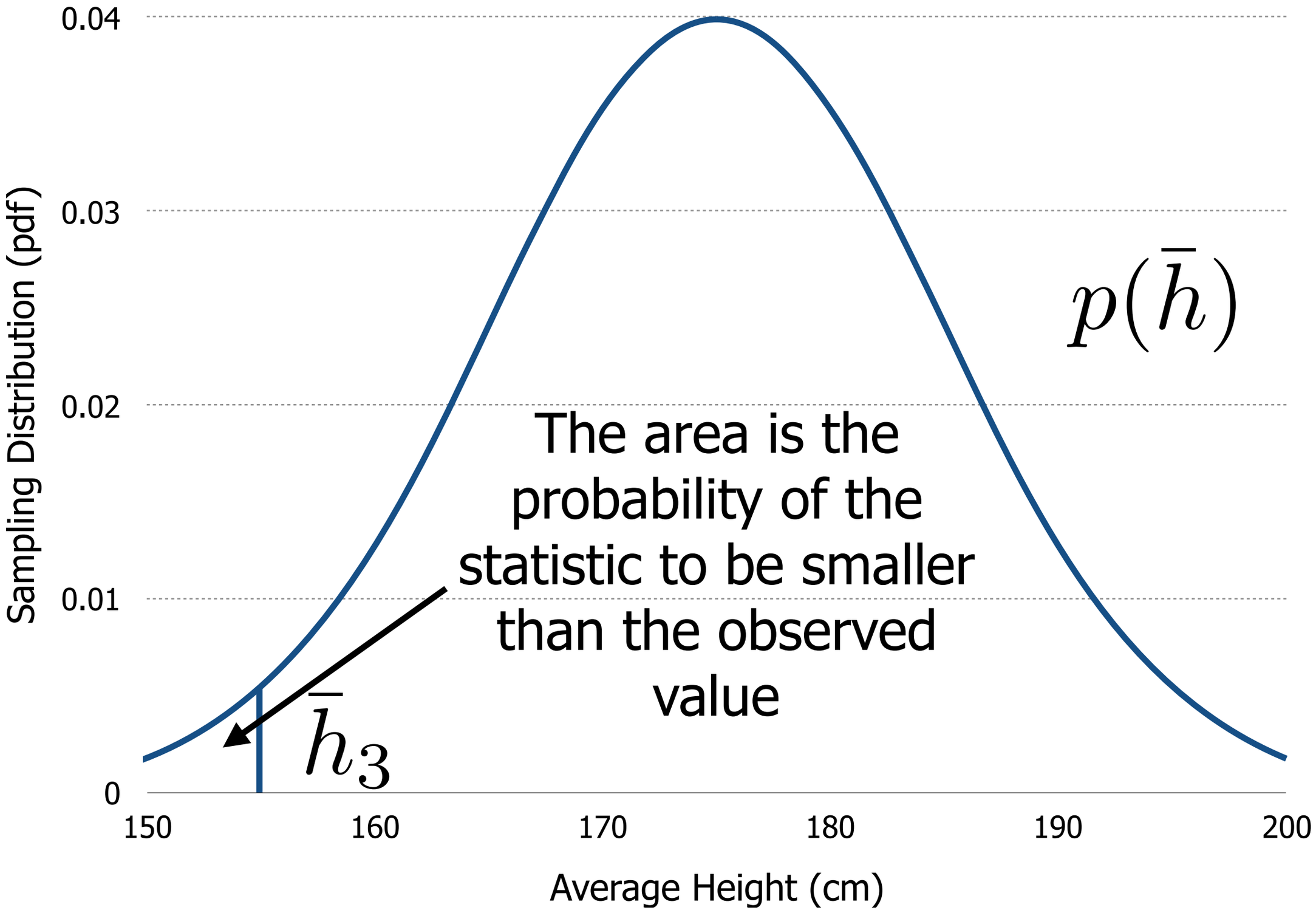
Reject the Null Hypothesis

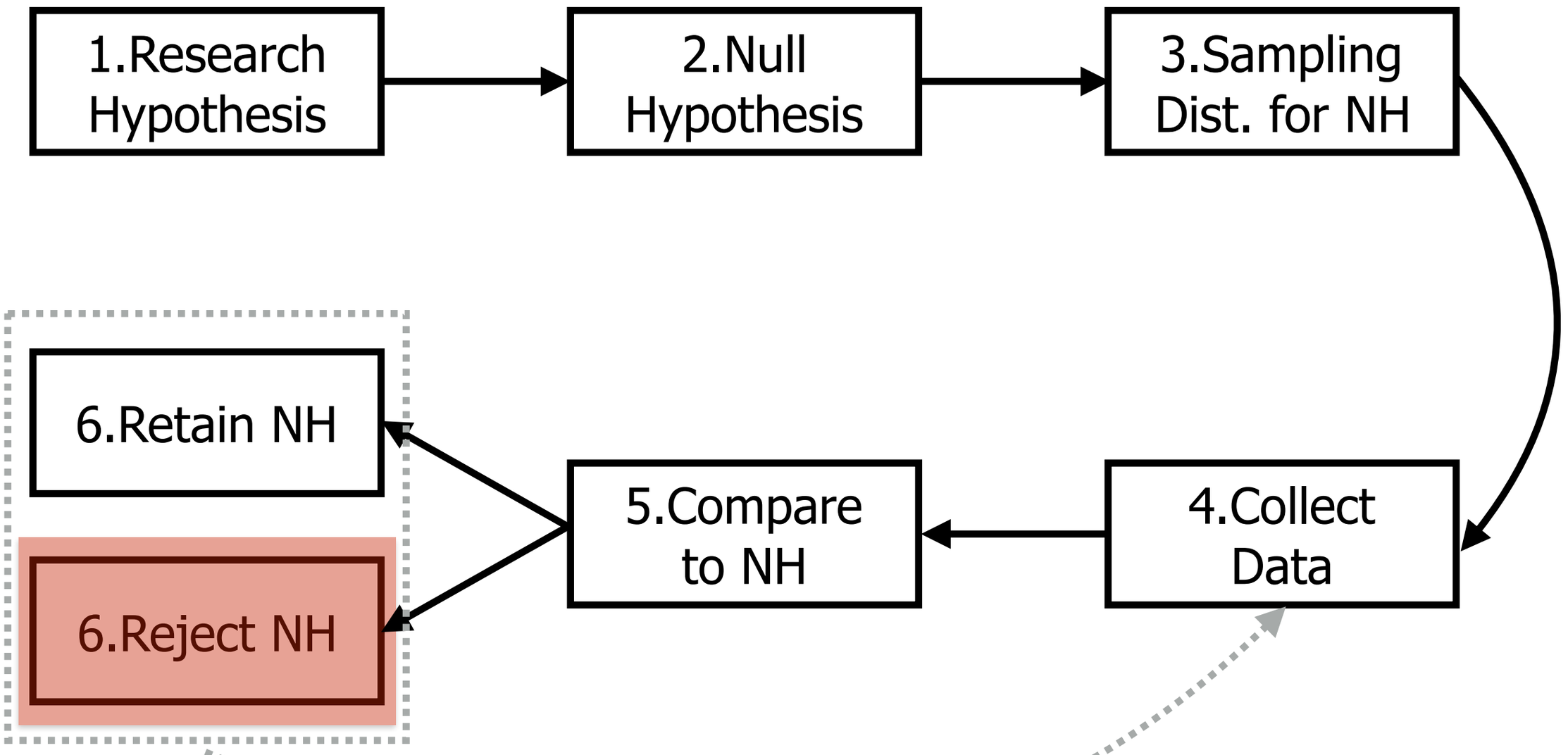
- When the probability to observe a average height higher than or equal to the one I observe is small enough, the null hypothesis can be rejected;
- What "small enough" means exactly will be explained later.


1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

$\bar{h}_3$


1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH


0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
$p(\bar{h})$
The area is the probability of the statistic to be smaller than the observed value
$\bar{h}_3$
150
160
170
180
190
200
Average Height (cm)


1.Research Hypothesis
2.Null Hypothesis
3.Sampling Dist. for NH
4.Collect Data
5.Compare to NH
6.Retain NH
6.Reject NH

Reject the Null Hypothesis

- When the probability to observe a average height lower than or equal to the one I observe is small enough, the null hypothesis can be rejected;
- What "small enough" means exactly will be explained later.

Outline

- Sampling Distributions
- Hypothesis Testing
- The Null Hypothesis
- Error Types
- Directionality

The Null Hypothesis?

"[…] we can never prove something to be true, but we can prove something to be false. Observing 3000 people with two arms does not prove the statement 'Everyone has two arms'. However, finding one person with three arms does disprove the original statement […] "

D.C.Howell, "Statistical Methods for Psychology", Chapter 4, Cengage Learning, 2009.

