Hypothesis Testing

Prof. Wells

STA 209, 3/15/23

Outline

In this lecture, we will. . .

Outline

In this lecture, we will. . .

- Introduce hypothesis tests as too for assessing strength of statistical evidence
- Discuss hypothesis testing framework
- Implement the hypothesis testing framework in a specific example

Section 1

Coin Flipping

Heads Up

- In the long run, a fair coin should land heads about 50% of the time

Heads Up

- In the long run, 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.

Heads Up

- In the long run, 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.

Heads Up

- In the long run, 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 5 heads in a row is 3.125%, while 8 heads in a rows is 0.39%

Heads Up

- In the long run, 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 5 heads in a row is 3.125%, while 8 heads in a rows is 0.39%
- If I flip a coin 8 times in each of 256 classes over the next several years, I expect to get 8 heads in a row in 1 one of them.

Heads Up

- In the long run, 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 5 heads in a row is 3.125%, while 8 heads in a rows is 0.39%
- If I flip a coin 8 times in each of 256 classes over the next several years, I expect to get 8 heads in a row in 1 one of them.
- But, I have spent **many** hours practicing flipping coins, and have perfected a technique to flip heads every time.

Heads Up

- In the long run, 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 5 heads in a row is 3.125%, while 8 heads in a rows is 0.39%
- If I flip a coin 8 times in each of 256 classes over the next several years, I expect to get 8 heads in a row in 1 one of them.
- But, I have spent **many** hours practicing flipping coins, and have perfected a technique to flip heads every time.

Let's do an experiment. I'll flip a coin 8 times and count how many heads I get in a row.

- If and when you believe me that I have a coin-flipping technique, raise your hand.

Heads Up

So. . .

Heads Up

So. . .

- What are some possible explanations?

Heads Up

So. . .

- What are some possible explanations?
    - I have a special coin flipping technique
    - I lied about the result
    - The coin was not fair

Heads Up

So. . .

- What are some possible explanations?
  - I have a special coin flipping technique
  - I lied about the result
  - The coin was not fair
  - We witnessed an unlikely event for a fair coin, and the result is due to chance

Heads Up

So. . .

- What are some possible explanations?
  - I have a special coin flipping technique
  - I lied about the result
  - The coin was not fair
  - We witnessed an unlikely event for a fair coin, and the result is due to chance
- The guiding principle of hypothesis testing is:

  The more unlikely an event is under one hypothesis, the more credence we should give to alternative hypotheses

Section 2

Hypothesis Testing Framework

Scientific Method

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

Scientific Method

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

1. Present research question

Scientific Method

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

1. Present research question
2. Identify hypotheses

Scientific Method

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 data

Scientific Method

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 data
4. Calculate relevant statistics

Scientific Method

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 data
4. Calculate relevant statistics
5. Compute likelihood of observing statistic under original hypothesis

Scientific Method

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 data
4. Calculate relevant statistics
5. Compute likelihood of observing statistic under original hypothesis
6. Determine statistical significance and make conclusion on research question

Informal vs. Formal Hypotheses

- Before the coin flipping experiment, we may have several (informal) hypotheses:

Informal vs. Formal Hypotheses

- Before the coin flipping experiment, we may have several (informal) hypotheses:
  - The coin is fair
  - Prof. Wells can always flip heads
  - The coin is unfair
  - Prof. Wells will lie about the results

Informal vs. Formal Hypotheses

- Before the coin flipping experiment, we may have several (informal) hypotheses:
  - The coin is fair
  - Prof. Wells can always flip heads
  - The coin is unfair
  - Prof. Wells will lie about the results
- But in order to compare these, it would be helpful to consider a set of hypotheses that:
  1. Are mutually exclusive
  2. Make specific statements about a parameter
  3. Do not discuss the specific outcome of the experiment

Informal vs. Formal Hypotheses

- Before the coin flipping experiment, we may have several (informal) hypotheses:
  - The coin is fair
  - Prof. Wells can always flip heads
  - The coin is unfair
  - Prof. Wells will lie about the results
- But in order to compare these, it would be helpful to consider a set of hypotheses that:
  1. Are mutually exclusive
  2. Make specific statements about a parameter
  3. Do not discuss the specific outcome of the experiment
- Let p denote the true probability that the coin flips heads.

Informal vs. Formal Hypotheses

- Before the coin flipping experiment, we may have several (informal) hypotheses:
  - The coin is fair
  - Prof. Wells can always flip heads
  - The coin is unfair
  - Prof. Wells will lie about the results
- But in order to compare these, it would be helpful to consider a set of hypotheses that:
  1. Are mutually exclusive
  2. Make specific statements about a parameter
  3. Do not discuss the specific outcome of the experiment
- Let p denote the true probability that the coin flips heads.

$$\text{Hypothesis 1:} \quad p = 0.5 \qquad\qquad \text{Hypothesis 2:} \quad p > 0.5$$

Informal vs. Formal Hypotheses

- Before the coin flipping experiment, we may have several (informal) hypotheses:
  - The coin is fair
  - Prof. Wells can always flip heads
  - The coin is unfair
  - Prof. Wells will lie about the results
- But in order to compare these, it would be helpful to consider a set of hypotheses that:
  1. Are mutually exclusive
  2. Make specific statements about a parameter
  3. Do not discuss the specific outcome of the experiment
- Let p denote the true probability that the coin flips heads.

$$\text{Hypothesis 1:} \quad p = 0.5 \qquad\qquad \text{Hypothesis 2:} \quad p > 0.5$$

- The first informal hypothesis is represented by Hypothesis 1. The other three are represented by Hypothesis 2.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective 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 perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to 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 perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.

