Improved Margin Of Error Estimates For Proportions In Business: An Educational Example

George Arzumanyan, Rule26.com & Walnut Oak Capital, LLC, USA
Dennis Halcoussis, California State University, Northridge, USA
G. Michael Phillips, California State University, Northridge, USA

ABSTRACT

This paper presents the Agresti & Coull "Adjusted Wald" method for computing confidence intervals and margins of error for common proportion estimates. The presented method is easily implementable by business students and practitioners and provides more accurate estimates of proportions particularly in extreme samples and small sample situations. The proposed method may have particular applications to focus group analysis, industry benchmarking, and destructive testing sampling. The paper discusses a computational strategy and several comparison examples.

Keywords: Confidence Intervals; Margin Of Error; Wald; Sampling; Econometrics Education

INTRODUCTION

Estimating proportions is a common research activity in almost all areas of business. Some examples, drawing across disciplines, might include estimating percentages of loan documents that meet with compliance standards, the proportion of accounting records that meet audit standards, the percentages of firms offering dividends, the proportion of mutual funds that outperformed their benchmarks, the portion of women executives in an industry, the percent of sales calls leading to purchases, proportion of goods meeting a quality standard, the probability of new graduates being hired within three months of college, and so on. The applications of proportion estimates are almost endless. To interpret the estimated proportions, it is necessary to know the "plus or minus" range associated with the estimate by computing a confidence interval or its half-width, the "margin of error". Students are taught to perform these computations throughout the business curriculum (e.g., business stats, marketing, accounting, managerial economics) using the binomial probability distribution and a readily computable formula for the confidence interval[1].

Unfortunately, in too many cases, the tried-and-true formula may be wrong and can lead to misleading results. Some of the potential problems include identifying confidence intervals that do not include the "true" population value or that are associated with the wrong probability level, rejecting groups that may actually meet a threshold, and accepting others that don't. Although the mathematics and statistics literature has been aware of these difficulties for decades, and has offered several solutions, it is our observation that many business students and practitioners have been unaware of this research[2]. One of these methods is particularly elegant and appropriate for business applications, the Agresti – Coull method (Agresti & Coull (1998), Agresti & Caffo(2000), Sauro & Lewis(2005)). This method has been shown to work substantially better than the traditional approach while still being relatively easy to implement in practice. Application of the Agresti-Coull method is demonstrated in this paper to several practical problems from different areas of business.

---

[1] The traditional estimate of confidence intervals is based on an estimate of the standard deviation for a proportion computed as $s = (p*(1-p)/n)^{.5}$ where p is the proportion of the sample, n is the sample size, and s is the estimated standard deviation of the proportion p. This equation is based on the binomial probability distribution described, for instance, in Cochrane (1977).
[2] A statistics textbook describing many of these solutions is Agresti (2013).

      185   *The Clute Institute*

We begin by reviewing the traditional calculations for estimating proportions and their confidence intervals and then discuss some of the challenges faced when using the traditional approach. We next present the Agresti-Coull "Adjusted Wald" method, demonstrating how it can be easily calculated. We finish the paper with several examples illustrating differences between the traditional "binomial" method and the "Adjusted Wald" method.

COMPUTING CONFIDENCE INTERVALS THE TRADITIONAL WAY: SOME VOCABULARY AND EQUATIONS

While there are many kinds of data measurements in business, this paper addresses a specific kind in which a measured characteristic is found or not. When the attribute is present in an observation, sometimes called a "trial" or "subject", the presence of the attribute is called a "success" and absence of the attribute is called a "failure". These are traditional terms and do not convey any particular value judgment; the terms might become confusing when describing a bad outcome as a "success" and a good outcome as a "failure". In practice it may be better to use more case-specific terminology but for this discussion we will use the historical terminology.

