TESTING HYPOTHESES WITH DATA

ST101 – DR. ARIC LABARR

HYPOTHESIS TESTING

- A **hypothesis test** uses data to help evaluate an initial claim about a parameter from the population.

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25

Do you still think the coin is fair?

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125

Do you still think the coin is fair?

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125
4	Heads	0.0625

Do you still think the coin is fair?

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125
4	Heads	0.0625
5	Heads	0.03125

Do you still think the coin is fair?

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125
4	Heads	0.0625
5	Heads	0.03125

- No longer believe the coin is fair.

HYPOTHESIS TESTING THROUGH EXAMPLE

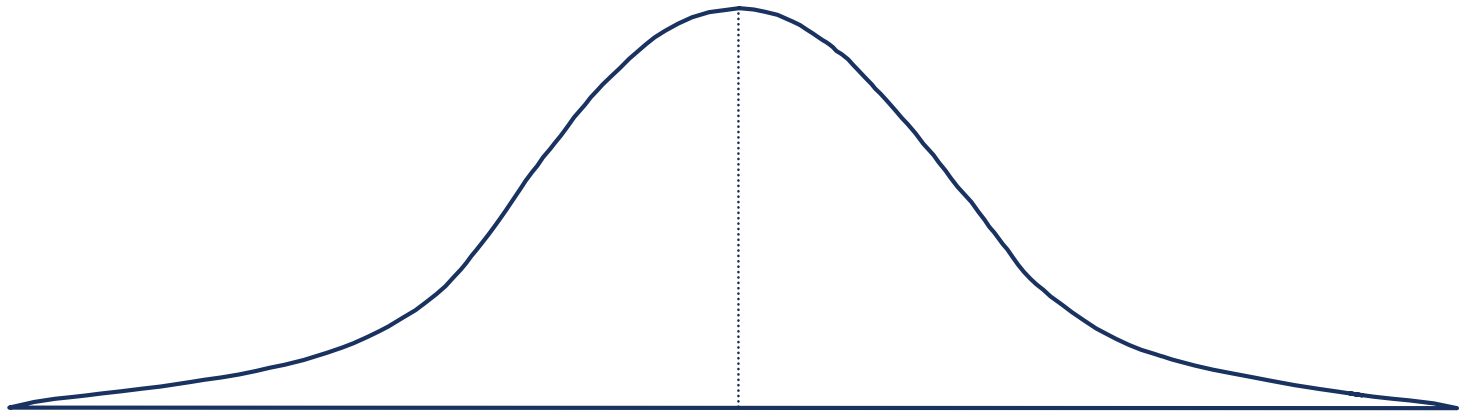
- I have a coin that you believe is fair to start. NULL Hypothesis
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results. Test Statistic

Flip Number	Result	P-value
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125
4	Heads	0.0625
5	Heads	0.03125

- No longer believe the coin is fair. Decision on NULL Hypothesis

BIKE DATA EXAMPLE WITH MEANS

- According to the CLT, sample means follow a Normal distribution as long as the sample size is big enough.

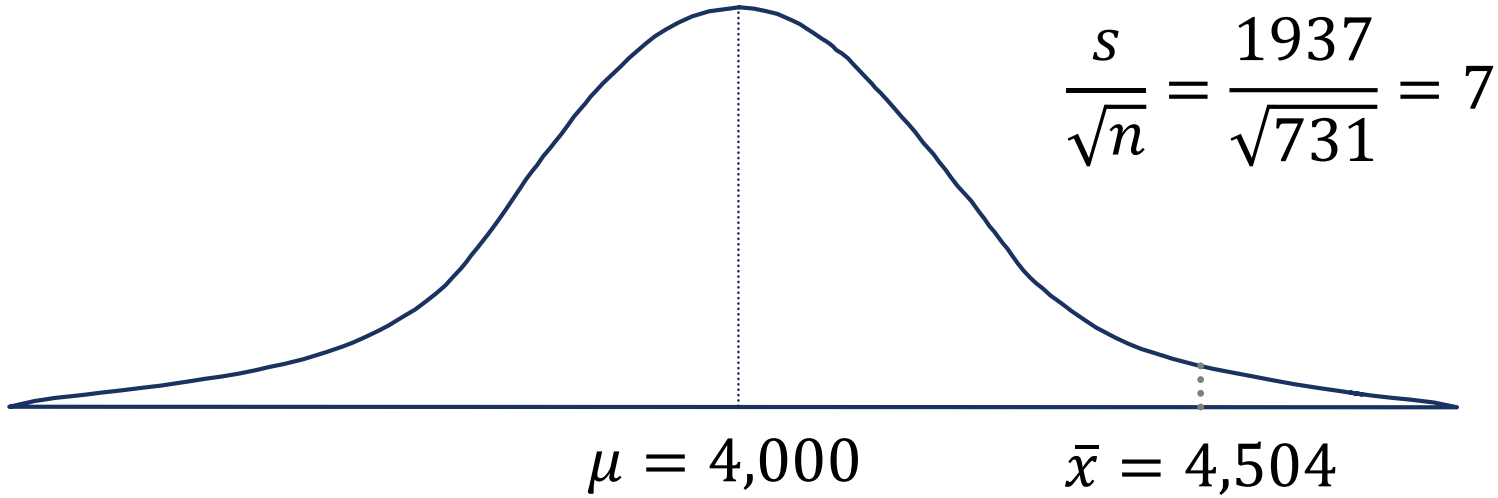


BIKE DATA EXAMPLE WITH MEANS

- You believe the average daily number of total users is 4,000, but you want to know if there is more than that. You collect a sample of 731 days with an average daily number of total users at 4,504 with a standard deviation of 1,937.

