Lab 10 – MATH 240 – Computational Statistics

Cristian Palmer
Student
Mathematics
cpalmer@colgate.edu

Abstract

In this lab we were interested in studying margin of error, specifically the accuracy and methodology of margins of error in research. For this lab, we looked at a February 2025 Gallup poll which was conducted to gauge how happy American adults are with the current state of the country. We focused on Gallup's published guidelines, stating that doubling the sample size from 1000 to 2000 would lower the margin of error from plus or minus 4% to plus or minus 2%. Subsequently, we used multiple statistical approaches including simulation, resampling, and theoretical calculations to examine how margins of error behave across different sample sizes and population proportions. Ultimately, through graphing and statistical analysis, we found that the Gallup guidelines on margin of error are accurate for the most part, although slightly conservative in certain scenarios. We also were able to find evidence suggesting that simulations can provide us with nearly identical results to mathematical calculations.

1 Introduction

Our goal throughout this lab was to investigate the statistical principles behind polling margins of error and test the validity of Gallup's published guidelines regarding sample size and accuracy. In particular, we examined how changes in sample size and population proportions affect margin of error calculations in practice. We approached this question in multiple ways including simulations, resampling, and mathematical calculations. For each approach, we were able to graph and analyze our findings in ways which allowed us to compare our results to each other and to the Gallup results. Specifically, in this lab we graphically focused on creating histograms and raster plots to analyze and compare our results from each approach. Throughout this lab, we were able to systematically evaluate each aspect of margin of error estimation and develop a comprehensive understanding of how simulation compares to resampling compares to mathematical calculation. We also discovered when Gallup's guidelines are most applicable and when they might overstate the uncertainty in polling results.

2 Methods

Our first task in this lab was to conduct a basic simulation study and compare the margin of error of our study to the Gallup margin of error. Assuming that the true probability that someone is satisfied with the position of the United States in the world today is 0.39, we were tasked with using `rbinom()` to generate 10k polls, each with a sample size of 1004 which is the sample size that Gallup used. Subsequently, we plotted a histogram of the resulting sample proportions with a superimposed density to provide a visual approximation of the sampling distribution for the sample proportion. To calculate the margin of error, we used the `quantile()` function to calculate one half of the middle range of the middle 95 % of the data. After finishing our first simulation, we repeated the process, doubling the sample size to 2008, twice as much as the original Gallup poll.

After completing our simulations, our next task was to perform resampling on the actual Gallup data specified to us in the lab's directions. First, we had to create a tibble of the data based on what the lab's directions specify. Once we had the data, we used the `sample()` function, specifying `replace = T`. Specifying `replace = T` is what makes this a resampling. It is important to note that to perform this resampling, we had to keep the sample size the same as the data, which was 1004. Same to how we graphed the simulations, we subsequently plotted a histogram of the resulting resampling proportions with a superimposed density.

Next, we completed another simulation, similar to before except this time the simulation was over n and p. Specifically, we performed 10000 simulations for n in {100, 110, 120, ..., 3000} and p in {0.01, 0.02, ..., 0.99}. For each iteration, we stored half of the range between the 2.5th and 97.5th percentiles which we calculated the same was as we did in our previous simulations. To summarize these results, we created a raster plot with `geom_raster()`, comparing the population proportion to the sample size.

Finally, our last task was to perform the Wilson margin of error calculation to determine what the true margin of error is for the simulation which we performed in the previous step. In our lab directions the algebra was all done out for us, leaving us to compute only the final part of the equation. By breaking up the equation into a numerator and denominator, we were able to write calculations for each part in terms of n, p, and z. Both n, p are the sample size and proportions which we are varying. On the other hand, z is the z-score corresponding to 95% confidence, which we calculated using `qnorm(0.975)`, yielding a value of approximately 1.96. We ran the simulation inputting the various n and p values into the equation. As we did for the resampling data collection, we also stored half of the range between the 2.5th and 97.5th percentiles and graphed the results raster plot comparing the population proportion to the sample size.

3 Results

Pictured below is the histogram showing the simulation we ran with a sample size of 1004.