Suppose that n subjects are randomly selected from a population of N. The population might include companies in an industry, files to be audited, experimental subjects, consumers in a marketplace, or almost any grouping of potential subjects for measurement. Suppose that x of the n subjects are classified as "success" and (n-x) are classified as "failure". Then, following Cochrane (1977), the following relationships define the estimated rates and confidence intervals for those estimated rates:

(1) $p = x/n$ defines the "point estimate" (the specific numerical estimate) of the success rate of the sample

(2) $V = p*(1-p)/n$ defines the commonly used "population variance" formula

(3) $s^2 = ((N-n)/N)*(p*(1-p)/(n-1))$ defines the "unbiased estimator" of population variance from a sample, and

(4) $s^2 = p*(1-p)/(n-1)$ defines the "unbiased estimate" of the variance for a large population when the sample size n is relatively small (so the term $(N-n)/N$ is close to 1 in value; this term is also known as a "finite population correction (fpc)" term)

(5) $p \pm 1.96*s$ is a confidence interval for p at the 95% confidence level (two sided); more generally, $p \pm z * s$ is a confidence interval for p at the 1- $F(z/2)$ confidence level (two sided) where $F(z/2)$ refers to the cumulative normal probability distribution

(6)   Margin of Error $= 1.96*s$ at the 95% confidence level.

When equation (2) is used to estimate variance (V), the square root of V is then used as the value of s, the sample standard deviation.

In practice in business applications, it is our experience that equations (1) and either (2) or (4) are most frequently used (for example, Aaker and Day (1986)) even in situations when the sample size is large relative to the population size and so the fpc would not be essentially equal to 1.

Equations (1) – (4) are based on the Binomial distribution, which implicitly assumes either sampling from a very large population or sampling with replacement. (Sampling with replacement means that there is a small but real possibility that you could draw the same lucky person every time for your sample.) This method of computing confidence intervals shown in (5) is also known as the "Wald" method in honor of the statistician who derived many of its properties.

An alternative is to assume that once a subject is selected and tested, that subject is no longer able to be considered for future sampling. This approach is a better match to how many real world sampling activities are conducted in that once a unit is tested it may not be available for additional testing, which is particularly well suited for "destructive testing" or assessing the impact of an advertising exposure on consumer behavior. The Hypergeometric distribution is similar to the Binomial but is based on the premise that subjects are not sampled with replacement. The Hypergeometric provides an "Exact" solution for the probability of a given outcome but it is computationally cumbersome even with modern spreadsheets to use exact methods for computing confidence bands

around an estimated proportion. In the following section we present an alternative from the statistics literature that is easily computed but has superior properties to either the Binomial or to using exact approaches.

COMPUTING BETTER CONFIDENCE INTERVALS:
AGRESTI – COULL "ADJUSTED WALD INTERVALS"

In 1998, two statisticians (Agresti & Coull(1998)) published an elegant alternative to the Hypergeometric and other methods that allows business professors, students, and practitioners to compute approximate confidence bands that work better than those based on the Binomial distribution. By better, we mean that in numerous tests the real population probability of success is included in a 95% confidence interval closer to 95% of the time compared to when the traditional or exact methods are used[3]. This method was further developed by Agresti (Agresti & Caffo (2000)) and it is that form that we follow here.

The Adjusted Wald Interval is based on the following equations for computing the 95% confidence level and margin of error:

(1a) $p' = (x+2)/(n+4)$

(2a) $V' = p'*(1-p')/(n+4)$

from which we obtain

(3a) $p \pm 1.96*(V')^{1/2}$ is a confidence interval for p at the 95% confidence level (two sided); more generally, $p \pm z * s$ is a confidence interval for p at the 1- F(z/2) confidence level (two sided) where F(z/2) refers to the cumulative normal probability distribution

(4a)   Margin of Error = $1.96*(V')^{1/2}$ at the 95% confidence level.

In equation (1a), the (x+2) term reflects the x successes plus an additional 2 for purposes of computation and the (n+4) term reflects the n trials plus the additional successes and an equal number of failures.

Describing this approach, Agresti and Coull (1998) write:

> *...an instructor will not go far wrong in giving the following advice: "Add two successes and two failures and then use the Wald formula (1)*

which is presented as equations (1a) and (2a) above.