Identify Hypotheses

- The **null hypothesis** H_0 is the claim we are testing. It often represents a skeptical perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- 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 perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- 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 perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- 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 perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- 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 perspective or that there is no relationship among several variables.
  - H_0: The probability of heads is 50%, or $p = 0.5$.
- The **alternative hypothesis** H_a is contrary to the null hypothesis. It is often the theory we would like to prove.
  - H_a: The probability of heads is greater than 50%, or $p > 0.5$.
- 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 $\hat{p} = 1 > 0.5$.

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class. . .

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class. . .
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class. . .
    - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
    - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
    - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class. . .
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
  - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)
- Because of the logic of hypothesis testing, the null hypothesis should represent the *status quo* belief about the parameter.

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class...
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
  - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)
- Because of the logic of hypothesis testing, the null hypothesis should represent the *status quo* belief about the parameter.
  - It is the default that would be assumed if no statistical investigation were conducted, and will be the position maintained if the study is inconclusive.

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class. . .
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
  - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)
- Because of the logic of hypothesis testing, the null hypothesis should represent the *status quo* belief about the parameter.
  - It is the default that would be assumed if no statistical investigation were conducted, and will be the position maintained if the study is inconclusive.
- The alternative hypothesis often represents the research goal, or the claim for which we seek evidence.

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class...
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
  - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)
- Because of the logic of hypothesis testing, the null hypothesis should represent the *status quo* belief about the parameter.
  - It is the default that would be assumed if no statistical investigation were conducted, and will be the position maintained if the study is inconclusive.
- The alternative hypothesis often represents the research goal, or the claim for which we seek evidence.
  - It is the only statement we will be able to provide evidence for after our test.

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class...
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
  - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)
- Because of the logic of hypothesis testing, the null hypothesis should represent the *status quo* belief about the parameter.
  - It is the default that would be assumed if no statistical investigation were conducted, and will be the position maintained if the study is inconclusive.
- The alternative hypothesis often represents the research goal, or the claim for which we seek evidence.
  - It is the only statement we will be able to provide evidence for after our test.
- In the coin flipping experiment, all else equal, we assume that a coin is fair. But I claimed that I had a technique for producing heads.

Determining the Null Hypothesis

- The Null and Alternative Hypothesis statements are *not* interchangeable. In our class. . .
  - The null hypothesis will (mostly) be a statement of equality for a parameter (i.e $p = .5$)
  - The alternative hypothesis will be a statement of inequality for a parameter (i.e. $p > .5$)
  - Other types of hypotheses are explored in further statistics classes (STA 310 / STA 336)
- Because of the logic of hypothesis testing, the null hypothesis should represent the *status quo* belief about the parameter.
  - It is the default that would be assumed if no statistical investigation were conducted, and will be the position maintained if the study is inconclusive.
- The alternative hypothesis often represents the research goal, or the claim for which we seek evidence.
  - It is the only statement we will be able to provide evidence for after our test.
- In the coin flipping experiment, all else equal, we assume that a coin is fair. But I claimed that I had a technique for producing heads.
  - The null hypothesis is that the coin is fair. The alternative is that coin flips heads more often than not.

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. (i.e. $p \neq .5$)

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. (i.e. $p \neq .5$)
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value (i.e. one of $p > .5$ or $p < .5$)

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. (i.e. $p \neq .5$)
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value (i.e. one of $p > .5$ or $p < .5$)
- Default to using two-sided hypothesis tests. Only use one-sided tests when you are truly interested in only a single direction of effect.

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. (i.e. $p \neq .5$)
- The alternate hypothesis in a **one-sided hypothesis test** proposes that the population parameter is less than (or greater than) the null value (i.e. one of $p > .5$ or $p < .5$)
- Default to using two-sided hypothesis tests. Only use one-sided tests when you are truly interested in only a single direction of effect.
  - In the coin flipping experiment, we *were* interested in verifying my claim that I could flip heads consistently, so we did use a one-sided hypothesis ($p > .5$)

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
  - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
  - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)
  - What is the greatest number of heads you would plausibly expect to see?

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
  - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)
  - What is the greatest number of heads you would plausibly expect to see?
  - If I had 7 out of 8 heads, would you still believe the coin was fair?

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
    - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)
    - What is the greatest number of heads you would plausibly expect to see?
    - If I had 7 out of 8 heads, would you still believe the coin was fair?
    - How likely is is that 7 or more heads occur, if the coin were fair?

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
  - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)
  - What is the greatest number of heads you would plausibly expect to see?
  - If I had 7 out of 8 heads, would you still believe the coin was fair?
  - How likely is is that 7 or more heads occur, if the coin were fair?
- To answer questions like these, we need to know the distribution of the statistic of interest, *if the null hypothesis were true*.

