5-1-2009

Approximations to Power When Comparing Two Small Independent Proportions

Michael Vorburger

JMP Division, SAS Institute, mike.vorburger@jmp.com

Breda Munoz

JMP Division, SAS Institute, breda@rti.org

Approximations to Power When Comparing Two Small Independent Proportions

Michael Vorburger   Breda Munoz
JMP Division, SAS Institute

Researchers often face the problem of accurately calculating power for tests of differences between two independent proportions. Four commonly used and accepted approximations are the arc sine, the Chi-squared, and the continuity-corrected versions of each. Comparisons of these are discussed for various sample sizes, ultimately focusing on small proportions.

Key words: Fisher's exact test, power calculation, power approximation, arc sine approximation, Chi-squared approximation, small proportions.

Introduction

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 a 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). 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. Researchers often indiscriminately apply some of these formulas without questioning the reliability of the results obtained.

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),

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

and the Chi-squared approximation, provided by Fleiss (1973),

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

$$(2)$$

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

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

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

Michael Vorburger is a Systems Engineer at JMP Division, SAS Institute. Email: mike.vorburger@jmp.com. Breda Munoz is a Research Statistician at RTI International. Email: breda@rti.org.

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

Each of these corrected approximations offers advantages and drawbacks, depending on the sample size and 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 $2/n$.

Ury (1981) and Dobson and Gebski (1986) showed 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, when the 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.

Power calculations for detecting differences between smaller proportions, using Fisher's exact test, all the approximations and a sample size of 30, are presented and discussed. Also presented is a discussion of power results comparisons 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, power results are compared for detecting differences of relatively small proportions for sample sizes of 300, 750, and 1,500 using both corrected approximations, and the accuracy of these approximations is discussed.

Methodology

In a preliminary analysis, Fisher's exact test, corrected and uncorrected approximations were used to calculate the power needed to detect the differences between smaller proportions (ranging between 0.001 and 0.15). It was found that all of these approximations overestimate power for small proportions when the sample size is small, but the corrected approximations can be very accurate when the sample size is 300 or greater.

The power needed to detect differences of relatively small proportions using all four approximations, as well as Fisher's exact test, were calculated and compared. Table 1 shows the power, as calculated using Fisher's exact method and the two uncorrected approximations (Equations 1 and 2), 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.

Table 2 is a replication of Table 1, substituting the two corrected approximations (Equations 3 and 4) for the uncorrected approximations. Tables 3 and 4 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 3 compares exact vs. arc sine (Equations 1 and 3), and Table 4 compares exact vs. Chi-squared (Equations 2 and 4). Tables 5, 6, and 7 compare the power, calculated using the same methods as in Table 2, 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.

Results

All four approximations overestimate power, sometimes by as much as 1,000% when P_1 is less than 0.2, P_2 is less than 0.1, and $n = 30$ (see Tables 1 and 2). 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

Table 1: Power of Fisher's Exact Test, with Both Uncorrected Approximations
($n = 30$, $\alpha = 0.05$)

Approximation	Larger Proportion (P_1)	Smaller Proportion (P_2)				
		0.001	0.003	0.005	0.007	0.008
Exact Power	0.075	0.06	0.06	0.06	0.05	0.05
Corrected Arc Sine Approximation		0.60	0.53	0.48	0.44	0.42
Corrected Chi-Squared Approximation		0.44	0.42	0.40	0.37	0.36
Exact Power	0.100	0.17	0.16	0.15	0.14	0.14
Corrected Arc Sine Approximation		0.73	0.66	0.62	0.58	0.56
Corrected Chi-Squared Approximation		0.54	0.52	0.50	0.48	0.47
Exact Power	0.150	0.46	0.40	0.42	0.41	0.40
Corrected Arc Sine Approximation		0.88	0.84	0.81	0.78	0.77
Corrected Chi-Squared Approximation		0.71	0.70	0.68	0.67	0.66

Table 2: Power of Fisher's Exact Test, with Both Corrected Approximations
($n = 30$, $\alpha = 0.05$)

Approximation	Larger Proportion (P_1)	Smaller Proportion (P_2)				
		0.001	0.003	0.005	0.007	0.008
Exact Power	0.075	0.06	0.06	0.06	0.05	0.05
Corrected Arc Sine Approximation		0.22	0.20	0.18	0.17	0.16
Corrected Chi-Squared Approximation		0.12	0.10	0.09	0.07	0.06
Exact Power	0.100	0.17	0.16	0.15	0.14	0.14
Corrected Arc Sine Approximation		0.34	0.32	0.30	0.28	0.27
Corrected Chi-Squared Approximation		0.25	0.24	0.22	0.21	0.20
Exact Power	0.150	0.46	0.40	0.42	0.41	0.40
Corrected Arc Sine Approximation		0.59	0.56	0.54	0.52	0.51
Corrected Chi-Squared Approximation		0.49	0.47	0.46	0.44	0.43