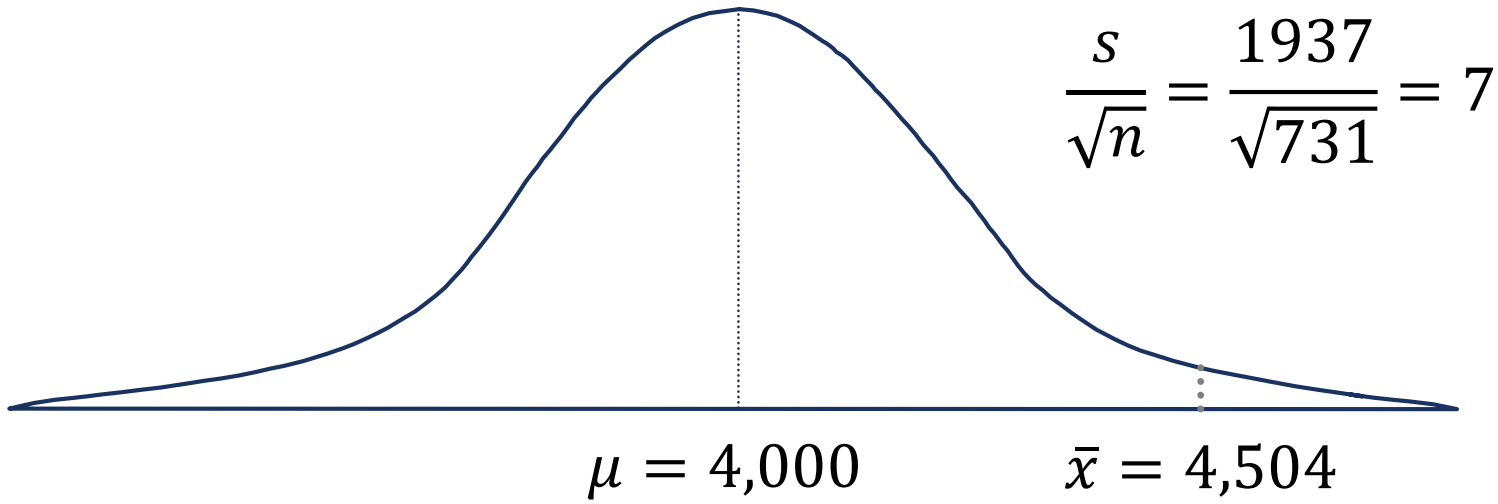
BIKE DATA EXAMPLE WITH MEANS

- You believe the average daily number of total users is 4,000, but you want to know if there is more than that. You collect a sample of 731 days with an average daily number of total users at 4,504 with a standard deviation of 1,937.
- What is the probability you see this under the initial thought of 4,000 for an average?



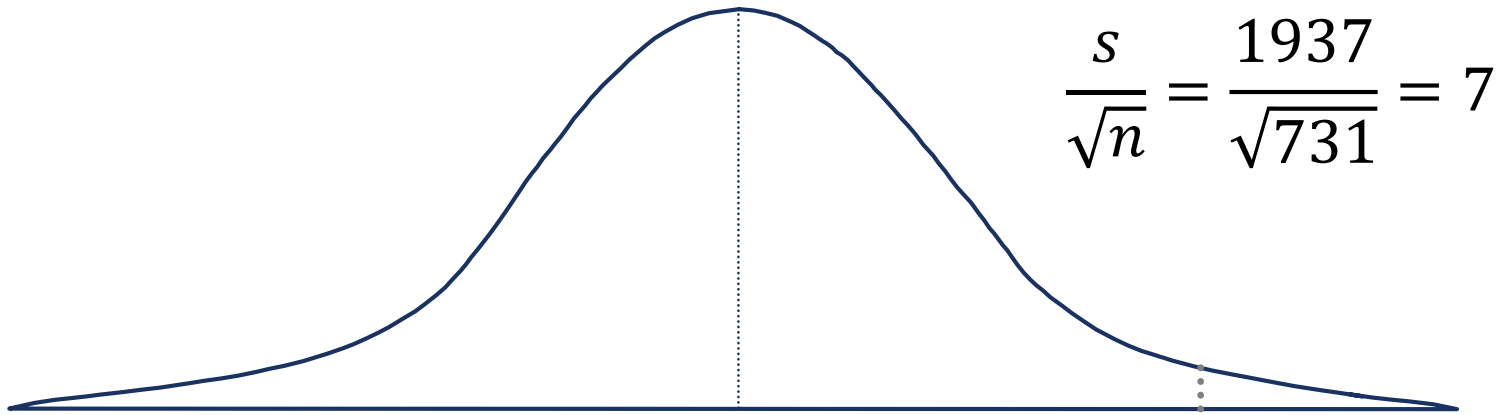
BIKE DATA EXAMPLE WITH MEANS

- What is the probability you see this under the initial thought of 4,000 for an average?



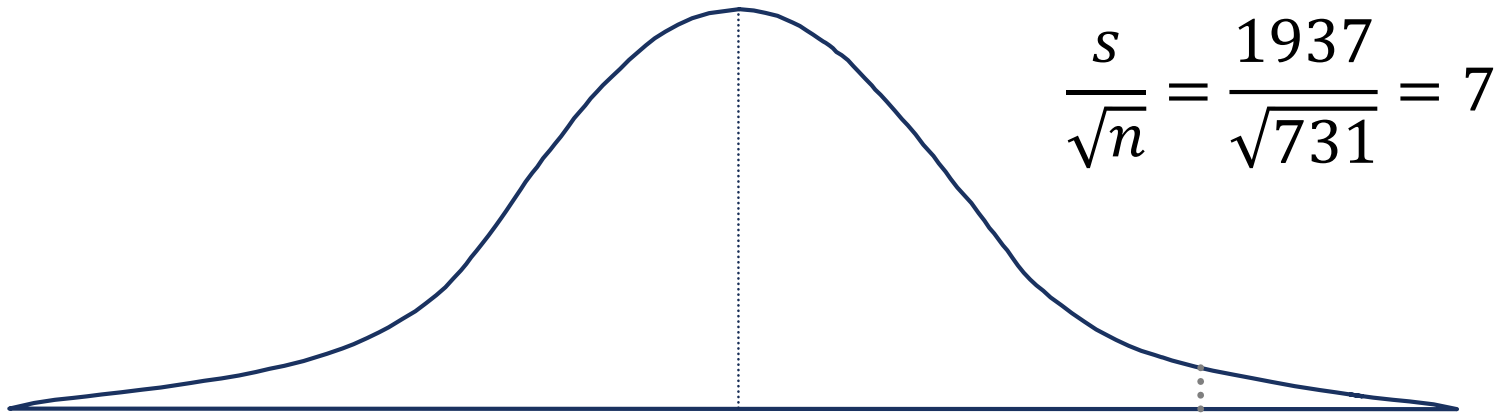
BIKE DATA EXAMPLE WITH MEANS

- What is the probability you see this under the initial thought of 4,000 for an average?



BIKE DATA EXAMPLE WITH MEANS

- What is the probability you see this under the initial thought of 4,000 for an average? < 0.0001



BIKE DATA EXAMPLE WITH MEANS

- You believe the average daily number of total users is 4,000, but you want to know if there is more than that.
- You collect a sample of 731 days with an average daily number of total users at 4,504 with a standard deviation of 1,937.
- What is the probability you see this under the initial thought of 4,000 for an average? < 0.0001!
- Do you still believe your original hypothesis?

