Lab 10 – MATH 240 – Computational Statistics

Jack Schaeffer
Professor Cipolli
Math 240
jschaeffer@colgate.edu

April 8, 2025

Abstract

Focused on analyzing effects on the margin of error, we calculated differing approximate margins of errors based on Gallup polls. We first determined how this value changes based on sample size before comparing the effect within a binomial distribution in comparison to sample size and p.

Keywords: Binomial distribution, margin of error, resampling

1 Introduction

Questions concerning margin of error stemmed from the Gallup polls article "*How Are Polls Conducted?*" that described the formation of polls and how statistical calculations are made. The article asserted that doubling their 1000 person sample size would serve to halve the margin of error, and my work was testing this with my own calculations. My initial work was testing margin of error after doubling sample size of a known p value before moving on to compare these values with resampling of the Gallup polls data. The final test on margin of error was comparing the margin of error of varying sample size and p values through both simulations of the binomial distribution and calculation of the Wilson Estimate.

2 Methods

My work was mainly focused on simulating or estimating data that could then be summarized with `tidyverse` and `patchwork` to obtain more useful information (Wickham et al., 2019; Pedersen, 2024). To first answer the question of doubling sample size, I ran two separate simulations of the binomial distribution under the assumption that Gallup polls produced an accurate p value $p = 0.39$. Although I could not use resampling to test a larger sample size than the given 1004, resampling was an effective way to compare the simulated margin of error to ensure that the given p value is a reasonable assumption.

I tested the margin of error over numerous combinations of sample size and p value in the same way as my initial simulations, but scaled to include far more combinations of p and n values. I used a similar combination fo p and n values to test the margin of error calculated through the Wilson Estimate, so that I could compare the effect of sample size and p value on margin of error through both methods.

3 Results

To compare the margin of error of different sample sizes, I constructed Figure 1, which demonstrates that a higher sample size reduces the range of estimates p values within my simulations. We can see this in Figure 1 through the noticeably higher kurtosis in the plot with a sample size of 2008. Figure 1 contains data made with the assumption that $p = .39$ and simulating this data for different sample sizes. Calculation on the data reveals that a sample size of 1004 leads to a 3.0% margin of error while a sample size of 2008 has a 2.1% margin of error.

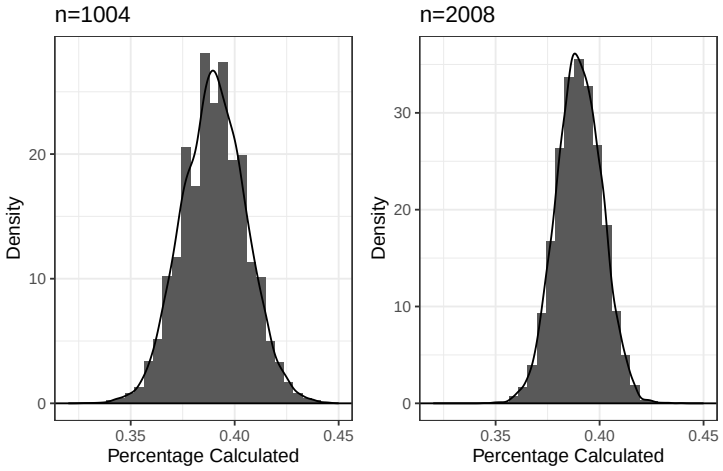

Figure 1: A comparison of simulations for different sample sizes

Calculating the same values using resampling of a sample size $n = 1004$ yielded similar results in Figure 2. When resampling, I calculated a 2.9% margin of error, nearly the same as the 3.0% under the assumption that $p = .39$ is correct.

