Ch. 3: Estimation: How Large Is the Effect?

Navigation

By Date

- February 18: start - end
- February 20: start - end
- February 25: start - end
- February 27: start - end

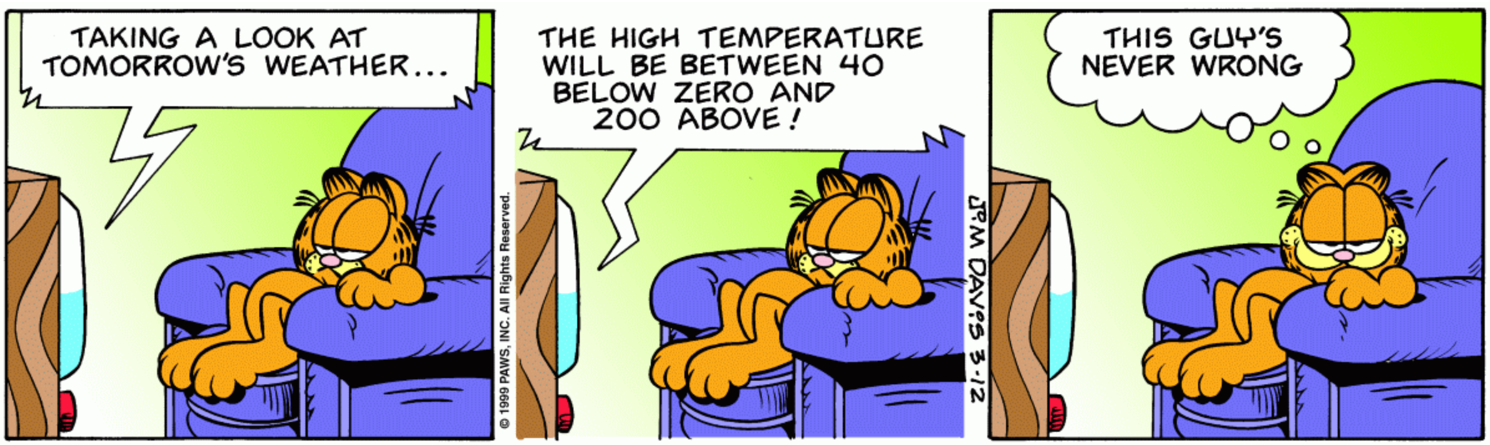
By Section

- 3.1: start - end
- 3.2: start - end
- 3.3: start - end
- 3.4: start - end

3.1: Statistical Inference: Confidence Intervals

Introduction

- So far, we have been looking at whether our parameter may be equal to a value or if it is equal to something else
- Other ways to find the parameter?
  - Range of values?



Confidence Intervals

When we fail to reject our null hypothesis…
Conclusion: Null hypothesis value is __________

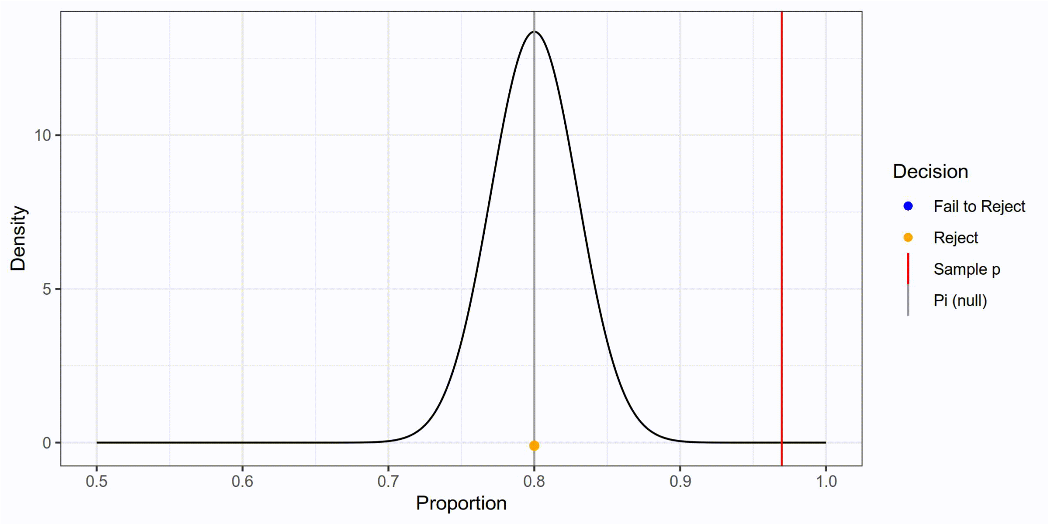
- We could do multiple two-sided tests to find which null hypothesis values are plausible and which are unlikely

