Lab 03 – Voter Turnout

Mrs. Lovett

2025-01-28

The goal of this lab is to demonstrate understanding of material from Chapters 11-13. In particular, we will do an inference on a single population proportion and

- Critique data-based claims
- Compute and interpret confidence intervals.
- Articulate the basic steps in a hypothesis test.

1 Introduction

In a survey conducted by Survey USA between September 30, 2023 and October 3, 2023, 2759 registered voters from all 50 US states were asked:

> America will hold an election for President of the United States next November. Not everyone makes the time to vote in every election. Which best describes you? Are you certain to vote? Will you probably vote? Are the chances you will vote about 50/50? Or will you probably not vote?

Based on these responses, we can ask whether we expect voter turnout to be greater than it was in the 2020 presidental election. That year, 66% of the voting-eligible population turned out for the presidential election (the highest rate for any national election since 1900).

Packages and Data

For this analysis, we'll use the **tidyverse** and **tidymodels** packages.

```
library(tidyverse)
library(tidymodels)
```

The data from this survey is in the file `voting-survey.csv` saved in your posit.cloud workspace.

```
voting_survey <- read_csv("voting-survey.csv")
```

2 Preliminary Analysis

> Task 1
>
> Make a bar chart to visualize the distribution of responses. Then calculate the proportion of respondents who are certain to vote in the next presidential election.

```
# insert code for bar chart here
```

```
# insert code to get the information that you need to calculate the proportion,
# or else do the calcuation outright.
# You can see an example of how to do this in AE 02.
```

> Task 1 Solution
>
> Report on your results for Task 1 here.

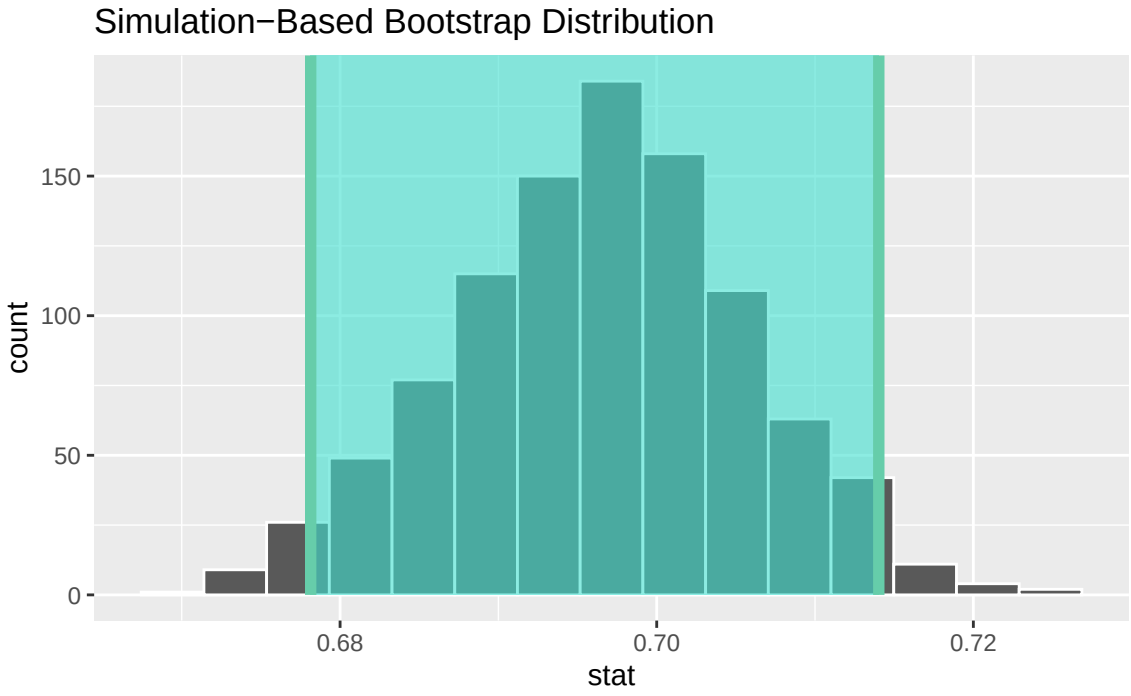
3 Estimation with Confidence Interval

Our goal in this section is to estimate the true proportion of registered US voters who are **certain to vote** in the next presidential election, based on these data.

