Lab 10 – MATH 240 – Computational Statistics

Harrison Wolfe
Colgate University
Math Department
hwolfe@colgate.edu

4/10/2025

Abstract

Confidence intervals are one of the most important statistical tools when trying to present data. They provide a range of values for which a certain percentage (typically around 95) of confidence intervals contain the true proportion or mean. There are several ways to create said intervals and some will be discussed below.

Keywords: Confidence Intervals; Resampling; Point Estimates

1 Introduction

In this lab we will be discussing different ways of calcualting margin of error and therefore creating a confidence interval. For the first two methods, a Basic Simulation and Resampling, we will be calculating the margin of error given an example of a poll. This poll surveyed 1004 American's on the current state of the nation. They either said they were satisfied, dissatisfied, or that they did not have an opinion. In this survey it was found that 39 percent of participants were satisfied with the country however this number is much more useful when associated with a confidence interval to know how precise the estimate is. The other 2 methods, a more complicated simulation and the formula for margin of error, were done more generally for many different n and p values.

2 Basic Simulation

One of the ways to conduct a confidence interval under the context of the polls example is using a basic simulation. In this simulation we ran 10,000 iterations of polling 1004 people with the condition that the true parameter was 0.39. Using this data we could create a confidence interval using the point estimate (mean of all proportions) from each sample giving us a margin of error of around 0.02988. If we double the sample size we were able to see that the margin of error decreases to 0.02117. This shows that margin of error is inversely correlated with sample size. Below we see histograms of each simulation and we can see that the second histogram has less variance than the first contributing to the smaller margin of error. The data that made these plots were made using `tidyverse` and the plots themselves were made using `ggplot2` (Wickham et al., 2019) (Wickham, 2016).
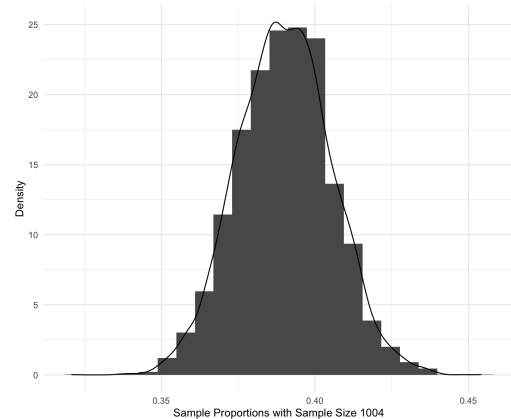

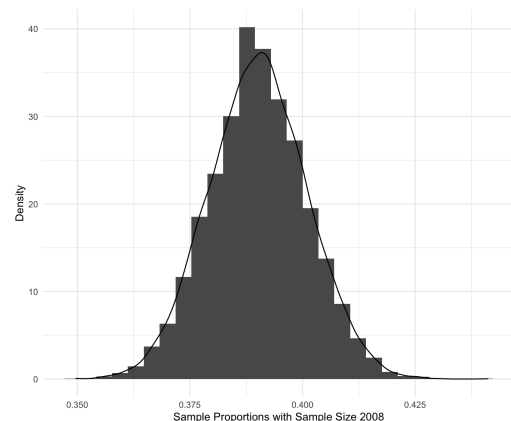Figure 1: Basic Simulation with Sample Size 1004

Figure 2: Basic Simulation with Sample Size 2008

3 Resampling

The second way we found margin of error was through resampling. Resampling is the process of taking a sample and taking random samples from the original sample with replacement and then calculaitng the statistics of interest. In this example we took the survey data regarding the satisfaction with the United States and resampled it 10,000 times. This process then gave us the ability to find a confidence interval with the middle 95 percent of the sample proportions which

in turn gave us a margin of error. This margin of error was 0.02988. In Figure 3 below we can see the variance of the distribution and the point estimate which was able to give us this margin of error found.
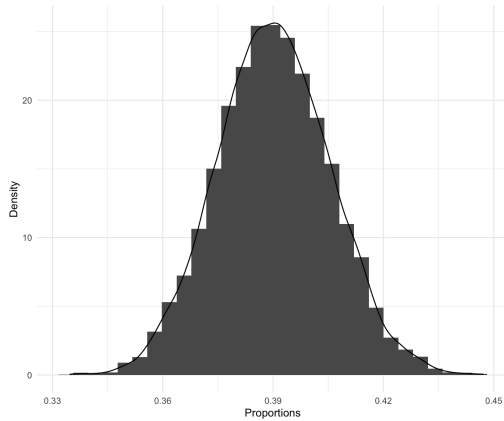

Figure 3: 10,000 Resampled Sample Proportions

4 Simulation over n and p

In this simulation we simulated values with sample sizes going from 100 to 3000 by 10 for each proportion from 0.01 to 0.99 by 0.01. Using this information we were able to calculate a margin of error for each point. We found when the sample size was 1000 and the proportion was 0.39 (the survey data) the margin of error is 0.0305 which is similar to the resampling method. We can see in the Figure 4 the way that margin of error changes with sample size and proportion. Margin of error grows when there is a low sample size and a proportion near 50 percent because that is given the most variance. However as we move in any direction away from these points we see that the margin of error decreases.
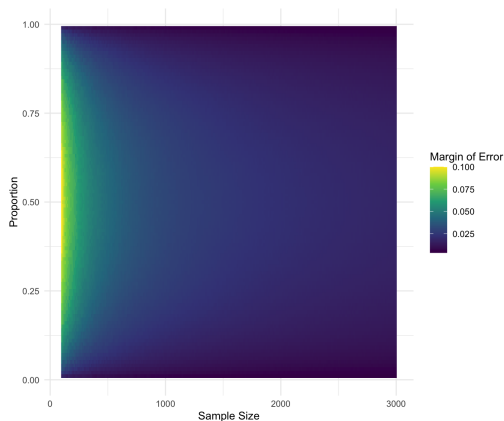

Figure 4: Margin of Error Based on Sample Proportion and Sample Size through Simulation

5 Actual Margin of Error Calculation

The final way we calculated margin of error was through a formula. This formula depends on our values for sample size, sample proportion, and the width of the confidence interval. Again for this method we calculated every margin of error for the same sample size and proportion values as the simulation. Through this we found that the margin of error for sample size 1000 and proportion 0.39 (like in the survey) was 0.03015. This value is again very similar to the resampling method and the simulation method. Again for this we plotted each point with each sample size and proportion in a heat map. Figure 5 yields similar conclusions as 4; it proves that margin of error generally is highest when sample size is low nad the sample proportion is around 0.5 and as we escape from these points in the yellow region margin of error decreases.
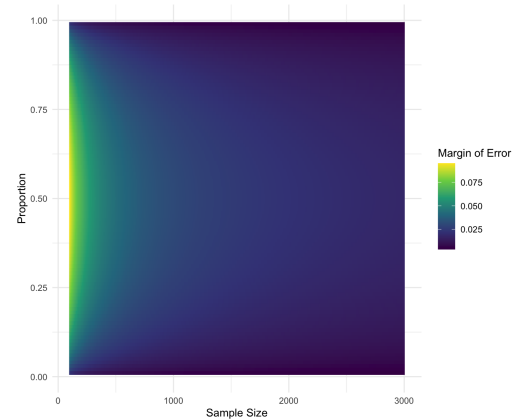

Figure 5: Margin of Error Based on Sample Proportion and Sample Size through Formula

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.