MINIMUM SAMPLE SIZES FOR COCHRAN'S TEST

Patricia P. Ramsey, Fordham University
and
Philip H. Ramsey, Queens College of CUNY

Cochran (1950) presented a statistical test, $\underline{Q}$, for the significance of the difference between $\underline{c}$ correlated proportions. The statistic, $\underline{Q}$, is

$$Q = \frac{c(c - 1)\Sigma(T_j - \bar{T})^2}{c\Sigma u_i - \Sigma u_i^2},$$

where T_j is the number of successes in proportion j, $\bar{T}$ is the mean of the T_j, and u_i is the number of successes in the ith case. Cochran noted that the value of $\underline{Q}$ is not changed by dropping any cases with either 0 or $\underline{c}$ successes. Thus the effective sample size, N^*, is the number of cases with more than 0 but less than $\underline{c}$ successes. Cochran further noted that as N^* gets large, the distribution of $\underline{Q}$ approaches a chi-square distribution with $\underline{c}$ - 1 degrees of freedom.

The special case, $c = 2$, was first considered by McNemar (1947) and has a simple method for calculating $\underline{Q}$ as

$$Q = \frac{(D - A)^2}{D + A},$$

where $\underline{A}$ is the number of cases with successes only on the first observation and $\underline{D}$ is the number of cases with successes only on the second. In the McNemar case an exact solution is provided by the well-known sign test. Cochran (1950) described an exact test which is applicable for any value of $\underline{c}$. However, the computational simplicity of $\underline{Q}$ as a chi square test makes it highly desirable for practical applications.

To apply $\underline{Q}$ with the chi square distribution requires some knowledge of the accuracy of that distribution as an approximation to the exact distribution. Tate and Brown (1970) noted that even the distinction between the total sample size, N, and the effective sample size, N^*, had often been omitted by texts which present the procedure. That oversight has been continued in more recent versions of some texts even for those which are generally given quite favorable reviews such as the texts by Hays (1973) and Winer (1971). These same texts mention the importance of "large sample sizes" but do not specify the meaning of "large" nor its application to N^* rather than to $\underline{N}$.

In his original presentation, McNemar (1947) suggested the use of the chi square test provided N^* is 10 or greater. In later discussions McNemar (1969) advocated the use of the exact sign test for N^* less than 10, a continuity corrected chi square test for N^* from 10 to 20, and the original chi square test for N^* greater than 20. The continuity corrected chi square test, $\underline{Q}'$, is found from

$$Q' = \frac{(|D - A| - 1)^2}{D + A}.$$

Cochran (1950) considered a different approach to correcting $\underline{Q}$. He simply averaged $\underline{Q}$ with the next lower possible value of $\underline{Q}$. In the present study that corrected chi square procedure is designated CCS. Cochran reported an investigation of the average bias in $\underline{Q}$ and CCS in comparison to the exact test. For alpha levels near .05 he found $\underline{Q}$ to to have a negative bias with chi square probabilities less than those of the exact test. At similar alpha levels CCS was found to have a positive bias. At alpha levels near .01 both tests were found to have positive average bias. Since the average bias in the uncorrected chi square test of $\underline{Q}$ was only found to be about 14%, Cochran concluded that, "these results appear close enough for routine decisions" (p. 263). The sample sizes investigated by Cochran ranged from 10 to 16 cases at $c = 3$, 6 to 10 cases at $c = 4$ and only 8 cases at $c = 5$.

Since Cochran had recommended the $\underline{Q}$ test without correction over CCS, Tate and Brown (1970) investigated only the uncorrected version. They reported that "the chi-square approximation to $\underline{Q}$ seems good enough, on the average, for practical work with samples yielding tables of 24 or more scores" (p. 159). Their recommendation is therefore to use the uncorrected chi square test for $\underline{Q}$ whenever the product of $\underline{c}$ and N^* is 24 or more. Applying their guideline to the percent error in chi square in comparison to the exact probabilities, Tate and Brown show results with median percent errors of 20, 16, 12, and 15 for $\underline{c}$ of 3, 4, 5, and 6 respectively. These percent errors are close to the 14% value reported by Cochran and therefore their recommendations can be considered to be consistent with those of Cochran. It had been noted by Cochran that an average error of 14% would correspond to an actual alpha level of 0.043 to 0.057 when the nominal alpha is .05 and an actual level of 0.008 to 0.012 when the nominal alpha is .01.