BIKE DATA EXAMPLE WITH MEANS

- You believe the average daily number of total users is 4,000 (NULL Hypothesis), but you want to know if there is more than that.
- You collect a sample of 731 days with an average daily number of total users at 4,504 with a standard deviation of 1,937. Test Statistic
- What is the probability you see this under the initial thought of 4,000 for an average? P-value
- Do you still believe your original hypothesis? Decision on NULL Hypothesis

SUMMARY

- A hypothesis test uses data to help evaluate an initial claim about a parameter from the population.
- There are 4 main steps to hypothesis testing:
  1. State the hypotheses
  2. Test statistic
  3. P-value
  4. Decision on null hypothesis

NULL AND ALTERNATIVE HYPOTHESIS

TESTING HYPOTHESES WITH DATA

HYPOTHESIS TESTING

- **Hypothesis Testing** uses data to help evaluate an initial claim about a parameter from the population.
- The **null hypothesis**, denoted by H_0, is a tentative assumption about a population parameter.
- The **alternative hypothesis**, denoted by H_a, is the opposite of what is stated in the null hypothesis.
- The hypothesis testing procedure uses data from a sample to test these two competing hypotheses.

DEVELOPING NULL AND ALTERNATIVE

- It is not always obvious how the null and alternative hypotheses should be formulated.
- The context of the situation is very important in determining how the hypotheses should be stated.
- In some cases it is easier to identify the alternative hypothesis first!
- Typically, the alternative is what we are trying to test and want to collect evidence for.

NULL HYPOTHESIS, H_0

- The **null hypothesis** is the status quo, or the initial claim about the data.
- For example, the average daily number of total users is 4,000.

$$H_0: \mu = 4{,}000$$

NULL HYPOTHESIS, H_0

- The **null hypothesis** is the status quo, or the initial claim about the data.
- For example, the average daily number of total users is 4,000.

$$H_0: \mu = 4{,}000$$

- The null hypothesis is about the population parameter of interest, NOT sample statistics.
- Parameters are unknown, while statistics are known.

$H_0: \mu = 4{,}000$ $\quad$ ~~$H_0: \bar{x} = 4{,}000$~~

NULL HYPOTHESIS, H_0

- The **null hypothesis** is the status quo, or the initial claim about the data.
- For example, the average daily number of total users is 4,000.

$$H_0: \mu = 4{,}000$$

- This is the truth until you can prove otherwise – innocent until proven guilty.
- Always contains **one** of the following: $=, \geq, \leq$
- May reject the null (guilty) or fail to reject the null (not guilty).

ALTERNATIVE HYPOTHESIS, H_a

- The **alternative hypothesis** is the opposite of the null hypothesis.
- For example, the average daily number of total users is greater than 4,000.

$$H_a: \mu > 4{,}000$$

- This is typically what we are trying to prove.
- Always contains **one** of the following: $\neq$, $<$, $>$
- Never say we prove it – it isn't the claim on trial.

NULL VS. ALTERNATIVE

- Equality piece of the hypothesis is contained in the null hypothesis.
- Hypotheses are about a population parameter like μ.
- General Forms:

$$H_0: \mu \leq \mu_0 \quad H_a: \mu > \mu_0$$

$$H_0: \mu \geq \mu_0 \quad H_a: \mu < \mu_0$$

One-Sided Tests

$$H_0: \mu = \mu_0 \quad H_a: \mu \neq \mu_0$$

Two-Sided Test

SUMMARY

- The null hypothesis, denoted by H_0, is a tentative assumption about a population parameter.
- The alternative hypothesis, denoted by H_a, is the opposite of what is stated in the null hypothesis.
- A one-sided test happens when the alternative hypothesis only points in one direction (either only > or only <), while a two-sided test allows both directions to be possible.

TEST STATISTIC

TESTING HYPOTHESES WITH DATA

TEST STATISTIC

- The **test statistic** summarizes the amount of information provided in the sample.
- Imagine this like evidence in a court case.
- Test statistics have a common form:

$$\text{Test Statistic} = \frac{\text{Statistic} - \text{Null Value}}{\text{Standard Error}}$$

TEST STATISTIC

- The **test statistic** summarizes the amount of information provided in the sample.
- Imagine this like evidence in a court case.
- Test statistics have a common form:

$$\text{Test Statistic} = \frac{\text{Statistic} - \text{Null Value}}{\text{Standard Error}}$$

Sample Information

TEST STATISTIC

- The **test statistic** summarizes the amount of information provided in the sample.
- Imagine this like evidence in a court case.
- Test statistics have a common form:

$$\text{Test Statistic} = \frac{\text{Statistic} - \text{Null Value}}{\text{Standard Error}}$$

Null Hypothesis Information

TEST STATISTIC

- The **test statistic** summarizes the amount of information provided in the sample.
- Imagine this like evidence in a court case.
- Test statistics have a common form:

$$\text{Test Statistic} = \frac{\text{Statistic} - \text{Null Value}}{\text{Standard Error}}$$

Estimated Variability from Sampling Distribution of Statistic

TEST STATISTIC FOR MEANS

- The **test statistic** summarizes the amount of information provided in the sample.
- Sample means need the t-distribution because of the unknown values of the population standard deviation.

$$t = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$$

TEST STATISTIC FOR PROPORTIONS

- The **test statistic** summarizes the amount of information provided in the sample.
- Sample proportions use the Normal distribution.

$$z = \frac{\hat{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)}$$

SUMMARY

- The test statistic summarizes the amount of information provided in the sample.
- The test statistic calculation typically requires 3 pieces of information:
  1. Statistic – information obtained from the sample.
  2. Null value – information about the null hypothesis.
  3. Standard error – measure of variability for the sampling distribution of the statistic.

P-VALUE AND SIGNIFICANCE LEVEL

TESTING HYPOTHESES WITH DATA

P-VALUES

- Once the test statistic has been determined, we can calculate the probability that we got the information we did from our sample, **assuming that the null hypothesis is true**.
- The **p-value** is the probability we got our sample, or a sample more extreme, under the null hypothesis.

SIGNIFICANCE LEVEL VS. P-VALUE

- If the p-value is low, this implies that the sample we obtained from the population is **extremely rare** IF we assume that the null hypothesis is true.
- This leads us to question the validity of the null hypothesis – rejecting the null hypothesis if the p-value is low enough.
- How low is low enough?

