PSC 202

SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

HYPOTHESIS TESTING WHEN USING SAMPLES, PART 2

LAST TIME

Biden Approval Ratings Diverge by Gender, Education, Race

Job Approval Ratings of President Biden, by Subgroup

	Approve %	Disapprove %	N
All U.S. adults	56	39	2,937
Gender			
Men	49	45	1,643
Women	62	34	1,294

PROBLEM

- We have a *random sample*
  - Men: 49% approval
  - Women: 62% approval
- Want to know: is mean approval rating of men and women in the *population* the same or not?

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘
We Conclude There Is No Relation	✘	✔

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘
We Conclude There Is No Relation	✘	✔

TYPE I ERROR

- We conclude there is a relationship between X and Y when in reality there is not
  - "Type I error"
  - We falsely reject H_0
  - Example: There is no difference between men and women in approval rating in the population, but we conclude there is

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘
We Conclude There Is No Relation	✘	✔

TYPE II ERROR

- We conclude there is no relationship between X and Y when in reality there is
  - “Type II error”
  - We falsely do not reject H_0
  - Example: There is a difference between men and women in approval rating in the population, but we conclude there is none

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘ Type I
We Conclude There Is No Relation	✘ Type II	✔

DECISION

- It's really bad if we conclude there is a relationship when in reality there is not (Type I error)
  - Type II error is also not great, but not as bad
- We privilege H_0

DECISION

- By default: We start out with assumption that there is no relationship in population (so H_0 is true)
  - No difference between men and women in Biden approval in population

DECISION

- Ask: Is there enough evidence in the *sample* to reject H_0?
  - Is the observed difference between men and women in *sample* large enough to reject null hypothesis that no difference between them in population?

DECISION

Job Approval Ratings of President Biden, by Subgroup

	Approve	Disapprove	N
	%	%	
All U.S. adults	56	39	2,937
Gender			
Men	49	45	1,643
Women	62	34	1,294

- **The larger the difference in approval ratings between men and women in our samples, the less likely it is that the mean in the population is the same**

P-VALUE

- Q: When do we decide that we have “enough” evidence?
- A: When the chance of falsely rejecting H_0 is 5% or less
  - Equivalent: Change of Type I error less than 5%
  - Probability of falsely rejecting H_0 is called the “p-value”

IDEA

- We start out thinking H_0 is true
  - No difference between men and women in population
- We have a sample that shows some difference
  - Do we reject H_0?
- Ask: If H_0 is true, what is the probability (p) of observing a difference at least as large as we did in our sample?

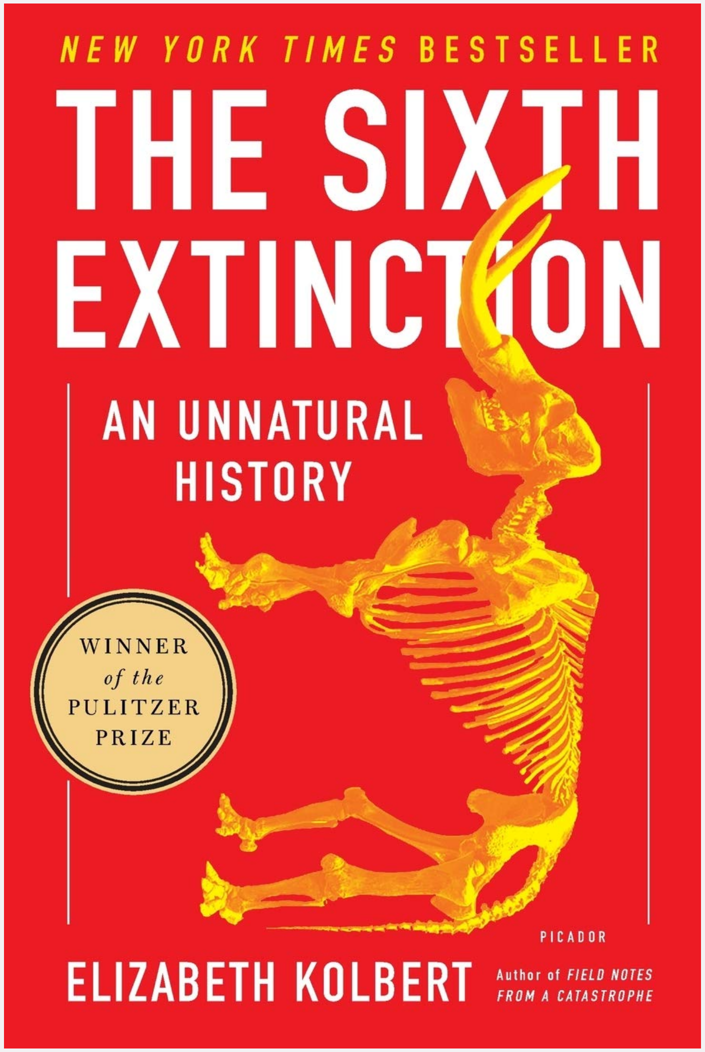
SIGNIFICANCE

- If less than 5% ($p<0.05$): we reject H_0
  - "Statistically significant difference between men and women in support for Biden"
- If more than 5% ($p>0.05$): we don't reject H_0
  - "Difference between men and women in support for Biden is not statistically significant"

REJECTING H0

- So: High bar before we reject H_0 that X has no effect on Y
  - We are conservative and need a lot of evidence before we are willing to reject H_0

REJECTING H0



PODCAST

Longform Podcast

#315: Elizabeth Kolbert

00:00 00:00

PLAY IN NEW WINDOW DOWNLOAD

Elizabeth Kolbert, author of *Field Notes from a Catastrophe: Man, Nature, and Climate Change* and *The Sixth Extinction: An Unnatural History*, is a staff writer at *The New Yorker*.