Data Simulation (Sample Size = 1004)
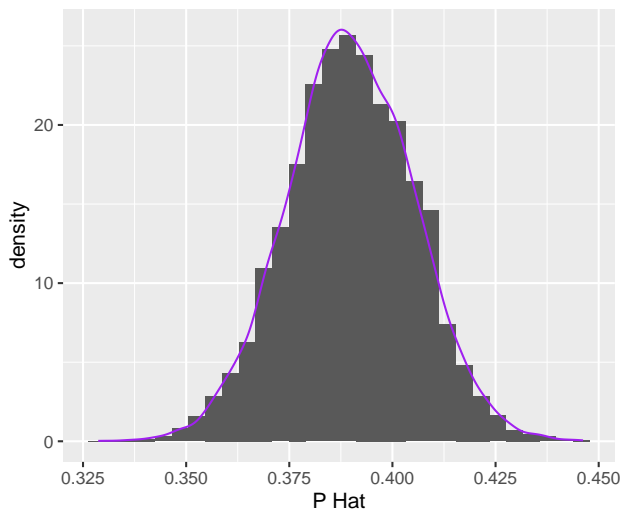
This histogram is unimodal and is centered around p-hat=0.39. The margin of error comes out to about 3%, which is actually slightly less than Gallup's margin of error of 4&. Since our margin of error came out to be a little less than Gallup's, it is possible that they were slightly conservative when reporting their margin of error for their actual sample size.

Now, we can take a look at the histogram showing the sample size of 2008, twice the original size.

Data Simulation (Sample Size = 2008)
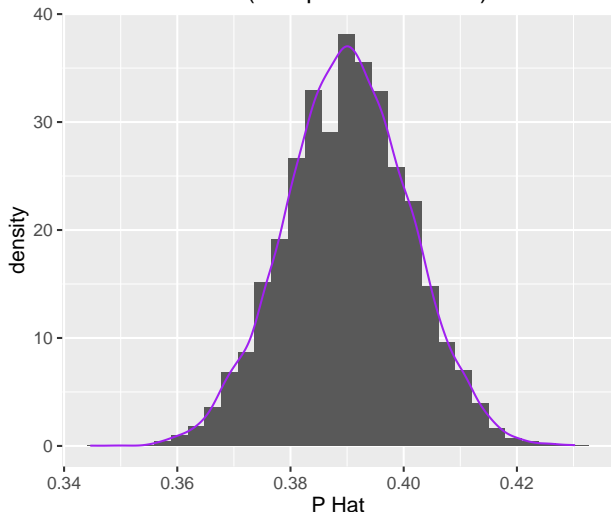
This histogram looks extremely similiar to the one we just saw. It is also unimodal, and is also centered around p-hat=0.39. For this sample size, the margin of error comes out to just around 2%, which is what Gallup predicted it would be if they were able to have double the sample size.

Unsurprisingly, the resampling histogram seen below looks extremely similiar to the previous two graphs.

Resampling Data
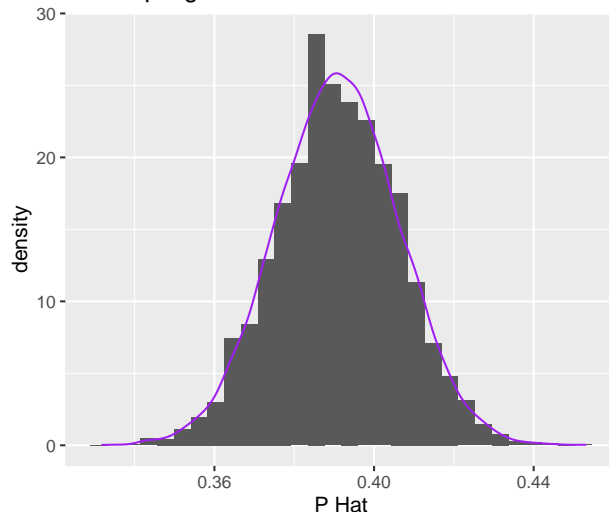
Just like the previous histograms, this one is also unimodal and centered around p-hat=0.39. Similarily to our first simulations graph, the margin of error found by resampling is just about 3%, which as we saw is slightly lower than Gallup's reported 4%.

Finally, we have the two raster plots which have been combined using `patchwork` (Pedersen, 2024), and placed in the Appendix so that the nuanced differences between the plots can be more easily recognized. Graphically, these two plots appear to be nearly identical, which is because they are in fact nearly identical. For each plot, the margin of error changes in exactly the same was as the sample size and population proportion's each increase. Together, these plots show how running a simulation enough times with a large enough sample size provides us with nearly identical results to the true values found through mathematical calculations. The only difference in the plots is the pixelation which can be seen around the base of the simulation graph. The pixelation is a visual artifact of limited data in the simulation. The math-based plot, however, is built from our theoretical model so it is completely smooth.