- This sets up an interval of values that could be our parameter rather than just a single estimate

Confidence Intervals

When we fail to reject our null hypothesis…
Conclusion: Null hypothesis value is plausible

- We could do multiple two-sided tests to find which null hypothesis values are plausible and which are unlikely
- This sets up an interval of values that could be our parameter rather than just a single estimate



Example 3.1 - Dogs Smell Cancer

Can dogs sniff out cancer? Marine the labrador was presented with a bag breathed into by a colorectal cancer patient. Then, she was shown 5 bags, one from a colorectal cancer patient and 4 from controls who did not have cancer. This procedure was repeated 33 times.

- What is the experimental unit?
- What is n?
- What are the hypotheses?

Example 3.1 - Dogs Smell Cancer

Marine guessed correctly in 32/33 trials. $\hat{p} = .969$

- Test the hypothesis that $\pi = 0.2$ (random chance)

- Test the hypothesis that $\pi = 0.7$

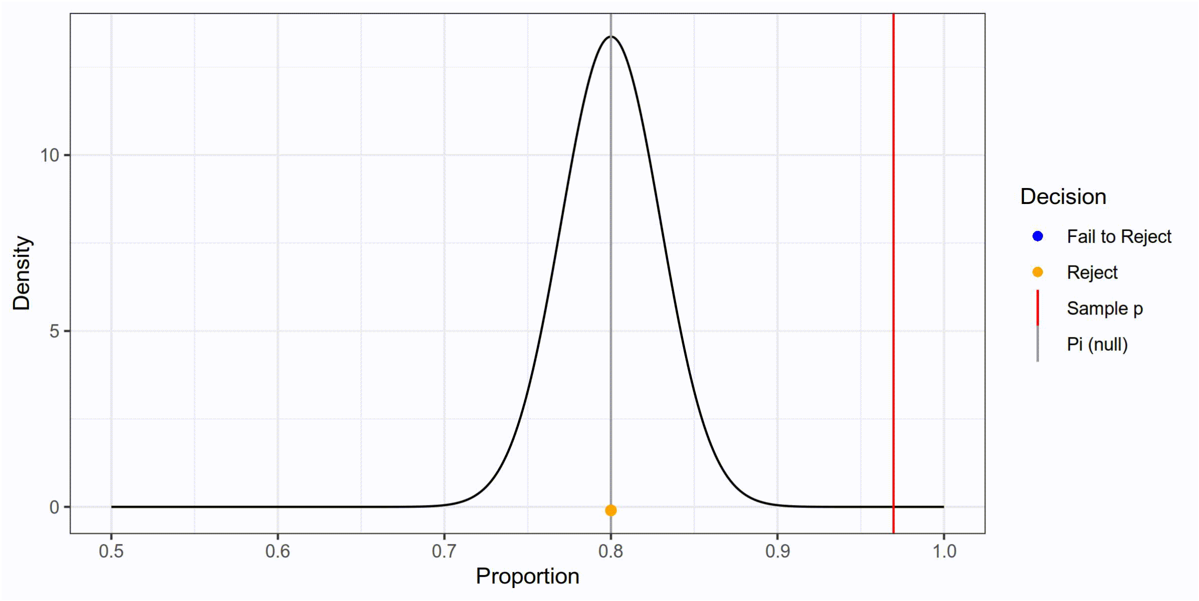
- Test the hypothesis that $\pi = 0.85$

Example 3.1 - Dogs Smell Cancer

By row, each group should test the following value:

Row	Hypothesized value	Accept/reject
A	0.75	
B	0.78	
C	0.81	
D	0.84	
E	0.87	
F	0.90	
G	0.93	
H	0.96	
	1.0	

Visual Confidence Interval



- Wherever we fail to reject is an interval of plausible values for π
- 95% confidence interval for π interpretation:
  We are 95% confident that our true parameter value falls within our confidence interval
- Found by obtaining parameter values that fail to be rejected by a two-sided test of significance with 0.05 level

Changing Significance Level

- We've been rejecting tests with significance level 0.05
- If we change and reject at 0.01 instead, what happens?
  - we reject H_0 when it is true 1% of the time
  - p-values have to be smaller before we reject

    Thus, we're rejecting H_0 less frequently
    If our p-value is 0.03, we would have rejected at $\alpha = 0.05$ but now we fail to reject

  - What happens to the interval?

    It gets wider

Changing Significance Level

- As significance level α goes down…
  - We reject less often
  - We fail to reject more often
  - We have more plausible values
  - Our confidence interval is wider
- As significance level α increases…
  - We reject more often
  - We fail to reject less often
  - We have fewer plausible values
  - Our confidence interval is narrower
- *Confidence level* **increases** when *significance level* **decreases**

Example: Lottery tickets

We want to try to win the lottery and have to decide how many tickets to buy.

To be more confident that we will win, should we buy more tickets or less tickets?

More confidence -> need to buy more tickets to have a higher chance of winning

When we're constructing a confidence interval, we need to include more values in the interval to have higher confidence.

Section 3.1 Summary

- "Simulating" confidence intervals
  - Trying many different parameter values
  - Testing: which ones do we reject (not plausible), and which ones do we fail to reject? (plausible value)
- What level we reject at dictates our "Confidence Level"
  - Significance level + Confidence level = 1
  - Reject less than 0.05, 95% Confidence level
  - Reject less than 0.10, 90% Confidence level
  - Which one is wider?
- Connection between p-value and confidence interval
  - Significance level + Confidence level = 100%

Exploration 3.1

Most people are right-dominant. In a study reported in Nature (2003), researchers showed that couples tend to lean their heads to the right while kissing. The researchers observed kissing couples in various public places and recorded which direction their heads tilted, excluding any couples who were holding objects that might affect the direction they turned. They observed that 80 of 124 couples leaned their heads to the right.

In groups, identify the observational units, variable of interest, and whether the variable is categorical or quantitative.

Work through questions 3-13 from the exploration in groups, then discuss questions 14-18. Upload the pdf of your responses as a group.

3.2: 2SD and Theory-Based Confidence Intervals For a Single Proportion

Theory Based Method

- We do not want to plug in those parameter values every time we want an interval

- Method boils down to one idea:
  We're finding a set of plausible values for the parameter centered around the ______________ and adding some wiggle room around it.

- The wiggle room is known as the ________________

Theory Based Method

- We do not want to plug in those parameter values every time we want an interval

- Method boils down to one idea:
  We're finding a set of plausible values for the parameter centered around the sample statistic and adding some wiggle room around it.

- The wiggle room is known as the margin of error

Confidence Interval

So the basic set up for our confidence interval is:

$$\text{Sample statistic} \pm \text{margin of error} = (lower, upper)$$

Margin of error is based on two values:

1. The ____________________ of the sample statistic

2. A __________ reflecting the confidence level we want

Thus the margin of error is…

Confidence Interval

So the basic set up for our confidence interval is:

$$\text{Sample statistic} \pm \text{margin of error}$$

Margin of error is based on two values:

1. The standard deviation of the sample statistic

2. A multiplier reflecting the confidence level we want

Thus the margin of error is...

