STAT 102: Week 6

Ricky’s Section

Introductions and Attendance

Introduction: Name

Question of the Week: What is your favorite "thing" at the moment? Snack, activity, book, etc.

Important Reminders

Midterm

- **Written Component:** Wed, 03/12 from 6 to 9 PM in Science Center 705 and 706
- **Oral Component:** Over Zoom afterwards on 03/13 and 03/14 (10 minute sessions)
- No class/section on Thur, Mar 13
- **You all got this!** 🙂

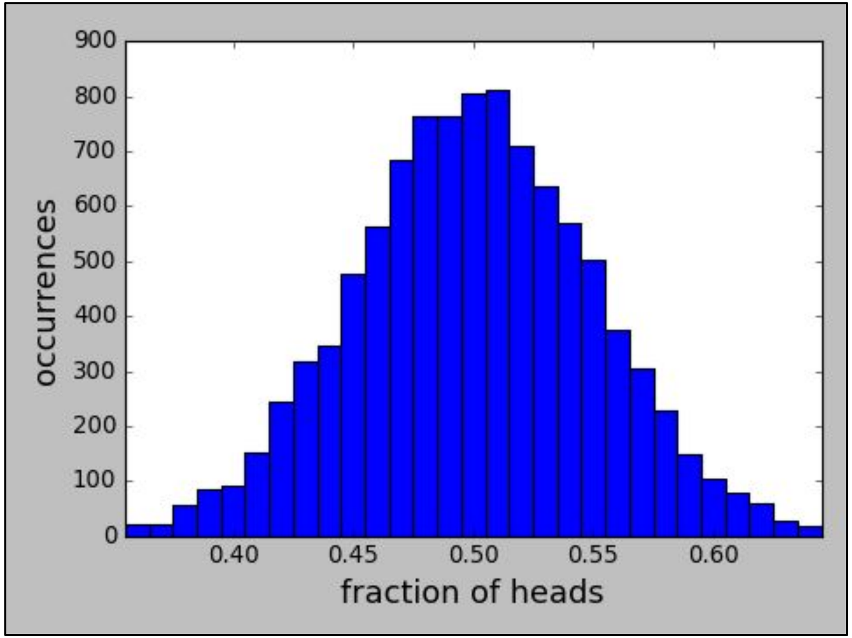
Review

- **Review Session** with me and Sarah on Sunday, 03/09 from 7 to 9 PM in Science Center 316!
- Sign up for practice oral exams if you haven't already!
  - https://canvas.harvard.edu/courses/143378/modules/items/1874820

Content Review: Week 5

“Theoretical” Distributions

- Theoretically, **Sample Mean** is approximately normally distributed with a **SD of $\sqrt{(p(1 - p)/n)}$**
- Theoretically, **Sample Proportion** is approximately normally distributed with a **SD of $\sqrt{(\sigma^2/n)}$**
- This is why we don’t use **Sample Standard Deviation** σ-hat as our estimate of the **Sample Statistic’s SD**
- Instead, we use the **SD of our Bootstrap Distribution**, which is theoretically a good approximation

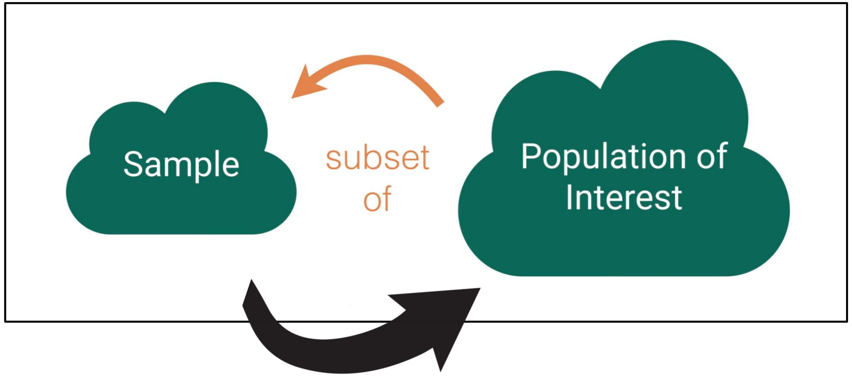


Content Review: Week 6

Recap of Inference

- Last week, we started **inference** with **confidence intervals**
- Now, we'll continue with **hypothesis testing**
- Though complementary, they are different
  - **Confidence intervals** estimate the **parameter**
  - **Hypotheses** test a certain "conjecture" about the **parameter**