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
  - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)
  - What is the greatest number of heads you would plausibly expect to see?
  - If I had 7 out of 8 heads, would you still believe the coin was fair?
  - How likely is is that 7 or more heads occur, if the coin were fair?
- To answer questions like these, we need to know the distribution of the statistic of interest, *if the null hypothesis were true*.
  - This distribution is called the **Null Distribution** and is the theoretical sampling distribution for the statistic if the null hypothesis were true.

Likelihood of Observing Sample Statistic

- To compare Null and Alternate Hypotheses, we need to quantify how likely it is to observe a particular sample statistic, *if the null hypothesis were true*.
  - If I flip a *fair* coin 8 times, do you expect me to get exactly 4 heads? (Why / Why not?)
  - What is the greatest number of heads you would plausibly expect to see?
  - If I had 7 out of 8 heads, would you still believe the coin was fair?
  - How likely is is that 7 or more heads occur, if the coin were fair?
- To answer questions like these, we need to know the distribution of the statistic of interest, *if the null hypothesis were true*.
  - This distribution is called the **Null Distribution** and is the theoretical sampling distribution for the statistic if the null hypothesis were true.
  - We can approximate the Null Distribution using simulation, randomization and bootstrapping.

A Model of Coin Flipping

We can use R to simulate one experiment of 8 coin flips by

A Model of Coin Flipping

We can use R to simulate one experiment of 8 coin flips by

- Creating a data frame consisting of Heads and Tails

```
coin <- data.frame(face = c("Heads", "Tails"))
```

A Model of Coin Flipping

We can use R to simulate one experiment of 8 coin flips by

- Creating a data frame consisting of Heads and Tails

```
coin <- data.frame(face = c("Heads", "Tails"))
```

- Sampling from this data frame *with replacement* 8 times

```
coin %>%rep_sample_n(coin, size = 8, replace = T)
```

```
##   replicate  face
## 1         1 Tails
## 2         1 Tails
## 3         1 Tails
## 4         1 Heads
## 5         1 Tails
## 6         1 Heads
## 7         1 Heads
## 8         1 Tails
```

A Model of Coin Flipping

We can use R to simulate one experiment of 8 coin flips by

- Creating a data frame consisting of Heads and Tails

```
coin <- data.frame(face = c("Heads", "Tails"))
```

- Sampling from this data frame *with replacement* 8 times

```
coin %>%rep_sample_n(coin, size = 8, replace = T)
```

```
##   replicate  face
## 1         1 Tails
## 2         1 Tails
## 3         1 Tails
## 4         1 Heads
## 5         1 Tails
## 6         1 Heads
## 7         1 Heads
## 8         1 Tails
```

- Computing the number and proportion of heads obtained in this one experiment

```
coin %>% rep_sample_n(size = 8, replace = T) %>% summarize(n_heads = sum(face == "Heads")) %>%
  mutate(p_hat = n_heads/8)
```

```
##   n_heads p_hat
## 1       3 0.375
```

A Model of Coin Flipping

We can use R to simulate 2000 experiments of 8 coin flips by changing `reps = 1` to `reps = 2000`

A Model of Coin Flipping

We can use R to simulate 2000 experiments of 8 coin flips by changing `reps = 1` to `reps = 2000`

```
coin %>% rep_sample_n(size = 8, replace = T, reps = 2000) %>%
  summarize(n_heads = sum(face == "Heads")) %>% mutate(p_hat = n_heads/8)
```

```
## # A tibble: 2,000 x 3
##    replicate n_heads p_hat
##        <int>   <int> <dbl>
##  1         1       5 0.625
##  2         2       5 0.625
##  3         3       4 0.5
##  4         4       4 0.5
##  5         5       3 0.375
##  6         6       3 0.375
##  7         7       3 0.375
##  8         8       2 0.25
##  9         9       3 0.375
## 10        10       2 0.25
## # ... with 1,990 more rows
```

- Note that `rep_sample_n` automatically adds `group_by(replicate)` in preparation for `summarize`.

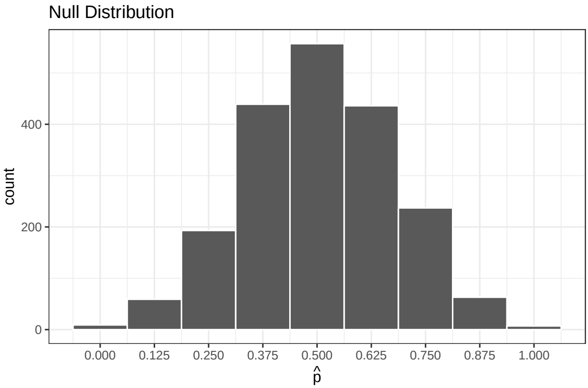
Visualizing the Null Distribution

- We can use a histogram to visualize the Null Distribution of the sample proportion $\hat{p}$

Visualizing the Null Distribution

- We can use a histogram to visualize the Null Distribution of the sample proportion $\hat{p}$

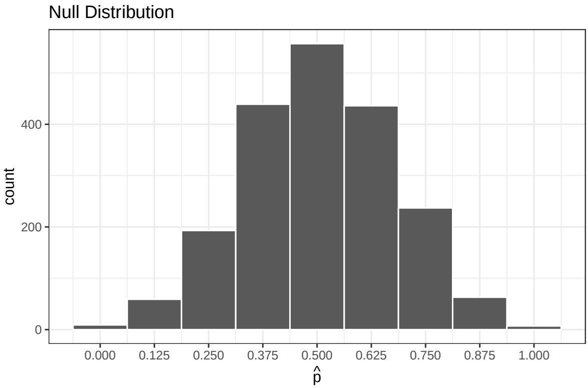
```
null_stats %>% ggplot(aes(x = p_hat))+geom_histogram(bins = 9, color = "white")
```



