Hypothesis Testing I

Nate Wells

Math 141, 3/22/21

Outline

In this lecture, we will. . .

Outline

In this lecture, we will. . .

- Review hypothesis testing activity from Friday
- Discuss hypothesis testing framework

Section 1

Hypothesis Testing Activity

Is Yawning Contagious?

- In an episode of the US show *Mythbusters*, the hosts conducted an experiment to determine whether yawning is contagious.

Is Yawning Contagious?

- In an episode of the US show *Mythbusters*, the hosts conducted an experiment to determine whether yawning is contagious.
- 50 participants were divided into two groups:
  - 34 participants in the `seed` group were exposed to a yawn. Of these, 10 later yawned.
  - 16 participants in the `control` group were not exposed to a yawn. Of these 4 later yawned.

Is Yawning Contagious?

- In an episode of the US show *Mythbusters*, the hosts conducted an experiment to determine whether yawning is contagious.
- 50 participants were divided into two groups:
  - 34 participants in the `seed` group were exposed to a yawn. Of these, 10 later yawned.
  - 16 participants in the `control` group were not exposed to a yawn. Of these 4 later yawned.

	`yawned`		
`group`	No	Yes	Total
Control	12	4	16
Seed	24	10	34
Total	36	14	50

Is Yawning Contagious?

- In an episode of the US show *Mythbusters*, the hosts conducted an experiment to determine whether yawning is contagious.
- 50 participants were divided into two groups:
  - 34 participants in the `seed` group were exposed to a yawn. Of these, 10 later yawned.
  - 16 participants in the `control` group were not exposed to a yawn. Of these 4 later yawned.

	`yawned`		
`group`	No	Yes	Total
Control	12	4	16
Seed	24	10	34
Total	36	14	50

The sample statistics were:

$$\hat{p}_s = 0.29 \quad \hat{p}_c = 0.25 \quad \hat{p} = 0.28 \qquad \hat{p}_s - \hat{p}_c = 0.04$$

Is Yawning Contagious?

- In an episode of the US show *Mythbusters*, the hosts conducted an experiment to determine whether yawning is contagious.
- 50 participants were divided into two groups:
  - 34 participants in the `seed` group were exposed to a yawn. Of these, 10 later yawned.
  - 16 participants in the `control` group were not exposed to a yawn. Of these 4 later yawned.

`group`	`yawned` No	Yes	Total
Control	12	4	16
Seed	24	10	34
Total	36	14	50

The sample statistics were:

$$\hat{p}_s = 0.29 \quad \hat{p}_c = 0.25 \quad \hat{p} = 0.28 \qquad \hat{p}_s - \hat{p}_c = 0.04$$

- If yawning is not contagious, how likely is it such a difference in proportion would be observed just due to chance?

Sampling Variability

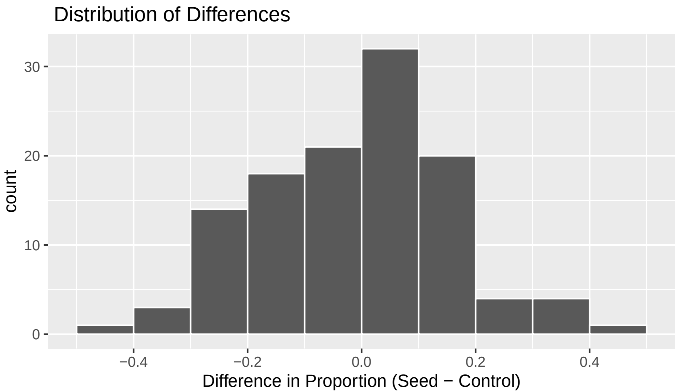
- We assume that exactly 14 / 50 participants would yawn, and randomly assigned them to groups of size 16 (`control`) and 34 (`seed`).

Sampling Variability

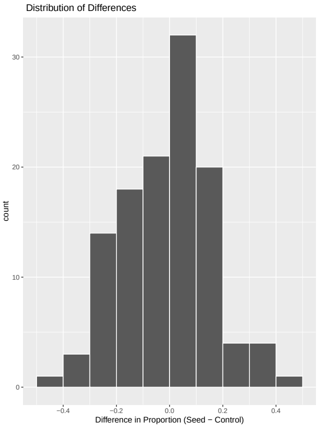
- We assume that exactly 14 / 50 participants would yawn, and randomly assigned them to groups of size 16 (`control`) and 34 (`seed`).
  - Compute proportion of yawns in each group and difference in proportion between groups.
  - Replicate this procedure many times:

Sampling Variability

- We assume that exactly 14 / 50 participants would yawn, and randomly assigned them to groups of size 16 (`control`) and 34 (`seed`).
    - Compute proportion of yawns in each group and difference in proportion between groups.
    - Replicate this procedure many times:

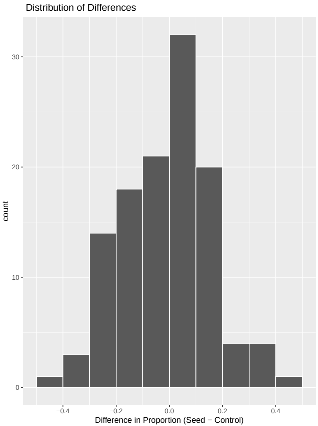


Follow-up Questions:



1. How does our observed proportion compare?
   $\hat{p}_s - \hat{p}_c = 0.044$

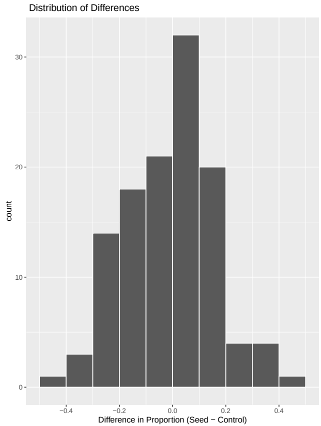
Follow-up Questions:



1. How does our observed proportion compare?
   $\hat{p}_s - \hat{p}_c = 0.044$

2. If there is truly no relationship between being exposed to a yawn and yawning, is it plausible to observe the difference in proportion we did?

Follow-up Questions:



1. How does our observed proportion compare? $\hat{p}_s - \hat{p}_c = 0.044$
2. If there is truly no relationship between being exposed to a yawn and yawning, is it plausible to observe the difference in proportion we did?
3. Based on the data, do we have strong enough evidence to conclude yawning is contagious?

Section 2

Hypothesis Testing Framework

A Fair Coin

- A fair coin should land heads about 50% of the time

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.
- The probability a coin flips heads n times in a row is 0.5^n.

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.
- The probability a coin flips heads n times in a row is 0.5^n.
  - i.e. The probability of 4 heads in a row is 6.25%, while 5 heads in a rows is 3.125%

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.
- The probability a coin flips heads n times in a row is 0.5^n.
  - i.e. The probability of 4 heads in a row is 6.25%, while 5 heads in a rows is 3.125%

Let's do an experiment. I'll flip a coin several times and count how many heads I get in a row.

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.
- The probability a coin flips heads n times in a row is 0.5^n.
  - i.e. The probability of 4 heads in a row is 6.25%, while 5 heads in a rows is 3.125%

Let's do an experiment. I'll flip a coin several times and count how many heads I get in a row.

- If you think the coin is unfair, choose the No button on the participants panel of Zoom.

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.
- The probability a coin flips heads n times in a row is 0.5^n.
  - i.e. The probability of 4 heads in a row is 6.25%, while 5 heads in a rows is 3.125%

Let's do an experiment. I'll flip a coin several times and count how many heads I get in a row.

- If you think the coin is unfair, choose the No button on the participants panel of Zoom.
- If I repeat this experiment in 30 classes over the next several years, I expect to get 5 heads in a row about 1 one of them.

A Fair Coin

- A fair coin should land heads about 50% of the time
- But coin flips are also random events, so it is possible for unlikely events to occur.
- The probability a coin flips heads n times in a row is 0.5^n.
  - i.e. The probability of 4 heads in a row is 6.25%, while 5 heads in a rows is 3.125%

Let's do an experiment. I'll flip a coin several times and count how many heads I get in a row.

- If you think the coin is unfair, choose the No button on the participants panel of Zoom.
- If I repeat this experiment in 30 classes over the next several years, I expect to get 5 heads in a row about 1 one of them.
- How willing are you to observe an improbable event and mistakenly conclude your original belief was incorrect?

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question
2. Identify hypotheses

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question
2. Identify hypotheses
3. Obtain samples and generate bootstrap samples

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question
2. Identify hypotheses
3. Obtain samples and generate bootstrap samples
4. Calculate sample statistics

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question
2. Identify hypotheses
3. Obtain samples and generate bootstrap samples
4. Calculate sample statistics
5. Compute liklihood of observing statistic under original hypothesis

Framework for Hypothesis Testing

Hypothesis Testing represents a type of scientific experiment, and so should follow the general scientific method.

