Simple Power Calculations: How Do We Know We Are Doing Them the Right Way?

Michael Vorburger[1], Breda Munoz[1]
RTI International[1]

Abstract

Researchers often face the problem of accurately calculating the power for tests of differences between two independent proportions. The method for calculating the exact power of these tests requires an extremely time-consuming, iterative process using 2 X 2 contingency tables. A common approach to circumventing this arduous process is to use an approximation of the power. Four commonly used, and historically accepted, approximations are the arc sine, the chi-squared, and the two continuity-corrected versions of each of these approximations. We will discuss comparisons of all of these approximations for detecting differences between relatively large proportions, when sample size is 30, that have been provided by previous publications; and we will present data for and discuss these same comparisons for smaller proportions. We will also present data for and discuss comparisons of all of these approximations for detecting differences between relatively small proportions, where the larger proportion is between 0.01 and 0.05 and the smaller proportion ranges from 0.001 to 0.007, for a sample size of 300. Finally, we will present data for and discuss the accuracy of the two corrected approximations for detecting differences of relatively small proportions for sample sizes of 300, 750, and 1,500.

Keywords: Fisher's Exact Test, Power Calculation, Power Approximation, Arc Sine Approximation, Chi-Squared Approximation

1. Background

The conditional probability of rejecting the null hypothesis, in an accept-reject test of hypothesis, given that the alternative hypothesis is true, is called the power of the test. Determining the power of the test is referred to as power calculation. For the purposes of this discussion, the alternative hypothesis is $P_1 > P_2$, where P_1 and P_2 are the larger and smaller proportions being compared, respectively. Many researchers use these hypothesis tests to determine the minimum detectable differences between two proportions, given desired power level (1-β), sample size (n), and significance level (α). Researchers often indiscriminately apply some of the formulas without questioning the reliability of the results obtained.

The two standard approximations used to calculate the power of a test of difference between two independent proportions are the arc sine approximation, provided by Cochran and Cox (1957),

$$Z_\beta = Z_{1-\alpha} - \sqrt{2n}\left(Sin^{-1}\sqrt{P_1} - Sin^{-1}\sqrt{P_2}\right), \quad (1)$$

and the chi-squared approximation, provided by Fleiss (1973), as follows:

$$Z_\beta = \frac{Z_{1-\alpha}\sqrt{\left(P_1+P_2\right)\left(1-\frac{P_1+P_2}{2}\right)} - \left(P_1-P_2\right)\sqrt{n}}{\sqrt{P_1\left(1-P_1\right)+P_2\left(1-P_2\right)}}. \quad (2)$$

A continuity-corrected version of the arc sine approximation has been provided by Walters (1979),

$$Z_\beta = Z_{1-\alpha} - \sqrt{2n}\left(Sin^{-1}\sqrt{\left(P_1-\frac{1}{2n}\right)} - Sin^{-1}\sqrt{\left(P_2+\frac{1}{2n}\right)}\right), \quad (3)$$

and a continuity-corrected version of the chi-squared approximation has been provided by Fleiss, Tytun, and Ury (1980), as follows:

$$Z_\beta = \frac{Z_{1-\alpha}\sqrt{2\left(\frac{P_1+P_2}{2}\right)\left(1-\frac{P_1+P_2}{2}\right)} - \sqrt{n\left(P_1-P_2\right)^2 - 2\left(P_1-P_2\right)}}{\sqrt{P_1\left(1-P_1\right)+P_2\left(1-P_2\right)}}. \quad (4)$$

Ury (1981) and Dobson and Gebski (1986) have shown that the corrected approximations [Equations (3) and (4)] yield a substantial improvement in the accuracy of the uncorrected approximations, as compared with Fisher's "exact" test for a 2 X 2 contingency table, where sample size is equal to 30, and the proportions are relatively large (i.e., P_1 of 0.6-0.9, P_2 of 0.1-0.8, with minimum difference of 0.1). To the best of our knowledge, the accuracy of results from these corrected approximations when testing differences between smaller proportions has not been previously evaluated. Additionally, each of these corrected approximations offers advantages and drawbacks, depending on the sample size and the magnitude of the proportions. The corrected arc sine formula (Equation 3) is a simpler formula but requires the use of the arc sine function for ($P_1 - 1/2n$), so P_1 must be greater than $1/2n$. Additionally, the corrected chi-squared formula is invalid when ($P_1 - P_2$) is less than $2/n$.

2. Introduction

In using these (corrected and uncorrected) approximations to test differences between smaller proportions, we found that the corrected approximations overestimate power for small proportions when sample size is small, but they can be very accurate for estimating power for small proportions when sample size is ~300 or greater.

