“Plus-four” or Wilson Estimate for the Confidence Interval of a proportion

Last time, I showed you that the confidence interval for a population proportion obtained by the normal approximation to the binomial differed from the exact confidence interval calculated from the binomial itself. Here is a trick, invented by Edwin Wilson in 1927, that brings the normally approximated interval closer to the true interval.

Add four “extra” observations to the data, two successes and two failures.

\[ \hat{p} = \frac{X + 2}{n + 4} \]

our confidence interval then becomes

\[ \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n + 4}} \]

Example: compare the exam problem 2b. considering people in favor of the euro.

Sample size estimation for a desired margin of error:

Recall that the margin of error is related to the sample size

\[ m = z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]

for the standard CI or

\[ m = z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n + 4}} \]

for the Wilson estimate

We can use these equations to calculate the sample size necessary to achieve a desired margin of error.

\[ n = \left( \frac{z^*}{m} \right)^2 \]

\[ n = \left( \frac{z^*}{m} \right)^2 p^*(1 - p^*) \]

Here, \( p^* \) is some hypothesized value for the population proportion, either guessed from a preliminary study or (in the simplest case) simply set to 0.5. In this last case the required sample size becomes

\[ n = \left( \frac{z^*}{m} \right)^2 \]