Multiplier

- The multiplier is based on the confidence level (Sec 3.1)
- For a _______ confidence level, the multiplier must get bigger
    - Want a ______ interval to cover more plausible values
- For a ______ confidence level, the multiplier will be smaller
    - Leads to a __________ interval
    - But we don't have as much confidence

Multiplier

- The multiplier is based on the confidence level (Sec 3.1)
- For a higher confidence level, the multiplier must get bigger
  - Want a wider interval to cover more plausible values
- For a lower confidence level, the multiplier will be smaller
  - Leads to a narrower interval
  - But we don't have as much confidence

Multiplier

- We will use a multiplier of ± 2 in this class
- Why?
    - Bell shaped distribution
    - 95% of the distribution captured within 2 standard deviations of the center of the null distribution
    - True parameter will be within two standard deviations of the observed sample statistic in about 95% of samples
- Other confidence levels lead to different multipliers

Standard Error in Confidence Intervals

- What standard error did we use for **proportions** within our standardized statistic?
    - But that was when we were testing a hypothesis that gave us a value for π
    - With a confidence interval, we are estimating where π is between 0 and 1
    - What is our best guess of what π is?

Standard Error

The **standard error** is the best guess for how much **variability** we have

- SE is the measure of spread of many samples
- must account for how big of a sample we have
- Different from the standard deviation of individual observations within a sample

for proportion data with sample proportion $\hat{p}$, the SE is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Standard Error

- Larger standard error ->
    - Less precise estimate
    - Harder to pin down "truth"
    - Wider confidence intervals
- Standard error also reflects sample size
    - as n increases, we have more information about the parameter
    - narrower confidence interval
    - SE goes down as n increases

Confidence Interval for a Proportion

$$\text{Sample statistic} \pm (Multiplier) \times (\text{Standard Error})$$

- Plug in our sample statistic and standard error for the type of data we have

$$\hat{p} \pm 2 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} =$$

Interpretation:

Conclusions

- Different interpretation for confidence intervals:
  We are 95% confident that the true value of … lies between … and …
- We can figure out what our p-value would be if we have a confidence interval and a set of hypotheses
  - If the null hypothesis value is INSIDE the interval…
  - If the null hypothesis value is OUTSIDE the interval…

Kissing Couples: Exploration 3.1

$\hat{p} = 0.645$

Our manual interval: (0.56, 0.72)

Standard error:

Using the 2SE method, calculate the confidence interval:

Applet

Use the Theory-Based Inference applet to check that your answer is approximately correct

Comparing other Confidence Levels

Conf Level	Interval
90%	(0.5745, 0.7159)
95%	(0.5610, 0.7294)
99%	(0.5345, 0.7559)

- Center: sample proportion $\hat{p}$
- Larger CI -> wider interval
- Notice the 95% interval is close to what we found with the 2 SE method
  - 2 is quick and easy as a multiplier
  - 95% actually uses a multiplier of 1.96

Review

- Confidence Intervals
    - Purpose: to find a range of **plausible** values for our parameter
    - General Formula breakdown:
- Concepts:
    - Where does the CI come from?
    - What affects the width?
    - How do we know what the p-value will be just from the confidence interval and a set of hypotheses?

Revivew

- Confidence Intervals
  - Purpose: to find a range of **plausible** values for our parameter
  - General Formula breakdown:

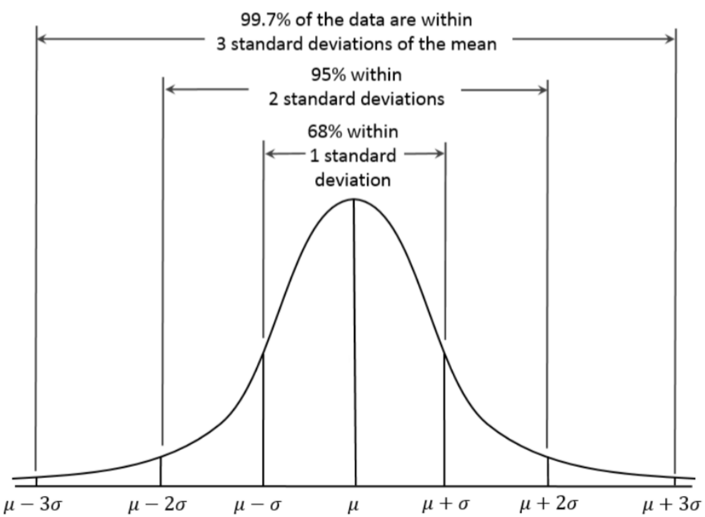
$$\text{statistic} \pm \text{multiplier} \times \text{margin of error}$$

- Concepts:
  - Where does the CI come from?
    many separate hypothesis tests of different values
  - What affects the width?
    the confidence level and the margin of error
  - How do we know what the p-value will be just from the confidence interval and a set of hypotheses?
    If the interval contains the null hypothesis value, the p-value will be above our significance level. If not, the p-value will be below our significance level.

Multipliers

The multiplier is a constant that indicates the number of standard deviations in a normal curve.

The larger the multiplier, the higher the confidence level.



Confidence	Multiplier
80%	1.28
90%	1.64
95%	1.96
99%	2.58

The confidence level is the percentage of area under the curve between the two values.

Proportions

- Validity conditions:
- Standard error:
- Confidence interval: (multiplier of 2 for 95% confidence intervals)
- Interpretation:

Proportions

- Validity conditions: 10 observations in each category
- Standard error:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Confidence interval: (multiplier of 2 for 95% confidence intervals)

$$\text{statistic} \pm \text{multiplier} \times \text{standard error} = \hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Interpretation: We are 95% confident that the long run proportion of **(context)** falls between **(lower bound)** and **(upper bound)**.

3.3: 2SD and Theory-based Confidence Intervals for a Single Mean

Means

- Validity conditions:

- Standard error:

- Confidence interval: (multiplier of 2 for 95% confidence intervals)

- Interpretation:

Means

- Validity conditions: symmetric distribution OR
  20 observations + not strongly skewed distribution

- Standard error:

$$\frac{s}{\sqrt{n}}$$

- Confidence interval: (multiplier of 2 for 95% confidence intervals)

$$\text{statistic} \pm \text{multiplier} \times \text{standard error} = \bar{x} \pm 2\frac{s}{\sqrt{n}}$$

- Interpretation: We are 95% confident that the long run average of **(context)** falls between **(lower bound)** and **(upper bound)**.

What affects the width of a CI

- Confidence Level
  - How often do we reject/fail to reject
  - Lottery example
  - In the formula as the __________
- Sample size
  - As you know more about the ___________, you get closer to knowing the true parameter value
  - Larger sample -> ________________
    Smaller sample -> _______________
- Standard error
  - Reflects ______________ of different samples
  - Highly variable (large SE) = harder to pin down what is "true"
  - Makes sense that **large SEs** lead to ______ confidence intervals

What affects the width of a CI

- Confidence Level
    - How often do we reject/fail to reject
    - Lottery example
    - In the formula as the multiplier
- Sample size
    - As you know more about the population, you get closer to knowing the true parameter value
    - Larger sample -> smaller interval
      Smaller sample -> larger interval
- Standard error
    - Reflects the variability of different samples
    - Highly variable (large SE) = harder to pin down what is "true"
    - Makes sense that **large SEs** lead to wider confidence intervals

Conclusions about hypotheses

- If the hypothesized mean is inside the interval
    - value is plausible
    - we would fail to reject H_0
    - p-value would be greater than significance level
- If the hypothesized mean is outside the interval
    - value is **not** plausible
    - reject H_0
    - p-value would be less than significance level

Example - Textbook prices

Two Cal Poly students gathered data on prices for a random sample of 30 textbooks from the campus bookstore. They found the average price was \$65.02, and the standard deviation of prices was \$51.42.

