

Lab 10 – MATH 240 – Computational Statistics

Camilo Granada Cossio
Colgate University
Department of Mathematics
cgranadacossio@colgate.edu

Abstract

In this lab, we explored how sample size and population proportion affect the margin of error in survey sampling. We conducted simulation studies and apply the Wilson margin of error formula to compute and visualize margins of error. Our findings demonstrate that while larger sample sizes generally reduce margin of error, the true proportion play a crucial role, especially at the extremes. This analysis provide better context for interpreting polling results like those reported by Gallup.

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: What topics does the lab cover concerning class? List 3-4 key terms here, separated by semicolons.

1 Introduction

Public opinion polls often report a fixed or rounded margin of error, such as $\pm 4\%$, but this simplification hides important nuances. The margin of error depends not only on the sample size but also on the estimated proportion of success. In this lab, we investigate how both factors contribute to uncertainty in polling estimates.

We begin by using simulated polling data to visualize the sampling distribution of the sample proportion under assumed population parameters. Then, we use resampling techniques to examine variation. Finally, we compute margins of error using the Wilson margin of error formula across a range of values to generalize our findings.

Provide an overarching summary of what you’re talking about. In this section, you introduce the idea to the reader, and your goal is to pull them in. What’s the mystery you aim to solve?

You want to provide enough background to understand the context of the work. Specifically, what is the question you are addressing? If it applies, describe what information currently exists about this problem, including citations, and explain how the question you’re answering complements this work.

Provide a roadmap of the structure of the paper.

1.1 Intro Subsection

Gallup states a single margin of error for their polls. However, the sampling distribution’s spread varies depending on the true proportion being estimated. When the true proportion is close to 0 or 1, variability is constrained, and the margin of error is smaller. This lab aims to make that relationship clear through simulations and mathematical reasoning.

species	Mean	Median	SD	IQR
Adelie	3700.66	3700.00	458.57	650.00
Chinstrap	3733.09	3700.00	384.34	462.50
Gentoo	5076.02	5000.00	504.12	800.00

Table 1: This is a table.

2 Methods

We conducted three main analyses:

1. Simulation study: We assumed a true proportion of 0.39 and generated 10,000 samples for various sample sizes using `rbinom()`. We visualized the sampling distributions and calculated the middle 95% range of simulated sample proportions.
2. Resampling: We simulated an original sample of 1004 individuals with 39% satisfaction and resampled 10,000 times with replacement to estimate sampling variability.
3. Wilson margin of error formula: We computed the margin of error using the Wilson formula across values of sample size (100 to 2000) and true proportion (0.01 to 0.99). This allowed us to create a heatmap of amrgin of error as a function of both variables.

R packages used include `tidyverse` and `ggplot2` for data manipulation and visualization.

2.1 Methods Subsection

We used the Wilson margin of error formula:

$$\text{MOE}_{\text{Wilson}} = z_{1-\alpha/2} \times \frac{\sqrt{n\hat{p}(1-\hat{p}) + \frac{z_{1-\alpha/2}^2}{4}}}{n + z_{1-\alpha/2}^2}$$

where $z_{1-\alpha/2}^2$ is the critical value from the standard normal distribution (1.96 for 95% confidenc).

3 Results

We began with a simulation assuming $p = 0.39$ and $n = 1000$, and found that the middle 95% range of sample proportions was approximately $[0.361, 0.419]$, giving a margin of error of about 2.9%. When we doubled the sample size to 2000, the margin of error shrank to about 2.1%, matching our theoretical expectations.

In the resam

3.1 Results Subsection

Subsections can be helpful for the Results section, too. This can be particularly helpful if you have different questions to answer.

4 Discussion

You should objectively evaluate the evidence you found in the data. Do not embellish or wish-terpet (my made-up phrase 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.

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.