Comparison of Statistical Tests of Independence for Sparse $I \times J$ Contingency Tables.

Contingency tables, and their associated statistical tests, are frequently used in educational and social research. Popular statistical tests used in contingency table analyses include the Pearson chi-square test and the likelihood ratio chi-square test. These two tests are chi-square distributed under large sample conditions. However, when a given contingency table includes small cell frequencies the obtained chi-square statistic may not be asymptotically valid, and the associated probability values will be inappropriate. An alternative statistic which has been recommended for small samples or sparse conditions is the power-divergence statistic (Read & Cressie, 1988). This study was an investigation of the performance of three tests of independence for $I \times J$ contingency tables under small sample conditions. Specifically, the objectives of the research were to investigate the power and Type I error rates under small sample or sparse conditions of the Pearson chi-square test, the likelihood ratio chi-square test, and the Read and Cressie power-divergence statistic with lambda = 2/3 (Read & Cressie, 1988). The power and Type I error rates were estimated for a variety of table dimensions, marginal distributions, sample sizes, and effect sizes. (Contains 16 references, 4 figures, and 4 tables.) (Author)

***********************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document.                                      *
***********************************************************************
Comparison of Statistical Tests of Independence for Sparse I x J Contingency Tables

Cynthia G. Parshall
Support, Technological Applications, and Research Department
American College Testing

Jeffrey D. Kromrey
Department of Educational Measurement and Research
University of South Florida

Ronald Dailey
Department of Educational Measurement and Research
University of South Florida


BEST COPY AVAILABLE

2
Contingency Tables

Comparison of Statistical Tests of Independence for Sparse I x J Contingency Tables

Abstract

Contingency tables, and their associated statistical tests, are frequently used in educational and social research. Popular statistical tests used in contingency table analyses include the Pearson chi-square test and the likelihood ratio chi-square test. These two tests are chi-square distributed, under large-sample conditions. However, when a given contingency table includes small cell frequencies the obtained chi-square statistic may not be asymptotically valid, and the associated probability values will be inappropriate. An alternative statistic which has been recommended for small samples or sparse conditions is the power-divergence statistic (Read & Cressie, 1988).

This study was an investigation of the performance of three tests of independence for I x J contingency tables under small sample conditions. Specifically, the objectives of the research were to investigate the power and Type I error rates under small sample or sparse conditions of the Pearson chi-square test, the likelihood ratio chi-square test, and the Read and Cressie power-divergence statistic with $\lambda = 2/3$ (Read & Cressie, 1988). The power and Type I error rates were estimated for a variety of table dimensions, marginal distributions, sample sizes, and effect sizes.
Introduction

Contingency tables, and their associated statistical tests, are frequently used in educational and social research. Popular statistical tests used in contingency table analyses include the Pearson chi-square test of independence and the likelihood ratio chi-square test of independence. These two tests are asymptotically equivalent and, under large-sample conditions, are chi-square distributed. However, when a given contingency table includes small cell frequencies, the obtained chi-square statistic may not be chi-square distributed. Under these conditions, when the statistic is not asymptotically valid, the associated probability values will be inappropriate. If an obtained statistic is not distributed according to the chi-square distribution, researchers may mistakenly base decisions on inaccurate information. They may reject a null hypothesis when it is true, or fail to reject a null hypothesis when it is false.

Several statistics have been proposed for testing independence between variables in a two dimensional contingency table. Three of these statistics were investigated in this research.

The Pearson chi-square test is given below, where \( O \) refers to the observed frequency of each cell and \( E \) to the expected frequency. The summation is over the cells in the contingency table.

\[
X^2 = \sum_{i} (O_i - E_i)^2 / E_i
\]
The likelihood ratio chi-square test is asymptotically equivalent to Pearson's chi-square test (Fienberg, 1977), and is often used with larger contingency tables, because of its facility with model testing and chi-square partitioning (Upton, 1978). This statistic is computed as:

\[ G^2 = 2 \sum_{i=1}^{c} 0_i \ln \left( \frac{0_i}{E_i} \right) \]