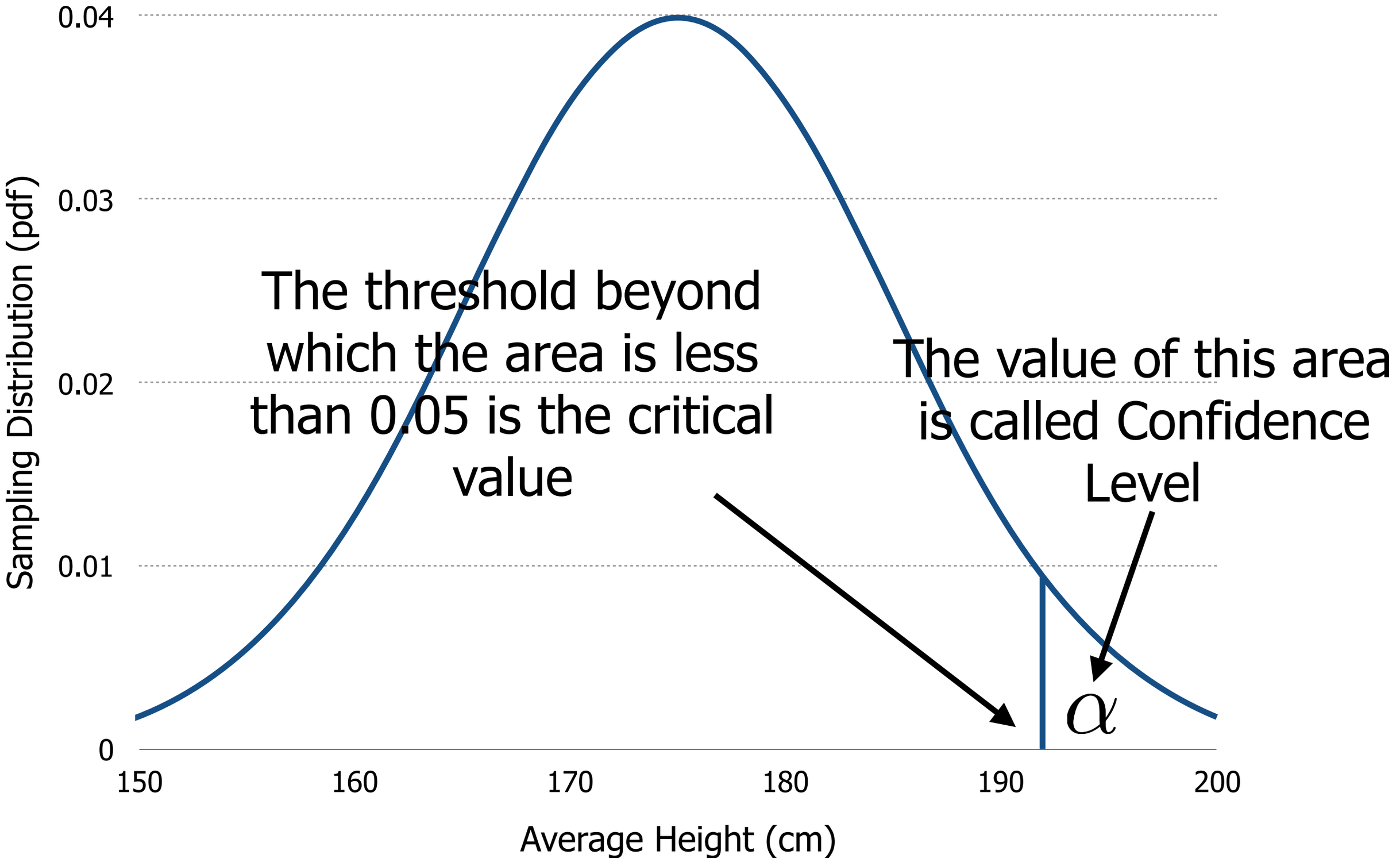
The Null Hypothesis (II)

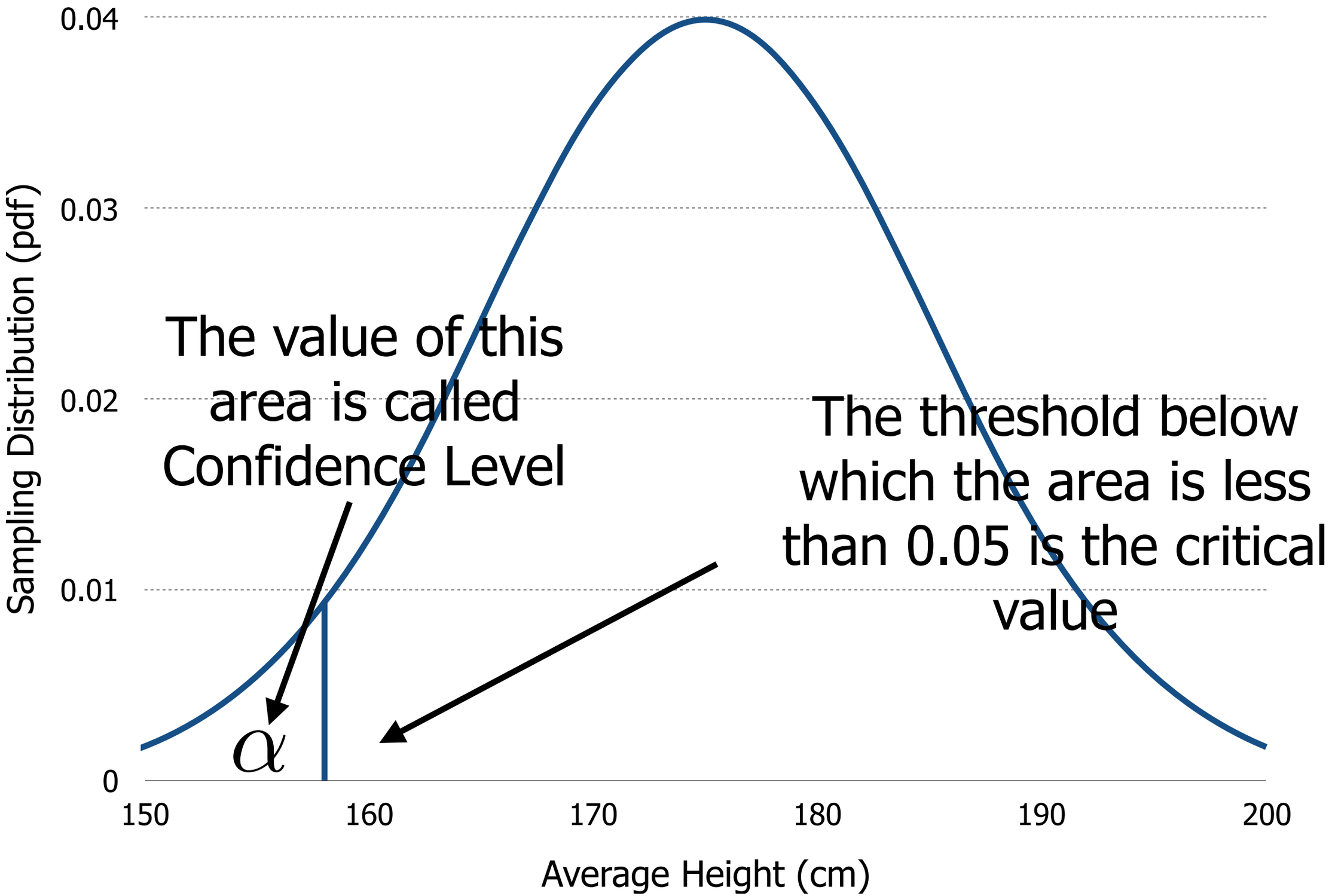
"The one thing on which all statisticians agree is that we can never claim to have 'proved' the null hypothesis."

