

Lab 10 – MATH 240 – Computational Statistics

Ben Horner
Colgate University
Math Department
bhorner@colgate.edu

Abstract

This document provides a basic template for the 2-page labs we will complete each week. Here, briefly summarize what you did and why it might be helpful. Provide all the top-line conclusions, but avoid providing *all* the details. Results should be limited to “we show X, Y, and Z.”

Keywords: Resampling; margin of error; sampling distribution; sample proportion

1 Introduction

Gallup polls is high profile polling, analysis, and consulting organization which often tries to track the opinions of Americans. To give further detail on their methodology, Gallup polls published a document called “*How Are Polls Conducted?*” that describes how Gallup selects which people to include among other details. Towards the end of their document, they mention how using a sample of 1000 adults “*derived using careful random selection procedures*”, their results are “*highly likely*” to be accurate within a $\pm 4\%$ margin of error. However, they also claim that increasing the sample size to 2000 reduces the margin of error to within $\pm 2\%$.

To examine the veracity of these statements, we reexamine and recalculate the margin of error in their February 3-16, 2025 poll of 1004 adults from their representative sample which revealed that 39% of respondents were satisfied with the position of the United States in the world today, compared to 59% who were dissatisfied (2% had no opinion). Gallup reported the same $\pm 4\%$ margin of error.

2 Methods

Often, statisticians and quantitative researchers will report a margin of error that provides 95% confidence. We will use the same confidence interval when examining the Gallup poll data.

2.1 Basic Simulation

Before delving into the real data, we want to conduct a basic simulation study which assumes the true probability that someone is satisfied with the position of the United States in the world today is 0.39. After using `rbinom()` to generate 10k polls of the same 1004 person sample size, we can calculate the range of the middle 95% of the resulting sampling distribution. Additionally, we can approximate the margin of

error by halving that range to compare it to the 4% reported by Gallup.

We also want to address Gallup’s claims that doubling their sample size to around 2000 will resultingly half their margin of error from 4 to 2%. This can be done by simply repeating the same basic simulation and adjusting the sample size to 2008 (double the original 1004) and repeating the same calculations.

2.2 Resampling

After simulating what the sampling distribution would look like after making an assumption about the actual population proportion, we examine how resampling can be used to approximate the sampling distribution for p . After importing the data from the original Gallup survey, we resampled 10k times to reflect our simulation. The range of the middle 95% and margin of error are calculated the same as with the basic simulation.

2.3 Simulation over n and p

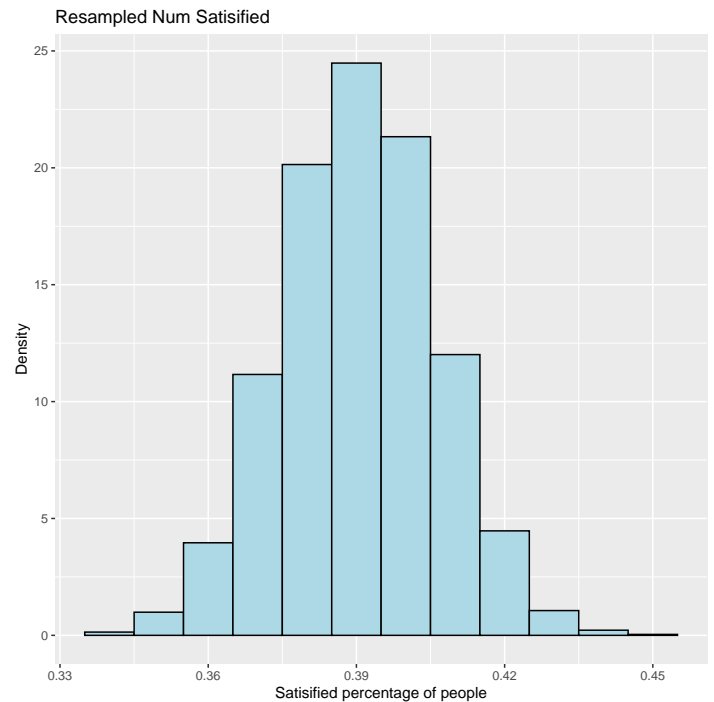
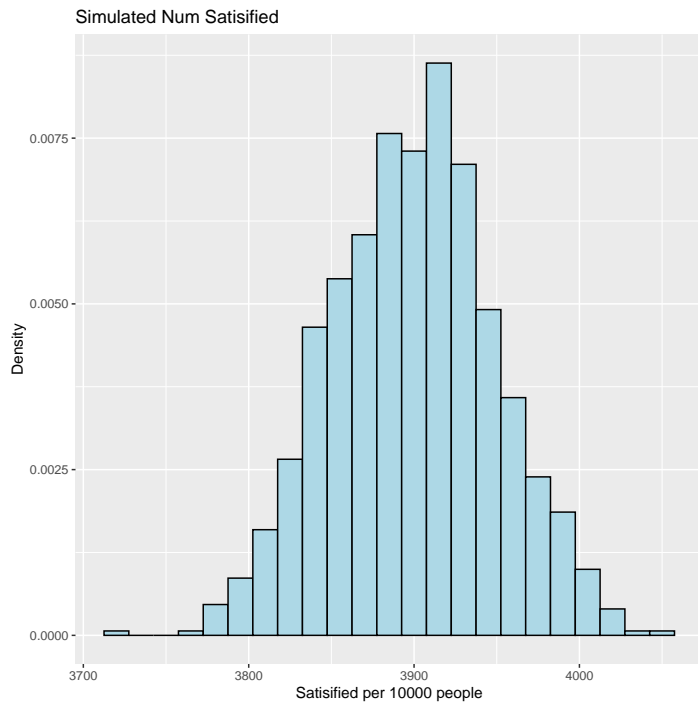
To provide better guidance for Gallup readers, simulating over sample sizes n in $\{100, 110, 130, \dots, 3000\}$ and p in $\{0.01, 0.02, \dots, 0.99\}$ can provide us with a better understanding of how the margin of error changes with polls.

2.4 Actual Margin of Error

Finally, as we have been estimating the margin of error as half the range of the middle 95%, we now compute the Wilson margin of error for the same set of n and p values. We use the `geom_raster()` function to visualize these two simulations over n and p .

3 Results

When we conduct the basic simulation study, assuming the population level values, the resulting sampling proportions follow a normal distribution.



In this case, our margin of error as a percentage is 3.04%, which is within Gallup's claims that they are accurate withing $\pm 4\%$.

4 Discussion

You should objectively evaluate the evidence you found in the data. Do not embellish or wishterpet (my made-up phase for making an interpretation you, or the researcher, wants to be true without the data *actually* supporting it). Connect your findings to the existing information you provided in the Introduction.

Finally, provide some concluding remarks that tie together the entire paper. Think of the last part of the results as abstract-like. Tell the reader what they just consumed – what's the takeaway message?

Bibliography: Note that when you add citations to your bib.bib file *and* you cite them in your document, the bibliography section will automatically populate here.

```
## 97.5%
## 188
## 97.5%
## 94
```

The resulting middle 95% of the data falls in range 192.925 with the estimated margin of error of 96.4625 people satisfied. Doubling the sample size reduces the middle range to 184.825 and the margin of error to 92.4125 people satisfied.

Using the true data and the resampling method also outputs a normally distributed sampling distribution.

```
## [1] 392
```

5 Appendix

If you have anything extra, you can add it here in the appendix. This can include images or tables that don't work well in the two-page setup, code snippets you might want to share, etc.