Lab 10 – MATH 240 – Computational Statistics

Jeremy Artiga
Colgate University
jartiga@colgate.edu

Abstract

In this lab, we will simulate data based on statistics provided by Gallup Polls concerning the satisfaction of the US politically to figure out how exactly Gallup derived their confidence interval and margin of errors while I've. This will involve the use of re-sampling and `R` simulations to illustrate how the confidence interval and margin of error changes with differences in probability and sample size.

1 Introduction

Statistical analysis of proportions, such as those found in public opinion polls, relies heavily on understanding the margin of error and confidence intervals. The margin of error reflects the uncertainty in a sample estimate, and is influenced by numerous factors. In this lab, we will explore how changes in sample size and probability affect the margin of error in statistical analyses of proportions and how these statistics are modeled in real life using data from Gallup's (Gallup, 2025) polls on the satisfaction of the current US well-being with emphasis on the use of simulations and re-sampling with `R`'s tidyverse (Wickham et al., 2019)!

2 Methodology

2.1 Simulations

Assuming the probability calculated by the Gallup polls to be true (concerning the satisfactions of the US political state), we will start by conducting 10000 simulations of the Gallup polls under the assumptions that $p = 0.39$ and $n = 1004$ (statistics derived from the Gallup study). We will take the range of this data and calculate a margin of error. We will also repeat this with $n = 2008$ to see how these statistics change.

As seen in Figure 1, the distribution of the proportions calculated in the simulations appear to be normal, with a peak at the probability 0.39. This is also supported by the calculated range (middle 95%) of the data, which is at 0.42-0.361. The margin of error calculated for this simulations sits at $\approx 3\%$, which is one lower than the margin of error calculated by the Gallup.
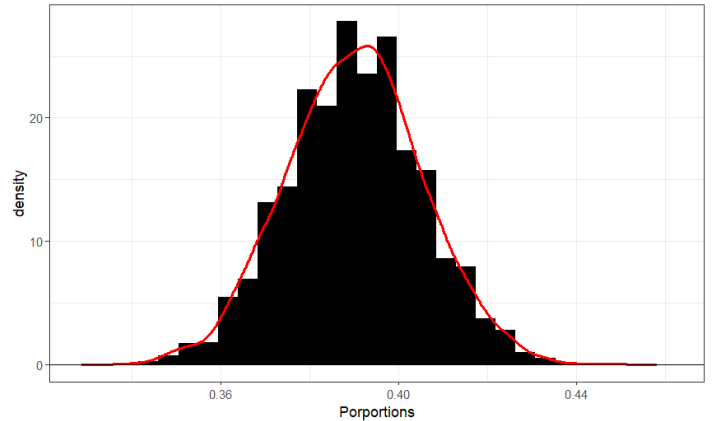

Figure 1: Distribution of 10000 Poll Simulations with Sample Size $n = 1004$, and Probability $p = 0.39$

As seen in Figure 2, when we double the sample size under the same assumptions (i.e. probability), we get a much cleaner normal distribution with a peak also centered at 0.39. The range in this data spans from 0.411 to 0.369, which is shorter than the $n = 1004$ distribution. Since our range is shorter, our margin of error was also reduced down to $\approx 2\%$, which matches what Gallup approximated.
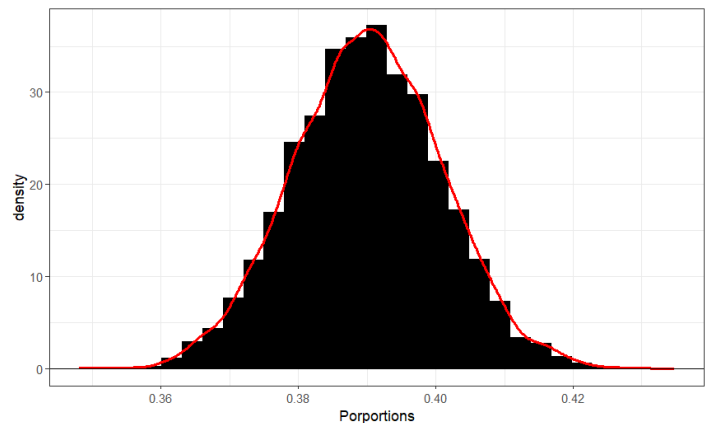

Figure 2

As we can see with simulations, by increasing the sample size, we are able to obtain a much narrower range with a lower margin of error for given data. This essentially illustrates the power of the sample size effect!

