COMS BC1016

Introduction to Computational Thinking and Data Science

Lecture 15: Confidence Intervals

Nov 10, 2025

Reminders

- Final Project Proposal Due **Friday**
  - Worth **10%** of the final project grade
  - Template is on the 1017 Courseworks
- HW 6 due next week Monday, Nov 17
- HW 7: Skip Question 4 about the survey
- Extra Credit (HW 5 Question 3) Due Monday, Nov 17
  - Completely optional, no late submissions

Final Project Groups

- Please check the info on the 1017 Courseworks is right
- If your final project group can attend one of the Thursday labs, consider switching
    - Groups are overwhelming attending Wednesday labs Dec 3/4
    - Email me if you want to switch

COMSBC1017_001_2025_3 - Introduction to Computational Thinking a > People > Groups

Groups (34)

Section 1 - Anna Pan and Anna Freitag	2 students
Section 1 - Audrey Riskey, Isabella Diaz, and Caitlyn O'Shea	3 students
Section 1 - Caroline Villamin and Jinghan Wang	2 students
Section 1 - Celia Costa and Bea Creighton	2 students
Section 1 - Jahnavi Bolleddula and Camila Pierre	2 students
Section 1 - Jenna Silvera and Diego Putinati	2 students
Section 1 - Juliana Chavis and Maheen Asaf	2 students
Section 1 - Jullye Campos and Regina Rodriguez	2 students

Lecture Outline

- AB Testing Review
- Confidence intervals
  - Percentiles
  - Estimations
  - Bootstrapping
- Summary

Review: A/B Testing

Review: P-Value

P-value: Observed significance level

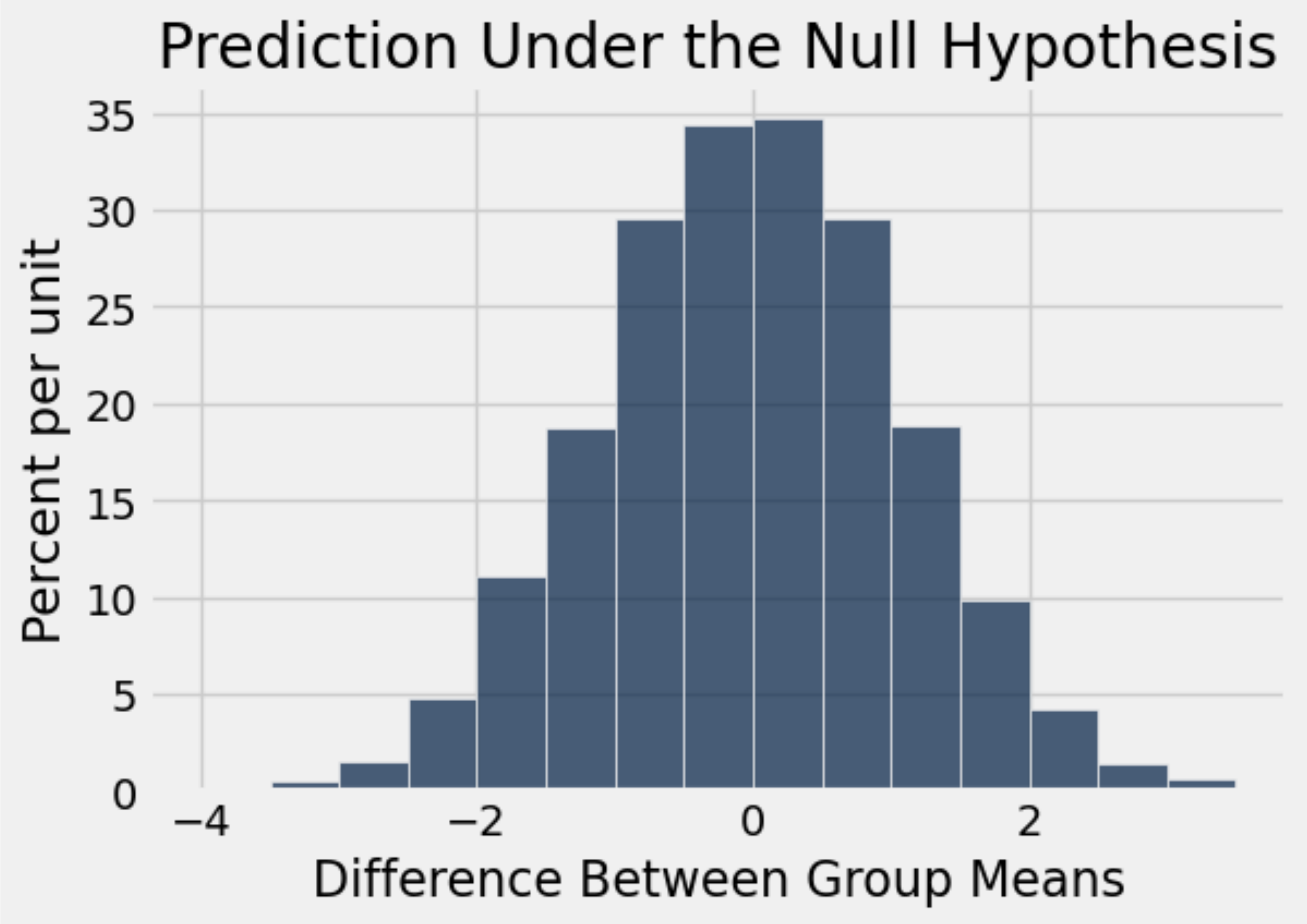
The **P-value** is the chance

under the null hypothesis

that the test statistic

is equal to the value that was observed in the data

or is even further in the direction of the alternative



A/B Testing

- Testing whether Group A and Group B have the same underlying distribution or not
  - <u>Null Hypothesis</u>: The distributions of [test statistic] from both groups are the same
    - Any differences we observe are due to chance
  - <u>Alternative Hypothesis</u>: The distributions are different
- If the distributions look different, it supports the alternative hypothesis

A/B Testing Process

- Under the null, if the Group A and Group B are the same, it doesn't matter what label (A or B) we assign each item
  - We expect the difference between the groups to be 0
- **Permutation Test**:
  - Shuffle all group labels using `shuttle(with_replacement = False)`
  - Compute the difference between the average of the two shuffled groups
  - Compare the difference to our observed difference

A/B Testing Process

1. Write a function that calculates the test static for one simulation

```
def one_simulated_difference(table, label, group_label):
    """Takes: name of table, column label of numerical variable,
    column label of group-label variable
    Returns: Difference of means of the two groups after shuffling labels"""

    # array of shuffled labels
    shuffled_labels = table.sample(with_replacement = False).column(group_label)

    # table of numerical variable and shuffled labels
    shuffled_table = table.with_column('Shuffled Label', shuffled_labels)

    return difference_of_means(shuffled_table, label, 'Shuffled Label')
```

2. Repeat that process in a for loop many times

```
differences = make_array()

for i in np.arange(2500):
    new_difference = one_simulated_difference(births, 'Birth Weight', 'Maternal Smoker')
    differences = np.append(differences, new_difference)
```

3. Plot the distribution and compare to our observed value

```
diff_tbl = Table().with_column('Difference Between Group Means', differences)
diff_tbl.hist()
```

Confidence Intervals

[Textbook Chapter on Confidence Intervals]
[Textbook Chapter on Using Confidence Intervals]

Confidence Interval

- Interval of estimates of a parameter
  - Statistic tells us the estimate of a single parameter, confidence interval tells us a range of values of the estimated parameter
    - "How often does our estimate capture the parameter"
- Based on the notion of percents, so helpful to understand what we mean by a percentile

Percentile

- The p-th percentile of a collection is the smallest value in the collection that is *at least* as large as $p\%$ of all the values

- Computing the p-th percentile of a list of n values:

  1. Sort the list

  2. Compute how many elements k the $p\%$ refers to

  $$k = \left\lceil \frac{p}{100} \times n \right\rceil$$

  3. Return the k-th smallest element (counting from 1)

Percentile

Example: What is the 80th percentile for [12, 17, 6, 9, 7]?

1. Sort the list

2. Compute how many elements k the $p\%$ refers to

3. Return the k-th smallest element (counting from 1)

Percentile

Example: What is the 80th percentile for [12, 17, 6, 9, 7]?