- longform.org/posts/longform-podcast-315-elizabeth-kolbert

NOW

- How exactly do we do this hypothesis testing?
    - How do we compute a p-value, etc.?

IN OUR CASE

Job Approval Ratings of President Biden, by Subgroup

	Approve %	Disapprove %	N
All U.S. adults	56	39	2,937
Gender			
Men	49	45	1,643
Women	62	34	1,294

- **H_0: No difference between men and women in population**
- **The survey does find a difference of 13 percentage points**

IN OUR CASE

Job Approval Ratings of President Biden, by Subgroup

	Approve	Disapprove	N
	%	%	
All U.S. adults	56	39	2,937
Gender			
Men	49	45	1,643
Women	62	34	1,294

- **Question: What is the probability of getting *sample* means that are at least 13 points different, if in the population there actually is no difference between men and women?**
  - **Equivalent: If we reject H_0 based on this survey, what is probability of committing Type I error?**

TEST STATISTIC

- **Test statistic t:**

$$t = \frac{H_A - H_0}{\text{Standard Error of Difference}}$$

- **H_A: observed difference between samples (here: 0.13)**
- **H_0: difference between samples if H_0 is true (0.00)**
- **Standard Error of Difference between the two samples (here 0.018)**

TEST STATISTIC

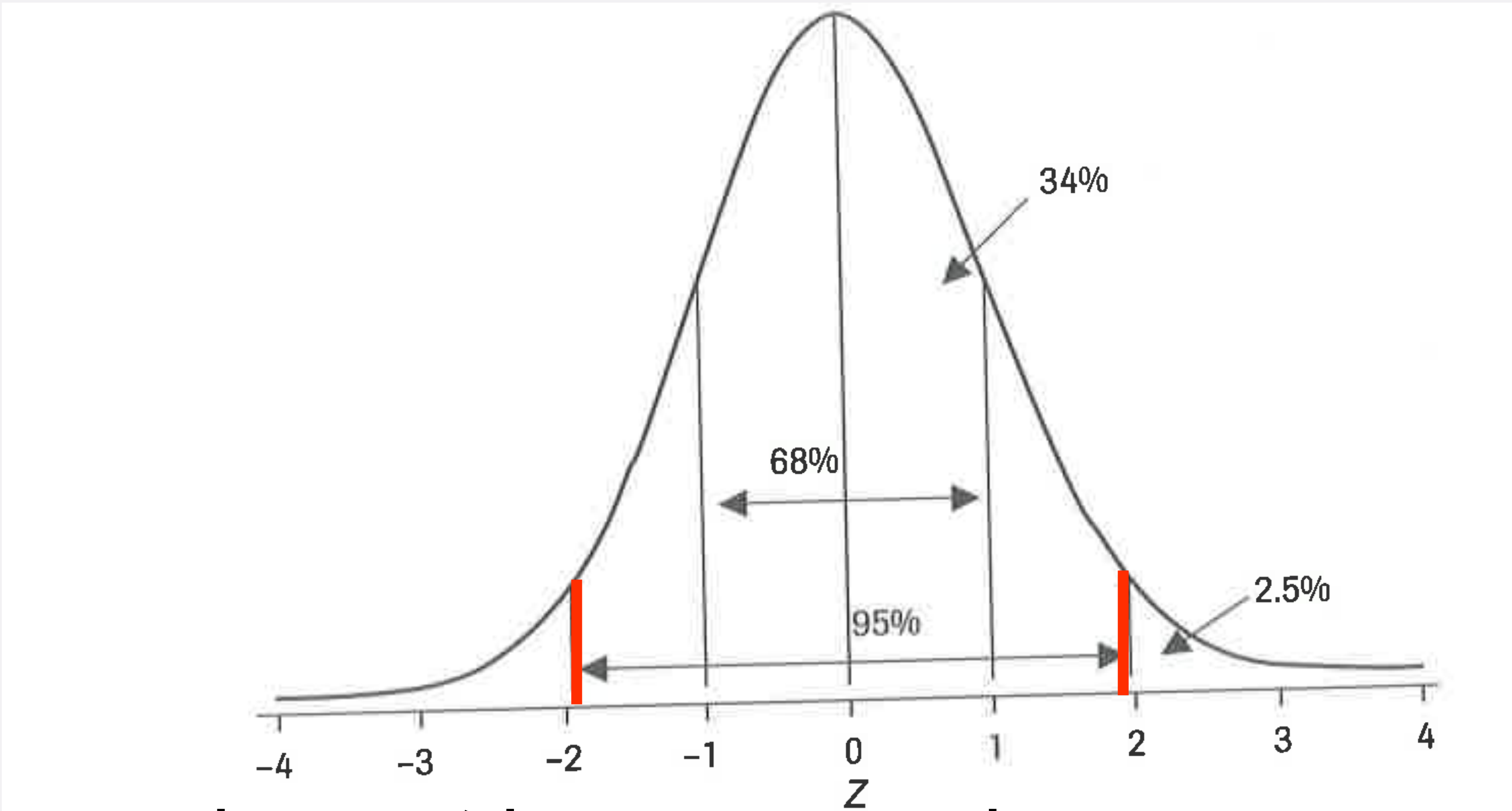
- H_A: 0.13
- H_0: 0
- Standard Error of Difference: 0.018

$$t = \frac{H_A - H_0}{\text{Standard Error of Difference}}$$

$$t = \frac{0.13 - 0.00}{0.018} = 7.22$$

- This is called the “t-statistic” or “t-ratio”