SIGNIFICANCE LEVEL VS. P-VALUE

- If the p-value is low, this implies that the sample we obtained from the population is **extremely rare** IF we assume that the null hypothesis is true.
- This leads us to question the validity of the null hypothesis – rejecting the null hypothesis if the p-value is low enough.
- How low is low enough? Significance level – α

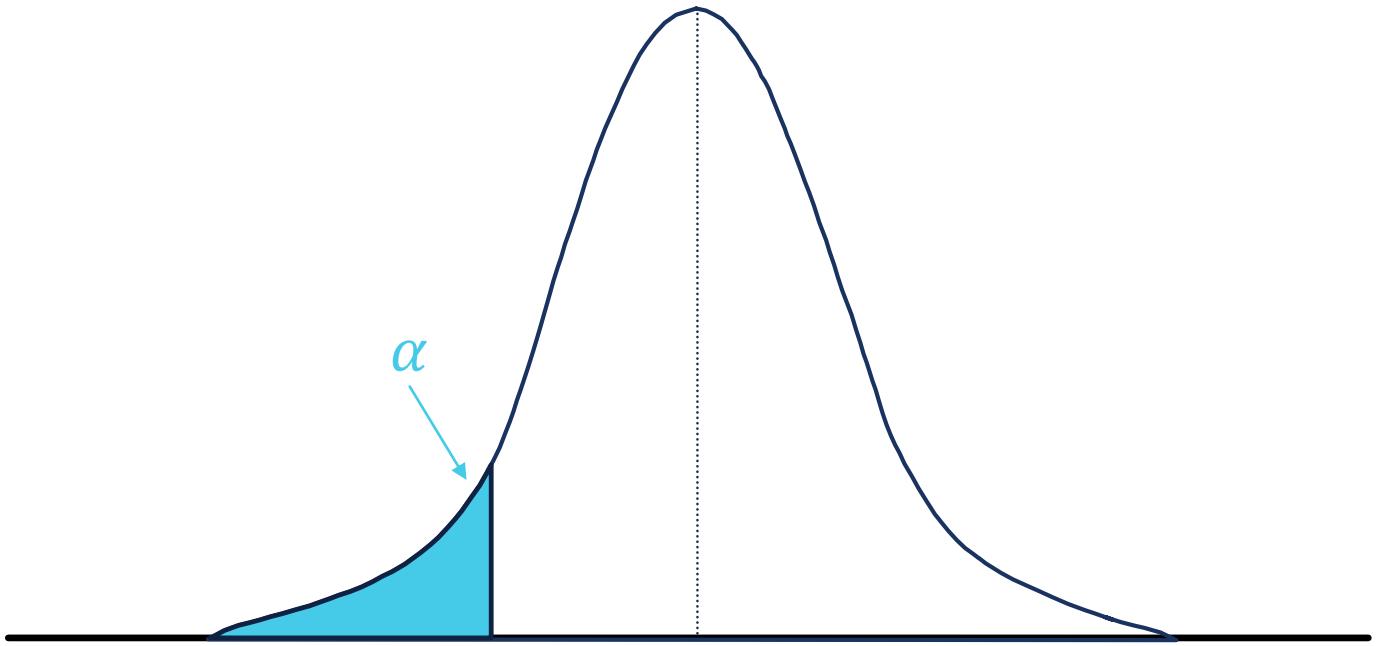
SIGNIFICANCE LEVEL VS. P-VALUE

- How low is low enough? Significance level – α
- If the p-value is less than or equal to the level of significance (α), the sample is so rare that we no longer believe the null hypothesis is true – called **rejecting the null hypothesis**.

SIGNIFICANCE LEVEL VS. P-VALUE

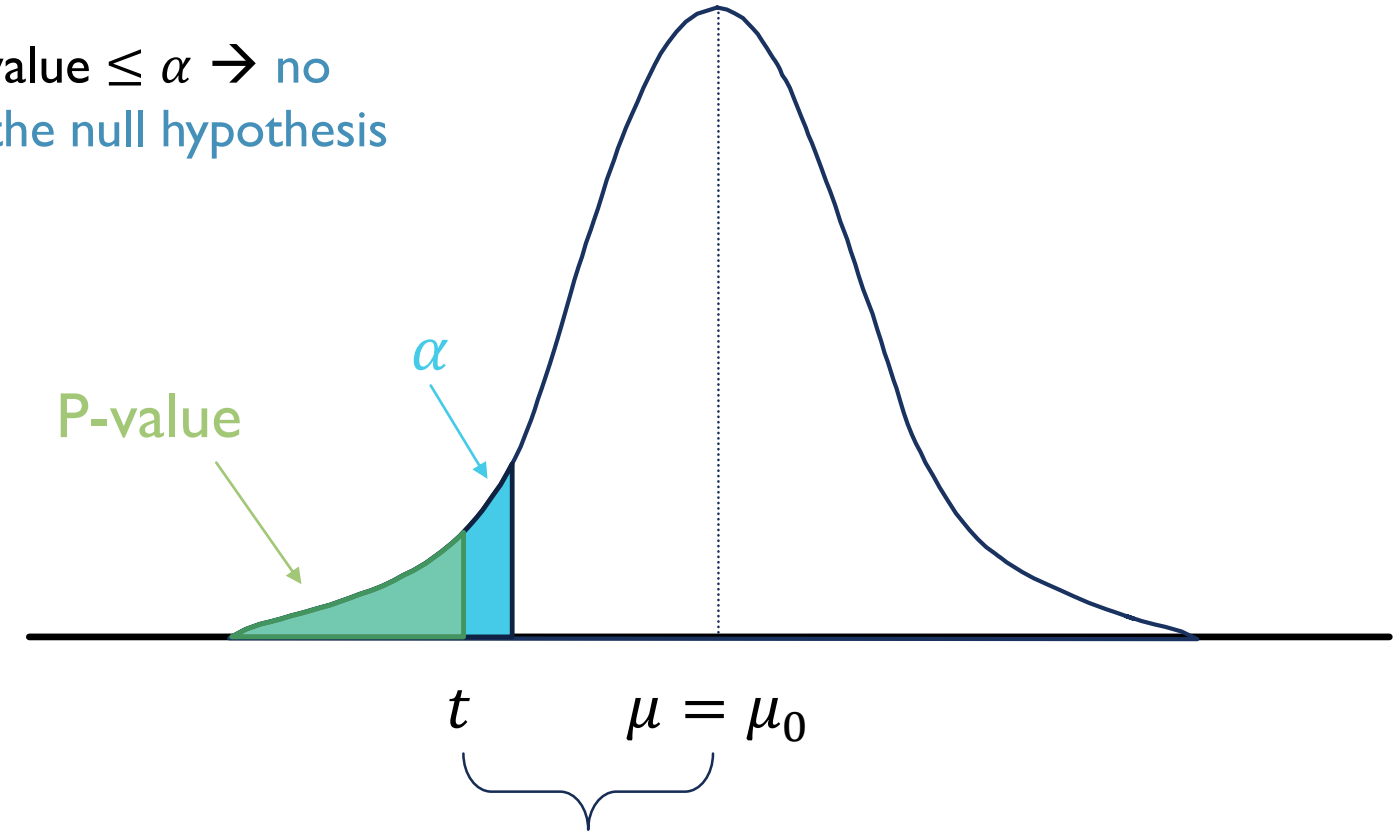
- How low is low enough? Significance level – α
- If the p-value is less than or equal to the level of significance (α), the sample is so rare that we no longer believe the null hypothesis is true – called **rejecting the null hypothesis**.
- 2 possible outcomes from a hypothesis test:
  - Reject H_0 if p-value $\leq \alpha$ → no longer believe the null hypothesis is true
  - Do not reject H_0 if p-value $> \alpha$ → not enough evidence to say null hypothesis isn't true

