

Lab 10 – MATH 240 – Computational Statistics

Jackson Colby
Colgate University
Mathematics
jcolby@colgate.edu

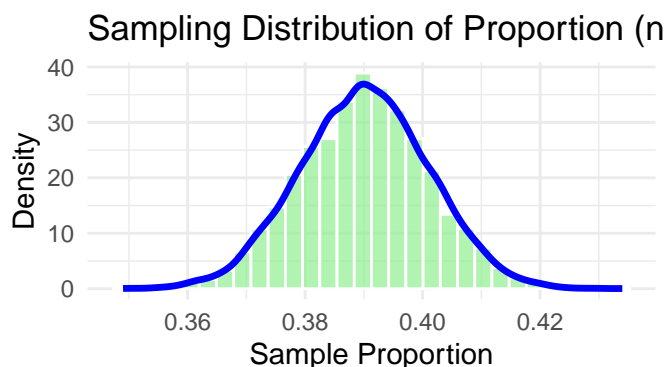
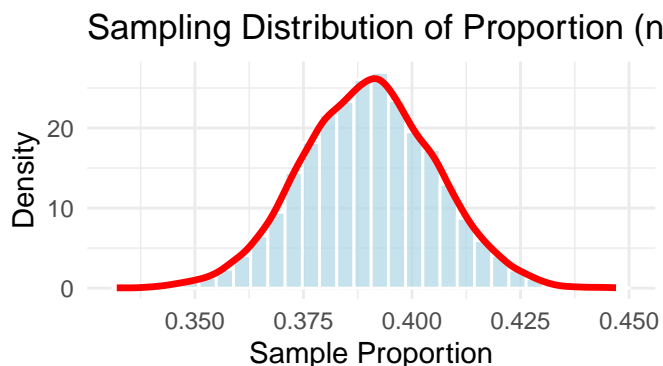
1 Introduction

A polling company Gallup conducted a study on whether or not 1004 adults were satisfied or not with the position of the United States. 39% of respondents said they were satisfied and Gallup reported a $\pm 4\%$ margin of error. Gallup mentions that by increasing the poll size to 2000, the results would be accurate within $\pm 2\%$. This claim by Gallup is what leads us to the question for the lab, testing the claim of Gallup that increasing the sample size by double would decrease the margin of error by a half.

Utilizing packages such as tidyverse (Wickham et al., 2019) and ggplot2 (Wickham, 2016), the following steps were conducted to test whether Gallup's method of calculating margin of error for their polls is correct. A basic simulation of two sample sizes $n=1000$ and $n=2000$ was conducted to see if the population value of 0.39 followed the claim of Gallup in their conducting of polls. Resampling was then used to find an approximation for \hat{p} to compare the margin of errors. To further test Gallup we then simulated of different n and p , and finally calculated the actual margin of error using Wilson's Estimate.

2 Basic Simulation

This analysis compares the sampling distribution of sample proportions from 10,000 simulated polls at two different sample sizes: 1000 and 2000, assuming a true satisfaction rate of 39%.

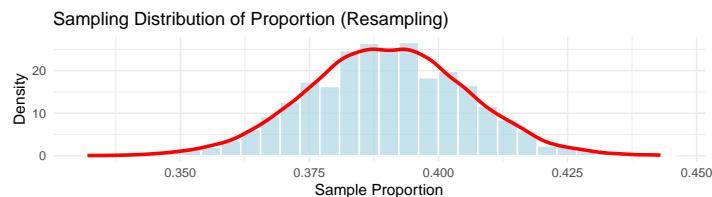


The shape of the sampling distribution for a sample size of 1000 forms an almost normal bell-shaped curve. The middle 95% of the data fell between 0.36 and 0.42, meaning it had a margin of error of 0.03, which is slightly less than the 0.04 (4%) that Gallup reported.

