

Lab 10 – MATH 240 – Computational Statistics

Pierce Leclerc
Colgate University
Department of Mathematics
pleclerc@colgate.edu

Abstract

We examine the effect of sample size and assumed true probability on margin of error. Data from Gallup is compared to random samples, and margins of errors of both are compared as well. Resampling is performed on the Gallup data to estimate margin of error. Margin of error is plotted as a function of n and p to demonstrate the pairs which maximize it, and these conclusions are verified by comparison to the Wilson Estimates.

Keywords: Margins of Error, Resampling, Simulation

1 Introduction

In Lab 10, we look to examine the effect that sample size and actual population proportion have on margin of error. Gallup reports that a sample size of 1000 national adults corresponds to accuracy with a margin of error of $\pm 4\%$, while doubling the sample size decreases the margin of error to $\pm 2\%$. Under an assumed true probability, we conduct a sample simulation study to determine margin of error under sample sizes of 1000 and 2000 to compare with Gallup's measurements. We then perform resampling on the Gallup data and conduct the same analyses. We perform 10,000 simulations under different values of n and p to observe the effects of both on margin of error and calculate the actual margin of error for verification.

2 Simulation Studies

Figures 1 and 2 illustrate the results of 10,000 generated polls of their respective sample sizes. They both appear approximately normally distributed.

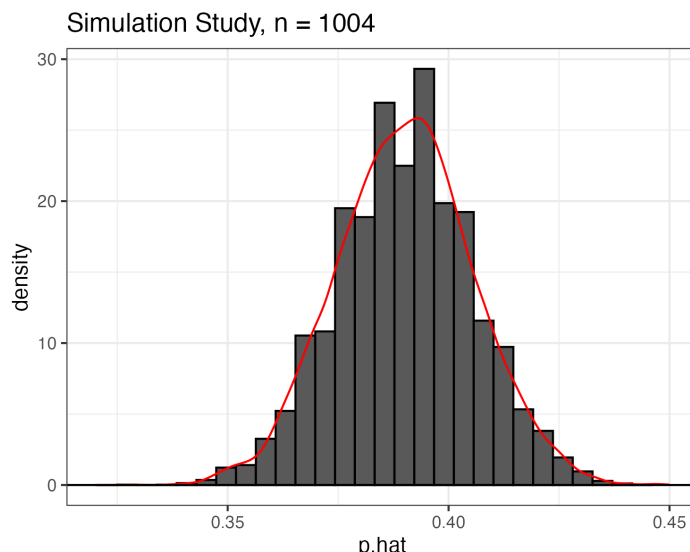


Figure 1: Simulation study with true probability 0.39 and $n = 1004$. Density superimposed in red.

With $n = 1004$, the margin of error is 0.02988048, or approximately 2.99%, slightly less than the 4% reported by Gallup.

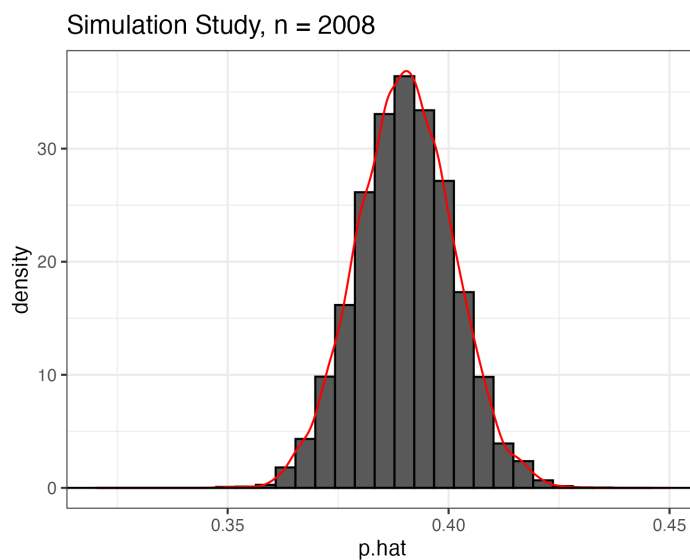


Figure 2: Simulation study with true probability 0.39 and $n = 2008$. Density superimposed in red.

With $n = 2008$, the margin of error is 0.02116534, or ap-

proximately 2.12%, similar to the estimate from Gallup of 2%.

3 Resampling

With sample data from the Gallup survey, resampling was performed and results are plotted in Figure 3. The data appears to be normally distributed once again, but with a mean below the previously assumed true probability of 0.39.