Sampling 😃



Inference 😱

A Tale of Two Hypotheses

- **Test statistic**: Numerical summary of the **sample data** (often, but not always, equal to our **observed sample statistic**)
- **Null hypothesis (H_o)**: World where **research conjecture is false** ("no change, status quo")
  - **Null distribution** is **sampling distribution** of **test statistic** assuming **null hypothesis is true**
- **Alternative hypothesis (H_A)**: World where **research conjecture is true**
  - **Alt. distribution** is **sampling distribution** of **test statistic** assuming **alt. hypothesis is true**
- **P-value**: Probability of getting the **observed test statistic OR MORE EXTREME** if **null hypothesis is true**, represented by area under curve of **null distribution**

A Note on Test Statistic

- **Test statistic**: Numerical summary of the **sample data** (often, but not always, equal to our **observed sample statistic**)
    - Last week, our **observed sample statistic** was $\hat{p} = 0.5$ from a random sample of 600 viewers
    - We note this as "**observed**" because if we took a different sample of 600 viewers, we probably would've gotten a different sample statistic, such as $\hat{p} = 0.4$ (i.e., **sampling variability**)
- As the name implies, this is the statistic we'll be using in our **(hypothesis) test**!
    - More on this later, but there will be other test statistics

Can the null and alternative hypotheses both be true?

Question:

Can the null and alternative hypotheses both be true?

No!

The **null hypothesis** and **alternative hypotheses** are mutually exclusive. That is, they CANNOT coexist.

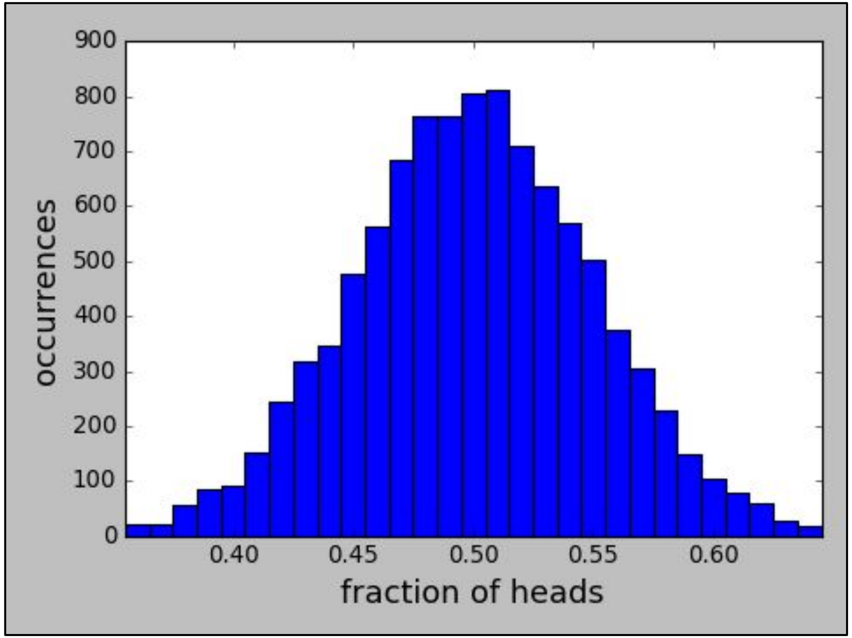
Only 1 can be true. Either this drug works, or it doesn't. Either this coin is rigged, or it's not. And so on.