The shape of the sampling distribution for a sample size of 2000 also formed an almost normal bell-shaped curve. The data in this sample was a little bit more centralized around the assumed population mean. The middle 95% of the data fell between 0.369 and 0.411, meaning it had a margin of error of 0.021, which is very close to the 0.02 (2%) that Gallup reported.

3 Resampling

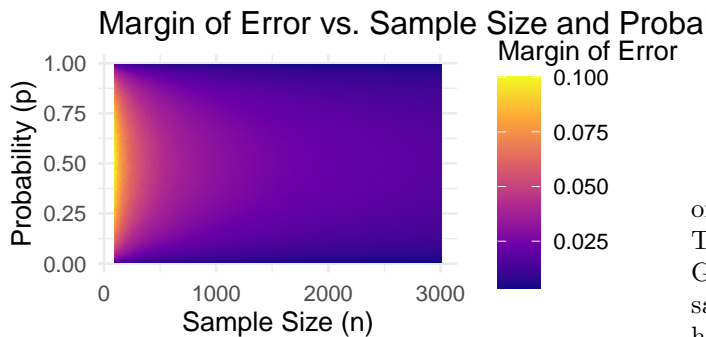
This analysis uses resampling to simulate the distribution of the sample proportion from a Gallup survey where 39% of respondents are satisfied.



The resampling approach produced a normal bell shaped curve like the samples above, however the data fits the curve a little better than the basic simulation. The middle 95% of the data fell between 0.36 and 0.42 leading to a margin of error of approximately 0.03, which is less than the 0.04 (4%) that Gallup reported. This number is approximately the same number calculated from the data in the basic simulation for sample size of 1000.

4 Simulation over n and p

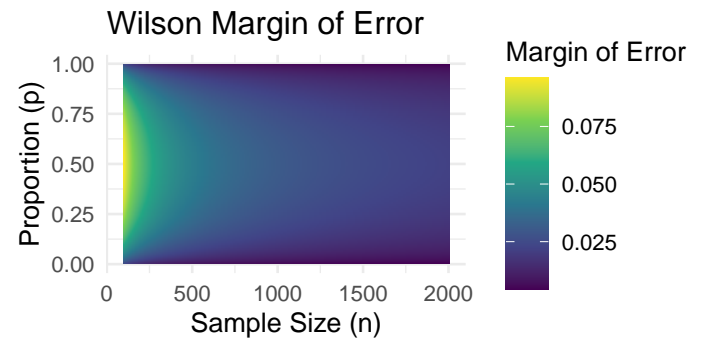
This plot uses simulation to estimate the margin of error for different combinations of sample size (n) and true proportion (p), based on 10,000 binomial trials per combination.



As seen in the figure above, Gallup mentioning the sample size as the sole factor in changing the margin of error is not completely true. The sample size does play a role in determining the margin of error, but the proportion p also plays a role. When p is close to the extremes (0 or 1) the margin of error is smaller for the same sample size. This is due to not being able to expand beyond the parameter space of $[0,1]$.

5 Actual Margin of Error

This plot visualizes the margin of error for different combinations of sample sizes (n) and proportions (p), using the Wilson Margin of Error.



Similarly to the simulation over n and p , the actual margin of error using the Wilson Estimate paints a similar picture. The plots follow a similar pattern, ultimately showing that Gallup's story is not as straightforward as they mention. The sample size does decrease the margin of error as n increases, however p also plays a factor in the margin of error. Similarly to before the values of p closer to the extremes of 0 and 1 have lower margin of errors since the values can not go outside the parameter space of $[0,1]$.

6 Results

Wilson's claim that doubling the sample size from 1000 to 2000 would decrease the margin of error from $\pm 4\%$ to $\pm 2\%$ is not as simple as they claim. This is primarily shown from the Simulation over n and p and the Actual Margin of Error Calculation utilizing Wilson's Estimate. The sample size is not the only factor in determining the margin of error, the proportion p also playing a role. This suggests that Gallup's claim is not true, or at a minimum it is not true for all values of p .

References

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
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.