LOWER-TAILED TEST WITH P-VALUE – $H_a: \mu < \mu_0$



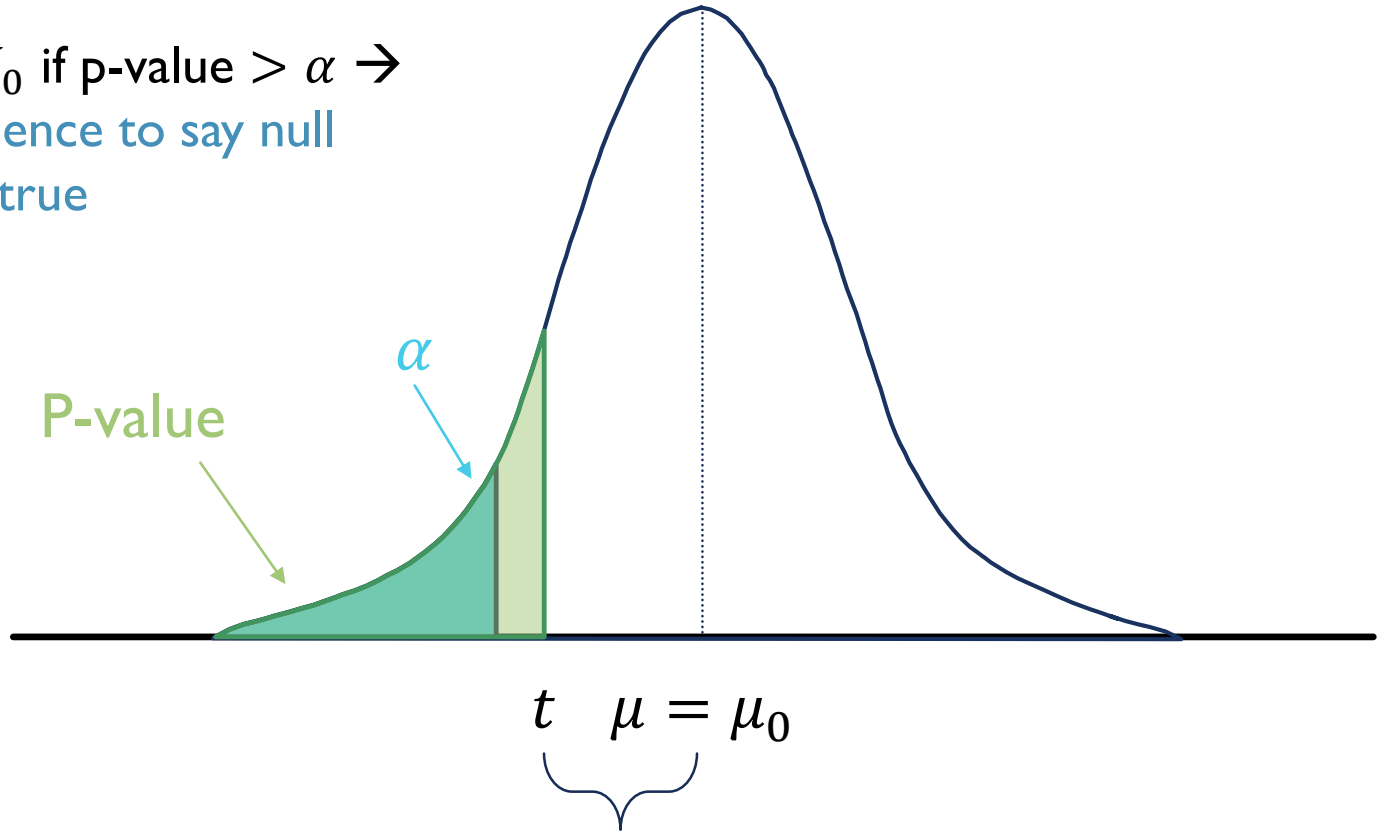
LOWER-TAILED TEST WITH P-VALUE – $H_a: \mu < \mu_0$

Reject H_0 if p-value $\leq \alpha$ → no longer believe the null hypothesis is true



LOWER-TAILED TEST WITH P-VALUE – $H_a: \mu < \mu_0$

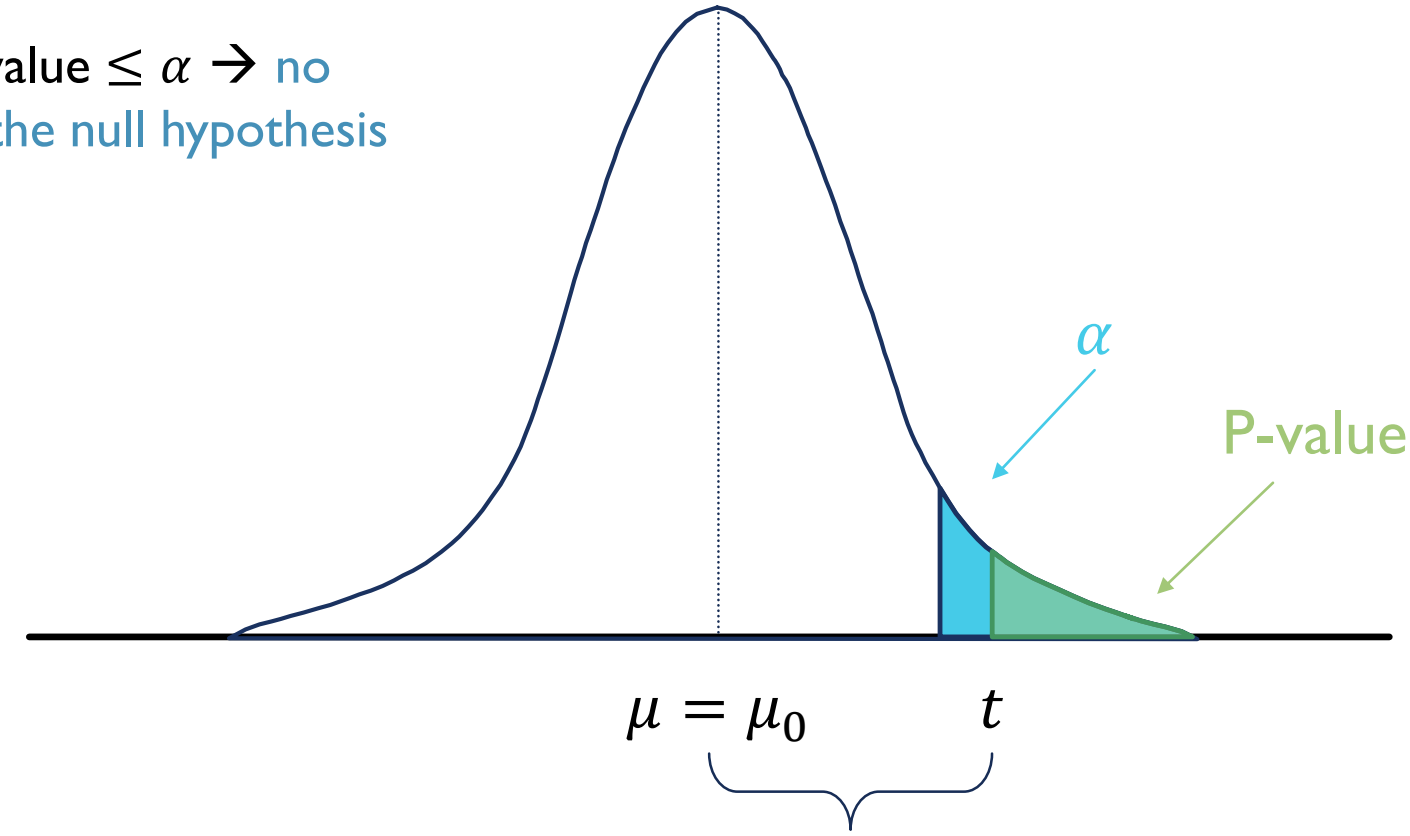
Do not reject H_0 if p-value $> \alpha$ →
not enough evidence to say null hypothesis isn't true