Essentials of Hypothesis Testing

- **Step 1**: State **hypotheses** (in terms of **population parameter**)
  - Null hypothesis posits the coin is normal. Alternative hypothesis argues it's rigged. $H_0: p = 0.5$, $H_A: p > 0.5$
- **Step 2**: Specify a **significance level**, α (usually $\alpha = 0.05$)
- **Step 3**: Generate **null distribution**
  - If I were to repeatedly sample under the null hypothesis (assuming the coin has a normal 50% chance of heads), what would my sampling distribution look like?
- **Step 4**: Compute **observed test statistic** and **compute p-value**
  - Let's say, with $n = 50$, I observe 30 heads, so $\hat{p} = 0.6$. Under our null distribution, this has a p-value of 0.103.
- **Step 5**: Draw conclusions **in the context of the problem**
  - The probability of seeing 30 or more heads when flipping a fair coin 50 times is equal to 0.103. Since our p-value is high ($0.103 > 0.05$), we fail to reject the null hypothesis. There is little evidence the coin is rigged.

Coin Flips: An Intuition behind Sampling Distributions

- Let's flip a FAIR coin 15 times and record the proportion of heads
- Will our sample statistic always be 0.5? No!
- The center is the "theoretical" population proportion ($p = 0.5$)
- We're graphing a bunch of sample proportions ($\hat{p}_1 = 0.4$, $\hat{p}_2 = 0.5$, $\hat{p}_3 = 0.6$, ...)



The “P-Value Formula”

- **“If {null hypothesis} were true, then the probability of observing {test statistic} or {more extreme} would be {p-value}.”**
  - This is “interpreting the p-value”
- **“Because {p-value} is a {high/low} probability compared to {alpha}, we reject {reject/fail to reject} the null hypothesis.”**
  - This is “drawing a relevant conclusion”

If I want to see whether or not the majority of Harvard students agree with a new bill, what should my hypotheses be (in terms of my pop. parameters)?

Other Parameters and Statistics

	Response Variable		Numeric Quantity	Sample Statistic	Population Parameter
1 variable	Numerical		Mean	$\bar{x}$	μ
	Binary Categorical		Proportion	$\hat{p}$	p
	Response variable	**Explanatory Variable**	**Numeric Quantity**	**Sample Statistic**	**Population Parameter**
2 variables	Numerical	Binary Categorical	Difference in Means	$\bar{x}_1 - \bar{x}_2$	$\mu_1 - \mu_2$
	Binary Categorical	Binary Categorical	Difference in Proportions	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$
	Numerical	Numerical	Correlation	r	ρ

Question:

If I want to see whether or not the majority of Harvard students agree with a new bill, what should my hypotheses be (in terms of my pop. parameters)?

We have a **binary categorical response variable** (fraction of students that agree). This is a **one-tailed proportion**.

H_0: $p = ½$ (There is no majority)

H_A: $p > ½$ (The majority agree)

If I want to see if Harvard students get less sleep than other college students, what should my hypotheses be (in terms of pop. parameters)?

Question:

If I want to see if Harvard students get less sleep than other college students, what should my hypotheses be (in terms of pop. parameters)?

We have a **binary categorical explanatory variable** (Harvard or not) and **numerical response variable** (hours of sleep). This is a **one-tailed difference of means**.

H_0: $\mu_{Harvard} - \mu_{Other} = 0$ (Harvard students get same amount of sleep)

H_A: $\mu_{Harvard} - \mu_{Other} < 0$ (Harvard students get less sleep)

If I observe a difference of means of -2.7 hours (and a p-value of 0.003), what is an interpretation of the p-value and a conclusion? Assume $\alpha = 5\%$.

Question:

If I observe a difference of means of -2.7 hours (and a p-value of 0.003), what is an interpretation of the p-value and a conclusion? Assume α = 5%.

Using the p-value formula...

If **there was no difference in mean hours of sleep between Harvard and non-Harvard students**, then the probability of observing our **test statistic, a difference of -2.7 hours**, or **less** would be 0.3%.