Example - Textbook prices

Two Cal Poly students gathered data on prices for a random sample of 30 textbooks from the campus bookstore. They found the average price was $65.02, and the standard deviation of prices was $51.42.

$\overline{x} = 65.02$

$s = 51.42$

$n = 30$

$$
\begin{aligned}
\text{statistic} \pm \text{multiplier} \times \text{standard error} &= \bar{x} \pm 2s/\sqrt{n} \\
&= 65.02 \pm 2 \times 51.42/\sqrt{30} \\
&= 65.02 \pm 18.776 \\
&= (46.24, 83.80)
\end{aligned}
$$

Example - UNL Sleep Data

A 2014 study of 66 UNL students found that on average, participants slept 5.9 hours the night before, with a sample standard deviation of 1.67.

Construct a confidence interval for the average hours all UNL students slept the night before. Is 7 a plausible value?

Example - UNL Sleep Data

A 2014 study of 66 UNL students found that on average, participants slept 5.9 hours the night before, with a sample standard deviation of 1.67.

Construct a confidence interval for the average hours all UNL students slept the night before. Is 7 a plausible value?

We are 95% confident that UNL students get, on average, between 5.49 and 6.31 hours of sleep per night.

7 is not a plausible value.

Example - Skittles

A sample of 154 students were blindly given a single **strawberry** Skittles candy to eat. They were told the 5 possible flavors and asked to identify the one they ate.

The responses were recorded as follows: 89 strawberry, 23 lime, 14 lemon, 17 orange, and 11 grape answer.

Find a 95% confidence interval for the parameter (wanting to see if students get it right).

Example - Body Temperature

Normal body temperature is often reported to be $98.6^\circ F$. Is that number accurate? Data from 65 healthy females between 18 and 40 years old was collected, and the mean body temperature was $98.39^\circ F$ with a standard deviation of 1.2. Construct a 95% confidence interval for the parameter.

Example - College students

A random sample of 97 male college students found that 12 students were left-handed, and the average height was 5.9 feet with a standard deviation of 0.75 feet. Find a 99% confidence interval for the proportion of all male college students that are left-handed.

3.4: Factors That Affect the Width of a Confidence Interval

What happens	Background Information	Wider or narrower?
Confidence level goes up Significance level goes down		
Confidence level goes down significance level goes up		
Sample size goes up		
Sample size goes down		
Sample standard deviation increases		
Sample standard deviation decreases		

What happens	Reject or Fail to Reject?	p-value is…
H_0 value inside the interval		
H_0 value outside the interval		

Confidence Intervals

- Width of a CI is affected by
  - Sample size (via standard error calculation)
  - Sample standard deviation (mean) or sample proportion (via standard error calculation)
  - Confidence level (via multiplier)
- **Focus**: What does the confidence level actually mean?

Exploration 3.4B

What does it mean to be 95% or 99% confident?

- Reeses candies come in orange, yellow, and brown
- We want to estimate the long-run proportion of orange candies
- **assuming** for now that this proportion, π, is 0.50

- **Question 1** - If you take a sample of 100 Reese's Pieces candies and find the sample proportion of orange, is there any guarantee that the sample proportion will equal 0.50?

- **Question 2** - Suppose we calculate a confidence interval from this sample proportion. Is there any guarantee that the interval will contain the value 0.50?

Exploration 3.4B

- **Question 3** - Suppose that you select another random sample of 100 Reese's pieces candies. Is there any guarantee that the sample proportion will be the same as for the first sample? Will the confidence interval based on the new sample necessarily be the same as the confidence interval based on the first sample? How do you think they will differ?