Values are "close together" according to p-value

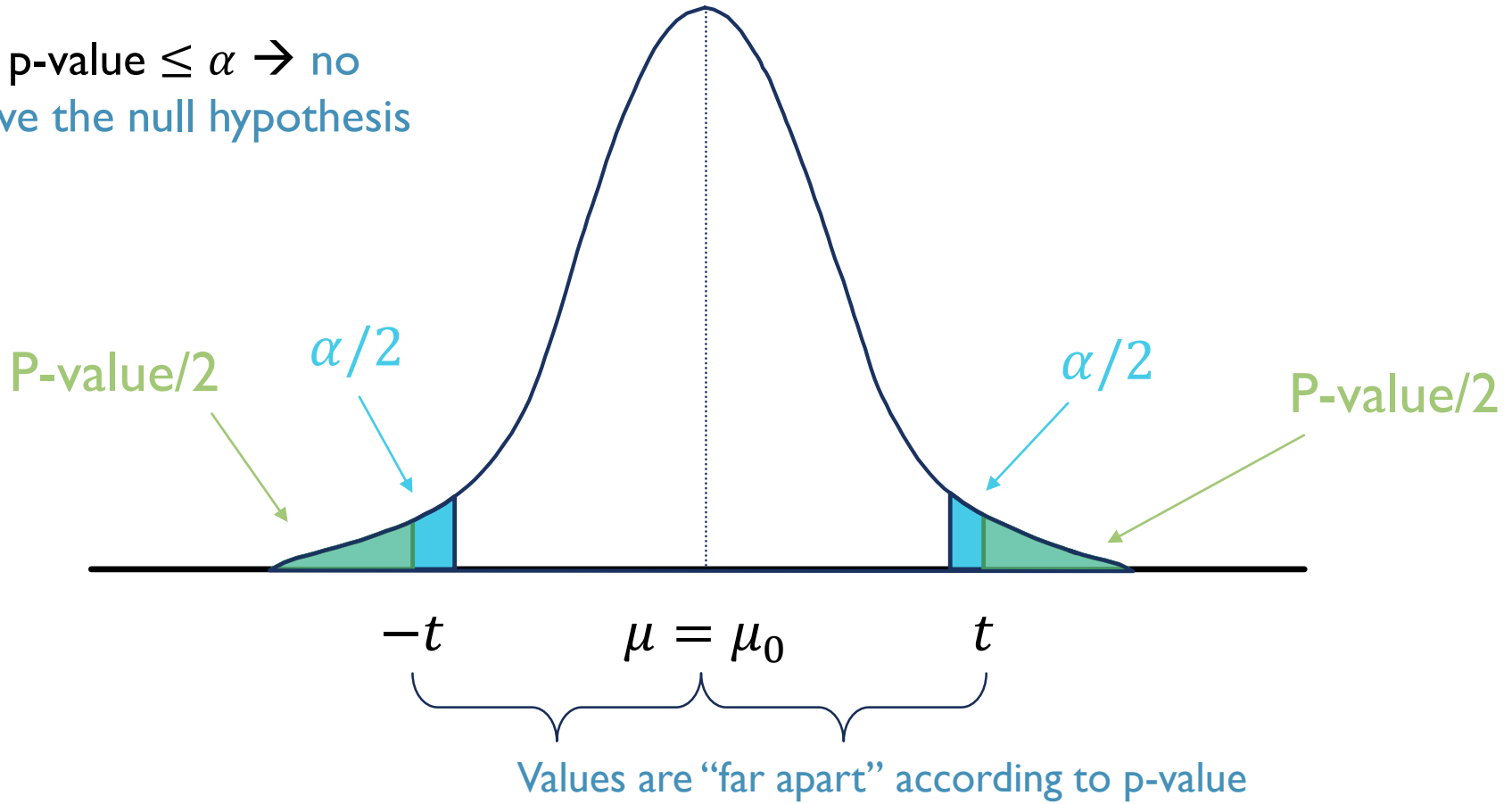
UPPER-TAILED TEST WITH P-VALUE – $H_a: \mu > \mu_0$

Reject H_0 if p-value $\leq \alpha$ → no longer believe the null hypothesis is true



TWO-TAILED TEST WITH P-VALUE – $H_a: \mu \neq \mu_0$

Reject H_0 if p-value $\leq \alpha$ → no longer believe the null hypothesis is true



HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125
4	Heads	0.0625
5	Heads	0.03125

- No longer believe the coin is fair – but could it be?

HYPOTHESIS TESTING THROUGH EXAMPLE

- I have a coin that you believe is fair to start.
- To test if this coin is fair, you ask me to flip the coin repeatedly and record the results.