1. Sort the list: [6, 7, 9, 12, 17]

2. Compute how many elements k the $p\%$ refers to

3. Return the k-th smallest element (counting from 1)

Percentile

Example: What is the 80th percentile for [12, 17, 6, 9, 7]?

1. Sort the list: [6, 7, 9, 12, 17]

2. Compute how many elements k the p% refers to

$$k = \left\lceil \frac{p}{100} \times n \right\rceil = \left\lceil \frac{80}{100} \times 5 \right\rceil = \left\lceil \frac{4}{5} \times 5 \right\rceil = 4$$

3. Return the k-th smallest element (counting from 1)

Percentile

Example: What is the 80th percentile for [12, 17, 6, 9, 7]?

1. Sort the list: [6, 7, 9, 12, 17]

2. Compute how many elements k the p% refers to

$$k = \left\lceil \frac{p}{100} \times n \right\rceil = \left\lceil \frac{80}{100} \times 5 \right\rceil = \left\lceil \frac{4}{5} \times 5 \right\rceil = 4$$

3. Return the k-th smallest element (counting from 1)

   4th smallest element of [6, 7, 9, 12, 17] is 12

`percentile` Function

- `percentile(p, values_array)`
    - Returns the smallest value in `values_array` that is at least as large as `p`% of the elements in the array
    - `p` is always between 0 and 100

Example

Let `s = [1, 3, 5, 7, 9]`

What would `percentile(20, s)` evaluate to?

`percentile(p,array)`
returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Example

`percentile(p,array)` returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Let `s = [1, 3, 5, 7, 9]`

What would `percentile(20, s)` evaluate to?

1. Sort the list

2. Compute how many elements k the $p\%$ refers to

$$k = \left\lceil \frac{p}{100} \times n \right\rceil$$

3. Return the k-th smallest element (counting from 1)

Example

`percentile(p,array)` returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Let `s = [1, 3, 5, 7, 9]`

What would `percentile(20, s)` evaluate to?

1. Sort the list

2. Compute how many elements k the $p\%$ refers to

$$k = \left\lceil \frac{p}{100} \times n \right\rceil = \left\lceil \frac{20}{100} \times 5 \right\rceil = 1$$

3. Return the k-th smallest element (counting from 1)

More Examples

`percentile(p,array)` returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Let `s = [1, 3, 5, 7, 9]`

Which of the following evaluate to `True`?

1. `percentile(10, s) == 0`

2. `percentile(39, s) == percentile(40, s)`

3. `percentile(40, s) == percentile(41, s)`

More Examples

`percentile(p,array)` returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Let `s = [1, 3, 5, 7, 9]`

Which of the following evaluate to `True`?

1. `percentile(10, s) == 0`

   False! Element is an element from the array

2. `percentile(39, s) == percentile(40, s)`

3. `percentile(40, s) == percentile(41, s)`

More Examples

`percentile(p,array)` returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Let `s = [1, 3, 5, 7, 9]`

Which of the following evaluate to `True`?

1. `percentile(10, s) == 0`

   False! Element is an element from the array

2. `percentile(39, s) == percentile(40, s)`

   True! Both evaluate to 3

3. `percentile(40, s) == percentile(41, s)`

More Examples

`percentile(p,array)` returns the smallest value in `array` that is at least as large as `p`% of elements in the array

Let `s = [1, 3, 5, 7, 9]`

Which of the following evaluate to `True`?

1. `percentile(10, s) == 0`

   False! Element is an element from the array

2. `percentile(39, s) == percentile(40, s)`

   True! Both evaluate to 3

3. `percentile(40, s) == percentile(41, s)`

   False. First value is 3 but the second one is 5

Estimation

Estimation

- We use estimation to figure out the value of an unknown parameter

If entire population is known

Calculate **parameter** directly

If we only have access to a random sample

We calculate a **statistic** as an **estimate** of the **parameter**

Notebook Demo: SF Gov’t Salaries

Quantifying Uncertainty

- Our estimate depends on the sample we collected. How different would the estimate be if the sample were different?
  - Can we determine how accurate our estimate is?
  - In theory, we could collect a different sample and check how similar the statistic we calculated is
- What if we can't go back and collect more samples?

Bootstrap Method

Bootstrap

- A technique for simulating repeated random sampling

- The term “bootstrap” means *to improve or lift yourself using your own resources without external help*

    - From the phrase “to pull oneself up by one’s bootstraps”

- **Bootstrap method** tries to **learn about the population** using **only the sample itself** without any outside (theoretical) distribution assumptions

The Bootstrap Method

- Suppose we have a large random sample from the population
    - By the Law of Averages, it probably resembles the population from which it’s drawn
    - We can treat it like a miniature version of the population
- We can replicate sampling from the population by *sampling from the sample*
    - To resample, draw at random with replacement the same number of times as the original sample size

The Bootstrap Method

- Important to resample with replacement the same number of times as the original sample
  - Suppose we computed the original statistic based on n samples. We need to compare it to another statistic also based on n samples
  - Drawing without replacement gives the same sample back, so you need to sample with replacement

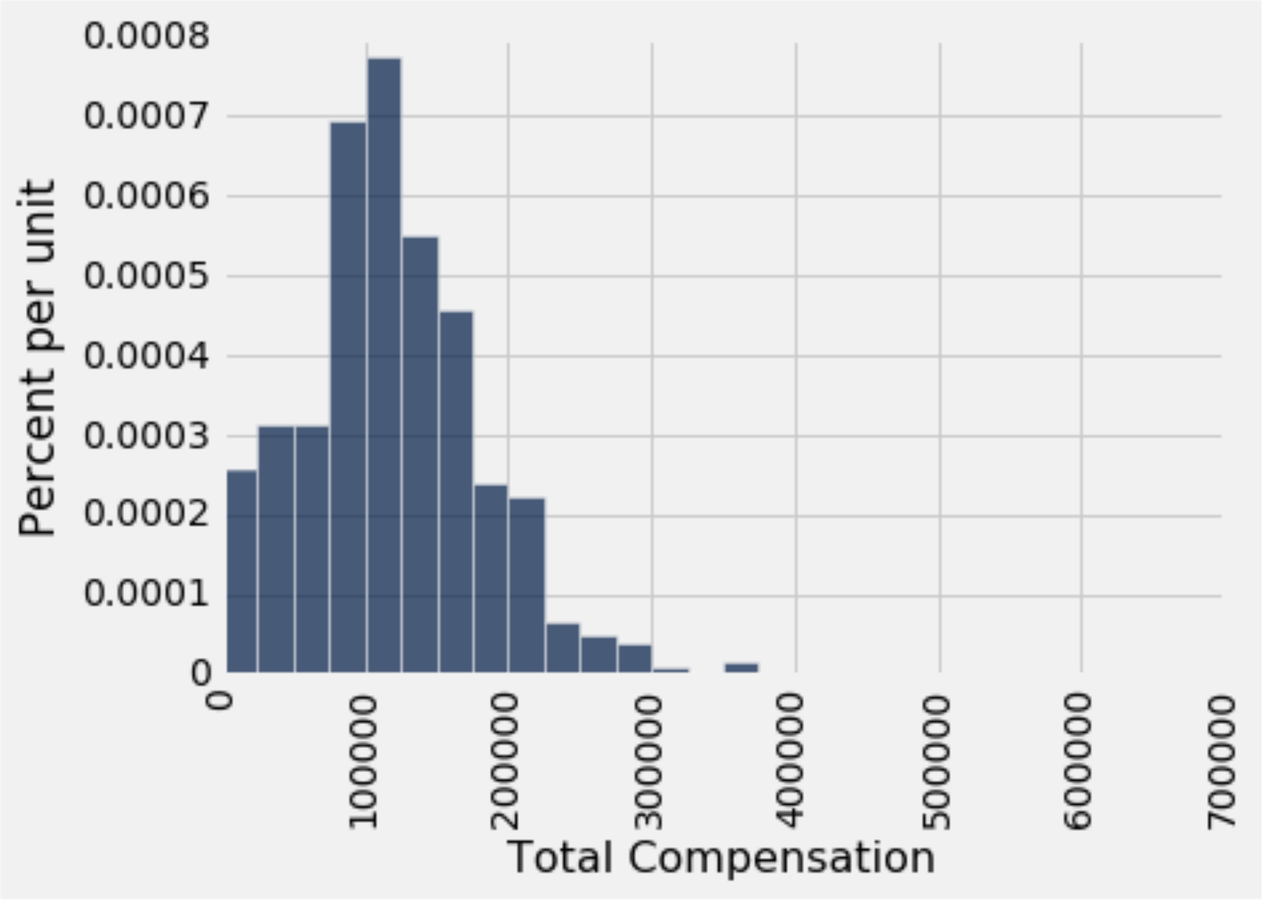
The Bootstrap

Population

?