NORMAL DISTRIBUTION

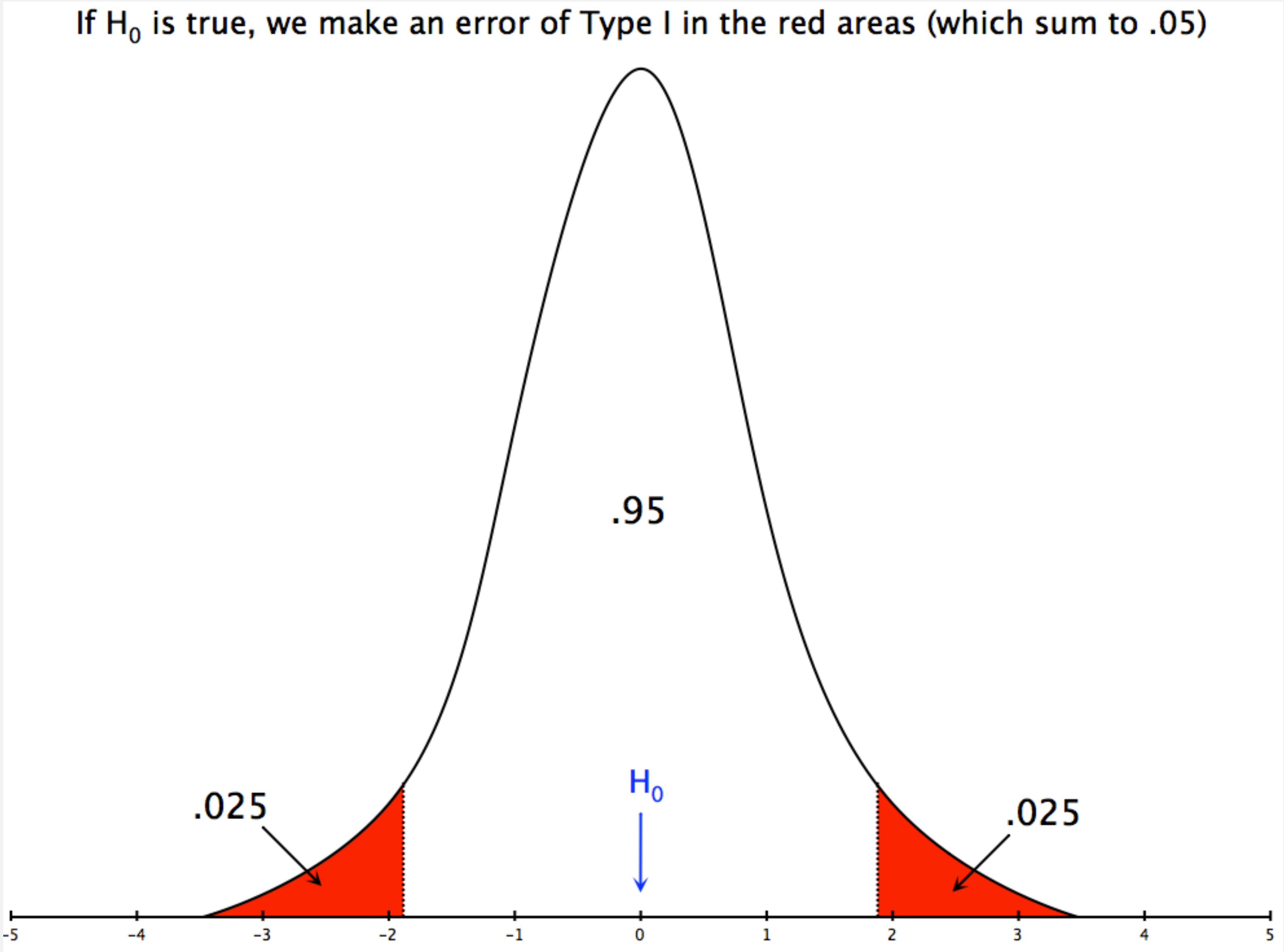


- Remember: 95% between scores between -1.96 and 1.96
- 5% of scores outside of those scores
- T-statistic is (basically) normally distributed