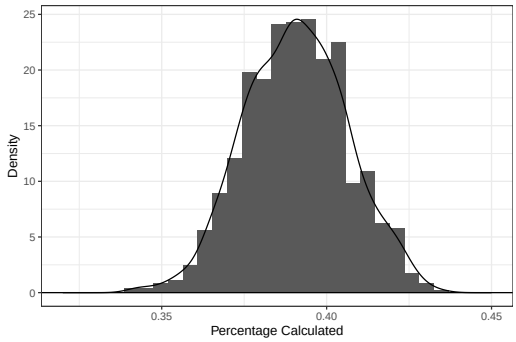

Figure 2: Calculated p values with resampling

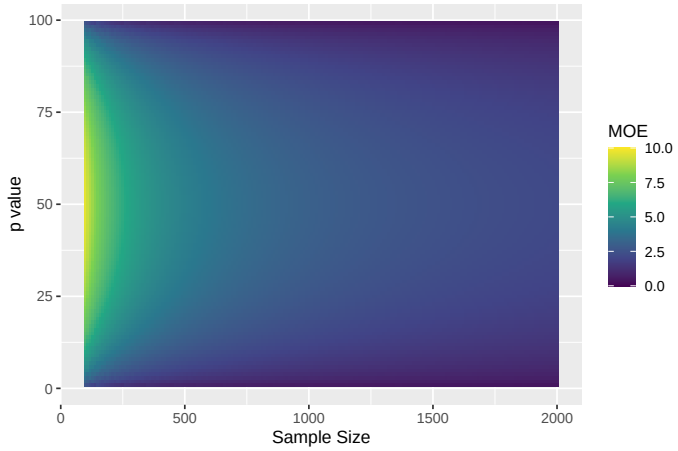

Figure 4: Margin of error calculated through Wilson Estimate

Figures 3 and 4 showcase how margin of error changes dependent on samples size and p. Figure 3 calculates MOE using simulations of the binomial distribution as was done for Figure 1. Figure 4 calculates MOE through the Wilson Estimate instead. Although calculated through different methods, the results of both plots appear similar.

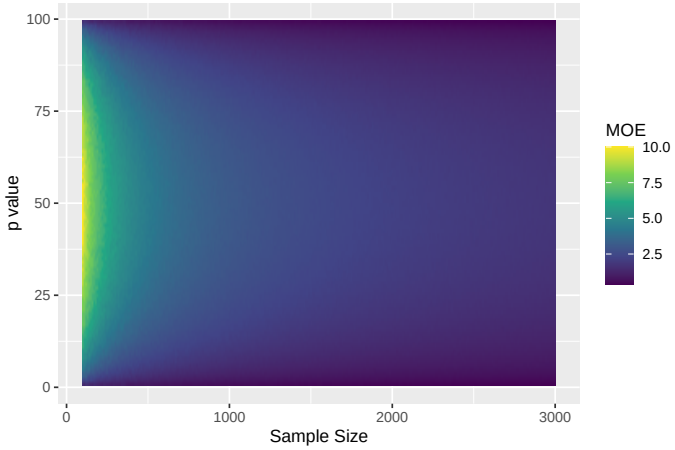

Figure 3: Simulated margin of error for binomial distribution

4 Discussion

Figures 1-4 demonstrate how margin of error changes within the binomial distribution. Importantly, Figure 1 indicates that Gallup polls was incorrect in their assertion that doubling their sample size would halve the margin of error, instead decreasing by approximately 50%. Figure 2 reflects this contrast as well by suggesting through resampling that the margin of error of Gallup polls' data is 3% instead of 4%.

Figures 3 and 4 give us an even better idea about how margin of error is affected by sample size and p value. Changes to sample size appear to have the greatest effect when the sample size is low, but increasing sample sizes past 1000 do not appear to have noticeable changes to margin of error in either figure. This reaffirms what we saw in Figure 1 that doubling sample size would not serve to halve margin of error. Further, Figures 3 and 4 demonstrate that margin of error changes more noticeably at p values close to 0.5, suggesting that there would be a relatively large change to margin of error for the given p value of 0.39. This suggests that the 1% decrease to margin of error seen in Figure 1 is large compared to how margin of error would change for a p value further from 0.5.

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., Grolemund, 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.