2.2 Resampling

To further validate our margin of error estimate, we will perform re-sampling with the same data used by the Gallup poll in their study. Specifically, we took 1000 re-samples (with replacement) from the original set of responses, each of size $n = 1004$. Focusing on only the responses with a "yes" output, we performed the same range and margin of erro calculations as we did in our simulations . This approach mirrors how poll results might vary in repeated sampling under the same conditions.

As illustrated in Figure 3, the resulting distribution of proportions is approximately normal, with a peak centered near 0.39. The 95% interval of these simulated proportions range from approximately 0.36 to 0.42. From this, we calculate a margin of error of $\approx 3\%$, which aligns closely with the margin of error reported in the original Gallup poll.
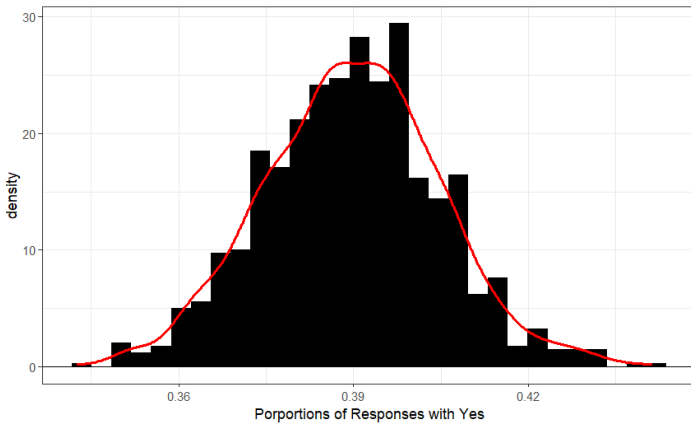

Figure 3: Distribution of Gallup Poll Respondents who Answered "Yes"

This method of re-sampling demonstrates how the margin of error originates in from sampling variability and reinforces the concept of larger spreading in sample proportions yield higher uncertainty in our estimate. The use of re-sampling thus offers a practical and intuitive approach to assessing confidence in survey data.

2.3 How Does Error Change Over Probability and Sample Size?

With the way we've seen the confidence interval (range) and margin of error change with alterations in our sample size, that begs an interesting question, how does confidence interval and margin of error change with alterations in probability and sample size?
To illustrate this, we will need to create a plot which illustrates the relationship between probability and sample size with respect to confidence interval and margin of error. Since margin of error is directly correlated to the confidence interval, we will use the margin of error to illustrate the relationship between probability and sample size. To accomplish this, we simulated various margin of errors across $p = \{0.01, 0.02, ..., 0.99\}$ and $n = \{100, 110, ..., 3000\}$ where p is defined as probability and n is defined as our sample

size. The results of our simulations our illustrated in Figure 4 below.
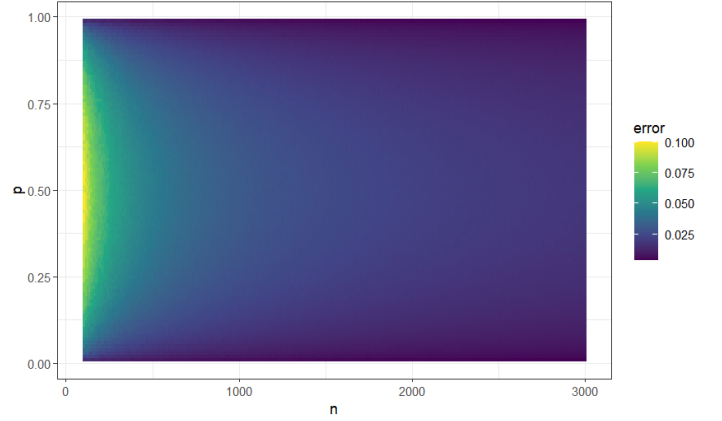