Because 0.3% is a **low** probability (0.3% < 5%), we **reject** the null hypothesis.

Decisions, Decisions

- There are 4 potential outcomes of a **hypothesis test** (shown below), depending on what we do and what's actually true
- **α** - Probability of Type I Error (rejecting H_0 when it's true)
- **β** - Probability of Type II Error (failing to reject H_0 when H_A is true)
  - As **α** decreases, **β** increases (but they DON'T add up to 1)
- **Power**: Probability of rejecting H_0 when H_A is true (best outcome 😁)
  - **Power = 1 - β**

	We Reject H_0	**We Fail to Reject H_0**
H_0 is true	Type I Error	Correct Decision 🙂
H_A is true	Correct Decision 😁	Type II Error

If we reject the null hypothesis, is it possible we committed a Type I Error? A Type II Error?

Question:

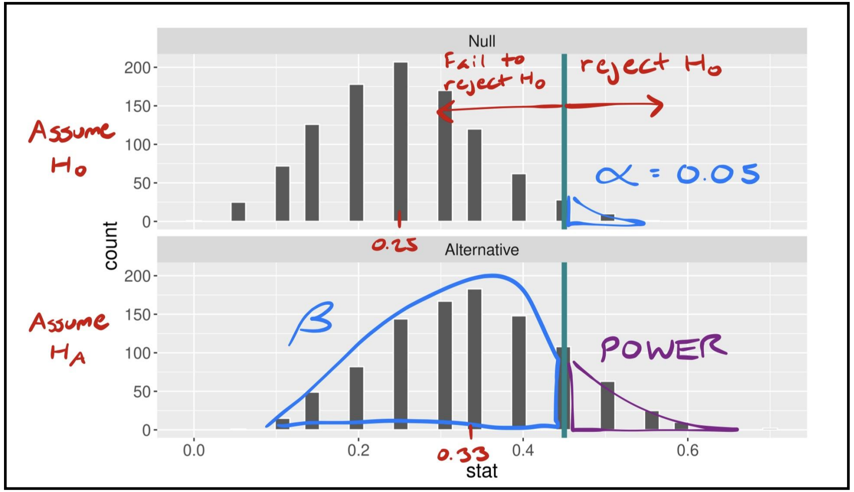
If we reject the null hypothesis, is it possible we committed a Type I Error? A Type II Error?

I remember Type I Error as a "delusional scientist" and Type II Error as a "missed opportunity."

If we reject the null hypothesis, there's a possibility we committed a Type I Error but no possibility we committed a Type II Error (by definition, this would require FAILING to reject the null hypothesis).

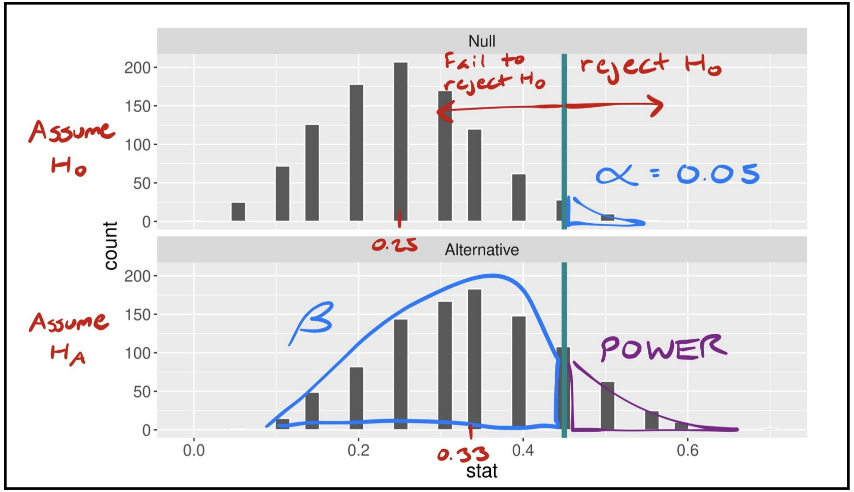
More on Power…

- **Power**: Assuming $\mathbf{H_A}$, what is the probability we reject $\mathbf{H_o}$?
- Think of **power** as a thought experiment—it helps us better understand **hypothesis testing**
  - In real life, we don't know if H_A is true… or where it's centered at!
  - There is an infinite number of alternative distributions that could exist… let's pick just one