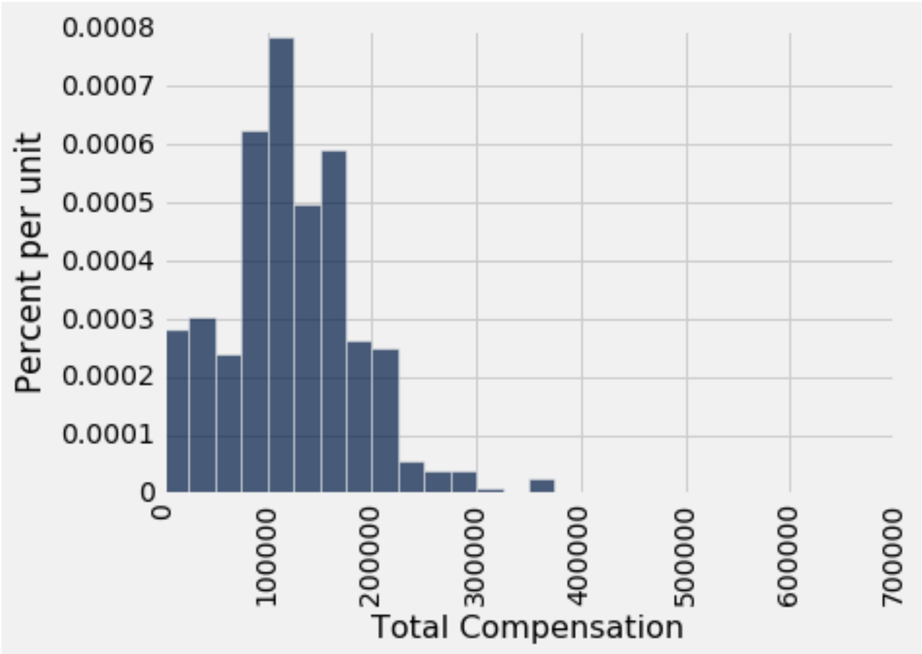
We don’t know the entire population and thus can’t calculate the parameter directly