D.C.Howell, "Statistical Methods for Psychology", Chapter 4, Cengage Learning, 2009.

Outline

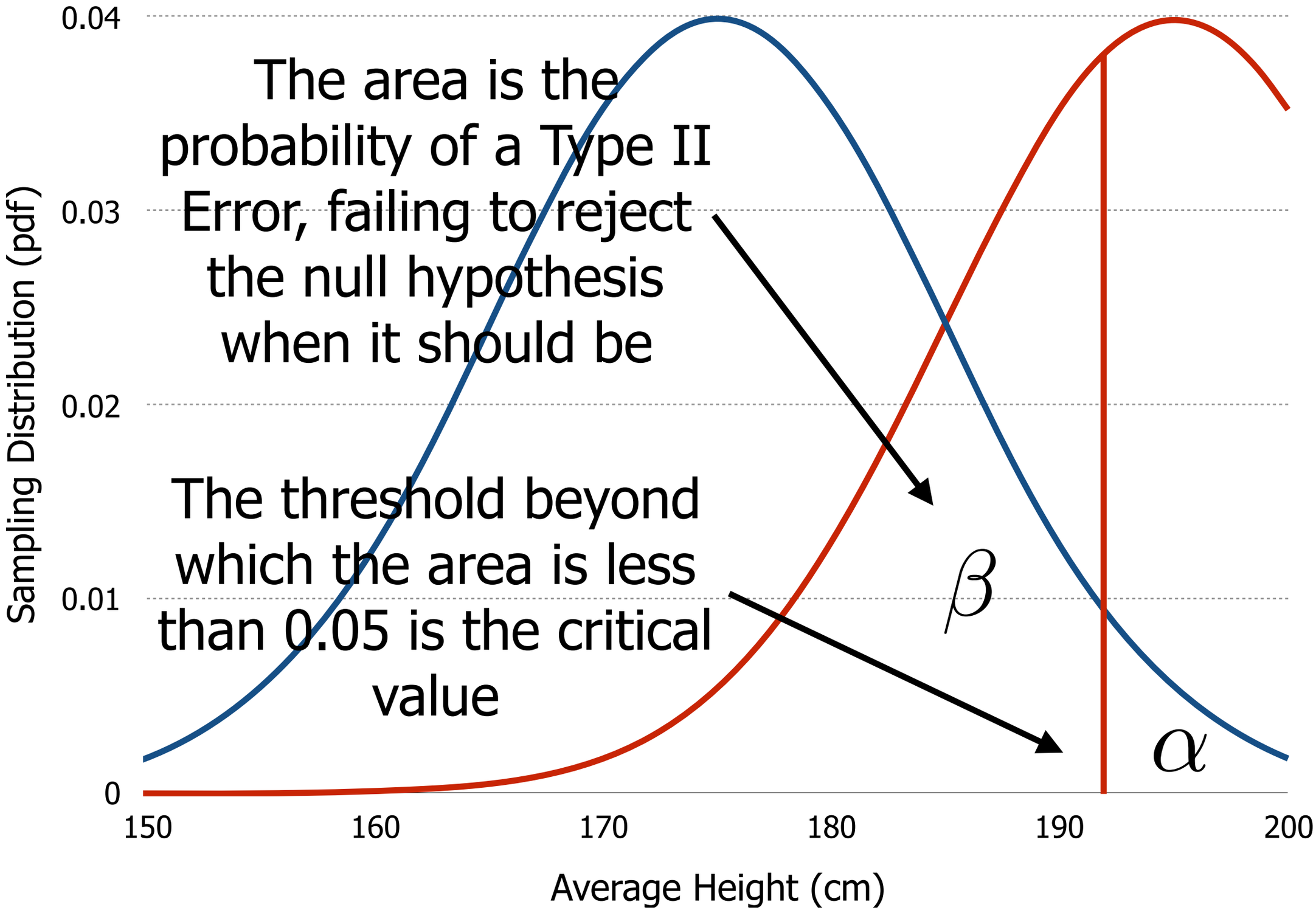
- Sampling Distributions
- Hypothesis Testing
- The Null Hypothesis
- Error Types
- Directionality


0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
The threshold beyond which the area is less than 0.05 is the critical value
The value of this area is called Confidence Level
α
150
160
170
180
190
200
Average Height (cm)


0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
The value of this area is called Confidence Level
The threshold below which the area is less than 0.05 is the critical value
α
150
160
170
180
190
200
Average Height (cm)