Intuition behind Power

- **Power**: Assuming $\mathbf{H_A}$, what is the probability we reject $\mathbf{H_o}$?
  - Given $\mathbf{H_A}$ is true, we look at the **alternative distribution** (which, now, is the true state of the world)
  - The **alpha level** is the probability of rejecting $\mathbf{H_o}$ in the **null distribution**
    - The **critical region** (to the right of **α**) is where we reject $\mathbf{H_o}$
  - Thus, in the **alternative distribution**, the region to the right of the **alpha level** is **power**

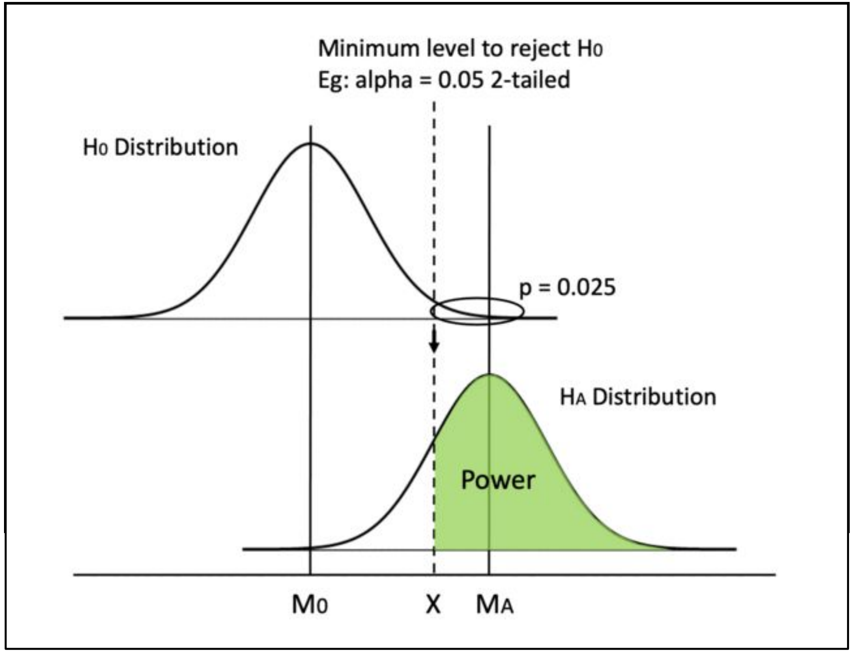


Example: Baseball

- Ricky, an avid baseball player, has been a **0.250** career hitter, but, magically, he improves to be a **0.333** hitter
- He wants a raise, but he has to convince his manager he genuinely improved
- The manager offers to examine his performance in **20 trials**
- **H_0: p = 0.250, H_A: p > 0.250 ($p \stackrel{?}{=} 0.333$)**
  - Because the **alternative hypothesis** is **p > 0.250**, there is an infinite amount of **alternative distributions** that could exist… specifically, I'm interested in the one centered at 0.333
- He wants his test to be "powerful"
  - When **α = 0.05**, he needs to get **9 or more hits** to get a small enough **p-value** to reject **H_0**
  - Unfortunately, at **α = 0.05**, the **power of this test** is **0.211** (only a 21% probability of being in the best outcome), so how can we improve the **power of this test**?