Figure 4: Simulated Margin of Error in terms of Probability and Sample Size

As seen by the raster plot in Figure 4, we can see that as n increases, our error becomes much smaller for a constant p. This means that as our sample size increases, our margin of error decreases under the assumption our probability doesn't change. In terms of p however, we see that as p approaches the mid point of the interval [0,1], our error becomes much larger (and vice-versa) for a constant n. In other words, our margin of error becomes much larger as we approach probability .50 under the assumption that sample size of constant.

This figure validates our previous findings of the sample size effect as it illustrates how increases in sample sizes with correlates with a decrease in the margin of error (i.e. the sample size effect) and decrease in variability . It also describes how probability has a hand in influencing the margin of error and confidence intervals (in the binomial distribution), as values that approach a probability of 0.50 end up maximizing the margin of error, which also implies a wider confidence interval.

2.4 The Actual Margin Of Error

Up until now, we've been using simulations and re-sampling to calculate our margin of errors, however, we are able to calculate the actual margin of errors using Wilson's margin of error formula. Using this formula yields greater accuracy in the calculations of the margin of errors for our data, if we can show that the raster plot for Wilson's margin of errors is very similar to our simulations raster plots, then our findings are validated!
By using the same process to generate the raster plots, we calculated Wilson's margin of error across $p = \{0.01, 0.02, ..., 0.99\}$ and $n = \{100, 110, ..., 3000\}$ where p is defined as probability and n is defined as our sample size. The results of our simulations our illustrated in Figure 5 below.
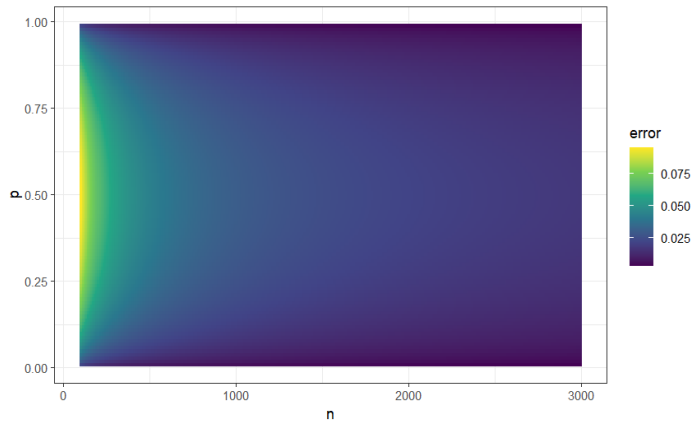
Figure 5: Wilson's Margin of Error in terms of Probability and Sample Size

As we can see, the raster plots in Figures 4 and 5 look practically identical, which reinforces our findings concerning the relationship between the confidence interval and margin of error with probability and sample size!

3 Discussion

This lab illustrated that sample size and probability play a crucial role in the accuracy of survey estimates. Increasing the sample size reduces the margin of error, making results more reliable. As we've seen in the simulations and use of the Gallup polls, the increase of the sample size has greatly reduced the margin of error and variance for the plotted distribution, which also allowed us to validate the Gallup's poll results.

Additionally, we also found that the probability of an outcome affects the margin of error. When the probability is around 0.5, the margin of error is larger, meaning results are more uncertain. This should be considered when designing surveys. Using Wilson's margin of error formula also helps ensure accuracy, validating the results of our simulations. These findings highlight the importance of sample size and probability in creating more reliable surveys and improving decision-making.

References

Gallup, I. (2025). Gallup poll: U.S. satisfaction with the political situation. Accessed: 2025-04-10.

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.