Exploration 3.4B

- **Question 4**: Simulating Confidence Interval Applet
  - Find one sample and the confidence interval
  - Record sample proportion, lower bound, and upper bound
  - Is the sample proportion in the interval?
  - Is the long-run proportion in the interval?
- **Question 5**: More sampling…
  - Simulate 99 more samples and intervals
  - At what value is the graph centered?
  - Click red dot, answer questions
  - Click green dot, answer questions
  - Watch what happens to the running total as the sample value goes up
- **Question 6**: What happens with 90% intervals? 99% intervals?

Confidence Intervals

- We can take lots of samples… say, 100
- For each sample, find the 95% confidence interval
- We expect 95 of the 100 intervals will contain the true parameter
- This does **NOT** mean the probability that the true parameter is in the confidence interval is 0.95.

3.5: Cautions When Conducting Inference

The "Bradley Effect"

- Discrepancies between polling numbers and election outcomes

Hypothesized to be caused by "social desirability bias" - people answering poll questions in the way they thought the interviewer would want them to answer (e.g. politically correct)

- Originally showed Tom Bradley was likely to win an election against George Deukmejian in the 1982 LA mayoral election
    - Bradley is black, Deukmejian is white
- Believed to be a factor in:
    - polling between H. Clinton/Obama in 2008 primary elections
    - in both directions - gender-based and race-based
    - may have cancelled each other out or affected different states
    - polling between H. Clinton and Trump in 2016

The "Bradley Effect"

New Hampshire polling in the 2008 Democratic primary: 778 likely voters, and only days before the primary, 319/778 (41%) said they were voting for Obama, compared to 218/778 (28%) who said they would vote for Clinton.

Polling methodology:

- Poll used random digit dialing
- People who answered the phone were asked if they were likely to vote in the upcoming primary
- Only 9% of people contacted participated in the survey

The "Bradley Effect"

New Hampshire polling in the 2008 Democratic primary: 778 likely voters, and only days before the primary, 319/778 (41%) said they were voting for Obama, compared to 218/778 (28%) who said they would vote for Clinton.

Assumptions:

1. Random digit dialing is a reasonable way to get ahold of voters
2. Self-identified likely voters are representative of the population of people who vote in the primary
3. The 9% who participate in the poll are similar to the 91% who don't participate
4. Respondents' answers match their voting behavior

Testing for the "Bradley Effect"

What is our research question?

Does the proportion of people who say they will vote for Obama in the poll differ from the proportion of people who actually voted for Obama in the primary?

Testing for the "Bradley Effect"

What is our parameter of interest?

π is the probability voters will claim they are voting for Obama in the poll

What is our parameter value under H_0?

0.3645, the proportion of people who voted for Obama in the primary

Testing for the "Bradley Effect"

What are the null and alternative hypotheses (in symbols and in words)?

$H_0:$

$H_A:$

Testing for the "Bradley Effect"

What are the null and alternative hypotheses (in symbols and in words)?

$$H_0 : \pi = 0.3645$$

$$H_A : \pi > 0.3645$$

The null hypothesis is that the proportion of voters who claim they will vote for Obama in the poll is 0.3645, the same as the proportion who voted for him in the primary

The alternative hypothesis is that the proportion of voters who claim they will vote for Obama in the poll is greater than 0.3645; that is, that there is some effect which makes individuals who answer the poll more likely to say they will vote for Obama compared to their actual behavior

Testing for the "Bradley Effect"

What is our n?