In essence, Agresti and Coull (1998) have adjusted the formula for an inverted Wald confidence interval in order to improve the performance of the interval. The goal is to have a coverage probability of around 95%. This means the confidence interval will contain the true value from the population 95% of the time. If the coverage probability is much higher than 95%, it means the interval will tend to be bigger or wider than it needs to be. If the coverage probability is much less than 95%, it means the confidence interval is too narrow or small and will not have a big enough chance of containing the true value.   Agresti and Coull (1998) show that the "add two successes and two failures" adjustment is more likely to give a coverage probability closer to 95% than the traditional inverted Wald confidence interval.[4]

---

[3] Sauro and Lewis (2005) present the results of an initial experiment in which they computed "average coverage by confidence intervals" created using several methods including an exact method akin to the Hypergeometric approach, the traditional method, and their proposed Adjusted Wald method. Their experiment was designed to test whether 95% confidence intervals actually included the "true" population parameter 95% of the time. Their experiment resulted in a "Goldilocks" solution: the traditional method was too restrictive and only covered the correct population proportion about 72% of the time, the exact approach they tested was too expansive and included the correct population proportion about 99.4% of the time, and the Agresti-Coull Adjusted Wald approach covered the correct population proportion about 96.7% of the time.
[4] See Agresti-Coull (1998) p. 120 and also Sauro and Lewis (2005) mentioned in the previous footnote.

SOME EXAMPLES COMPARING THE TRADITIONAL
AND AGRESTI-COULL ADJUSTED WALD APPROACH

The following are a series of applications and examples illustrating how the two methods might compare in practice.

A.   Supposed a human resources manager needs to estimate what percentage of workers on night shift take their regularly scheduled meal and rest breaks. The manager studies 40 night shift employees and determines that 20 workers regularly take these breaks and 20 workers do not.

Using the traditional method (equation (5)), the manager estimates that the percentage of workers taking their regularly scheduled breaks is between 34.3% and 65.7%:
From equation (1), $p = x/n = 20/40 = 0.50$.

Then from equation (4) the estimated variance is $s^2 = p*(1-p)/(n-1) = [0.50*(1-0.50)]/[40-1] = 0.00641$ so the estimated standard error, s, is the square root of 0.00641 or 0.0801.

Using equation (5), the lower end of the interval is $p - 1.96*s = 0.50 - 1.96*0.0801 = 34.3\%$ and the upper end of the confidence interval is $p + 1.96*s = 0.50 + 1.96*0.0801 = 65.7\%$.

Using the Adjusted Formula (equation (3a), the manager estimates that the percentage is between 35.2% and 64.8%:

Using equation 1a: $p' = (x+2)/(n+4) = (20 + 2)/(40 +4) = 0.50$ (the same as when using the traditional method, but this is an unusual case, as will be seen in the next example.)

Then use Equation 2a to get an estimate of the variance, $V' = p'*(1-p')/(n+4) = [0.50*(1-0.50)]/[40 +4] = 0.00568$
For use in Equation 3a below, the square root of 0.00568 is 0.0754
Finally, use Equation 3a to get the upper and lower ends of the adjusted Wald inverted confidence interval:
The lower end of the interval is $p - 1.96*(V')^{1/2} = 0.50 - 1.96*0.0754 = 35.2\%$
The upper end of the interval is $p + 1.96*(V')^{1/2} = 0.50 + 1.96*0.0754 = 64.8\%$

Now suppose that the human resources manager performs a similar study but of swing shift workers instead. The manager studies 40 swing shift employees and determines that 10 workers regularly take these breaks and 30 do not.

Using the traditional method, the manager estimates that the percentage of workers taking their regularly scheduled breaks is between 11.4% and 38.6%.

Using equation (1), $p = x/n = 10/40 = 0.25$.

From equation (4), the estimated variance is $s^2 = p*(1-p)/(n-1) = [0.25*(1-0.25)]/[40-1] = 0.00481$
Taking the square root to get the estimated standard error, s, gives 0.0694.

From equation (5), the lower end of the interval is $p - 1.96*s = 0.25 - 1.96*0.0694 = 11.4\%$.
The upper end of the confidence interval is $p + 1.96*s = 0.25 + 1.96*0.0694 = 38.6\%$.

Using the Adjusted Formula, the manager estimates that the percentage is between 14.0% and 40.4%.

From equation 1a: $p' = (x+2)/(n+4) = (10 + 2)/(40 +4) = 0.272$ (note it was 0.25 when using the traditional method.)

Then use Equation 2a to get an estimate of the variance, $V' = p'*(1-p')/(n+4) = [0.272*(1-0.272)]/[40 +4] = 0.00450$

For use in Equation 3a below, the square root of 0.00450 is 0.0671
Finally, use Equation 3a to get the upper and lower ends of the adjusted Wald inverted confidence interval:
The lower end of the interval is p - 1.96*(V')$^{1/2}$ = 0.272 - 1.96*0.0671 = 14.0%
The upper end of the interval is p + 1.96*(V')$^{1/2}$ = 0.272 + 1.96*0.0671 = 40.4%

Finally, suppose that the manager performs the same study of day shift workers and finds that 35 regularly take their breaks and 5 do not.