Visualizing the Null Distribution

- We can use a histogram to visualize the Null Distribution of the sample proportion $\hat{p}$

```
null_stats %>% ggplot(aes(x = p_hat))+geom_histogram(bins = 9, color = "white")
```

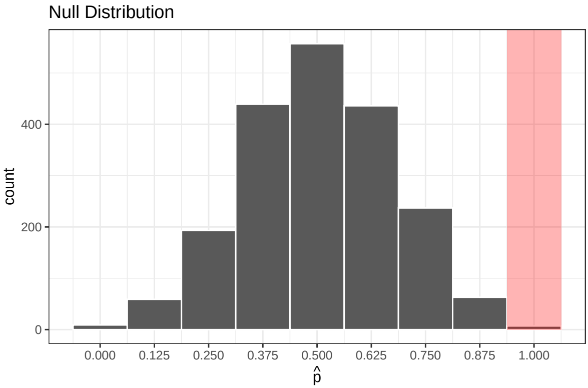


- How often would we have observed $\hat{p} = 1.0$?

Visualizing the Null Distribution

- We can use a histogram to visualize the Null Distribution of the sample proportion $\hat{p}$

```
null_stats %>% ggplot(aes(x = p_hat))+geom_histogram(bins = 9, color = "white")
```



- How often would we have observed $\hat{p} = 1.0$?

Section 3

Strength of Evidence

P-Values

- 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.

P-Values

- 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**

P-Values

- 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 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.

P-Values

- 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 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 8 heads in 8 flips} = 0.5^8 = 0.0039$$

P-Values

- 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 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 8 heads in 8 flips} = 0.5^8 = 0.0039$$

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

P-Values

- 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 8 times and obtained heads 100% of the time. The test statistic is $\hat{p} = 1.0$.
  - The p-value for this test statistic is

$$\text{Probability of at least 8 heads in 8 flips} = 0.5^8 = 0.0039$$

- The p-value quantifies the strength of evidence against the Null Hypothesis. Smaller p-values represent stronger evidence to reject H_0.
  - P-values very close to 0 represent statistics that were very unlikely to arise by chance, if the null hypothesis were true.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% filter(p_hat >=1.0) %>%
  summarize(n = n()) %>%
  mutate(proportion = n/2000)
```

```
## # A tibble: 1 x 2
##       n proportion
##   <int>      <dbl>
## 1     7     0.0035
```

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
    - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% filter(p_hat >=1.0) %>%
  summarize(n = n()) %>%
  mutate(proportion = n/2000)
```

```
## # A tibble: 1 x 2
##       n proportion
##   <int>      <dbl>
## 1     7     0.0035
```

- Method 2: We use theory-based tools to create the theoretical null distribution.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% filter(p_hat >=1.0) %>%
  summarize(n = n()) %>%
  mutate(proportion = n/2000)
```

```
## # A tibble: 1 x 2
##       n proportion
##   <int>      <dbl>
## 1     7     0.0035
```

- Method 2: We use theory-based tools to create the theoretical null distribution.
  - Then use the model to calculate the theoretical probability of observing a sample statistic as extreme as the test statistic.

Calculating P-Values

- Method 1: We can approximate the null distribution using simulation, bootstrapping, and randomization.
  - Then calculate the proportion of simulated statistics as extreme as the test statistic.

```
null_stats %>% filter(p_hat >=1.0) %>%
  summarize(n = n()) %>%
  mutate(proportion = n/2000)
```

```
## # A tibble: 1 x 2
##       n proportion
##   <int>      <dbl>
## 1     7     0.0035
```

- Method 2: We use theory-based tools to create the theoretical null distribution.
  - Then use the model to calculate the theoretical probability of observing a sample statistic as extreme as the test statistic.
  - Assuming that coin flips heads with probability 0.5 and that each flip is independent of the others, then the probability of 8 consecutive heads is