SIGNIFICANCE TEST

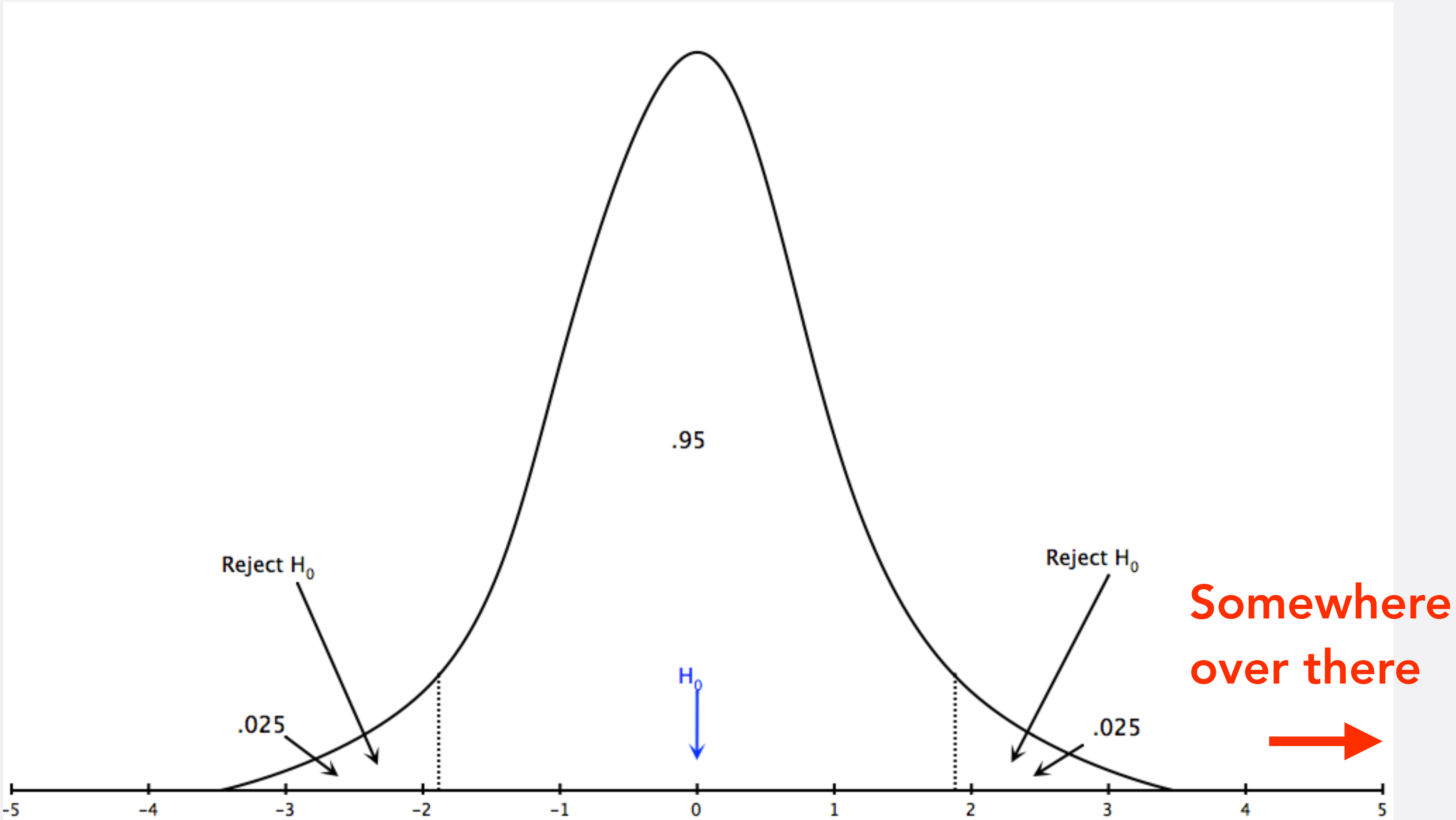
- We reject H_0 (no difference between men and women) if t-value is such that it is unlikely that we commit a Type I error
  - 5% chance that we falsely reject H_0

SIGNIFICANCE TEST



- We reject H_0 if $t<-1.96$ or $t>1.96$

SIGNIFICANCE TEST



- t-score: 7.22

SIGNIFICANCE TEST

Job Approval Ratings of President Biden, by Subgroup

	Approve %	Disapprove %	N
All U.S. adults	56	39	2,937
Gender			
Men	49	45	1,643
Women	62	34	1,294

- If there is no difference between men and women in population, chance that we find 13 percentage points difference in sample is less than 5 percent

SIGNIFICANCE TEST

Job Approval Ratings of President Biden, by Subgroup

	Approve %	Disapprove %	N
All U.S. adults	56	39	2,937
Gender			
Men	49	45	1,643
Women	62	34	1,294

- So we reject the null hypothesis that there is no difference between men and women in approval of J. Biden
- In favor of the alternative hypothesis that he has higher support among women

ANOTHER EXAMPLE

- Survey: ANES 2016
- DV: Opinion about Obamacare
    - 1=favor a great deal, 7=oppose a great deal
    - mean=4.09
    - n=1,606

OBAMACARE

Partisanship	Mean Evaluation	Frequency
Dem	2.92	924
Rep	5.69	682
Total	4.09	1606

- Difference: 5.69-2.92=2.77

HYPOTHESIS TEST

- Assuming H_0 is true
  - No difference between R and D
- What is the probability that we observe a difference of 2.77 between R and D (in a random sample of 1,606)?