Table 3: Power of Fisher's Exact Test, with Arc Sine Approximations
($n = 300$, $\alpha = 0.05$)

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

Table 4: Power of Fisher's Exact Test, with Chi-Squared Approximations
($n = 300$, $\alpha = 0.05$)

Approximation	Larger Proportion (P_1)	Smaller Proportion (P_2)				
		0.001	0.002	0.003	0.005	0.007
Exact Power	0.020	0.62	0.53	0.46	0.34	0.24
Corrected Chi-Squared Approximation		0.58	0.51	0.45	0.34	0.25
Uncorrected Chi-Squared Approximation		0.74	0.68	0.62	0.50	0.40
Exact Power	0.025	0.79	0.72	0.65	0.51	0.40
Corrected Chi-Squared Approximation		0.71	0.66	0.61	0.50	0.40
Uncorrected Chi-Squared Approximation		0.83	0.79	0.74	0.64	0.54
Exact Power	0.030	0.89	0.84	0.78	0.67	0.57
Corrected Chi-Squared Approximation		0.81	0.77	0.73	0.64	0.55
Uncorrected Chi-Squared Approximation		0.89	0.86	0.83	0.76	0.67

Table 5: Power of Fisher's Exact Test, with Both Corrected Approximations
($n = 300$, $\alpha = 0.05$)

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

Table 6: Power of Fisher's Exact Test, with Both Corrected Approximations
($n = 750$, $\alpha = 0.05$)

Approximation	Larger Proportion (P_1)	Smaller Proportion (P_2)				
		0.001	0.002	0.003	0.005	0.007
Exact Power	0.010	0.68	0.51	0.38	0.19	0.09
Corrected Arc Sine Approximation		0.70	0.54	0.40	0.21	0.10
Corrected Chi-Squared Approximation		0.63	0.50	0.38	0.19	0.08
Exact Power	0.015	0.92	0.84	0.74	0.53	0.34
Corrected Arc Sine Approximation		0.92	0.84	0.74	0.53	0.35
Corrected Chi-Squared Approximation		0.86	0.79	0.70	0.51	0.33
Exact Power	0.020	0.98	0.96	0.92	0.80	0.64
Corrected Arc Sine Approximation		0.99	0.96	0.92	0.80	0.64
Corrected Chi-Squared Approximation		0.96	0.93	0.88	0.77	0.62
Exact Power	0.025	0.99	0.99	0.98	0.93	0.85
Corrected Arc Sine Approximation		0.99	0.99	0.98	0.93	0.85
Corrected Chi-Squared Approximation		0.99	0.98	0.96	0.91	0.82

Table 7: Power of Fisher's Exact Test, with Both Corrected Approximations
($n = 1,500$, $\alpha = 0.05$)

Approximation	Larger Proportion (P_1)	Smaller Proportion (P_2)				
		0.001	0.002	0.003	0.005	0.007
Exact Power	0.010	0.96	0.87	0.72	0.40	0.17
Corrected Arc Sine Approximation		0.96	0.86	0.72	0.40	0.17
Corrected Chi-Squared Approximation		0.92	0.83	0.69	0.39	0.16
Exact Power	0.015	0.99	0.99	0.96	0.84	0.62
Corrected Arc Sine Approximation		0.99	0.99	0.97	0.84	0.62
Corrected Chi-Squared Approximation		0.99	0.98	0.95	0.82	0.61
Exact Power	0.020	0.99	0.99	0.99	0.98	0.91
Corrected Arc Sine Approximation		0.99	0.99	0.99	0.98	0.91
Corrected Chi-Squared Approximation		0.99	0.99	0.99	0.97	0.90
Exact Power	0.025	0.99	0.99	0.99	0.99	0.98
Corrected Arc Sine Approximation		0.99	0.99	0.99	0.99	0.99
Corrected Chi-Squared Approximation		0.99	0.99	0.99	0.99	0.98

when the proportions are small and $n = 300$ (see Tables 3 and 4).

When $n = 300$ (see Table 5), 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 6), 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 underestimates the power.

Furthermore, as n reaches 1,500 (see Table 7), 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.

Conclusion

Analysis of 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.

References

Cochran, W. G., & Cox, G. M. (1957). *Experimental design* (*2nd Ed.*). NY: Wiley.

Dobson, A. J, & 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*. NY: Wiley.

Fleiss, J. L., Tytun, A., & 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 defense of the arc sine approximation. *The Statistician*, *28,* 219-222.