Flip Number	Result	Probability
1	Heads	0.50
2	Heads	0.25
3	Heads	0.125
4	Heads	0.0625
5	Heads	0.03125

- No longer believe the coin is fair – but could it be? **YES!**

SIGNIFICANCE LEVEL, α

- Defines the unlikely values of the sample statistic **if the null hypothesis is true**.
- This area is typically called the **rejection region** of the sampling distribution.
- Selected before the hypothesis test is even run!
- Typical values are 0.01, 0.05, 0.10.

SUMMARY

- The p-value is the probability we got our sample, or a sample more extreme, under the null hypothesis.
- If the p-value is low, this implies that the sample we obtained from the population is extremely rare IF we assume that the null hypothesis is true.
- The significance level defines the unlikely values of the sample statistic if the null hypothesis is true.

HYPOTHESIS TEST FOR MEANS

TESTING HYPOTHESES WITH DATA

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

- You believe the average daily number of total users is 4,000, but you want to know if there is more than that so you can decide on orders for future bikes to be added.
- You collect a sample of 731 days with an average daily number of total users at 4,504 with a standard deviation of 1,937.
- With a significance level of 0.05, conduct a hypothesis test on this claim.

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

3. P-value

FINDING P-VALUE

One-Tail	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	.0005
Two-Tail	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
.	.	.	.	.	.	.	.	.	.	.
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
500	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
.	.	.	.	.	.	.	.	.	.	.

FINDING P-VALUE

One-Tail	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	.0005
Two-Tail	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
.	.	.	.	.	.	.	.	.	.	.
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
500	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
.	.	.	.	.	.	.	.	.	.	.

FINDING P-VALUE

One-Tail	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	.0005
Two-Tail	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
.	.	.	.	.	.	.	.	.	.	.
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
500	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
.	.	.	.	.	.	.	.	.	.	.

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

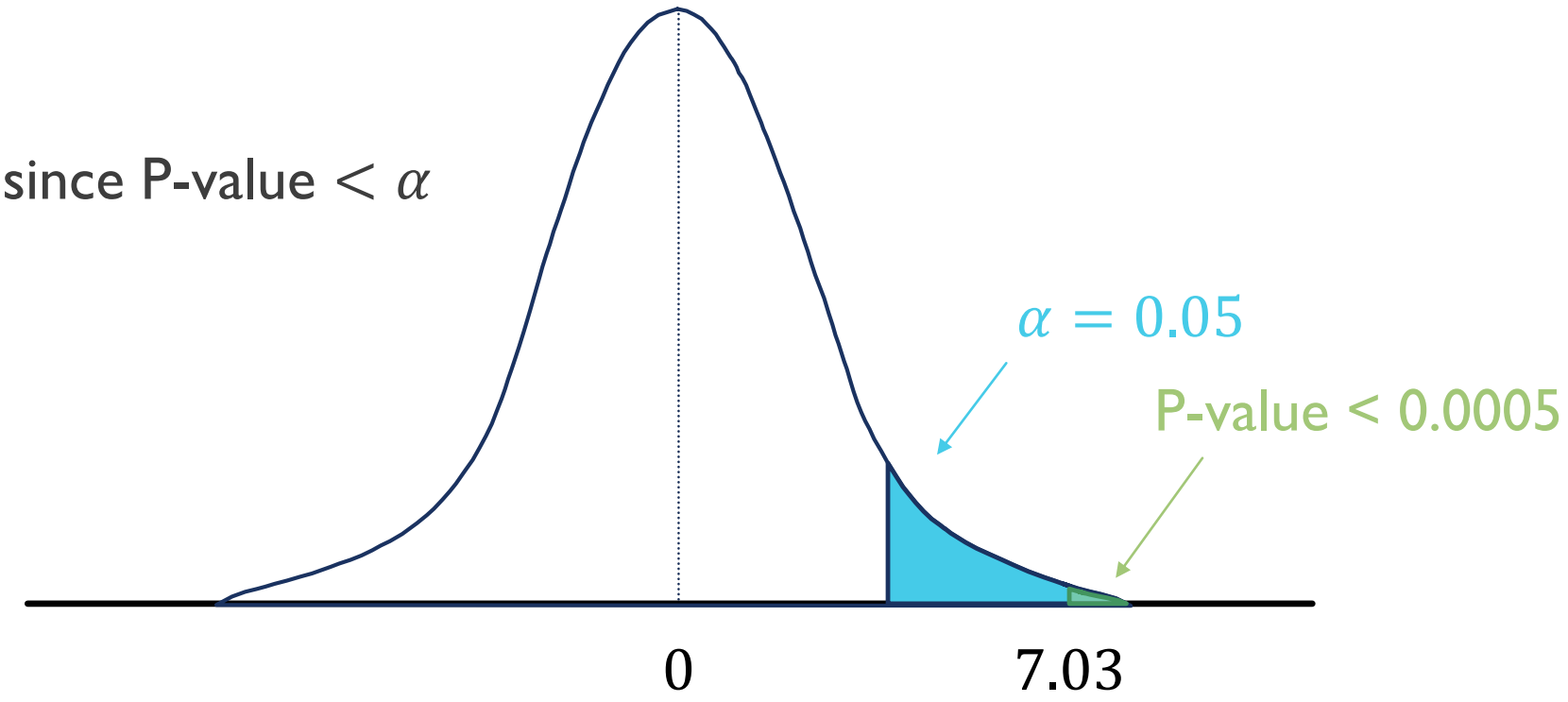
1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

3. P-value < 0.0005

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

Reject H_0 since P-value $< \alpha$



BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

3. P-value < 0.0005
4. Reject H_0

BIKE DATA EXAMPLE FOR TWO-TAIL HYPOTHESIS TEST

- You believe the average daily number of total users is 4,000, but you want to know if there is more than that so you can decide on orders for future bikes to be added OR less than 4,000 so you can pull stock from the streets, so bikes don't sit unused.
- You collect a sample of 731 days with an average daily number of total users at 4,504 with a standard deviation of 1,937.
- With a significance level of 0.05, conduct a hypothesis test on this claim.

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

3. P-value

FINDING P-VALUE

One-Tail	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	.0005
Two-Tail	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
.	.	.	.	.	.	.	.	.	.	.
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
500	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
.	.	.	.	.	.	.	.	.	.	.

FINDING P-VALUE

One-Tail	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	.0005
Two-Tail	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
.	.	.	.	.	.	.	.	.	.	.
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
500	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
.	.	.	.	.	.	.	.	.	.	.

FINDING P-VALUE

One-Tail	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	.0005
Two-Tail	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
.	.	.	.	.	.	.	.	.	.	.
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
500	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
.	.	.	.	.	.	.	.	.	.	.