TEST STATISTIC

- **Test statistic t:**

$$t = \frac{H_A - H_0}{\text{Standard Error of Difference}}$$

- **H_A: 2.77 (observed difference)**
- **H_0: 0 (difference if H_0 is true)**
- **Standard Error of Difference: 0.098**

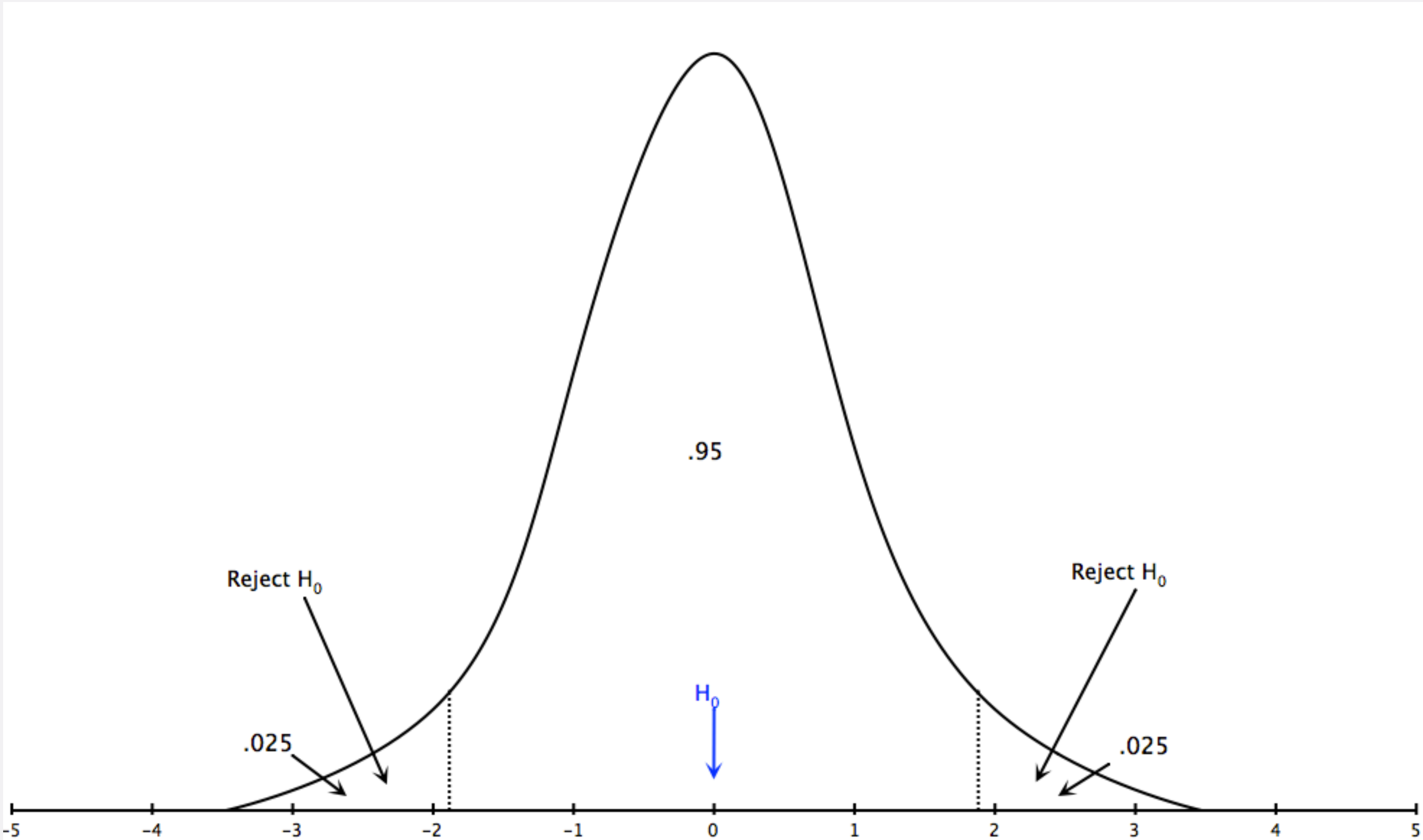
TEST STATISTIC

- **H_A: 2.77**
- **H_0: 0**
- **Standard Error of Difference: 0.098**

$$t = \frac{H_A - H_0}{\text{Standard Error of Difference}}$$

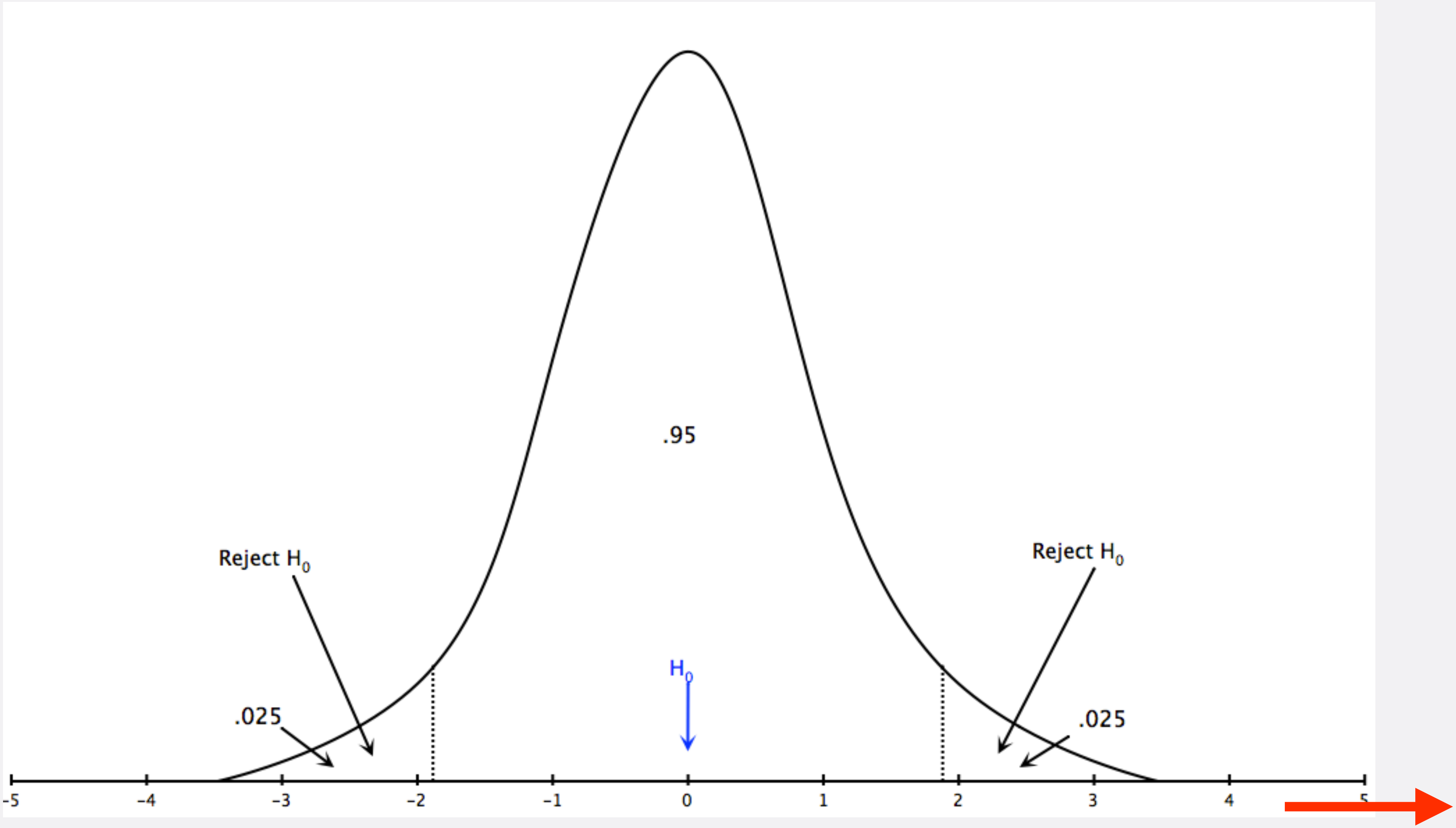
$$t = \frac{2.77 - 0.00}{0.098} = 28.26$$