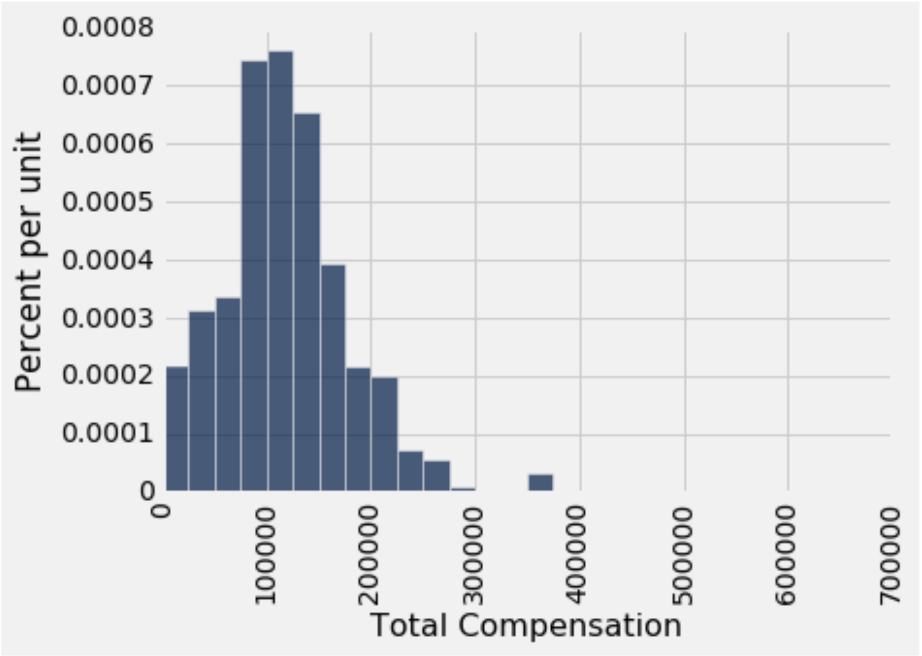
Sample

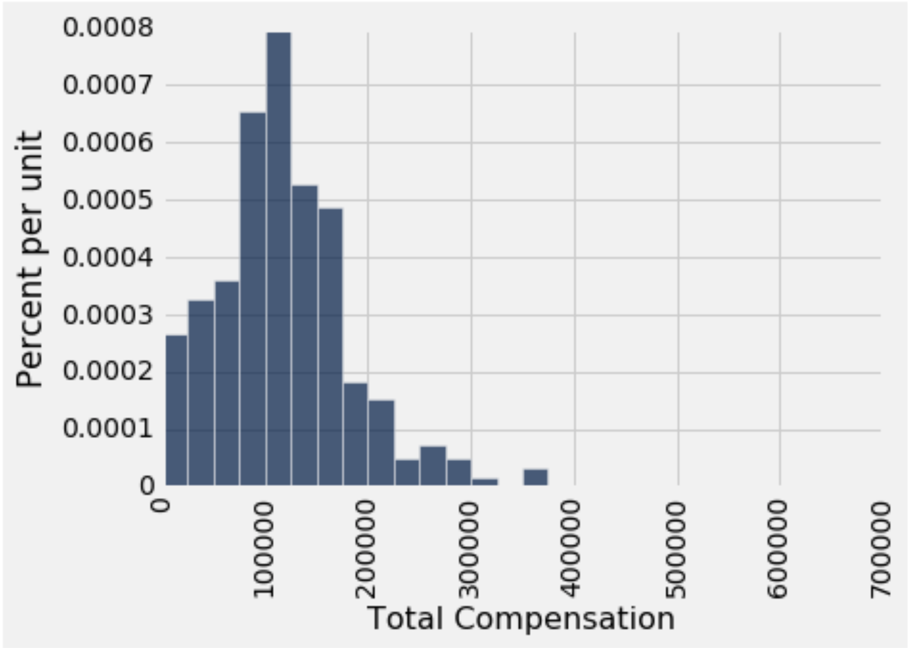


However, we can take a single sample…

Resamples

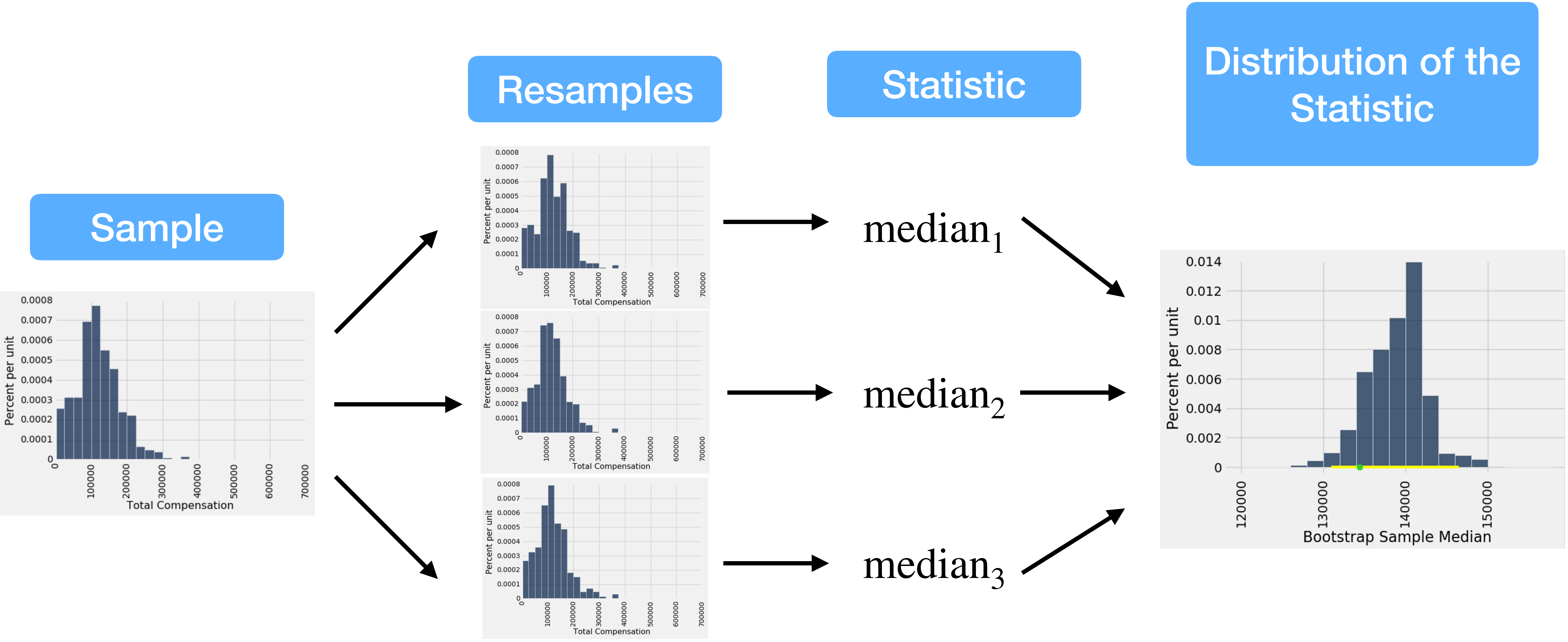






…and generate lots of resamples

Confidence Interval



Confidence Interval

- Interval of **estimates of a parameter**
  - How confident we are that the parameter (the real value calculated from the population) is likely to be within this interval
    - Good if the parameter is within the interval, bad if it's not
- Based on random sampling
  - 95% confidence interval is the middle 95% of the distribution of estimates
    - Can choose any percentage between 0 and 100
    - Higher numbers correspond to a wider interval

Interpreting Confidence Interval

- The confidence interval helps us state how confident we are that our sample statistic estimates (or contains) the true population parameter

- When we do bootstrapping and create, say, a 95% confidence interval for the median, what we are doing is **estimating the range of plausible values for the true population median based on the sample data**

- Confidence interval is a rough estimate of how large a parameter is. Common mistake to use it for other purposes

    - For example, a 95% confidence interval for an average age is **not** saying we think 95% of the population is within that age

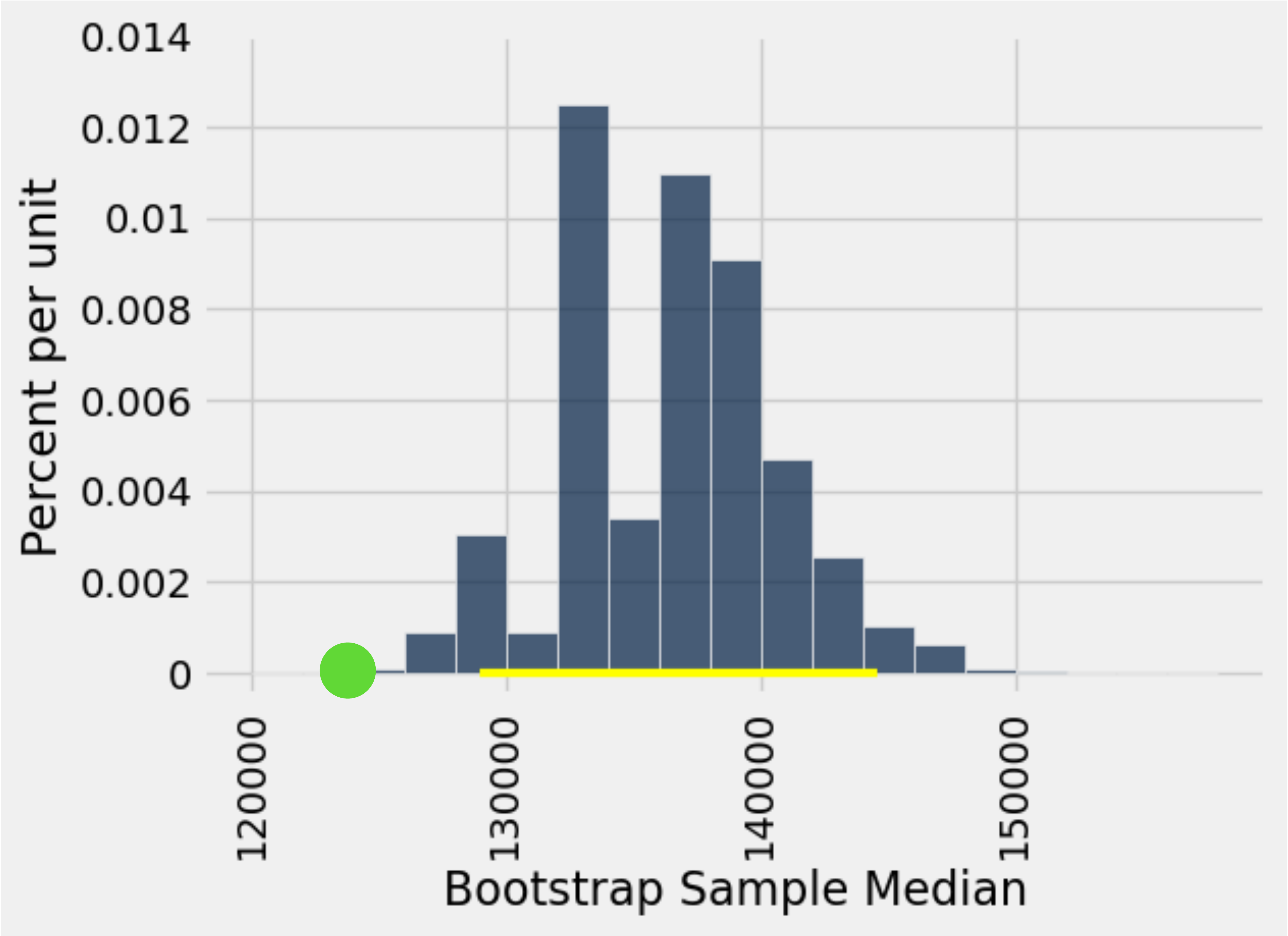
Relationship between Confidence Interval and P-value

- For a p-value cutoff of $p\%$, we reject the null hypothesis if the value is not within $(100 - p)\%$ confidence interval
  - If you use a $p\%$ cutoff for the p-value and the null hypothesis is true, then there is about a $p\%$ chance your test will conclude the alternative is true
    - If the null hypothesis is true, $(100 - p)\%$ chance test agrees with the null
  - $(100 - p)\%$ confidence interval says we're $(100 - p)\%$ certain the parameter is somewhere within the interval
    - If the value is outside of the interval, we reject the null hypothesis