4 Discussion

This lab explored margin of error in polling data using simulations, resampling, and mathematical calculations. Our findings largely confirmed Gallup's guideline that doubling the sample size from 1000 to 2000 reduces the margin of error from 4% to 2%, though our simulation with a sample size of 1004 showed a slightly lower margin of error when compared to Gallup's. Furthermore, our nearly identical side by side raster plots show that when a simulation is run enough times with a large enough sample size, the margin of error it produces is virtually just as accurate as mathematically calculating the true margin of error. These findings back up and reinforce our understanding of the central limit theorem, which tells us that as sample sizes increase, the sampling distribution of the sample proportion becomes approximately normal, allowing both simulation and theoretical methods to converge on the same result.

References

Pedersen, T. L. (2024). *patchwork: The Composer of Plots.* R package version 1.3.0.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grole-

mund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
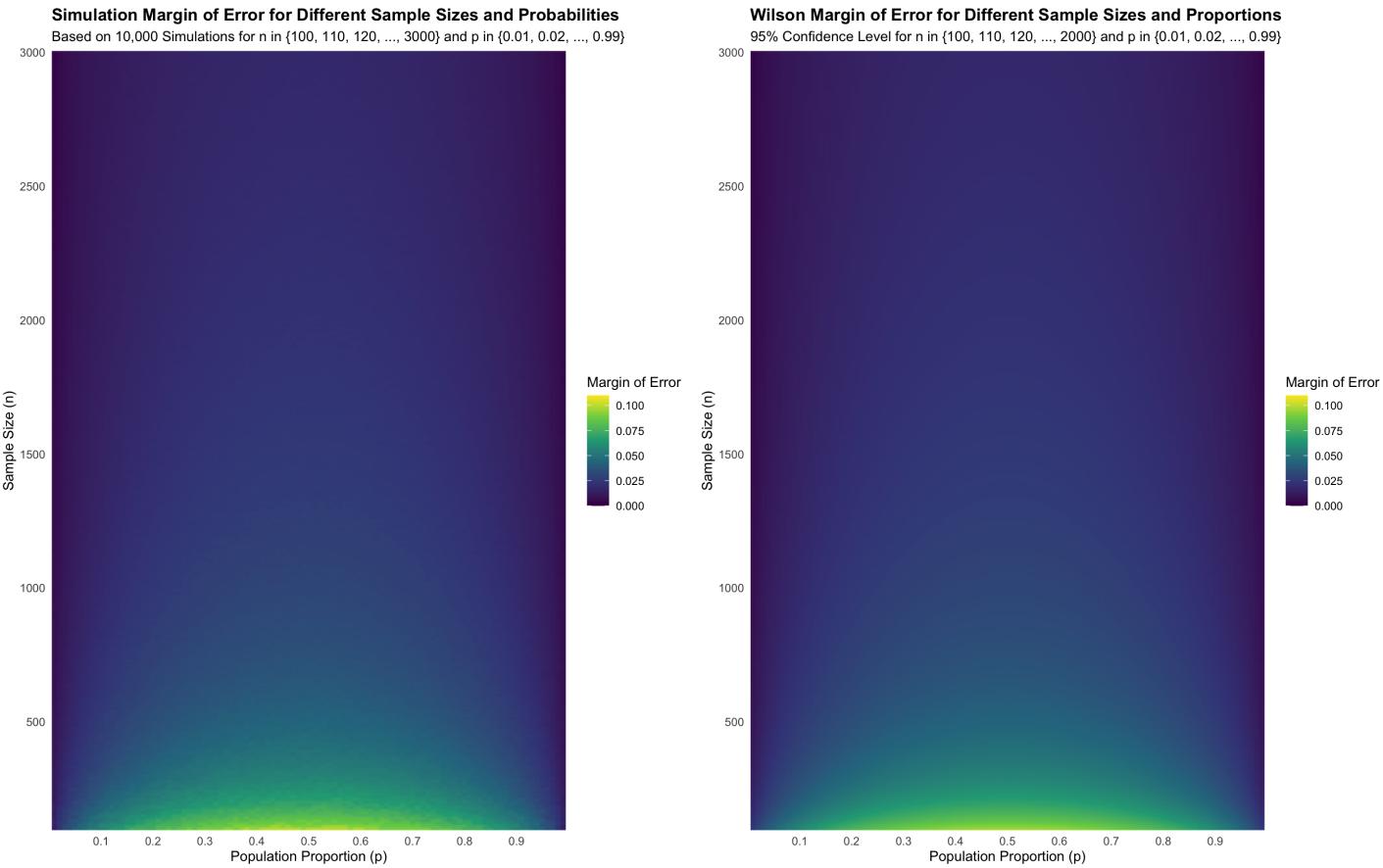
5   Appendix

Figure 1: Side by Side Raster Plots Comparing Margin of Error Estimates: Simulation vs. Wilson Formula