Researchers who are satisfied with the average percent errors reported by Tate and Brown can apply those results in determining a minimum sample size for $\underline{Q}$. However, the maximum error may be much larger than the average values reported by Tate and Brown. For the conditions meeting their guidelines Tate and Brown reported maximum percent errors of 33, 52, 50, and 62 at respective c of 3, 4, 5, and 6 all for alpha levels near .05. In discussing the chi square test for independent proportions Cochran (1954) suggested an upper limit of .06 or a maximum percent error of 20%. At the .01 level Cochran suggested an upper limit of .015 or a maximum percent error of 50%. It is clear from the above results that Tate and Brown's guidelines do not meet these limits on maximum error.

The present investigation examines the maximum error obtained in a chi square test of Q at alpha levels of .05 and .01. From these results new guidelines can be constructed.

Method

Patil's (1975) method was used to determine exact probabilities for specified values of c, N*, Q, and alpha. For respective c values of 2, 3, 4, 5, and 6, N* values were investigated up to 185, 35, 12, 7, and 5. At alpha = .05 and respective c values of 2, 3, 4, 5, and 6, critical chi square values of 3.841, 5.991, 7.815, 9.487, and 11.071 would be required for significance. At alpha = .01 the same critical values would be 6.635, 9.210, 11.345, 13.277, and 15.086. The accuracy of either corrected or uncorrected tests was determined by finding the exact probability of exceeding each of these critical values. Since Cochran had reported the uncorrected chi square test to be liberal (have positive bias) and CCS to have negative bias, a half corrected chi square test (HCS) was included in which Q was averaged with the value for CCS. This would be equivalent to averaging three values of Q with the next lower value.

Due to the discontinuous nature of the exact distribution, values of the true probability can fluctuate markedly from one value of N* to the next. To more clearly show the relationship between the true probability and N* it was often necessary to select the maximum true probability over a range of N*'s.

Conclusions

One method of applying the present results is to simply note the minimum value of N* required to limit the maximum value of the true probability to those values suggested by Cochran. Thus at alpha = .05 we need only determine the value of N* needed to limit the true probability to .06. Those values of N* were found to be 127, 20, 9, 6, and 6 for respective c of 2, 3, 4, 5, and 6.

A point of caution should be added to the above values. The fluctuation in values of the true probability can easily lead to inaccurate interpretations. For example, at c = 2 and N* = 104 a Q value of 3.846 exceeds the .05 critical value of 3.841 but has an exact probability of .0619. This is above Cochran's limit of .06 and would therefore not be acceptable by that standard. If we then look at all N* values from 105 to 125 we find that every Q value greater than 3.841 has an exact probability of less than .06. Thus a study investigating N* values only up to 125 might erroneously conclude that an N* of 126 would be sufficient. Since no value of Q in the N* range of 127 to 185 had an exact probability of more than .06 unless its value was less than 3.841, it seems safe to conclude that 127 is indeed the largest value of N* needed at c = 2 alpha = .05. The minimum values of N* seem stable up to about c = 5. However, at c = 6 the minimum N* of 6 is only suggestive since no cases above N* = 5 were examined at c = 6.

The general results of the present investigation are that at alpha = .01 even the uncorrected chi square test is conservative. This is in agreement with the results reported by Cochran (1950) and seems adequate to justify the use of Q at the .01 level if limiting the maximum error is the criterion for evaluation.

Another approach to the present investigation is to consider several different strategies which can be taken in applying some variation of the chi square test for Q.