How to Increase Power: Increase Alpha

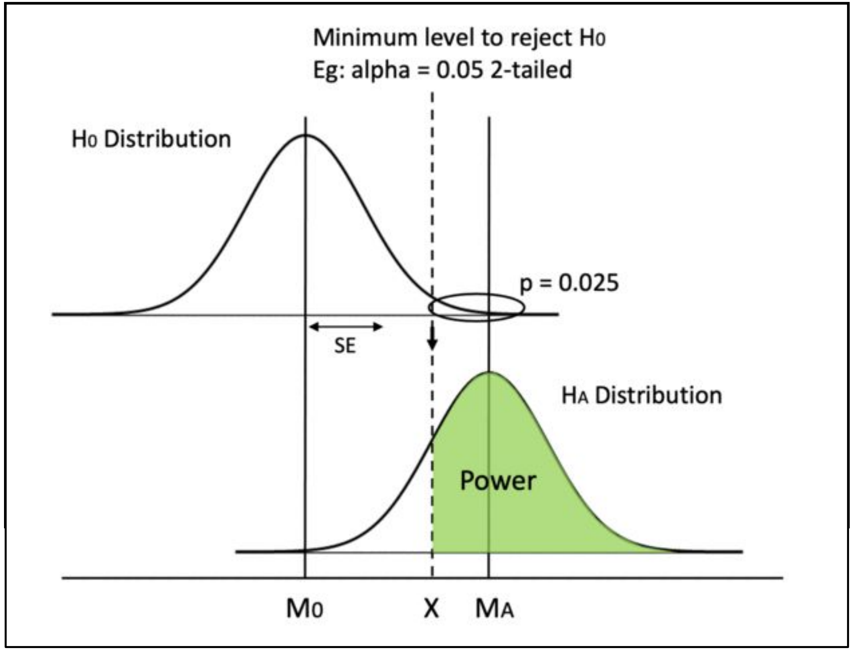
- This makes it easier to reject $\mathbf{H_o}$
- Also, this "shifts" the **critical line** to the left, leading to more area in the "**power region**" of the **alternative distribution**
- Intuitively, we now have a higher probability of rejecting $\mathbf{H_o}$, and **power** is probability of rejecting $\mathbf{H_o}$ when $\mathbf{H_A}$ is true



https://towardsdatascience.com/5-ways-to-increase-statistical-power-377c00dd0214

How to Increase Power: Increase Sample Size

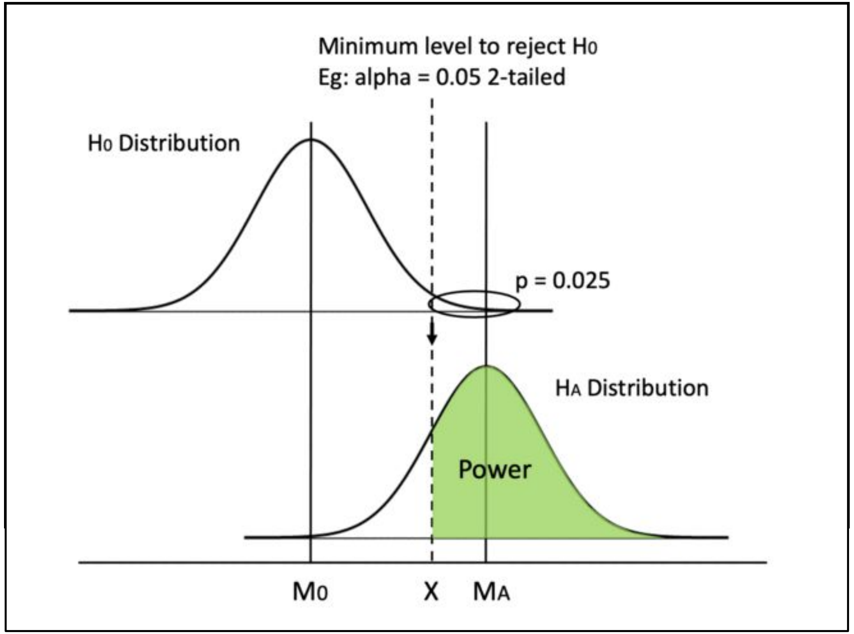
- This decreases **spread** of **histograms**, leading to less overlap between **null distribution** and **alternative distribution**



https://towardsdatascience.com/5-ways-to-increase-statistical-power-377c00dd0214

How to Increase Power: Increase Effect Size

- **Effect Size**: Difference between **true value of parameter** and **null value**
- This makes it easier for us to notice a difference
- Also, this "shifts" the **center of the alternative distribution** to the right, leading to more area in the "**power region**"



https://towardsdatascience.com/5-ways-to-increase-statistical-power-377c00dd0214

More on Effect Size...