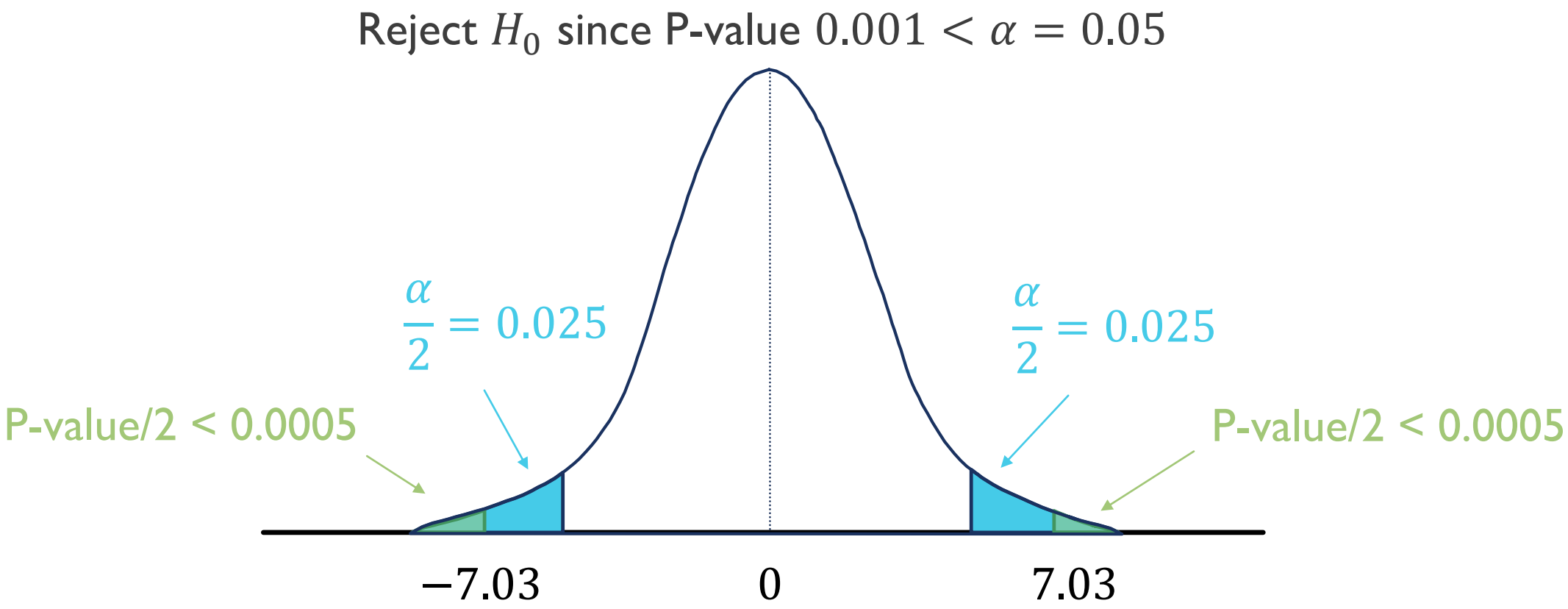
BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

3. P-value < 0.001 (Twice that of one-sided P-value)

BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST



BIKE DATA EXAMPLE FOR ONE-TAIL HYPOTHESIS TEST

1. $H_0: \mu \leq 4{,}000$
   $H_a: \mu > 4{,}000$
2. Sample data: $\bar{x} = 4{,}504, s = 1{,}937, n = 731$

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{4{,}504 - 4{,}000}{\left(\frac{1{,}937}{\sqrt{731}}\right)} = 7.03$$

3. P-value < 0.001 (Twice that of one-sided P-value)
4. Reject H_0

ETHICS AROUND INFERENCE WITH DATA

TESTING HYPOTHESES WITH DATA

ERRORS IN HYPOTHESIS TESTS

- Hypothesis tests depend on sample data.
- Therefore, hypothesis tests may be wrong!
- There are two types of errors in hypothesis testing – **Type I and Type II errors**.

TYPE I VS. TYPE II ERRORS

CHOICE	TRUTH: H_0 True	TRUTH: H_0 False
Do Not Reject H_0	Correct	Type II
Reject H_0	Type I	Correct

TYPE I ERROR

- A **Type I error** is rejecting the null hypothesis when the null hypothesis was actually true.
- In other words, you have a false rejection.
- The probability of making a Type I error in a hypothesis test is called the **significance level**.
- Most hypothesis tests are referred to as **significance tests** because they only control the Type I error.

TYPE II ERROR

- A **Type II error** is accepting the null hypothesis when the null hypothesis was actually false.
- In other words, you have falsely accepted.
- The probability of NOT making a Type II error in a hypothesis test is called the **power**.
- Difficult to control the Type II error.
- Can only control for Type I or Type II at a time.

CAREFUL WITH INFERENCE

- What if your sample of data happened to be drawn on data from only summer months with clear days?
  - Maybe the days would be estimated to have too many users.
  - This could lead to incorrect actions to be taken.

CAREFUL WITH INFERENCE

- What if your sample of data happened to be drawn on data from only summer months with clear days?
  - Maybe the days would be estimated to have too many users.
  - This could lead to incorrect actions to be taken.

- Hypothesis tests completely depend on the data they are built from.
  - Garbage in → Garbage out

CAREFUL WITH INFERENCE

- What if your sample of data happened to be drawn on data from only summer months with clear days?
  - Maybe the days would be estimated to have too many users.
  - This could lead to incorrect actions to be taken.
- Hypothesis tests completely depend on the data they are built from.
  - Garbage in → Garbage out
- Hypothesis tests results reveal something, but not everything!

CAREFUL ABOUT JUSTIFICATION

- Hypothesis tests results reveal something, but not everything!
- People sometimes forget the possibility of errors when making claims from a statistical test.
- For example:
  - "We know that more than 4,000 bikes per day are used on average."

CAREFUL ABOUT JUSTIFICATION

- Hypothesis tests results reveal something, but not everything!
- People sometimes forget the possibility of errors when making claims from a statistical test.
- For example:
  - "We have strong evidence that more than 4,000 bikes per day are used on average."

CAREFUL ABOUT JUSTIFICATION

- Hypothesis tests results reveal something, but not everything!
- People sometimes forget the possibility of errors when making claims from a statistical test.
- For example:
  - "We have strong evidence that more than 4,000 bikes per day are used on average."
- Remember the analogy of a court case → we incorrectly claim people are guilty sometimes. Careful about rushing to judgement!

SUMMARY

- A Type I error is rejecting the null hypothesis when the null hypothesis was actually true.
- A Type II error is accepting the null hypothesis when the null hypothesis was actually false.
- Hypothesis tests completely depend on the data they are built from.
- People sometimes forget the possibility of errors when making claims from a statistical test.