Relationship between Confidence Interval and P-value

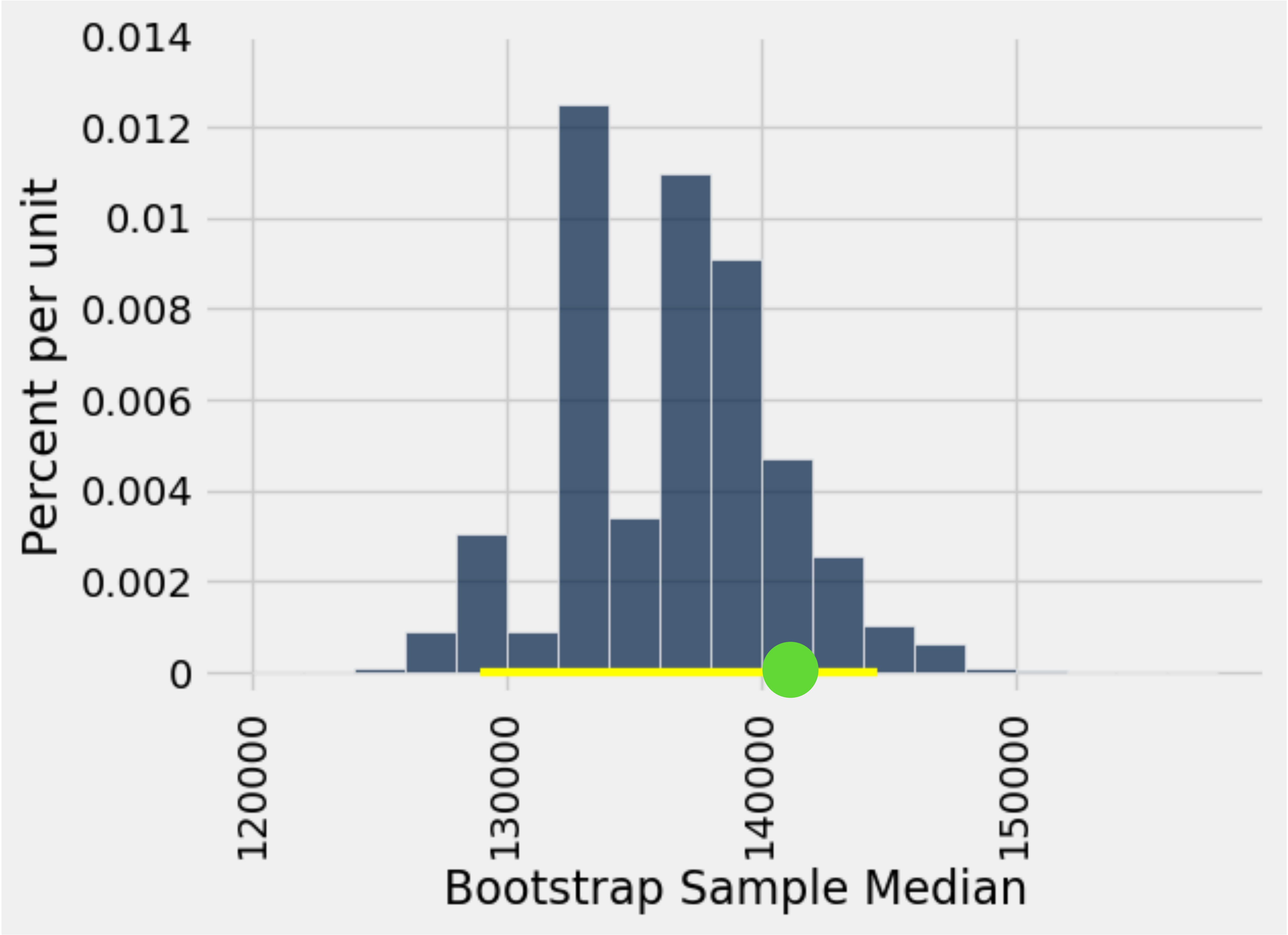
- Example:
  - Null hypothesis: Population average $= x$
  - Alternative Hypothesis: Population average $\neq x$
  - If x is not in our $(100 - p)\%$ interval, then we reject the null

Rejecting the null



Our x is outside the 95% confidence interval

Cannot reject null



Our x is inside the 95% confidence interval

Confidence Interval & P-Value Example

<u>Null Hypothesis</u>: You have a fair coin with 50% probability of getting heads or tails

<u>Alternative</u>: The coin is biased

Your observed value for **% heads** is 65%

Let's say your 95% confidence interval is [45, 60]

Confidence Interval & P-Value Example

Null Hypothesis: You have a fair coin with 50% probability of getting heads or tails

Alternative: The coin is biased

Your observed value for **% heads** is 65%

Let’s say your 95% confidence interval is [45, 60]

Questions:

1. For a 5% p-value cutoff, can we reject the null?

2. For a 10% p-value cutoff, can we reject the null?

Confidence Interval & P-Value Example

Null Hypothesis: You have a fair coin with 50% probability of getting heads or tails

Alternative: The coin is biased

Your observed value for **% heads** is 65%

Let's say your 95% confidence interval is [45, 60]

Questions:

1. For a 5% p-value cutoff, can we reject the null?
   - Yes, 65% is outside our confidence interval
2. For a 10% p-value cutoff, can we reject the null?

Confidence Interval & P-Value Example

<u>Null Hypothesis</u>: You have a fair coin with 50% probability of getting heads or tails

<u>Alternative</u>: The coin is biased

Your observed value for **% heads** is 65%

Let's say your 95% confidence interval is [45, 60]

Questions:

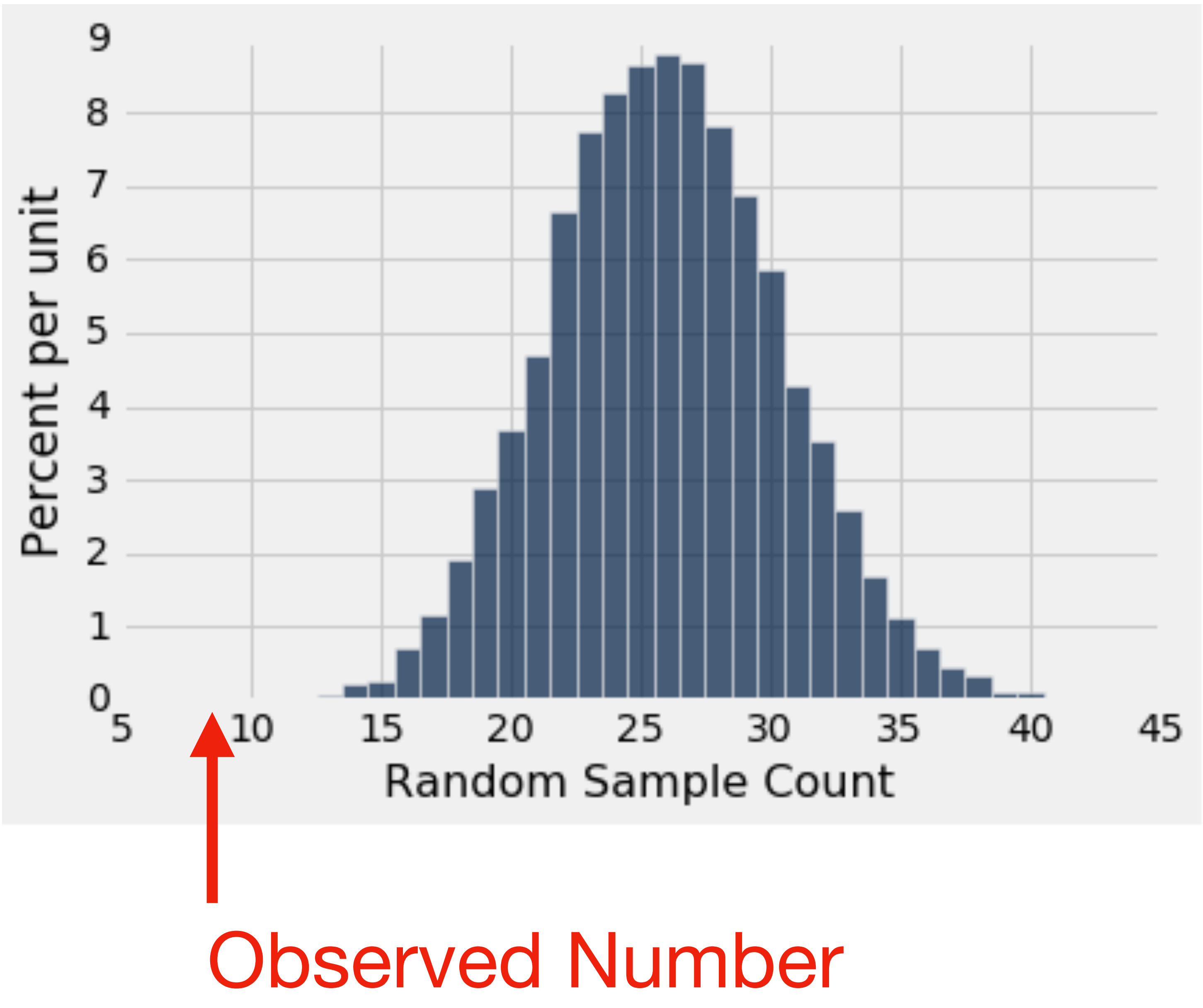
1. For a 5% p-value cutoff, can we reject the null?
   - Yes, 65% is outside our confidence interval
2. For a 10% p-value cutoff, can we reject the null?
   - Yes, we expect the confidence interval to be even narrower, so 65% would still be outside the confidence interval