```
0.5^8
```

```
## [1] 0.00390625
```

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?

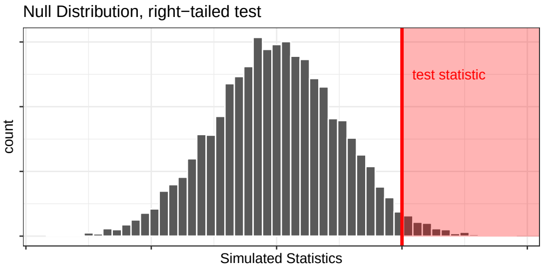
P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

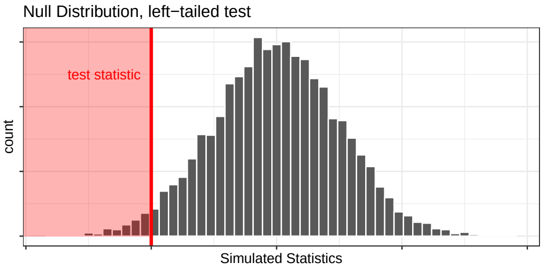
1. If H_a is of the form $\text{parameter} > \text{null value}$, then the p-value is the proportion of simulated statistics greater than or equal to the test statistic (i.e. the right tail)



P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

2. If H_a is of the form $\text{parameter} < \text{null value}$, then the p-value is the proportion of simulated statistics less than or equal to the test statistic (i.e. the left tail)



P-Values and the Alternative Hypothesis

- Does the specific alternative hypothesis play any role in making the null distribution?
  - No. The null distribution just depends on the null hypothesis. It describes the distribution of the statistic if the null hypothesis were true.
- Does the specific alternative hypothesis play any role in calculating the p-value?
  - Yes! The **direction** of the alternative hypotheses determines which "tail(s)" of the null distribution correspond to *extreme* values.

3. If H_a is of the form $\text{parameter} \neq \text{null value}$, then the p-value is twice the proportion of simulated statistics more extreme than the test statistic (i.e. both tails)

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

A Two-Tailed Example

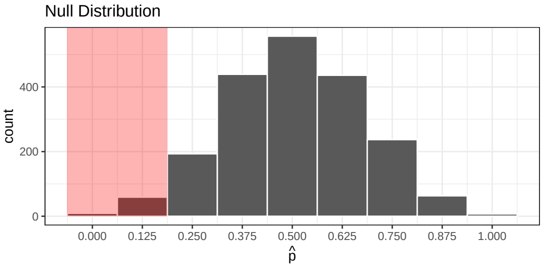
- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:
$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$
- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.
- Using the previous null-distribution, we shade values that are as extreme as our statistic:

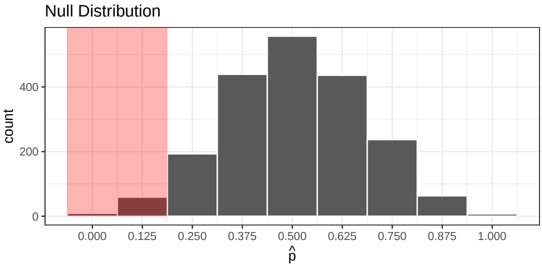


A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.
- Using the previous null-distribution, we shade values that are as extreme as our statistic:



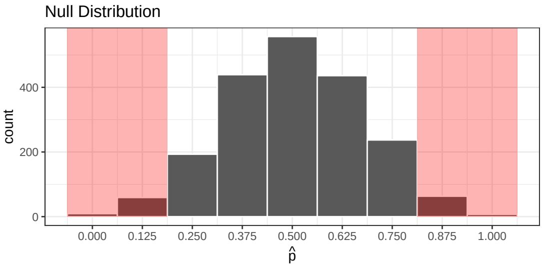
- We find the proportion of simulated statistics in the left tail is 0.034

A Two-Tailed Example

- Suppose we want to determine whether a coin is fair, but don't have any prior expectation that it is biased towards heads or tails.
- Our hypotheses are:

$$H_0 : p = 0.5 \qquad H_a : p \neq 0.5$$

- We flip the coin 8 times and obtain 1 heads, for a proportion $\hat{p} = 0.125$.
- Using the previous null-distribution, we shade values that are as extreme as our statistic:



- We double this to include the right-tail as well, and get a p-value of 0.068.

Section 4

Hypothesis Testing Example

Pregnancy Duration

Prof. Wells is expecting a baby in the next few weeks! (Due March 31st).

- How is this due date calculated?

Pregnancy Duration

Prof. Wells is expecting a baby in the next few weeks! (Due March 31st).

- How is this due date calculated?
- Historical medical records from the 19th and 20th century suggest that the average gestational length of a pregnancy (time from last menstrual period to live birth) is 40 weeks.
- A baby's predicted due date is defined as: 40 weeks from date of last period.

Pregnancy Duration

Prof. Wells is expecting a baby in the next few weeks! (Due March 31st).

- How is this due date calculated?
- Historical medical records from the 19th and 20th century suggest that the average gestational length of a pregnancy (time from last menstrual period to live birth) is 40 weeks.
- A baby's predicted due date is defined as: 40 weeks from date of last period.
- However, some contemporary research suggests that average gestational length in the US may differ from the conventional standard

Pregnancy Duration

Prof. Wells is expecting a baby in the next few weeks! (Due March 31st).

- How is this due date calculated?
- Historical medical records from the 19th and 20th century suggest that the average gestational length of a pregnancy (time from last menstrual period to live birth) is 40 weeks.
- A baby's predicted due date is defined as: 40 weeks from date of last period.
- However, some contemporary research suggests that average gestational length in the US may differ from the conventional standard
- We have data for 200 live births in the US in 2014, randomly sampled from a data set on all recorded live births in the US in 2014.

Pregnancy Duration

Prof. Wells is expecting a baby in the next few weeks! (Due March 31st).

- How is this due date calculated?