SIGNIFICANCE TEST



- We reject H_0 if $t<-1.96$ or $t>1.96$
- This is equivalent to $p<0.05$

SIGNIFICANCE TEST



way over there

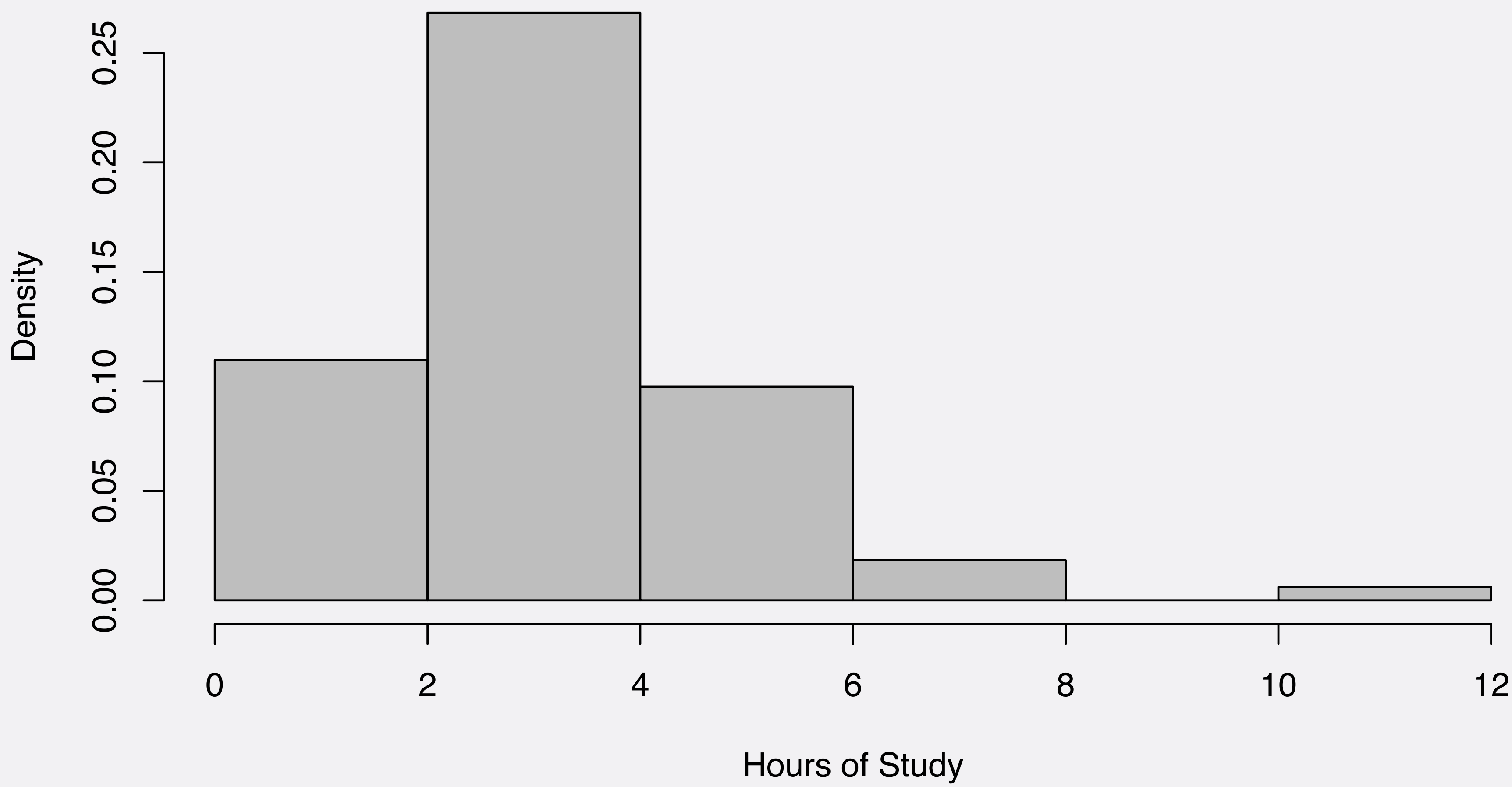
- **t-score: 28.26**

SIGNIFICANCE TEST

- With n=1,606, a mean difference of 2.77 (SE 0.098) produces a t-statistic of 28.26
- We reject H_0 if $t < -1.96$ or $t > 1.96$
    - It is extremely extremely unlikely to find such a large difference in a sample if there is no difference in the population
- So in this case, we reject null hypothesis that there is no difference between R and D in evaluation of Obamacare