Summary of Stats so far…

Hypothesis Testing

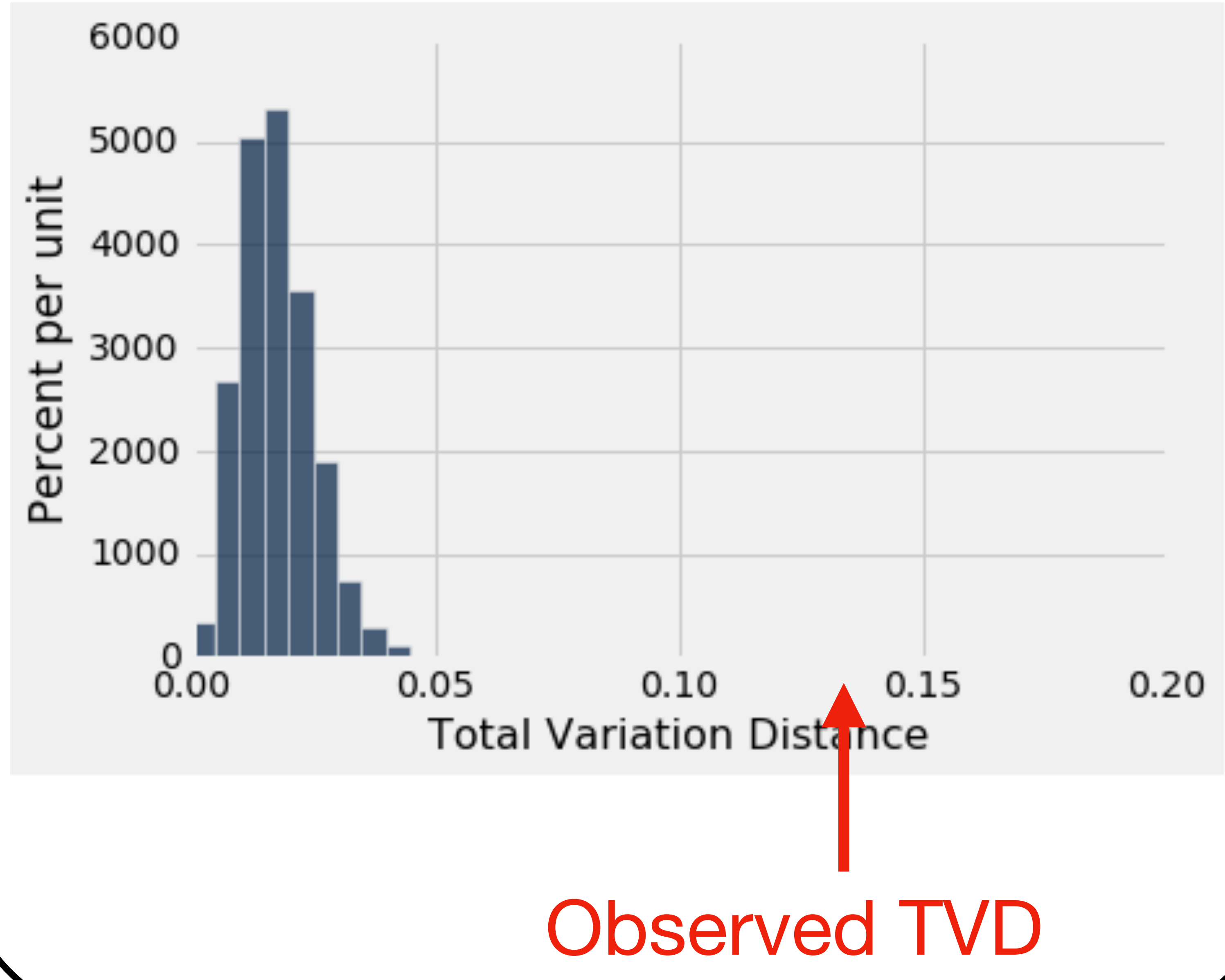
- Modeling expected outcomes under the null and comparing it to our observed outcome