Type I Errors

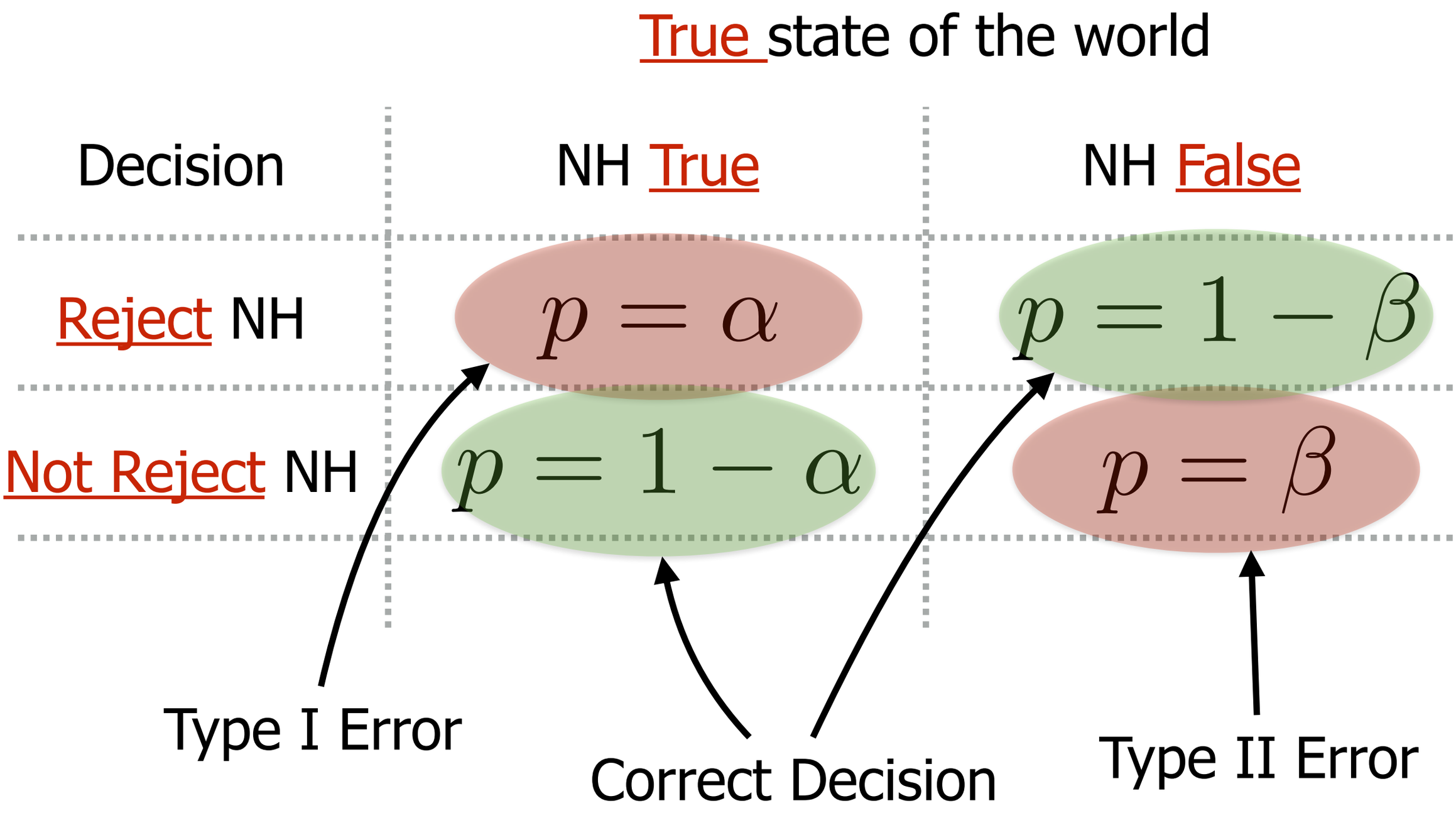
- The people below or beyond the critical values are "normal" (not necessarily primary school pupils or basketball players);
- The value p=0.05 is the probability of rejecting the null hypothesis when it is true, a Type I Error;
- It is a conventional confidence level widely accepted in experimental practice.


0.04
0.03
0.02
0.01
0
Sampling Distribution (pdf)
The area is the probability of a Type II Error, failing to reject the null hypothesis when it should be
The threshold beyond which the area is less than 0.05 is the critical value
β
α
150
160
170
180
190
200
Average Height (cm)