We calculated and compared the powers needed to detect differences of relatively small proportions using all four approximations as well as Fisher's exact test. Table 1 compares the power levels, as calculated using Fisher's exact method and the two corrected approximations [Equations (3) and (4)], associated with detectable differences where the larger proportion ranges from 0.075 to 0.15, the smaller proportion ranges from 0.001 to 0.008, and sample size is 30. Tables 2 and 3 compare the power levels, as calculated using Fisher's exact method and both the uncorrected and corrected versions of each approximation, associated with detectable differences where the larger proportion ranges from 0.02 to 0.03, the smaller proportion ranges from 0.001 to 0.007, and sample size is 300. Table 2 compares exact vs. arc sine [Equations (1) and (3)], and Table 3 compares exact vs. chi-squared [Equations (2) and (4)]. Tables 4, 5, and 6 compare the power levels, calculated using the same methods as in Table 1, associated with detectable differences where the larger proportion is between 0.01 and 0.05, and the smaller proportion ranges from 0.001 to 0.007, for sample sizes of 300, 750, and 1,500, respectively.

Table 1. Power of Fisher's "Exact" Test, with Both Corrected Approximations (n = 30, α = 0.05)

Larger	Smaller Prop. (P_2)				
prop. (P_1)	0.001	0.003	0.005	0.007	0.008
0.075	0.06	0.06	0.06	0.05	0.05
	0.22	0.20	0.18	0.17	0.16
	0.12	0.10	0.09	0.07	0.06
0.100	0.17	0.16	0.15	0.14	0.14
	0.34	0.32	0.30	0.28	0.27
	0.25	0.24	0.22	0.21	0.20
0.150	0.46	0.40	0.42	0.41	0.40
	0.59	0.56	0.54	0.52	0.51
	0.49	0.47	0.46	0.44	0.43

Upper figure = exact power
Middle figure = corrected arc sine approximation
Lower figure = corrected chi-squared approximation

Table 2. Power of Fisher's "Exact" Test, with Arc Sine Approximations (n = 300, α = 0.05)

Larger	Smaller Prop. (P_2)				
prop. (P_1)	0.001	0.002	0.003	0.005	0.007
0.020	0.62	0.53	0.46	0.34	0.24
	0.66	0.58	0.50	0.37	0.27
	0.86	0.77	0.69	0.54	0.41
0.025	0.79	0.72	0.65	0.51	0.40
	0.80	0.73	0.67	0.54	0.43
	0.93	0.87	0.82	0.70	0.58
0.030	0.89	0.84	0.78	0.67	0.57
	0.89	0.84	0.79	0.69	0.58
	0.97	0.94	0.90	0.81	0.71

Upper figure = exact power
Middle figure = corrected arc sine approximation
Lower figure = uncorrected arc sine approximation

Table 3. Power of Fisher's "Exact" Test, with Chi-Squared Approximations (n = 300, α = 0.05)

Larger	Smaller Prop. (P_2)				
prop. (P_1)	0.001	0.002	0.003	0.005	0.007
0.020	0.62	0.53	0.46	0.34	0.24
	0.58	0.51	0.45	0.34	0.25
	0.74	0.68	0.62	0.50	0.40
0.025	0.79	0.72	0.65	0.51	0.40
	0.71	0.66	0.61	0.50	0.40
	0.83	0.79	0.74	0.64	0.54
0.030	0.89	0.84	0.78	0.67	0.57
	0.81	0.77	0.73	0.64	0.55
	0.89	0.86	0.83	0.76	0.67

Upper figure = exact power
Middle figure = corrected chi-squared approximation
Lower figure = uncorrected chi-squared approximation

Table 4. Power of Fisher's "Exact" Test, with Both Corrected Approximations (n = 300, α = 0.05)

Larger	Smaller Prop. (P_2)				
prop. (P_1)	0.001	0.002	0.003	0.005	0.007
0.020	0.62	0.53	0.46	0.34	0.24
	0.66	0.58	0.50	0.37	0.27
	0.58	0.51	0.45	0.34	0.25
0.025	0.79	0.72	0.65	0.51	0.40
	0.80	0.73	0.67	0.54	0.43
	0.71	0.66	0.61	0.50	0.40
0.030	0.89	0.84	0.78	0.67	0.57
	0.89	0.84	0.79	0.69	0.58
	0.81	0.77	0.73	0.64	0.55
0.050	0.99	0.99	0.98	0.96	0.93
	0.99	0.99	0.98	0.96	0.93
	0.97	0.99	0.95	0.93	0.90

Upper figure = exact power
Middle figure = corrected arc sine approximation
Lower figure = corrected chi-squared approximation

Table 5. Power of Fisher's "Exact" Test, with Both Corrected Approximations (n = 750, α = 0.05)