1. Present research question
2. Identify hypotheses
3. Obtain samples and generate bootstrap samples
4. Calculate sample statistics
5. Compute liklihood of observing statistic under original hypothesis
6. Determine statistical significance and make conclusion on research question

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.
  - H_a: This coin is biased to produce heads more often than tail. The probability it lands heads is $p > 0.5$.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.
  - H_a: This coin is biased to produce heads more often than tail. The probability it lands heads is $p > 0.5$.
- The Null and Alternative hypotheses are **always** statements about populations, and *often* are statements about the particular values of population parameters.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.
  - H_a: This coin is biased to produce heads more often than tail. The probability it lands heads is $p > 0.5$.
- The Null and Alternative hypotheses are **always** statements about populations, and *often* are statements about the particular values of population parameters.
- The **null value** is the value of the population parameter under the Null Hypothesis.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.
  - H_a: This coin is biased to produce heads more often than tail. The probability it lands heads is $p > 0.5$.
- The Null and Alternative hypotheses are **always** statements about populations, and *often* are statements about the particular values of population parameters.
- The **null value** is the value of the population parameter under the Null Hypothesis.
- H_0 and H_a are **never** statements about particular values of sample statistics. They are **hypotheses** and should be able to be expressed before any observation of data.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.
  - H_a: This coin is biased to produce heads more often than tail. The probability it lands heads is $p > 0.5$.
- The Null and Alternative hypotheses are **always** statements about populations, and *often* are statements about the particular values of population parameters.
- The **null value** is the value of the population parameter under the Null Hypothesis.
- H_0 and H_a are **never** statements about particular values of sample statistics. They are **hypotheses** and should be able to be expressed before any observation of data.
  - Incorrect H_0: The proportion of heads in 5 flips of the coin is $\hat{p} = 0.5$.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical prespective or that there is no relationship among several variables.
  - H_0: This is a fair coin. The probability it lands heads is $p = 0.5$.
- The **alternative hypothesis** H_a is the negation of the null hypothesis. It is often the theory we would like to prove.
  - H_a: This coin is biased to produce heads more often than tail. The probability it lands heads is $p > 0.5$.
- The Null and Alternative hypotheses are **always** statements about populations, and *often* are statements about the particular values of population parameters.
- The **null value** is the value of the population parameter under the Null Hypothesis.
- H_0 and H_a are **never** statements about particular values of sample statistics. They are **hypotheses** and should be able to be expressed before any observation of data.
  - Incorrect H_0: The proportion of heads in 5 flips of the coin is $\hat{p} = 0.5$.
  - Incorrect H_a: The proportion of heads in 5 flips of the coin was $\hat{p} = 1 > 0.5$.

Types of Alternative Hypotheses

- While there is only one logical negation of the Null Hypothesis, there are several statements contrary to the Null Hypothesis.

Types of Alternative Hypotheses

- While there is only one logical negation of the Null Hypothesis, there are several statements contrary to the Null Hypothesis.
  - If $H_0 : p = 0.5$, the logical negation is $H_a : p \neq 0.5$.

Types of Alternative Hypotheses

- While there is only one logical negation of the Null Hypothesis, there are several statements contrary to the Null Hypothesis.
  - If $H_0 : p = 0.5$, the logical negation is $H_a : p \neq 0.5$.
  - But two contrary statements include:
    1. $H_a : p > 0.5$;
    2. $H_a : p < 0.5$

Types of Alternative Hypotheses

- While there is only one logical negation of the Null Hypothesis, there are several statements contrary to the Null Hypothesis.
  - If $H_0 : p = 0.5$, the logical negation is $H_a : p \neq 0.5$.
  - But two contrary statements include:
    1. $H_a : p > 0.5$;
    2. $H_a : p < 0.5$
- The alternate hypothesis in a **two-sided hypothesis test** proposes that the population parameter is not equal null value.

Types of Alternative Hypotheses

- While there is only one logical negation of the Null Hypothesis, there are several statements contrary to the Null Hypothesis.
  - If $H_0 : p = 0.5$, the logical negation is $H_a : p \neq 0.5$.
  - But two contrary statements include:
    1. $H_a : p > 0.5$;
    2. $H_a : p < 0.5$
- The alternate hypothesis in a **two-sided hypothesis test** proposes that the population parameter is not equal null value.
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value.

Types of Alternative Hypotheses

- While there is only one logical negation of the Null Hypothesis, there are several statements contrary to the Null Hypothesis.
  - If $H_0 : p = 0.5$, the logical negation is $H_a : p \neq 0.5$.
  - But two contrary statements include:
    1. $H_a : p > 0.5$;
    2. $H_a : p < 0.5$
- The alternate hypothesis in a **two-sided hypothesis test** proposes that the population parameter is not equal null value.
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value.
- Default to using two-sided hypothesis tests. Only use one-sided tests when you are truly interested in only a single direction of effect.