Type II Errors

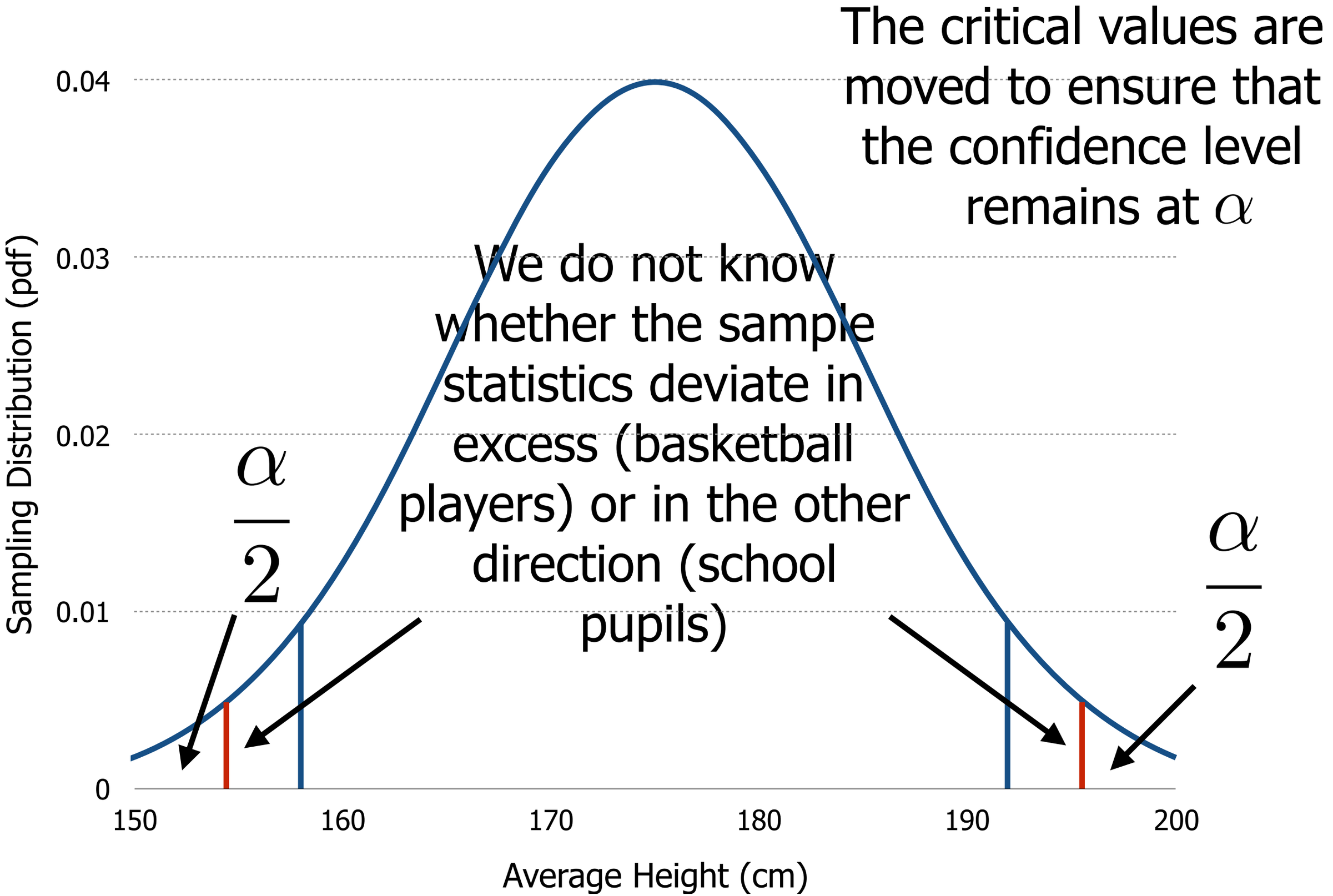
- A Type II Error takes place when the null hypothesis should be rejected, but it does not;
- Reducing the probability of a Type I Error automatically increases the probability of a Type II Error;
- A good tradeoff requires one to know the distribution of a statistic in the population where the null hypothesis is false (e.g., the basketball players).

