IMS 16 (ish): Inference for a single proportion

Libraries needed

```r
library(tidyverse)
library(openintro)
```

Two Sided Hypothesis Test with Proportions

Opinion on Fracking

We want to find out whether there is evidence that the proportion of people opposing fracking is different from 50%.

We will use data from a Pew Research survey conducted in the United States in November 2014, which used a random sample of $n = 1353$ people. Of these, 637 were opposed to fracking; the rest either favored it or had no definite opinion.

Hypothesis Testing

1. Frame the research question in terms of hypotheses.
2. Collect data (check conditions)
3. Model the randomness that would occur if the null hypothesis was true.
4. Analyze the data.
5. Form a conclusion.

Step 1: Frame the research question in terms of hypotheses.

How strong is the evidence that the proportion opposing fracking is different from 50%?

$$H_0: \quad p = 0.50$$
$$H_A: \quad p \neq 0.50$$

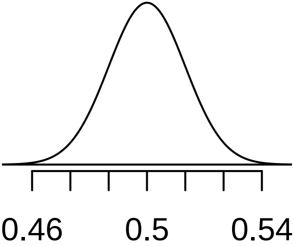
Step 2: Collect data (check conditions)

- ✓**Independence**: The people were selected by random-digit dialing of both landlines and cell phones. The sample can be treated as a random sample representative of the U.S. adult population.
- ✓**Success-Failure Condition**: Check condition under H_0.
  - $np_0 = 1353(0.5) = 676.5 \geq 10$
  - $n(1 - p_0) = 1353(1 - 0.5) = 676.5 \geq 10$

Step 3: Model the randomness that would occur if the null hypothesis was true.

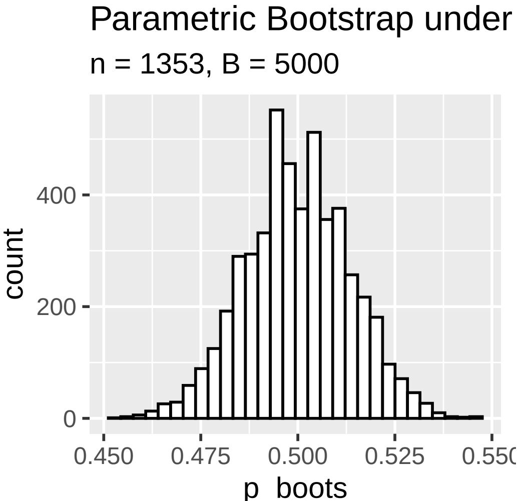
Use either a parametric bootstrap distribution, or the math approach (normal distribution approximation).

$$\begin{aligned} \hat{p} &\sim N(p_0, \sqrt{p_0(1-p_0)/n}) \\ &= N(0.5, 0.0136) \end{aligned}$$



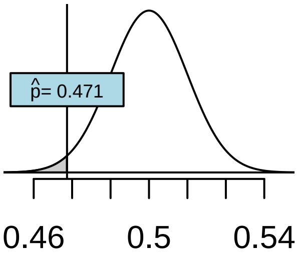
Step 3: Model the randomness that would occur if the null hypothesis was true.

Use either a parametric bootstrap distribution, or the math approach (normal distribution approximation).



Step 4: Analyze the data.

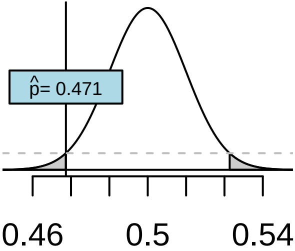
- $\hat{p} = 637/1353 \approx 0.471$
- Recall our one-sided hypothesis p-value:



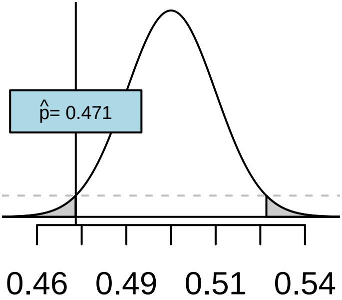
- The one-sided p-value is approximately 0.0165, but we need to consider *all* extreme values.

Step 4: Analyze the data.

- Recall the definition of the p-value: *the probability of observing our statistic or something more extreme* under H_0.
- Two-sided alternative $H_A : p \neq 0.5$ has a the two-tail p-value.
- Any value below the dashed line is considered *more extreme*. Since the normal distribution is symmetric, this means we multiply the tail area by 2.



Step 4: Analyze the data.

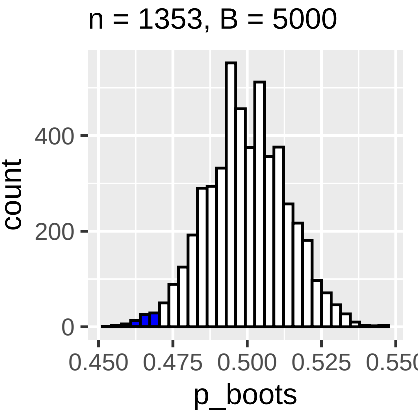



```
2*pnorm(.471, mean = .5, sd = 0.0136)
```

```
[1] 0.03297784
```

Step 4: Analyze the data.

- For the parametric bootstrap it is a similar process.
- Calculate the p-value as if it was a one-tail test.
- Multiply the one-sided p-value by 2.
- The blue tail area is 0.0174, so the p-value is 0.0348



Step 5: Form a conclusion.

- What did you check?
- Are you in favor or against H_0?
- State the conclusion in context in terms of H_0.

Check	Conclusion
p-value $\leq \alpha$	Against H_0
p-value $> \alpha$	Favor H_0

Step 5: Form a conclusion.

Our p-value was 0.33 which is larger than $\alpha = 0.05$. Therefore we conclude in favor in H_0. It is possible that the proportion of adults in the USA who oppose fracking is 50%.

Notes on Hypotheses for Single Proportions

- P-values can be used for all hypothesis tests.
- CIs can be used ONLY for two-sided hypothesis tests.
- Technically, the parametric bootstrap is still valid even if the success-failure condition is *not* met. However, CIs for proportions and the mathematical model still need this condition.
- Most of the time we are interested in a two-sided hypothesis test.

Practice Problem

[IMS 16.18 Adjusted] Elderly drivers. *The Marist Poll* published a report stating that from a random sample of 1018 adults nationally, 66% think licensed drivers should be required to retake their road test once they reach 65 years of age.

Suppose you saw another article stating that 65% adults think licensed drivers should be required to retake their road test once they reach 65 years of age. This article had no supporting data. Construct a two-sided hypothesis test to investigate this claim.

a) Describe the population parameter of interest. What is the value of the point estimate of this parameter?
b) Construct a two-sided hypothesis test to investigate the claim.
c) Check if the conditions are met for a hypotheses test.
d) Calculate the p-value using the mathematical model.
e) Make a conclusion using $\alpha = 0.10$.
f) Without doing any calculations, describe what would happen to the p-value if we had a larger sample but $\hat{p}$ did not change.

Practice Problem

Vegetarian college students. Suppose that a recent newspaper claimed that 1 in 10 college students are vegetarians. You want to test if this number is correct.

a) What are the null and alternative hypotheses to test the newspapers claim.
b) You decide to collect your own random sample of 200 students and observe that 16 are vegetarians. Check the conditions for using a mathematical model.
c) Compute the standard error using the mathematical model.
d) Draw the sampling distribution under H_0, and indicate the point estimate $(\hat{p})$ with a vertical line.
e) What code would you use in R to find the p-value using the mathematical model?