EXAMPLE

- **On a typical day, how many hours do you spend studying/revising/preparing for your classes, not counting time in class itself?**



GENDER AND STUDYING

Gender	Mean Hours	Frequency	Standard Error
Female	3.66	54	0.24
Male	3.05	27	0.30
Difference	0.61	81	0.39

- Do men really study less than women?

TEST STATISTIC

- H_A: 0.61
- H_0: 0
- Standard Error of Difference: 0.39

TEST STATISTIC

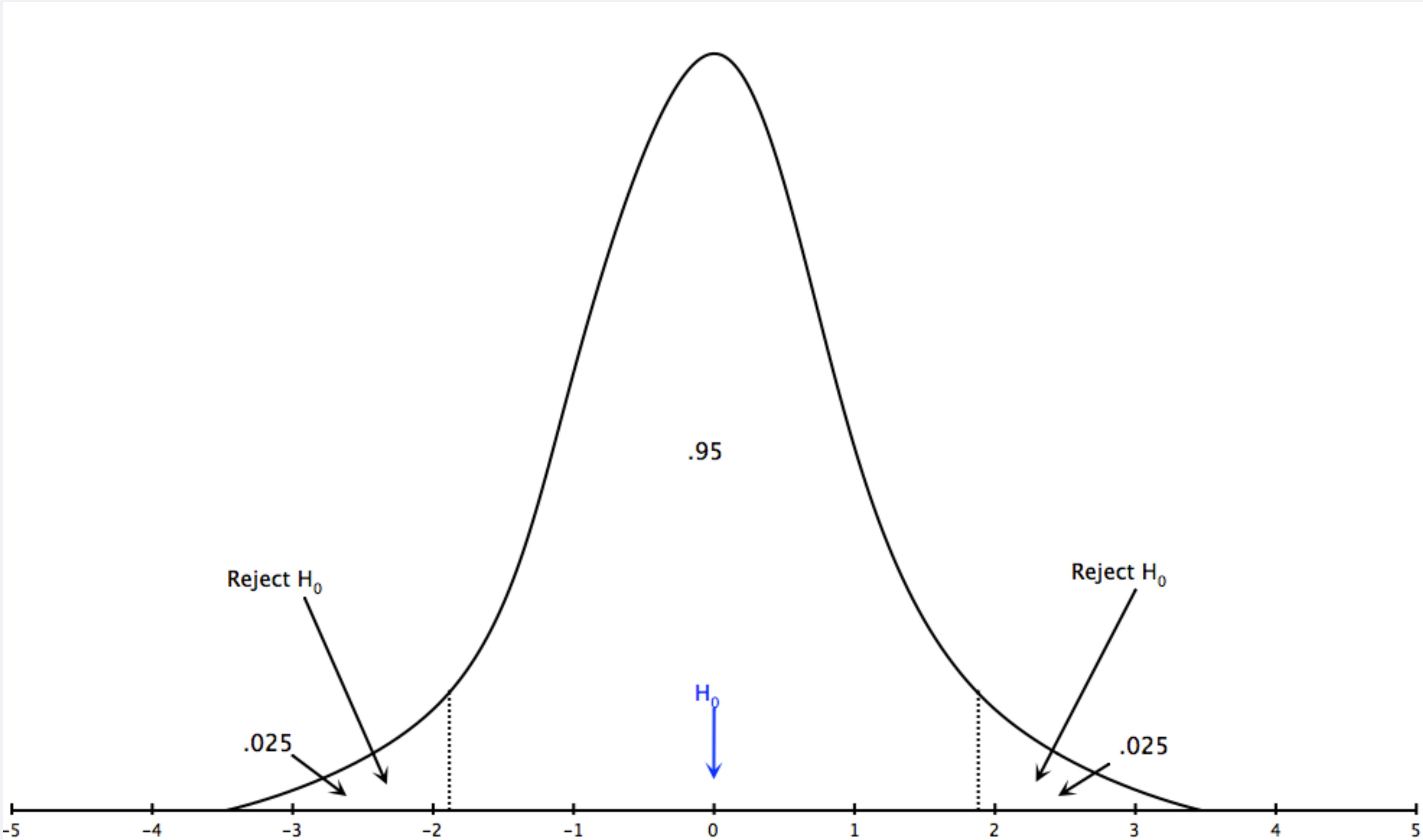
- **H_A: 0.61**
- **H_0: 0**
- **Standard Error of Difference: 0.39**