Using the traditional method, the manager estimates that the percentage of workers taking their regularly scheduled breaks is between 77.1% and 97.9%.

Using equation (1), p = x/n=35/40=0.875.

From equation (4), the estimated variance is s^2 = p*(1-p)/(n-1) = [0.875*(1-0.875)]/[40-1] = 0.00280
Taking the square root to get the estimated standard error, s, gives 0.0529.

From equation (5), the lower end of the interval is p - 1.96*s = 0.875- 1.96*0.0529 = 77.1%.
The upper end of the confidence interval is p + 1.96*s = 0.875 + 1.96*0.0529 = 97.9%.

Using the Adjusted Formula, the manager estimates that the percentage of workers taking their regularly scheduled breaks is between 73.3% and 94.9%.

From equation 1a: p' = (x+2)/(n+4) = (35 + 2)/(40 +4) = 0.841

Then use Equation 2a to get an estimate of the variance, V' = p'*(1-p')/(n+4) = 0.841 *(1-0.841)]/[40 +4] = 0.00304
For use in Equation 3a below, the square root of 0.00304 is 0.0551
Finally, use Equation 3a to get the upper and lower ends of the adjusted Wald inverted confidence interval:
The lower end of the interval is p - 1.96*(V')$^{1/2}$ = 0.841 - 1.96*0.0551= 73.3%
The upper end of the interval is p + 1.96*(V')$^{1/2}$ = 0.841 + 1.96*0.0551= 94.9%

Besides illustrating how differences in supervision might impact workers' compliance with break policies, the example illustrates how sometimes the traditional and adjusted methods provide similar interval estimates but sometimes one of the intervals is shifted higher or lower than the other.

B.   Consider an "acceptance study" to determine whether a pallet of imported goods meets the quality control standards of the customer.

Suppose that five units were randomly selected from the shipment and all five passed the quality test. In this case, p = 5/5 = 1 (or 100%).

Using the traditional method, the 95% confidence interval would be estimated as having a low value of 100% and a high value also of 100%. This is because if p=1, then 1 – p = 0, so p*(1-p)/n = 0 and the estimated variance is zero.

Using the Adjusted Wald formula, p' = (5+2)/(5+4) = 77.8%

Equation 2a gives an estimate of the variance, V' = p'*(1-p')/(n+4) = 0.778 *(1-0.778)]/[5 +4] = 0.0192
For use in Equation 3a below, the square root of 0.0192 is 0.139.
Using Equation 3a to get the upper and lower ends of the adjusted Wald inverted confidence interval:
The lower end of the interval is p - 1.96*(V')$^{1/2}$ = 0.778 - 1.96*0.139= 50.6%
The upper end of the interval is p + 1.96*(V')$^{1/2}$ = 0.778 + 1.96*0.139= 105%
Although the calculation for the upper end of the interval comes out to be 105%, it is impossible for a proportion to be more than 100%, so the confidence interval should be listed as 50.6% to 100%.

This example illustrates how the traditional Binomial version falls apart when sample results are either all successes or all failures. To help student intuition, consider flipping a coin five times and getting heads each time. Presumably students wouldn't conclude that flipping the coin repeatedly would always result in "heads" yet that would be the conclusion from traditional proportion estimates.

This example promotes exploring the minimum probability for the observed outcome consistent with the sample results. Certainly if the probability of "heads" was 100%, one would anticipate seeing five heads in a row in the sample. However, there are other lower true probabilities for the population which could also sometimes result in five "heads" in a row in the sample. The Adjusted Wald provides insight into the reasonable range of these probabilities that are consistent with the sample results. For the example used here, there could be a population where only 50.6% of the goods would pass the quality test, and that low percentage would still be consistent with drawing 5 out of 5 that pass the test.

C.   Consider the problem facing sponsors of a new stock investment strategy. Suppose that in the first six months, the investment strategy outperformed the S&P four times. Potential investors in the strategy would proceed only after the investment has demonstrated a "batting average" statistically greater than 50%. How long would the fund need to maintain the same four wins out of six to demonstrate a statistically significant percentage?