- Historical medical records from the 19th and 20th century suggest that the average gestational length of a pregnancy (time from last menstrual period to live birth) is 40 weeks.
- A baby's predicted due date is defined as: 40 weeks from date of last period.
- However, some contemporary research suggests that average gestational length in the US may differ from the conventional standard
- We have data for 200 live births in the US in 2014, randomly sampled from a data set on all recorded live births in the US in 2014.
- **Goal**: Use this data to assess the claim that the average length of pregnancy in the US is 40 weeks.

Understanding the Context

What sources of randomness are present in this investigation?

Understanding the Context

What sources of randomness are present in this investigation?

- **Randomness in the Population**:
    - Errors in gestational age estimation
    - "Natural" variation in pace of fetal maturation, as well as pace of natural delivery
    - Presence of other health factors that impact pregnancy length
    - Together, these three sources explain why pregnancy length can vary in the population

Understanding the Context

What sources of randomness are present in this investigation?

- **Randomness in the Population**:
  - Errors in gestational age estimation
  - "Natural" variation in pace of fetal maturation, as well as pace of natural delivery
  - Presence of other health factors that impact pregnancy length
  - Together, these three sources explain why pregnancy length can vary in the population
- **Randomness in the Sample**:
  - Variability due to random sampling
  - This is the only source of randomness in the sample that does not also exist in the population.

Understanding the Context

What sources of randomness are present in this investigation?

- **Randomness in the Population**:
    - Errors in gestational age estimation
    - "Natural" variation in pace of fetal maturation, as well as pace of natural delivery
    - Presence of other health factors that impact pregnancy length
    - Together, these three sources explain why pregnancy length can vary in the population
- **Randomness in the Sample**:
    - Variability due to random sampling
    - This is the only source of randomness in the sample that does not also exist in the population.
- Is this an observational study or a random experiment?

Understanding the Context

What sources of randomness are present in this investigation?

- **Randomness in the Population**:
    - Errors in gestational age estimation
    - "Natural" variation in pace of fetal maturation, as well as pace of natural delivery
    - Presence of other health factors that impact pregnancy length
    - Together, these three sources explain why pregnancy length can vary in the population
- **Randomness in the Sample**:
    - Variability due to random sampling
    - This is the only source of randomness in the sample that does not also exist in the population.
- Is this an observational study or a random experiment?
    - Observational study; we are not randomly assigning individuals to treatment and control groups.

Clarify Research Question

In groups of 2 or 3, answer the following questions about this investigation:

1. What is our research question?
2. What is the population of interest?
3. What parameter do we wish to estimate?
4. What is the sample?
5. What statistic could we calculate in the sample to estimate the parameter?
6. What are the formal statements of our null and alternative hypothesis (i.e. statements in symbols using the parameter values)?

Clarify Research Question

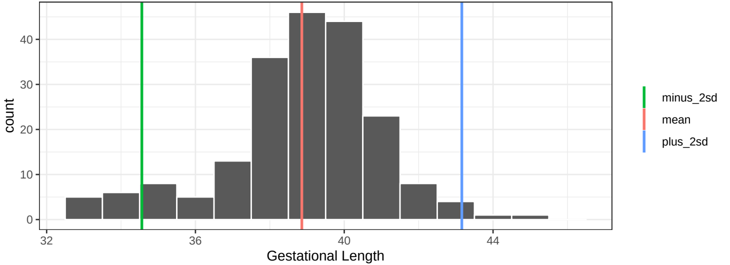
Answers:

1. What is our research question?
   Is the average gestational length in the US 40 weeks?
2. What is the population of interest?
   Contemporary births in the US
3. What parameter do we wish to estimate?
   The average gestational length μ
4. What is the sample?
   200 live births in the US from 2014
5. What statistic could we calculate in the sample to estimate the parameter?
   The average gestational length in these 200 live births $\bar{x}$
6. What are the formal statements of our null and alternative hypothesis?

$$H_0 : \mu = 40 \qquad H_a : \mu \neq 40$$

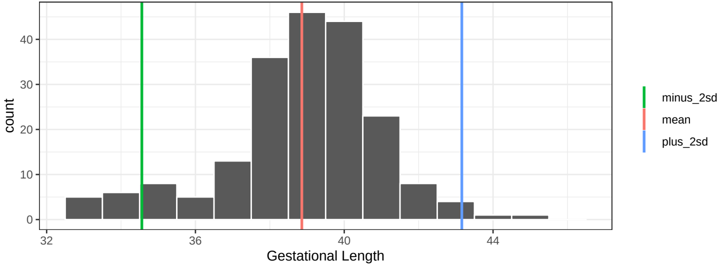
Investigate Sample

The graph below shows the distribution of gestational lengths among the 1000 births in the sample.



Investigate Sample

The graph below shows the distribution of gestational lengths among the 1000 births in the sample.



We can also calculate relevant summary statistics:

```
## # A tibble: 1 x 4
##   mean_length sd_length minus_2sd plus_2sd
##         <dbl>     <dbl>     <dbl>    <dbl>
## 1        38.9      2.15      34.6     43.2
```

The Null Distribution

If the true average gestational length were 40 weeks, how likely is it that a random sample of 1000 births would have mean of 38.9, or more extreme?

The Null Distribution

If the true average gestational length were 40 weeks, how likely is it that a random sample of 1000 births would have mean of 38.9, or more extreme?

- Based on the sample's distribution, we see that an individual difference in gestational length of 1 week or more is relatively common.

The Null Distribution

If the true average gestational length were 40 weeks, how likely is it that a random sample of 1000 births would have mean of 38.9, or more extreme?

- Based on the sample's distribution, we see that an individual difference in gestational length of 1 week or more is relatively common.
    - But is a difference of 1 week plausible in the **mean** of a sample of size 200?

The Null Distribution

If the true average gestational length were 40 weeks, how likely is it that a random sample of 1000 births would have mean of 38.9, or more extreme?

- Based on the sample's distribution, we see that an individual difference in gestational length of 1 week or more is relatively common.
    - But is a difference of 1 week plausible in the **mean** of a sample of size 200?
- To answer this question, we need to consider the distribution of sample means, if the null hypothesis were true.

The Null Distribution

If the true average gestational length were 40 weeks, how likely is it that a random sample of 1000 births would have mean of 38.9, or more extreme?

- Based on the sample's distribution, we see that an individual difference in gestational length of 1 week or more is relatively common.
  - But is a difference of 1 week plausible in the **mean** of a sample of size 200?
- To answer this question, we need to consider the distribution of sample means, if the null hypothesis were true.
- Previously, we were able to create the null distribution by simulating a large number of coin flips. But that won't work here (why?)

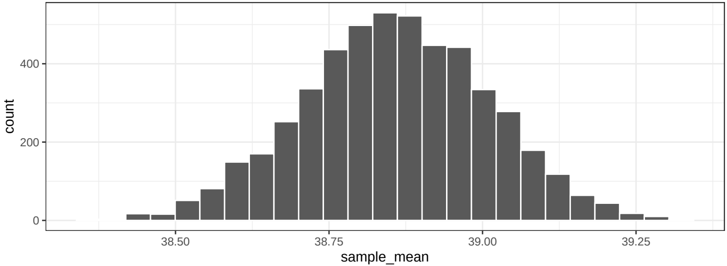
The Null Distribution

If the true average gestational length were 40 weeks, how likely is it that a random sample of 1000 births would have mean of 38.9, or more extreme?

- Based on the sample's distribution, we see that an individual difference in gestational length of 1 week or more is relatively common.
  - But is a difference of 1 week plausible in the **mean** of a sample of size 200?
- To answer this question, we need to consider the distribution of sample means, if the null hypothesis were true.
- Previously, we were able to create the null distribution by simulating a large number of coin flips. But that won't work here (why?)
- Are there any other ways to simulate new samples from a population?

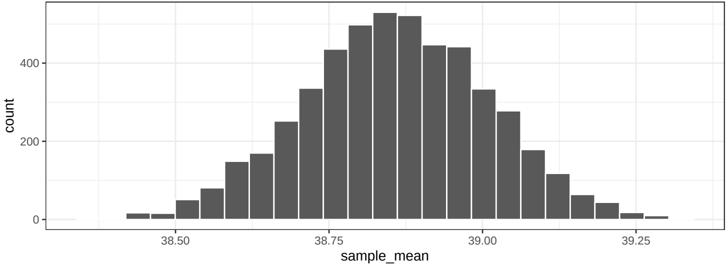
Bootstrapping the Null Distribution

- We can bootstrap from the original sample to create the bootstrap distribution:



Bootstrapping the Null Distribution

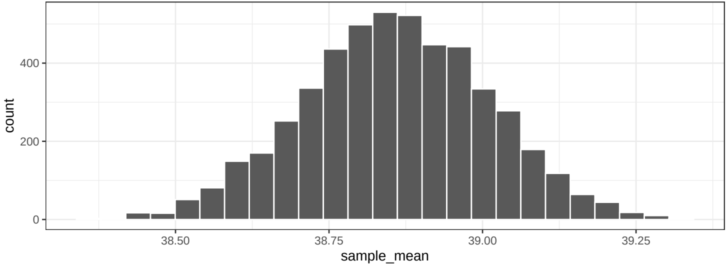
- We can bootstrap from the original sample to create the bootstrap distribution:



- The bootstrap distribution has the same shape and spread as the sampling distribution for the statistic.

Bootstrapping the Null Distribution

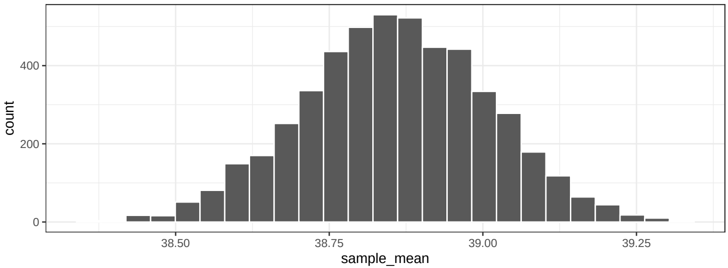
- We can bootstrap from the original sample to create the bootstrap distribution:



- The bootstrap distribution has the same shape and spread as the sampling distribution for the statistic.
- But there's one problem!

Bootstrapping the Null Distribution

- We can bootstrap from the original sample to create the bootstrap distribution:



- The bootstrap distribution has the same shape and spread as the sampling distribution for the statistic.
- But there's one problem!
  - The bootstrap distribution is centered at the *sample mean* ($\bar{x} = 38.9$), rather than the null value ($\mu = 40$)

Bootstrapping the Null Distribution

- But, we can compute the difference between the sample mean and the null value, and then add this amount to every statistic
  - This has the affect of centering the bootstrap distribution on the null value.

Bootstrapping the Null Distribution

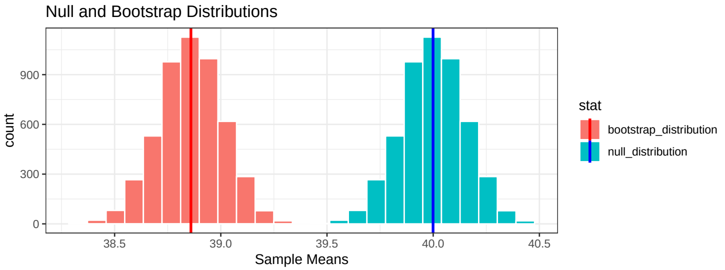
- But, we can compute the difference between the sample mean and the null value, and then add this amount to every statistic
  - This has the affect of centering the bootstrap distribution on the null value.