$n = 778$, the number of people in the poll

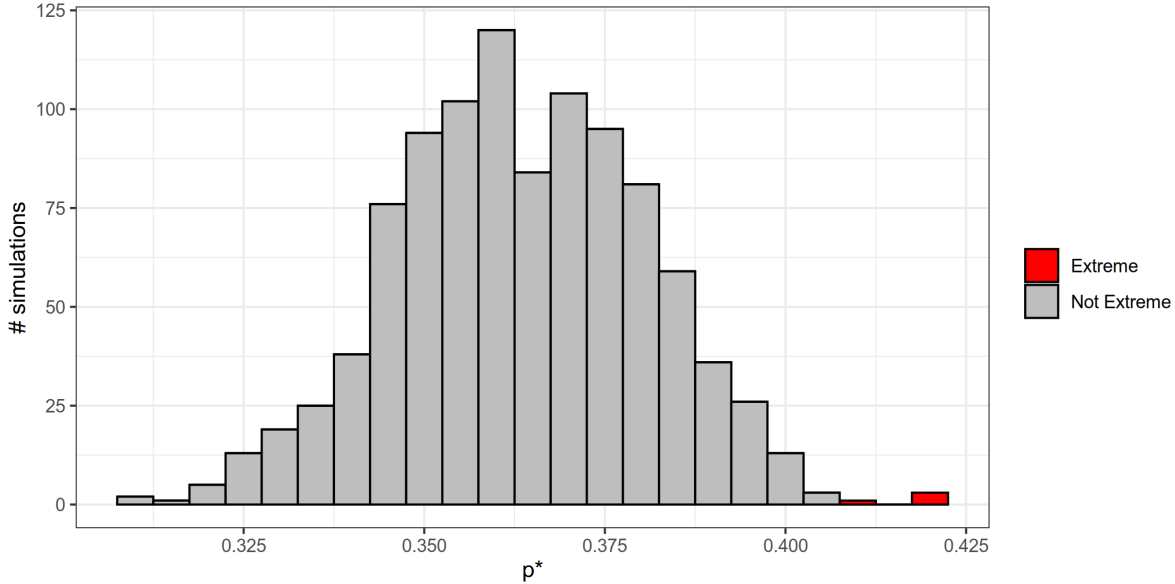
What is our sample Statistic?

$$\hat{p} = 319/778 = 0.41,$$

the proportion of people who indicated in the poll that they would vote for Obama

Testing for the "Bradley Effect"

Conduct a Simulation:



Our p-value is $P(p^* \geq \hat{p}) = P(p^* \geq 0.41) = 0.004$

Testing for the "Bradley Effect"

Strength of Evidence:

Evaluate the results of your simulation

It is unlikely that the proportion of people who claim they will vote for Obama in the poll would be 0.41 if the true proportion of people voting for Obama in the poll is 0.3645 ($p = 0.004$).

Evaluate the strength of your evidence

This provides strong evidence against our null hypothesis that $\pi = 0.3645$

Make a conclusion

We reject H_0 and conclude that the proportion of people who claim they will vote for Obama in the poll is higher than 0.3645.

Why?

Testing for the "Bradley Effect"

Assumptions:

1. Random digit dialing is a reasonable way to get ahold of voters
   Random digit dialing is effectively a simple random sample of residents who have a landline or cell phone (slightly overrepresenting people who have more than one phone line).

2. Self-identified likely voters are representative of the population of people who vote in the primary
   People are well-intentioned and may plan to vote, but may not actually make it to the polls. If these people are more likely to vote for one candidate over the other, this could bias results.

Testing for the "Bradley Effect"

Assumptions:

1. Random digit dialing is a reasonable way to get ahold of voters
2. Self-identified likely voters are representative of the population of people who vote in the primary
3. The 9% who participate in the poll are similar to the 91% who don't participate
   The **response rate** of a poll is the proportion of those asked who participate. Response rates are generally low for phone polls. Many other phone polls around the same time didn't show evidence of this bias, but that is no guarantee that this time the 9% who responded were actually a representative sample. Sometimes, you just get a bad random sample.
4. Respondents' answers match their voting behavior

People may change their minds between the poll and the election, or they may intentionally try to throw off the poll. If enough people do this in the same direction, the poll results may be off.

Testing for the "Bradley Effect"

We do not have any statistical evidence (yet) to decide whether these assumptions played a role in our results.

See Explanation 3.5.3 for a more thorough analysis of the post-election poll analysis findings

Nonrandom errors (systematic errors) are errors that affect the results of your measurement process (the system) consistently. Nonrandom errors would be a problem even if you could measure the whole population - they are not a failure of the statistical calculations, but of the measurement method.