Larger	Smaller Prop. (P_2)				
prop. (P_1)	0.001	0.002	0.003	0.005	0.007
0.010	0.68	0.51	0.38	0.19	0.09
	0.70	0.54	0.40	0.21	0.10
	0.63	0.50	0.38	0.19	0.08
0.015	0.92	0.84	0.74	0.53	0.34
	0.92	0.84	0.74	0.53	0.35
	0.86	0.79	0.70	0.51	0.33
0.020	0.98	0.96	0.92	0.80	0.64
	0.99	0.96	0.92	0.80	0.64
	0.96	0.93	0.88	0.77	0.62
0.025	0.99	0.99	0.98	0.93	0.85
	0.99	0.99	0.98	0.93	0.85
	0.99	0.98	0.96	0.91	0.82

Upper figure = exact power
Middle figure = corrected arc sine approximation
Lower figure = corrected chi-squared approximation

Table 6. Power of Fisher's "Exact" Test, with Both Corrected Approximations (n = 1,500, α = 0.05)

Larger	Smaller Prop. (P_2)				
prop. (P_1)	0.001	0.002	0.003	0.005	0.007
0.010	0.96	0.87	0.72	0.40	0.17
	0.96	0.86	0.72	0.40	0.17
	0.92	0.83	0.69	0.39	0.16
0.015	0.99	0.99	0.96	0.84	0.62
	0.99	0.99	0.97	0.84	0.62
	0.99	0.98	0.95	0.82	0.61
0.020	0.99	0.99	0.99	0.98	0.91
	0.99	0.99	0.99	0.98	0.91
	0.99	0.99	0.99	0.97	0.90
0.025	0.99	0.99	0.99	0.99	0.98
	0.99	0.99	0.99	0.99	0.99
	0.99	0.99	0.99	0.99	0.98

Upper figure = exact power
Middle figure = corrected arc sine approximation
Lower figure = corrected chi-squared approximation

3. Discussion

Both corrected approximations overestimate power, sometimes by as much as 267% when P_1 is less than 0.2, P_2 is less than 0.1, and n=30 (see Table 1). However, the corrected approximations can be very accurate in determining power when the proportions are small and the sample size approaches 300. Additionally, the corrected approximations are more accurate than the uncorrected versions when the proportions are small and n=300 (see Tables 2 and 3).

When n= 300 (see Table 4), the corrected chi-squared approximation (Equation 4) is more accurate for smaller proportions, whereas the corrected arc sine approximation (Equation 3) overestimates the exact power. As the proportions and differences become larger, the corrected arc sine approximation (Equation 3) becomes more accurate, although still slightly overestimating the exact power.

As n reaches 750 (see Table 5), the accuracy of both corrected approximations for calculating the power of tests of differences between relatively small proportions increases. Again, with smaller proportions the corrected chi-squared approximation (Equation 4) provides a more accurate and conservative calculation of power. However, once P_1 reaches 0.015, the corrected arc sine approximation (Equation 3) provides power calculations identical (to 2 decimal points) to Fisher's exact test, whereas the corrected chi-squared approximation (Equation 4) still slightly under-estimates the power.

Furthermore, as n reaches 1,500 (see Table 6), the corrected arc sine approximation (Equation 3) is more accurate regardless of the magnitude of the proportions considered, and it no longer overestimates the power for smaller proportions. Thus, these analysis results suggest that the corrected arc sine approximation (Equation 3) should be used exclusively to determine the power of tests of differences between two proportions once n reaches 1,500.

4. Summary

Analysis results suggest that the continuity-corrected approximations provided by Walters (1979) and Fleiss et al. (1980) result in more accurate power levels than the uncorrected versions previously provided by Cochran and Cox (1957), and Fleiss (1973), for determining the power of tests of differences between small proportions when sample size is at least 300. The uncorrected approximations greatly overestimate the power of these tests. Specifically, when n=300 or 750 the corrected chi-squared approximation (Equation 4) is more accurate for smaller proportions, whereas the corrected arc sine approximation (Equation 3) becomes more accurate as the size of the proportions increases. When n=1,500 the corrected arc sine approximation (Equation 3) is more accurate for all proportions presented above.

Acknowledgements

We would like to thank Dr. James Chromy for his valuable comments and suggestions.

References

Cochran, W.G., and Cox, G.M. (1957). *Experimental Design*, 2nd ed. New York: Wiley.

Dobson, A.J, and Gebski, V.J. (1986). "Sample Sizes for Comparing Two Independent Proportions Using the Continuity-Corrected Arc Sine Transformation." *The Statistician*, 35, 51-53.

Fleiss, J.L. (1973). *Statistical Methods for Rates and Proportions.* New York: Wiley.

Fleiss, J.L., Tytun, A., and Ury, H.K. (1980). A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions. *Biometrics*, 36, 343-346.

Ury, H.K. (1981). "Continuity-Corrected Approximations to Sample Size or Power When Comparing Two Proportions: Chi-Squared or Arc Sine?" *The Statistician*, 30, 199-203.

Walters, D.E. (1979). "In Defence of the Arc Sine Approximation." *The Statistician*, 28, 219-222.