$$t = \frac{H_A - H_0}{\text{Standard Error}}$$

$$= \frac{0.61 - 0.0}{0.39}$$

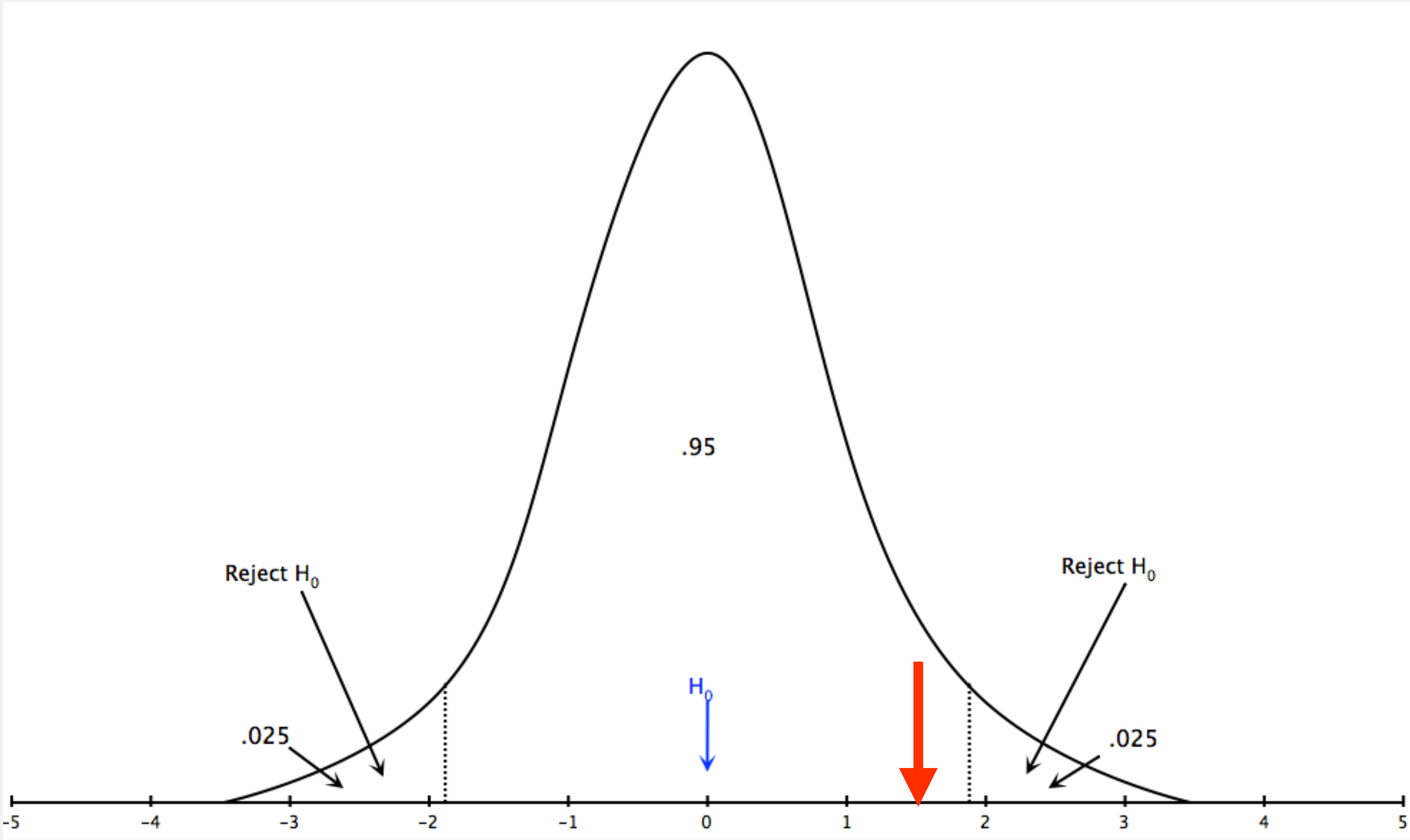
$$= 1.56$$

SIGNIFICANCE TEST



- We reject H_0 if $t<-1.96$ or $t>1.96$
- This is equivalent to $p<0.05$

SIGNIFICANCE TEST



- **t-score: 1.56**

SIGNIFICANCE TEST

- We *cannot* reject H_0
- Chance to get a difference of 0.61 hours (or larger) in sample of 81 students if there is no difference between men and women in population of students is larger than 5%