Decision Making Process

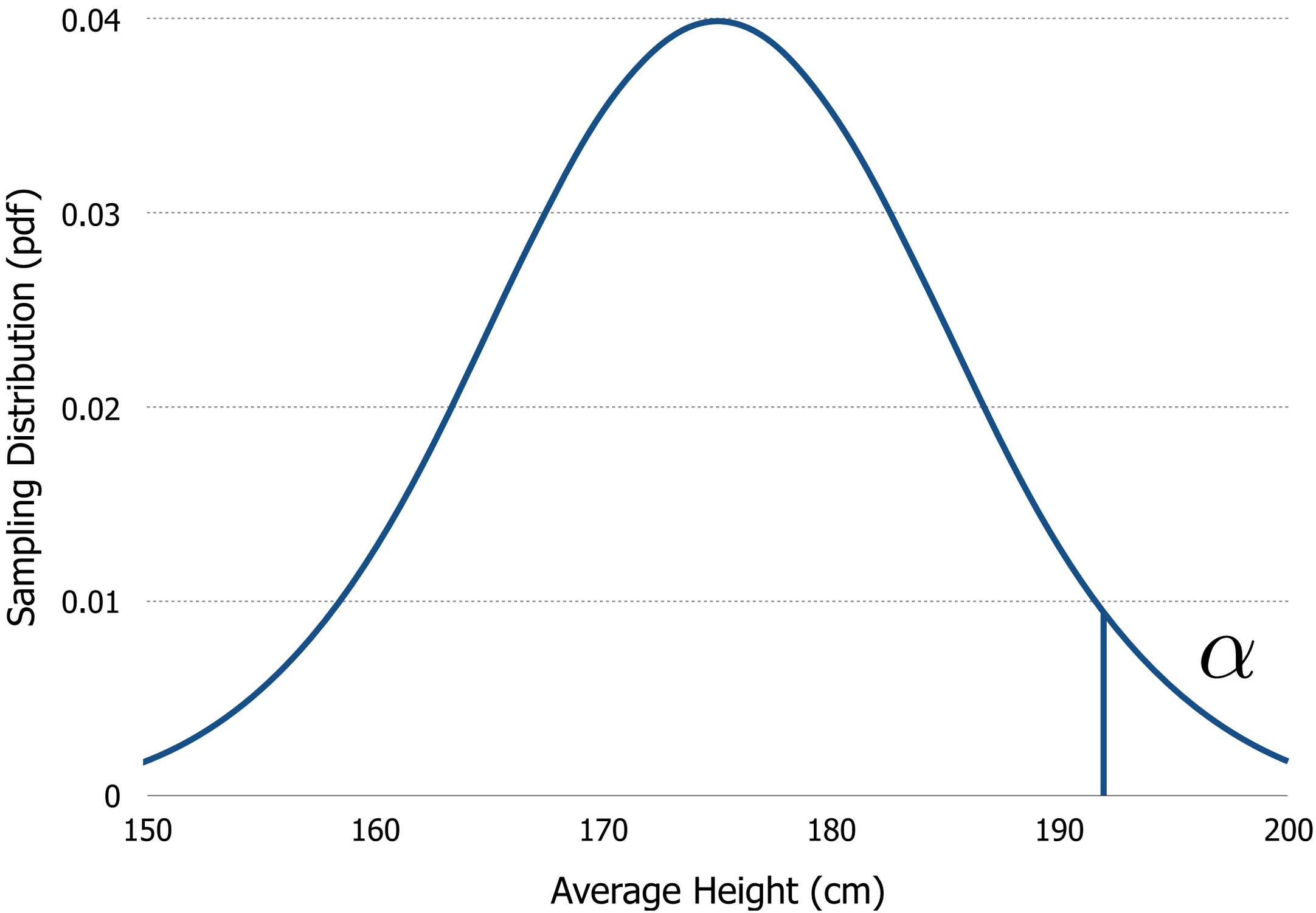


Outline

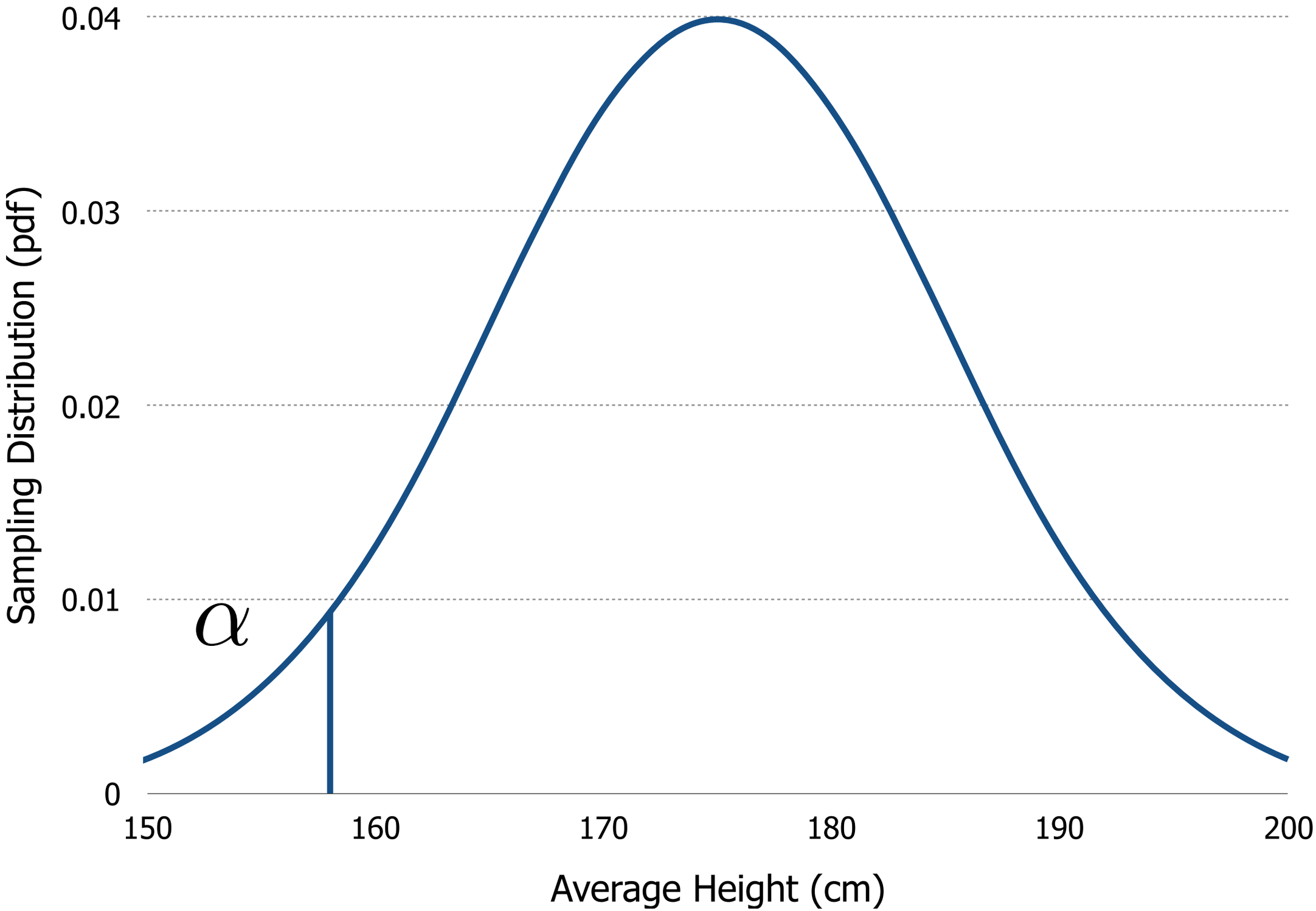
- Sampling Distributions
- Hypothesis Testing
- The Null Hypothesis
- Error Types
- Directionality