This is essentially the same computation as determining a sample size following a probe sample of a population. Using $p = 0.666667$ or $p' = .6$ along with the traditional and Adjusted Wald equations for the lower confidence limits, a minimum sample size can be computed such that the 95% confidence interval's lower bound equals 50%.

Using the formula for the lower bound, we find[5]

$$n^* = P*(1-P)/[((P - .5)/1.96)^2]$$

where P is either $p = 0.666667$ (from equation (1)) for the traditional method or $p' = 0.6$ (from equation (1a)) for the Adjusted Wald method. From this, the traditional estimate would suggest that 31 months would be sufficient but the Adjusted Wald would require a far longer 92 months of testing.

In contrast, were the batting average 5 months out of six, then the traditional estimate would indicate that just five months track record would be sufficient while the Adjusted Wald would require 20 months[6].

CONCLUSION

The computation of confidence intervals and margins of error is widespread across business disciplines. The traditional method of analyzing margins of error, based on using the sample proportion, the standard deviation estimate from a Binomial model, and then confidence bands using a normal probability distribution, are relatively easy for students to compute but may break down in some circumstances. Although the traditional method is a little easier to calculate, the applications above illustrate how a practitioner or student can implement the Agresti-Coull Adjusted Wald method, giving a better confidence interval compared to the traditional method. We also illustrate that in a situation where the sample contains all failures or all successes, the traditional method gives a misleading result. In these two specific cases, where the sample has either all failures or all successes, the traditional method will make it seem that that no confidence interval is necessary and that the population must contain either all failures or all successes. This is incorrect. The Agresti-Coull Adjusted Wald interval can be used instead to get a reasonable confidence interval.

---

[5] $.5 = p – 1.96*(V^{1/2}) \rightarrow V^{1/2} = (p - .5)/1.96 \rightarrow V = [(p - .5)/1.96]^2 \rightarrow n^* = p*(1-p)/[((p - .5)/1.96)^2]$
[6] As a classroom discussion topic, example C can be used to motivate the rule of thumb that investment managers don't invest client moneys in new hedge funds or similar products until there is at least a three year track record.

The Agresti-Coull Adjusted Wald Estimate discussed in recent statistics literature provides a readily computable approach that is consistent with the Wald approach that students may learn in prior statistics classes and that can be used to address a variety of sampling, confidence interval, and margin of error type problems in finance, accounting, marketing, management, and other business disciplines.

AUTHOR INFORMATION

George Arzumanyan is a research associate with Rule26.com and a portfolio manager for Walnut Oak Capital, LLC. He received his BS in Business from the David Nazarian College of Business and Economics, California State University, Northridge, with a quantitative business emphasis.

Dennis Halcoussis, Ph.D., is Professor of Economics at California State University, Northridge and a consulting econometrician with Rule26.com. He received his Ph.D. from the University of Pennsylvania with specializations in econometrics, finance, and cliometrics. He is the author of a leading econometrics textbook and a specialist in econometrics education.

G. Michael Phillips, Ph.D., is the Director of the Center for Financial Planning and Investment and a Professor of Finance, Financial Planning, and Insurance at California State University, Northridge. He also is a consulting econometrician with Rule26.com. He received his Ph.D. from the University of California, San Diego, with specializations in econometrics and applied economics. E-mail: mphillips@csun.edu (Corresponding author)

REFERENCES

Aaker, David A. & George S. Day (1986) *Marketing Research* (John Wiley & Sons: Hoboken)
Agresti, Alan (2013). *Categorical data analysis* (John Wiley & Sons: Hoboken)
Agresti, Alan & Brian Caffo (2000) Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures, *The American Statistician* 54:4, 280-288
Agresti, Alan & Brent A. Coull (1988) Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions, *The American Statistician* 52:2, 119-126
Cochran, William G. (2007) *Sampling techniques* (John Wiley & Sons: Hoboken)
Halcoussis, Dennis & G. Michael Phillips (2010) Interpreting Bivariate Regression Coefficients: Going Beyond the Average, *American Journal of Business Education* 3:4
Sauro, Jeff & James R. Lewis (2005) Estimating Completion Rates from Small Samples Using Binomial Confidence Intervals: Comparisons and Recommendations, *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting – 2005,* 2100 - 2104

NOTES