Various corrective actions may be taken by researchers when contingency tables have small overall samples or some sparse cells. One corrective action is to collapse across categories of one or more of the research variables. However, this may lead to categories of lesser meaning than the original variable held. A second action may be to use exact methods to compute the probability for a given table, rather than to base probability on the tabled chi-square distribution (Agresti, 1990). For larger 1 x J tables, this can be computationally prohibitive. Another recommended course of action has been the use of alternative statistics (e.g., Kroll, 1989; Richardson, 1990). Read and Cressie (1988) have suggested the following statistic as one which is less susceptible to the effects of sparseness than either \( X^2 \) or \( G^2 \).

\[ RC = \frac{2}{\lambda(\lambda + I)} \sum_{i=1}^{c} 0_i \left[ \left( \frac{0_i}{E_i} \right)^\lambda - I \right] \]
In their unified model, Cressie and Read (1984), and Read and Cressie (1988), suggest that in fact Pearson's chi-square (with λ set equal to 1) and the likelihood ratio chi-square (the limit as λ approaches 0) are just two members of a family of power-divergence statistics, all of which can be represented by the formula above. Read and Cressie (1988) have found the power-divergence statistic with λ = 2/3 to have some excellent properties, including optimal performance under small sample conditions.

Purpose

The Type I error estimates for small sample chi-square statistics (both $X^2$ and $G^2$) have been investigated for 2 x 2 tables (Camilli & Hopkins, 1978, 1979; Larntz, 1978; Roscoe & Byars, 1971; Thompson, 1988). Considerably less research has been conducted on relative power, or on tables of larger dimensions (Mehta & Hilton, 1993; Parshall & Kromrey, 1994). However, it is often the case that as table size increases, so does the probability of at least some very low frequency cells (Read & Cressie, 1988).

The purpose of this study was to investigate the performance of three tests of independence for $1 \times J$ contingency tables under small samples and sparse conditions. Specifically, the objectives of the research to be reported were to investigate the power and Type I error rates under small sample and sparse conditions of the Pearson chi-square test, the likelihood ratio chi-square test, and the Read and Cressie statistic with $\lambda = 2/3$. The power and Type I error rates were estimated for various table dimensions, marginal distributions, sample sizes, and effect sizes.
Educational Importance of the Study

Because of the frequent application in educational research of contingency table analyses, the operating characteristics of statistical tests of independence is an important area of inquiry. Similarly, the use of small samples is often necessary in educational research. Knowledge of the power and expected Type I error rates of tests of contingency tables in which small samples are involved (either for the total table or for some subset of cells) will help inform researchers who are planning studies in which contingency table analyses will be used.

Method

The research to be reported was a Monte Carlo study. Random samples were generated for a series of $I \times J$ contingency tables: $2 \times 5$, $3 \times 5$, $5 \times 5$, and $2 \times 7$, $3 \times 7$, $5 \times 7$. For each table dimension, three conditions of population marginal distributions were examined: equal marginals, slightly skewed, and highly skewed. Small, medium, and large population effect sizes were examined. In Cohen's (1988) power analysis text, these effect sizes correspond to $w$ values of 0.10, 0.30, and 0.50, respectively. Cohen's effect size $w$ is computed as:

$$w = \sqrt{\sum \left( \frac{P_{ij} - P_{ij0}}{P_{ij0}} \right)^2}.$$
Where \( P_0 \) is the expected proportion in cell \( i \), and \( P_1 \) is the observed proportion in cell \( i \). In addition to these effect sizes, a null model (no effect, or no dependency between the two marginal variables) was also examined, to evaluate the extent to which Type I error rates match nominal alpha levels.

**Programming for the Monte Carlo Study**

The programming for the Monte Carlo study was written in SAS version 6.06. The data were generated using uniform random numbers on the zero to one interval (the SAS RANUNI function). To simulate samples for a test of independence, a separate series of random numbers was generated for each row of the contingency table, with each row consisting of the same number of observations. The observations were then assigned to columns in the table based upon the value of the random number.

For example, with a 2 X 5 table with equal marginals and an effect size of zero, two series of random numbers were generated. Observations with random numbers between zero and .20 were assigned to the first column of the contingency table, those with random numbers between .20 and .40 were assigned to the second column, etc. This procedure yields tables in which the expected proportion in each cell is .10, each column marginal proportion is expected to be .10, and each row marginal proportion is fixed at .50.