Swain v Alabama



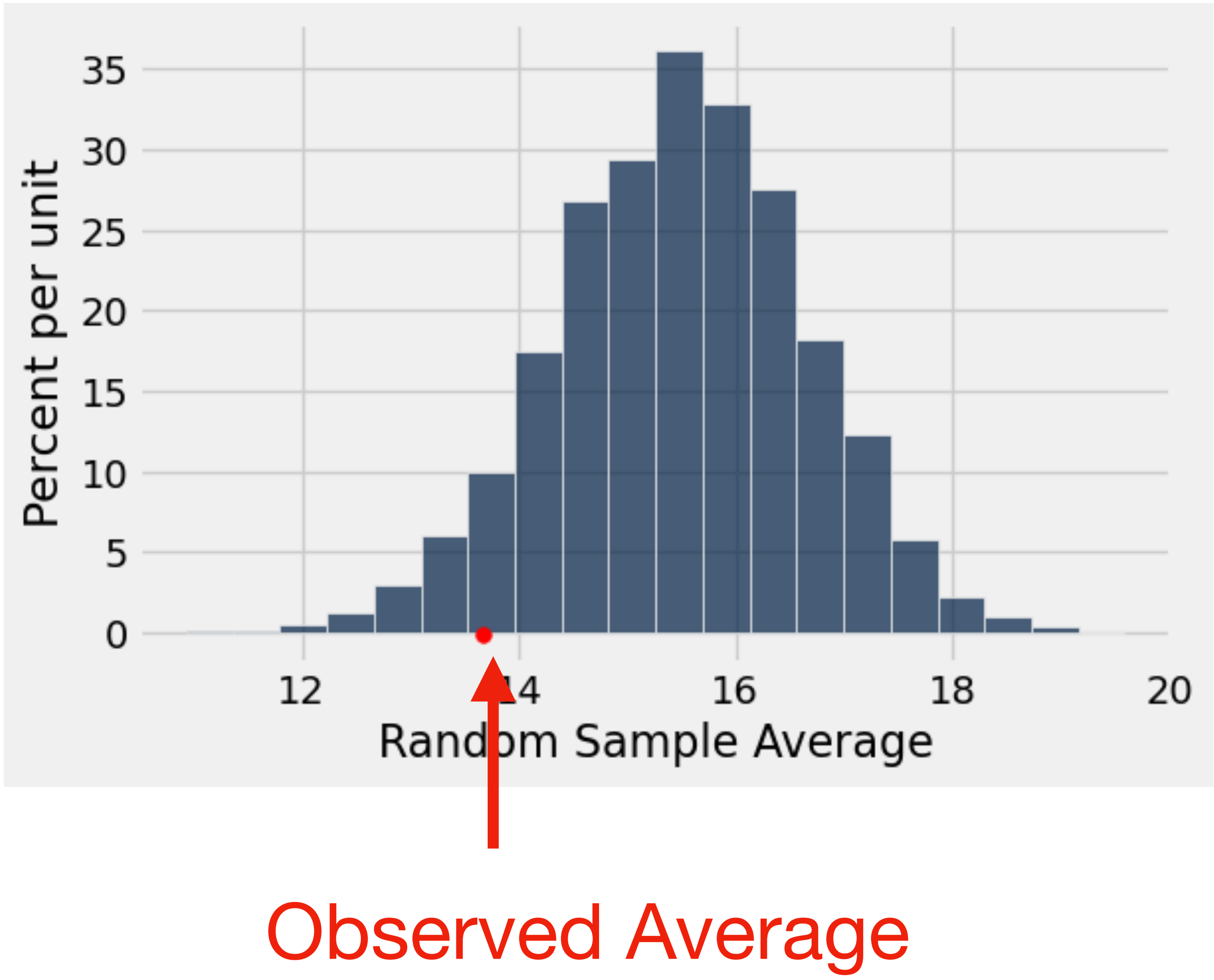
2 Categories

Alameda Jury



3+ Categories

Midterm Exam Scores

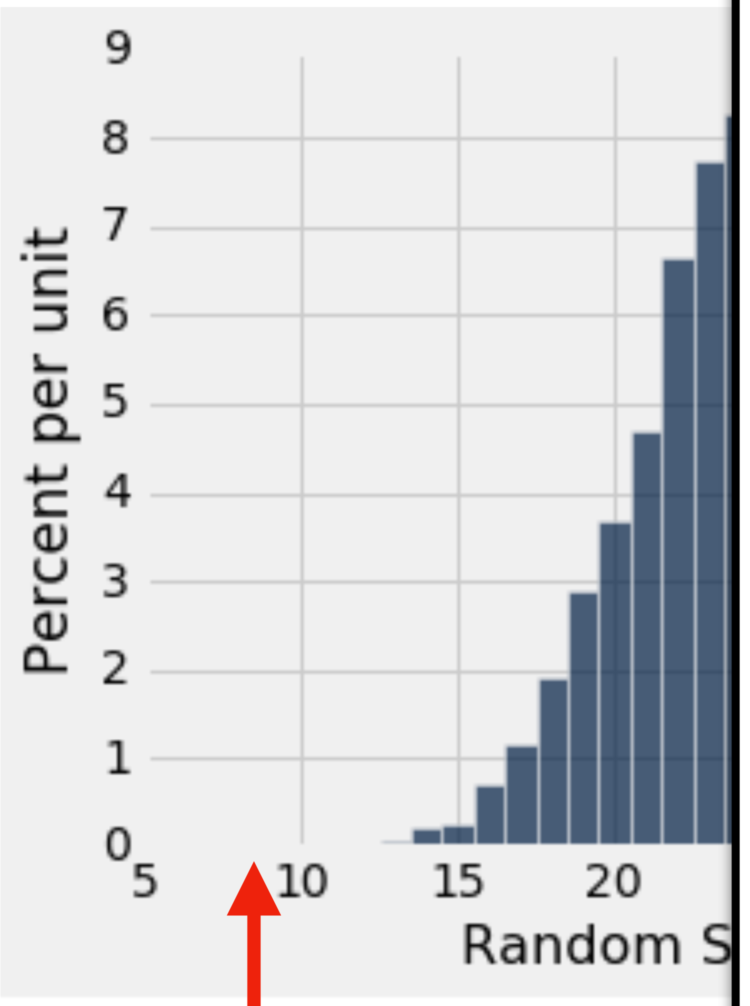


Numerical Data

Hypothesis Testing

- Modeling expected outcomes under the null and comparing it to our observed outcome

Swain v



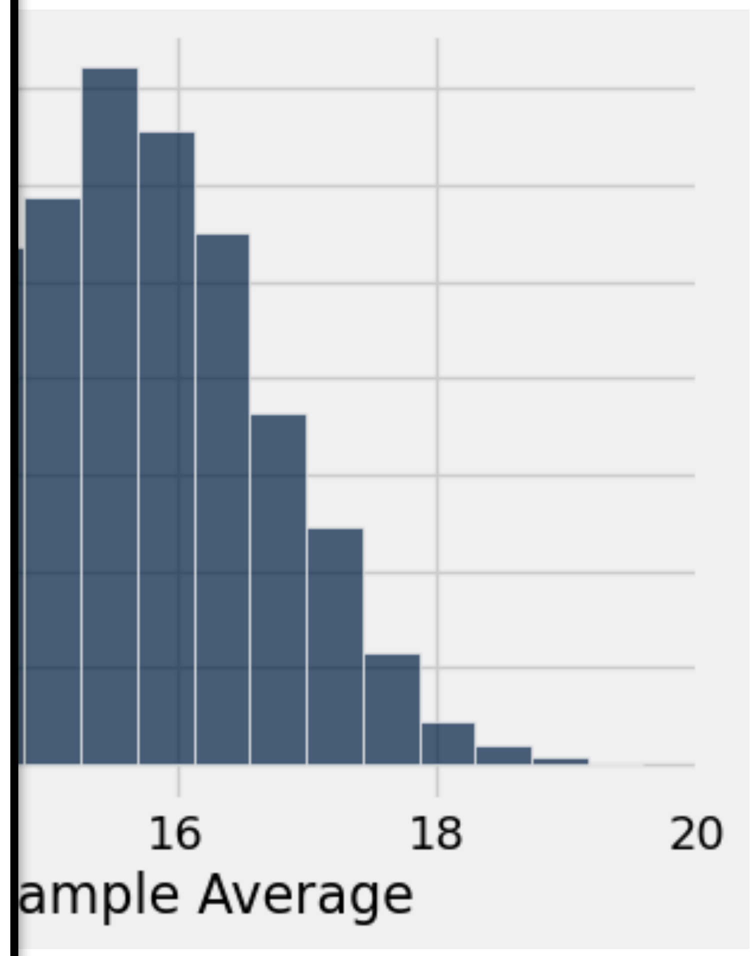
Observed N

It’s often not easy to say whether the observed outcome falls within our expectations…

How can we more precisely characterize the likelihood of observing an expected outcome?

p-value!

am Scores



Average

2 Categories

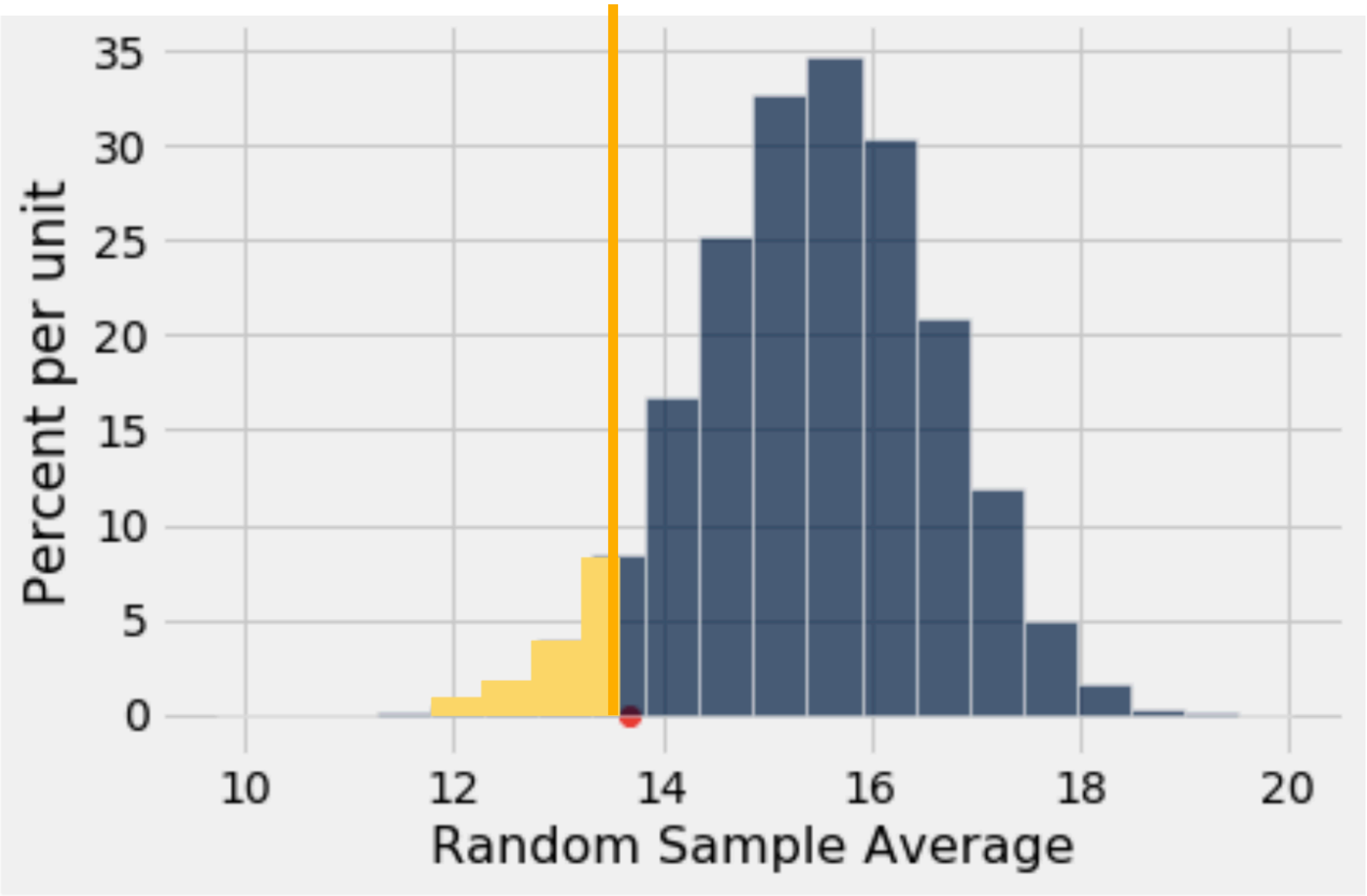
3+ Categories

Numerical Data

Hypothesis Testing

- Modeling expected outcomes under the null and comparing it to our observed outcome