Three strategies were devised from the present results. In each case tests were selected so as to produce the greatest power possible but subject to some constraint. The constraints were required to apply at both the .05 and .01 levels. The constraints were: (1) Choose a sequence of chi square tests so that the test is always conservative. In some cases this was not possible since even CCS had a true probability greater than the nominal alpha level. In those cases Bradley's (1978) definition of a negligible bias was used. Bradley uses an upper limit of .055 and .015 for .05 and .01 nominal alphas respectively. (2) Choose a combination of chi square tests which never exceed Cochran's limit. (3) Choose a single chi square test which never exceeds Cochran's limit.

The above three strategies would produce the following combinations at various values of c:

c = 2

1. Q'.
2. Q' if N* LE 126.
   Q if N* GE 127.
3. Q'.

c = 3

1. CCS (with Bradley's limit).
2. HCS if N* LE 19.
   Q if N* GE 20.
3. HCS.

c = 4

1. HCS if N* LE 5.
   CCS if N* GE 6 (with Bradley's lim)
2. HCS if N* LE 5.
   Q if N* GE 9.
3. HCS.

$\underline{c} = 5$

1. HCS if $\underline{N}^*$ LE 5.
   CCS if $\underline{N}^*$ GE 6.
2. HCS if $\underline{N}^*$ LE 5.
   $\underline{Q}$ if $\underline{N}^*$ GE 6.
3. HCS.

$\underline{c} = 6$

1. CCS if $\underline{N}^*$ LE 3.
   HCS if $\underline{N}^*$ GE 4.
2. Q if $\underline{N}^*$ LE 4.
   HCS if $\underline{N}^*$ GE 6.
3. HCS.

Other strategies can also be devised but these seem to cover the cases which are most likely to be considered. It must be noted that the exact test will still have an advantage over any chi square test or combination of tests such as those listed above.

In the McNemar case with $\underline{c} = 2$, $\underline{Q}'$ is almost identical to the exact test in that when $\underline{Q}'$ is significant the exact test is also significant. On the other hand when the exact test is significant $\underline{Q}'$ is almost always significant. In the present study with alpha = .05 there were only two cases in all $\underline{N}^*$ conditions up to 185 in which the exact test was significant when $\underline{Q}'$ was not. At alpha = .01 there were only 12 such cases. It seems likely that most practical applications will be well served by $\underline{Q}'$ when $\underline{c} = 2$.

For $\underline{c}$ of three or more there are some cases in which the exact test will be significant even when the uncorrected chi square test is not significant. This result occurs even in cases in which the uncorrected chi square test has an exact probability in excess of Cochran's limit. For example, it was noted above that a case was found in which $\underline{Q} = 11.207$ exceeds the .05 critical value of 11.071 required at the $\underline{c} = 6$ and $\underline{N}^* = 5$ condition even though the exact probability was .0648. However, in another case with $\underline{c} = 6$ and $\underline{N}^* = 5$ a $\underline{Q}$ value of 11.000 was found to have a true probability of .0430. This latter case would be nonsignificant by the chi square test but significant by the exact test. Thus, it is possible in applying the uncorrected chi square test to have an excess in both Type I and Type II errors in the same $\underline{c}$ and $\underline{N}^*$ condition. Anyone who wants to apply the exact test without requiring computer analysis can use the tables provided by Tate and Brown (1970).

References

Bradley, J. V. (1978), "Robustness?" British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Cochran, W. G. (1950), "The Comparison of Percentages in Matched Samples," Biometrika, 37, 256-266.

Cochran, W. G. (1954), "Some Methods for strengthening the common X^2 tests," Biometrics, 10, 417-451.

Hays, W. L. (1973), Statistics for the Social Sciences (2nd ed.), New York: McGraw-Hill.

McNemar, Q. (1947), "Note on the Sampling Error of Differences Between Correlated Proportions or Percentages," Psychometrika, 12, 153-157.

McNemar, Q. (1969), Psychological Statistics (4th ed.), New York: Wiley.

Patil, K. D. (1975), "Cochran's Q Test: Exact Distribution," Journal of the American Statistical Association, 70, 186-189.

Tate, M. W., and Brown, S. M. (1970), "Note on the Cochran Q Test," Journal of the American Statistical Association 65, 155-160.

Winer, B. J. (1971), Statistical Principles in Experimental Design, New York: McGraw-Hill.