The critical values are moved to ensure that the confidence level remains at α
We do not know whether the sample statistics deviate in excess (basketball players) or in the other direction (school pupils)
$\frac{\alpha}{2}$
$\frac{\alpha}{2}$
Sampling Distribution (pdf)
0.04
0.03
0.02
0.01
0
150
160
170
180
190
200
Average Height (cm)

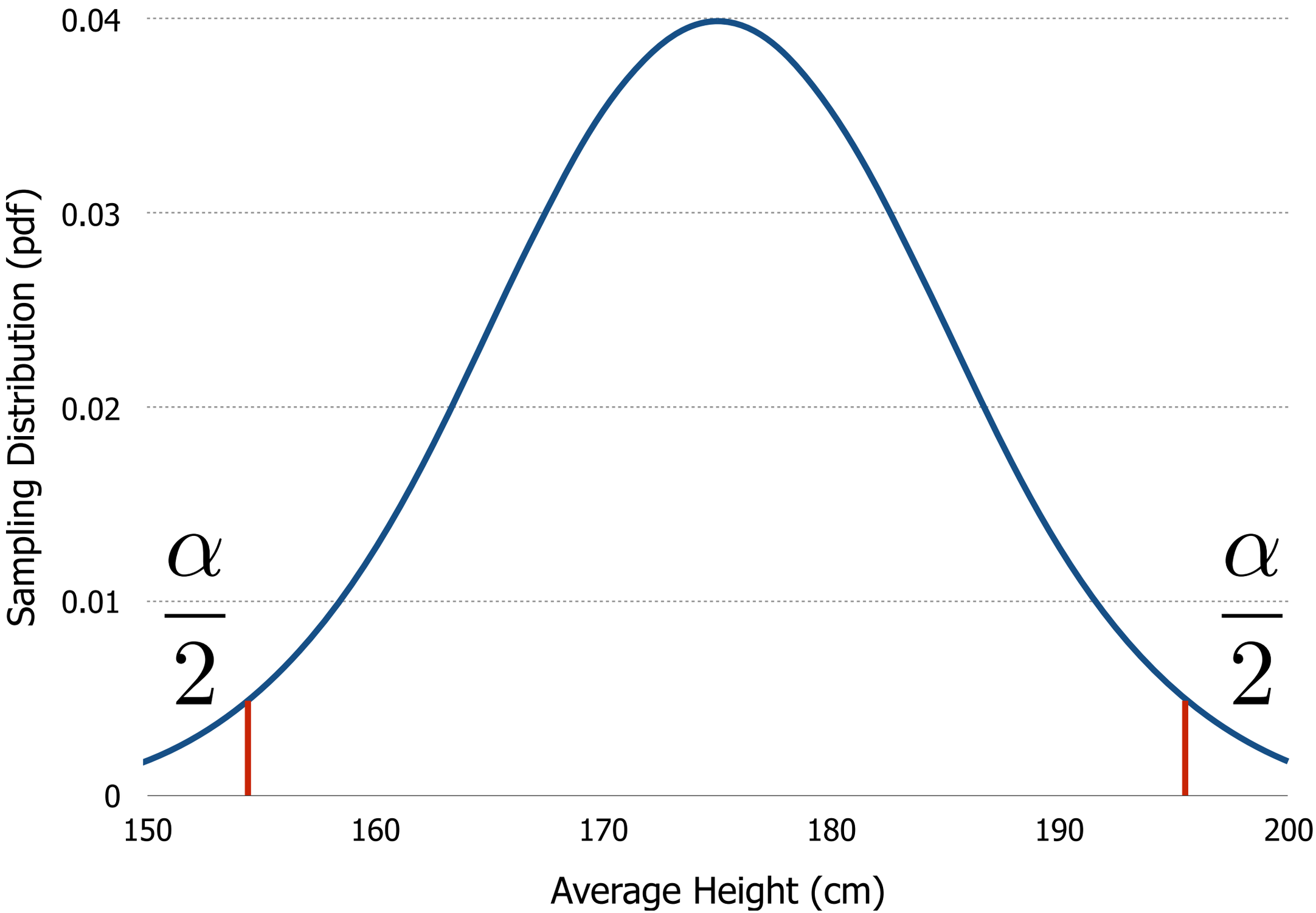
One-Tailed (Directional) Test



One-Tailed (Directional) Test



Two-Tailed (Nondirectional) Test



One vs Two-Tailed

- Two tailed tests ensure that deviations in both senses are taken into account;
- It rarely happens that a research hypothesis can be stated in directional terms with sufficient certainty;
- Both types of test are correct (they just tell a different story), but a two-tailed test is, on average, a safer choice.

Thank You!