- **Statistical Significance ≠ Practical Significance**
  - *Ex: A study concluded couples who met online are more likely to be satisfied (p-value < 0.001), but their happiness value of 5.64 isn't much more than the happiness value of 5.48 for couples who met in-person*
- Let's say, magically, Ricky actually improved to **0.400**
- Now, the **effect size** **(0.400 - 0.250)** is larger than before
  - Intuitively, it should now be more noticeable if he actually improved from before, so our **hypothesis test** is more "powerful"

What is the problem with increasing the alpha level?

Question:

What is the problem with increasing the alpha level?

Though increasing the **alpha level** leads to higher **power**, it also leads to more **false positives** (a higher probability of a **Type I Error**).

There are a lot of trade offs, so these important choices depend on the context of the study.

When Should I Know to Calculate Power?

- **Hint 1**: The problem is about a hypothesis test
  - *Ex: "Consider a scenario where at least 55% of voters must approve"*
  - *Here, we're interested in the population proportion of voters*
- **Hint 2**: The problem gives you a SPECIFIC value for the alternative hypothesis (in addition to a null value)
  - *Ex: "If 60% of U.S. adults actually think marijanua should be legal…"*
  - $H_0: p = 55\%,\ H_A: p > 55\%\ (p \overset{?}{=} 60\%)$
- **Hint 3**: You want to "test" something about your hypothesis test (e.g., if there is a sufficient sample size)
  - *Ex: "Would n = 400 be a reasonable sample size to demonstrate, with a one-sided test, that more than 55% of U.S. adults are in favor of legalization?"*

Important Code for Week 6

https://drive.google.com/file/d/1SD1xhBFjU2KxpbW_cXUNsOSrw74HCcr7/view?usp=drive_link

Questions?

Midterm Review (Weeks 1-6)

Week 2: Data Visualization

- **Grammar of graphics**: Dataset, geom, aesthetic
- **Color palettes**: Sequential, diverging, qualitative
- **Choosing the right graph**

Week 3: Data Wrangling

- **Data joins**: Left, (right), inner, full
- **Creating/modifying variables**
- **Grouping/filtering/selecting data**
- **Summary statistics**: Mean, median, SD, IQR
- **Handling missing values (NA)**
- **Interpreting code in English**

Week 4: Data Collection

- **Groups**: Sample, census, population
- **Observational study vs. experiment**
- **Two types of bias**: Sampling, nonresponse
- **Four sampling methods**: Simple, systematic, cluster, stratified

Week 5: Simulation-Based Inference

- **Parameter vs. statistic**
- **Distributions**: Sampling, bootstrap
- **Confidence intervals**: Constructing, interpreting

Week 6: Simulation-Based Hypothesis Testing

- **Hypotheses and Distributions**: Null and alternative
- **Hypothesis tests**: Constructing, interpreting
- **Pitfalls**: Power, p-value

Midterm Tips

- **PLEASE SET A TIMER FOR THE ORAL!** There should be 3 questions in 10 minutes, so try not to “ramble”
- If you haven’t already, make a **study guide**
- **Partial credit** counts (so don’t delete all your code)
- Remember to **load all relevant libraries**
- **Pace yourself**—if a question is taking too long, move on
- Sign up for **practice oral exams** (usually not 3 questions)

Debugging Practice

Let’s Practice Debugging!

https://posit.cloud/spaces/599255/content/9904652

Have a great rest of your week!