> Task 2
>
> What is the parameter of interest?

> Task 2 Solution
>
> Enter your Task n answer here.

We can use the bootstrap technique of Chapter 12 (and which we used in AE 07) to estimate a 95% confidence interval. Below is a histogram of the sample distribution, along with the confidence interval shaded in.



Task 3

Suppose the bounds of this confidence interval are L = 0.678 and U = 0.714. Your friend offers the following interpretation:

> 95% of the time, the true proportion of registered US voters who are certain to vote in the next presidential election is between L and U.

Comment on this interpretation. Is it correct? If not, how would you fix it?

Task 3 Solution

Enter your response to Task 3 here.

4 Hypothesis Testing

A newspaper claims that based on this study, the number of voters certain to vote in the 2024 election will be greater than it was in 2020. Do these data provide convincing evidence for this claim?

Task 4

What are the null and alternative hypotheses?

Task 4 Response

Enter your response to Task 4 here.

Task 5

Before we go any further, we should decide on a discernment level. What discernment level will you use and how did you decide? In your response, you should say something about Type I and Type II errors.

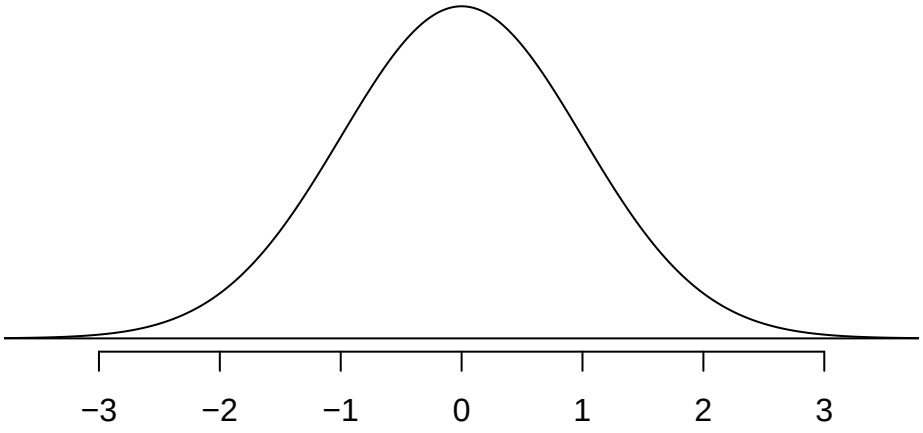
Task 5 Response

Enter your response to Task 5 here.

Task 6

Because the conditions of the CLT are met, we can use a normal distribution to find the p-value instead of using a simulation. Assuming the null hypothesis is true, where would this normal distribution be centered? Replace the 0 in the code chunk below with the correct value, and replace the 1 with 0.01 which is the appropriate standard error (SE) for this problem. We will see a formula that gives us this value in Chapter 17 but we'll just assume this for now.

```
openintro:: normTail( m = 0, s = 1)
```



Task 7

Using the SE value 0.01 and appropriate the values for $\hat{p}$ and p, find the Z-score for the observed statistic.
Then find the p-value corresponding to this Z-value.

Task 7 Response

Enter your responses to Task 7 here.

Interpreting the p-value

Task 8

Suppose the p-value you found is P, and your friends are in disagreement about the interpretation about this value. One friend claims:

> The probability that 66% of all of registered US voters are certain to vote in the next presidential election is approx P.

Another friend claims:

> The probability that more than 66% of all of registered US voters are certain to vote in the next presidential election is approx P.

Who is right? Explain your reasoning.

Task 8 Response

Enter your responses to Task 8 here.

Task 9

Summarize your findings in the context of this problem.

Task 9 Response

Enter your response to Task 9 here.