```
##   sample_mean null_value difference
## 1       38.86         40       1.14
```

Bootstrapping the Null Distribution

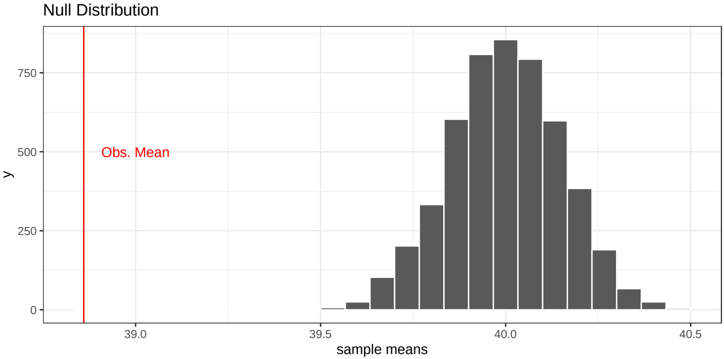
- But, we can compute the difference between the sample mean and the null value, and then add this amount to every statistic
  - This has the affect of centering the bootstrap distribution on the null value.

```
##   sample_mean null_value difference
## 1       38.86         40       1.14
```



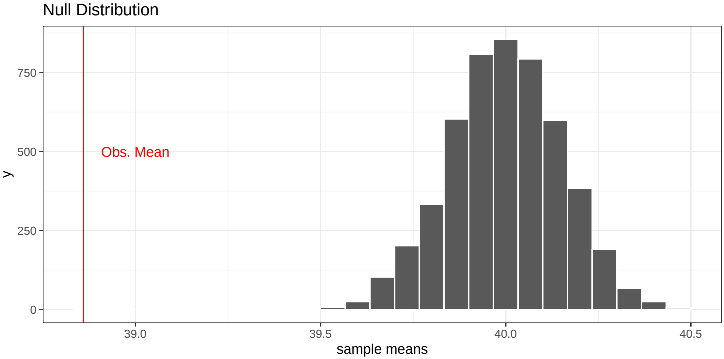
Calculate P-Value

- With the null distribution, we can now calculate the P-value:



Calculate P-Value

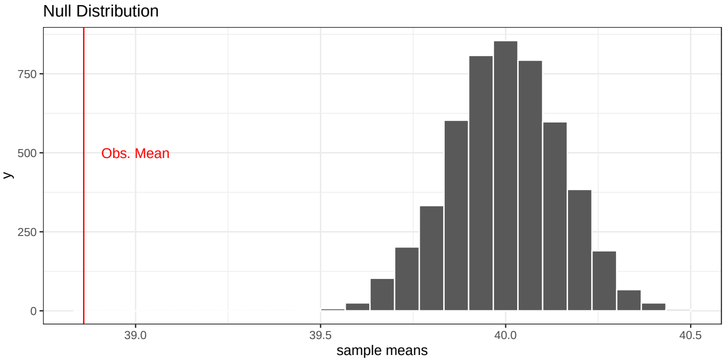
- With the null distribution, we can now calculate the P-value:



- Based on the simulated null distribution, none of the 5000 sample means were as extreme as the mean we observed in the original sample.

Calculate P-Value

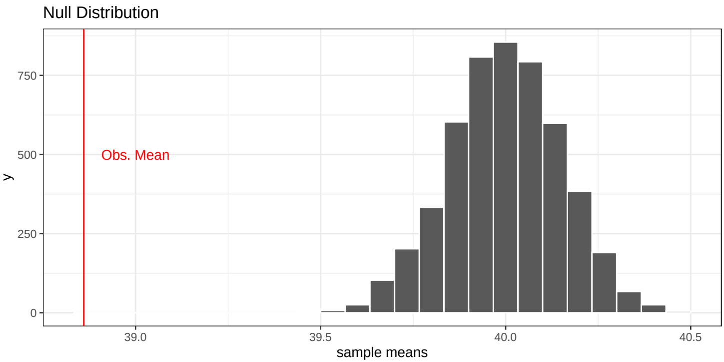
- With the null distribution, we can now calculate the P-value:



- Based on the simulated null distribution, none of the 5000 sample means were as extreme as the mean we observed in the original sample.
- This gives us a p-value of approximately 0.

Calculate P-Value

- With the null distribution, we can now calculate the P-value:



- Based on the simulated null distribution, none of the 5000 sample means were as extreme as the mean we observed in the original sample.
- This gives us a p-value of approximately 0.
- Thus, this sample provides relatively strong evidence that the true mean gestation