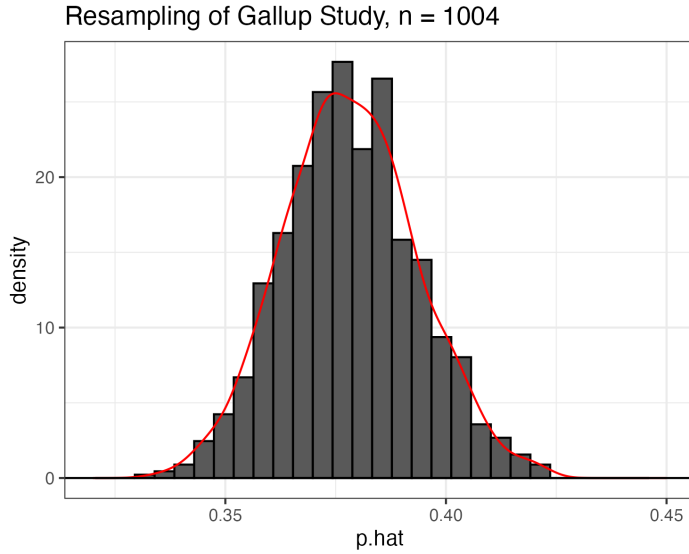


Figure 3: Histogram of resampled Gallup data ($n = 1004$) with density superimposed in red.

The margin of error in this case is 0.02889691, or approximately 2.89%, less than the reported estimate of 4% by Gallup.

4 Simulation over n and p

10,000 simulations were performed under different values of n and p and are plotted in Figure 4.

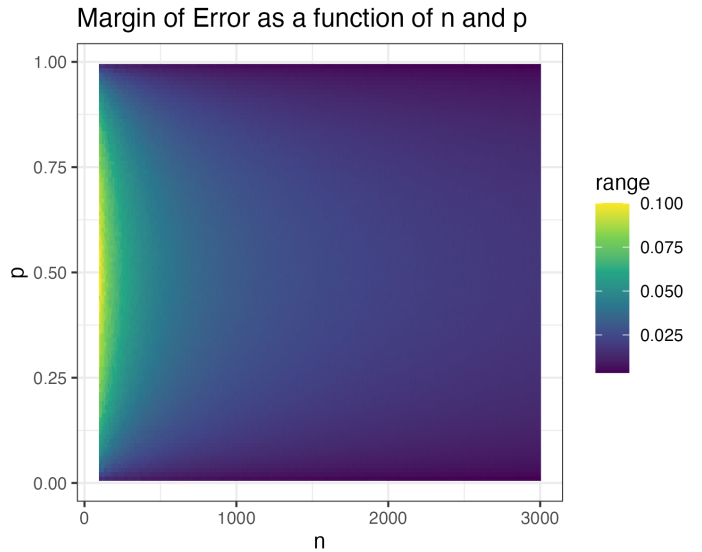


Figure 4: Raster plot of estimated margin of error as a function of n and p .

From this analysis, it is evident that margin of error is maximized when the sample size is small and the assumed true probability is near 50%. To verify this, we compute the actual margins of error via the Wilson Estimate, as shown in figure 5:

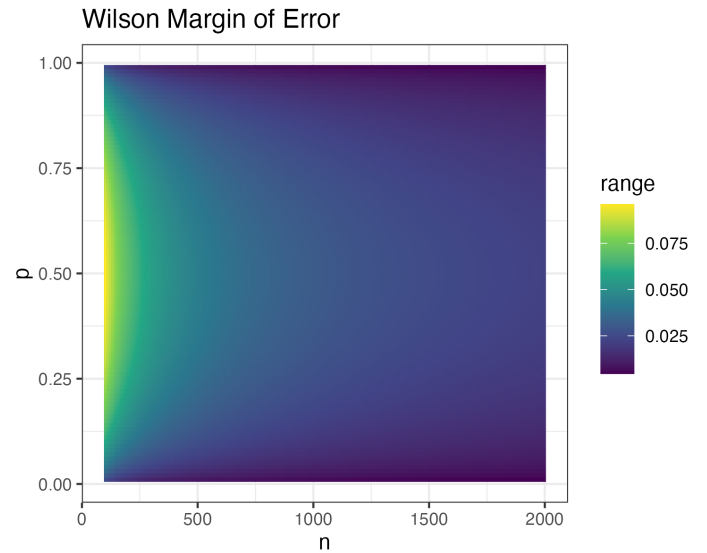


Figure 5: Wilson Margin of Error as a function of n and p .