Midterm Exam Scores



p-value = 0.058

p-value & **statistical significance**

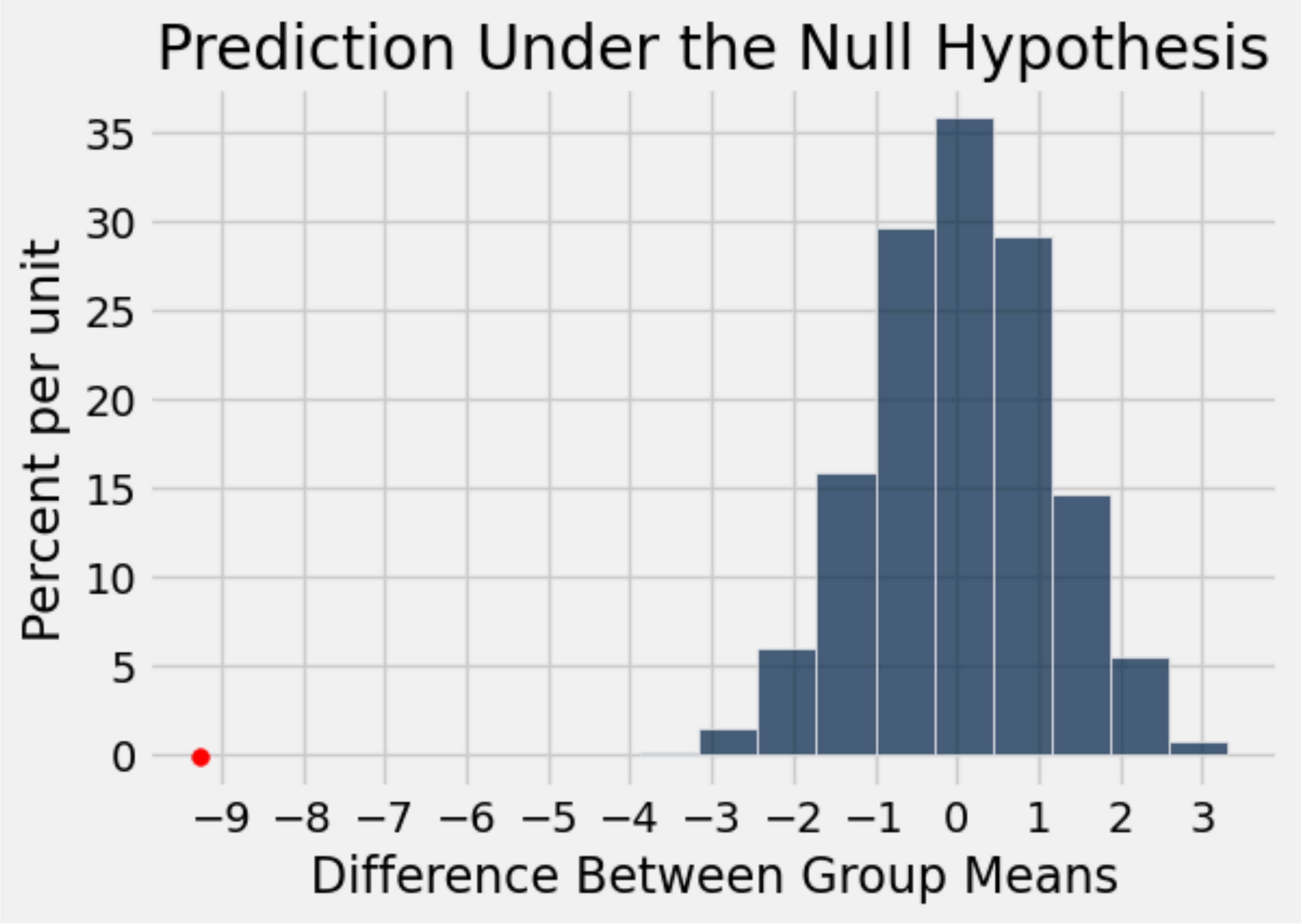
Process:

- Calculate the area of the tail (to the left/right of our observed value)

Hypothesis Testing

- Modeling expected outcomes under the null and comparing it to our observed outcome

Smoking vs Non-Smoking Mothers & Birthweight



p-value = 0

A/B Testing

Compare the difference between two groups

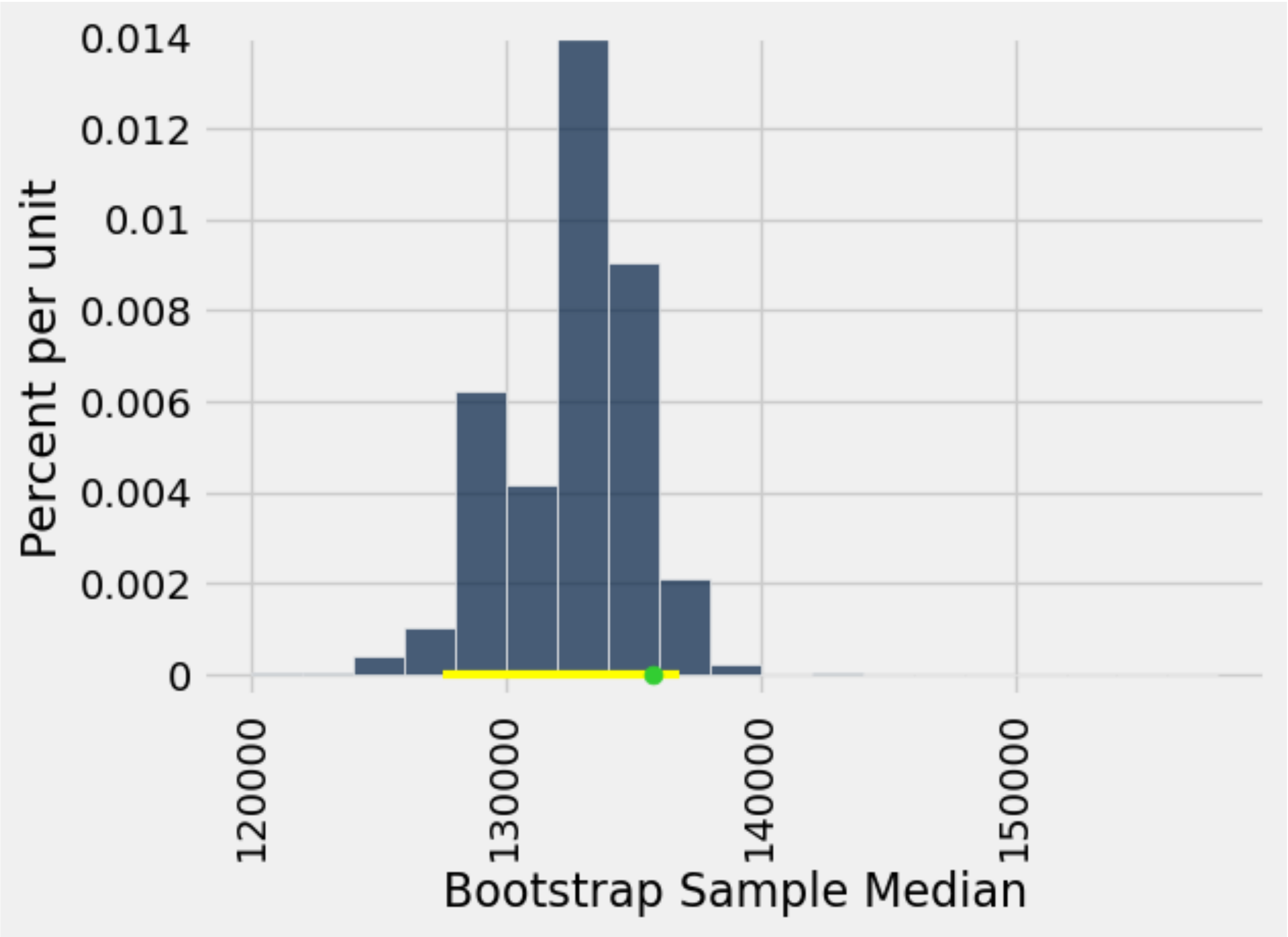
Process:

- Permutation test (shuffle labels)

Estimating a Parameter

- We want to estimate a population parameter from a sample statistic

Median Employee Salary



95% Confidence Interval: Median Salary between $125,745 and $140,318

Confidence Interval

Lets us estimate a range for what we think the parameter's value is

Process:

- Bootstrap

Next time

- Normal Distributions