Liklihood of Observing Sample Statistic

- In order to compare the Null and Alternate Hypotheses, we need to quantify how likely it was to observe a particular sample statistic.

Liklihood of Observing Sample Statistic

- In order to compare the Null and Alternate Hypotheses, we need to quantify how likely it was to observe a particular sample statistic.
- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.

Liklihood of Observing Sample Statistic

- In order to compare the Null and Alternate Hypotheses, we need to quantify how likely it was to observe a particular sample statistic.
- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distingish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**

Liklihood of Observing Sample Statistic

- In order to compare the Null and Alternate Hypotheses, we need to quantify how likely it was to observe a particular sample statistic.
- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distingish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**
  - In the prior experiment, we flipped a coin 5 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.

Liklihood of Observing Sample Statistic

- In order to compare the Null and Alternate Hypotheses, we need to quantify how likely it was to observe a particular sample statistic.
- The **p-value** of a sample is the probability of observing a sample statistic at least as favorable to the alternative hypothesis as the current statistic, if H_0 were true.
- To distinguish between sample statistics generally and the particular one obtained from the sample, we call the latter the **test statistic**
  - In the prior experiment, we flipped a coin 5 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 5 heads in 5 flips)} = 0.5^5 = 0.03125$$

- The P-Value quantifies the strength of evidence against the Null Hypothesis. Smaller P-values represent stronger evidence to reject H_0.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
    - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$
  - Our test was statistically significant and we reject the hypothesis that the coin is fair in favor of the hypothesis that the coin is more likely to flip heads.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$
  - Our test was statistically significant and we reject the hypothesis that the coin is fair in favor of the hypothesis that the coin is more likely to flip heads.
- We should **always** choose the value of α prior to conducting an experiment and observing data. *Usually* the choice is made for us depending on conventions in our field of study.

Statistical Significance

- How do we decide what counts as *sufficient evidence* to reject the null hypothesis in favor of the alternative?
- The threshold for an experiment is called the **significance level** (usually denote with α).
  - In many fields, the standard significance level is $\alpha = 0.05$. But this may vary widely depending on application.
- If the P-Value is less than the prescribed significance level of the test, we say the data is **statistically significant** and provides good evidence to reject H_0 in favor of H_a.
  - In the coin flip experiment, $\text{P-value} < \alpha$ since $0.03125 < 0.05$
  - Our test was statistically significant and we reject the hypothesis that the coin is fair in favor of the hypothesis that the coin is more likely to flip heads.
- We should **always** choose the value of α prior to conducting an experiment and observing data. *Usually* the choice is made for us depending on conventions in our field of study.
- Choosing a significance level of $\alpha = 0.05$ means that we treat any result that would have occurred by chance alone less than 5% of the time as good evidence that the null hypothesis is false.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.
  - The coin is actually fair. But we saw an unlikely event and claimed the coin was biased.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
  - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.
  - The coin is actually fair. But we saw an unlikely event and claimed the coin was biased.
- A **Type 2 Error** occurs when we fail to reject H_0 when it is in fact false.

Types of Errors

- Hypothesis Tests give framework for comparing uncertainty, but do not guarantee that our conclusion will never be in err.
    - Remember: Unlikely things happen. All of the time.
- There are four possible outcomes to a hypothesis test, summarized below:

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	Correct Decision	Type 1 Error
	H_A true	Type 2 Error	Correct Decision

- A **Type 1 Error** occurs when we reject H_0 when it is actually true.
    - The coin is actually fair. But we saw an unlikely event and claimed the coin was biased.
- A **Type 2 Error** occurs when we fail to reject H_0 when it is in fact false.
    - The coin was indeed biased. But we withheld judgment since unlikely events do happen from time to time.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

  - In general, computing power can be difficult, and requires we investigate the distribution of a sample statistic under the alternative hypothesis.

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

  - In general, computing power can be difficult, and requires we investigate the distribution of a sample statistic under the alternative hypothesis.

With great power comes...

Significance Level and Power

- The significance level of a hypothesis test corresponds to our willingness to make Type I errors.
- Decreasing the significance level decreases the number of Type I errors made across a large number of experiments.
- Is there a cost to decreasing significance level to ensure we do not make Type I errors?
  - Yes! Because decreasing the significance level also makes it less likely we will reject H_0, and so usually increases the chance of making a Type 2 error.
- The **power** of a statistical test is the probability of correctly rejecting the null hypothesis when it is false. That is

$$\text{Power} = 1 - \text{Probability of Type II Error}$$

  - In general, computing power can be difficult, and requires we investigate the distribution of a sample statistic under the alternative hypothesis.

With great power comes...greater chance of Type I error.