The column marginal proportions of the tables examined in the study were controlled by assigning larger or smaller ranges of the uniform random numbers to each column. For example, for a 2 X 5 table with 60:10:10:10:10 column marginals and an effect size of zero, observations with random numbers between zero and .60 were assigned to the first column of the contingency
Contingency Tables

8

table, those with random numbers between .60 and .70 were assigned to the second column, those with random numbers between .70 and .80 were assigned to the third column, those with random numbers between .80 and .90 were assigned to the fourth column, and those with random numbers higher than .90 were assigned to the fifth column. As with the equal marginals procedure described above, the tables produced with this procedure have cells with expected proportions that are equal to the products of the table's marginal proportions.

Three column marginal distributions were examined in this study. The equal marginal condition provided equal proportions at each level of the column variable. A slightly skewed marginal distribution was produced by generating tables in which the expected value of the first column of the contingency table was 60% of the data, and the remaining 40% was evenly dispersed over the other columns of the table (i.e., for a five-level column variable each level was expected to receive 40/4 or 10% of the observations; for a seven-level column variable each level was expected to receive 40/6 or approximately 6.67% of the data). Similarly, a more highly skewed column marginal was produced by generating tables in which the expected value of the first column was 80% of the data and the remaining 20% was evenly distributed over the remaining columns.

Finally, non-null effects were generated by assigning observations to table cells in proportions that differed from the products of the table's marginal proportions. For example, for a 2 X 5 table with equal marginals and an effect size of .50, observations from the first row with random numbers between zero and .312 were assigned to the first column of the contingency table, those with random numbers between .312 and .624 were assigned to the second column, those with random numbers between .624 and .712 were assigned to the third column, those with random
numbers between .712 and .800 were assigned to the fourth column, and those with random numbers larger than .800 were assigned to the fifth column. The procedure was reversed for the random numbers representing the second row of the table. For these data, random numbers between 0 and .088 were assigned to the first column, those with random numbers between .088 and .176 were assigned to the second column, those with random numbers between .176 and .488 were assigned to the third column, those with random numbers between .488 and .800 were assigned to the fourth column, and those with random numbers higher than .800 were assigned to the fifth column. This procedure yields tables in which the expected proportion in each cell of the first four columns deviates by .056 from the product of the marginal proportions, corresponding to Cohen's (1988) \( w \) of .50.

For each size of contingency table investigated, a total of five sample sizes were produced. For the 2 X J tables, overall sample sizes of 16, 32, 64, 128, and 256 were studied. For the 3 X J tables, sizes were 18, 33, 66, 129, and 258. For the 5 X J tables, the sample sizes investigated were 15, 30, 65, 130, and 255. For each table size and sample size, the total sample size was equally divided over the rows in the contingency table. Five thousand samples of each size were drawn under each of the experimental conditions examined. The use of five thousand replications provide maximum 95% confidence intervals of \( \pm .014 \) around the observed proportion of null hypotheses rejected (Robey & Barcikowski, 1992).

For each condition, three test statistics were computed: (a) Pearson's chi-square test of independence, (b) the likelihood ratio chi-square test, (c) and the Read and Cressie power-
divergence statistic with $\lambda = 2/3$. Estimates of the statistical power of each test were conducted at alpha levels of .01, .05, and .10. (However, to limit the number of figures included, only the results for the nominal alpha level of .05 are presented in figures.)

Results

The empirical estimates of the Type I error rates for the three statistical testing procedures are presented in Table 1 for the conditions examined in this study. At a .01 nominal alpha level, both the Pearson chi-square and the Read and Cressie statistic were consistently conservative in Type I error control for the smallest samples examined. For the samples of size 16, the estimated Type I error rates ranged from zero to only .003 for the Pearson chi-square and to .004 for the Read and Cressie statistic. As the sample sizes increased, the estimates of Type I error rates of both procedures approached the nominal alpha level. However, for the conditions involving asymmetric marginal distributions, these tests remained somewhat conservative even with the largest samples examined in this study. For example, with the 3 X 7 tables and samples of size 258, the estimates of Type I error rates were .005 for both of these tests in the extreme asymmetry condition. In contrast to the conservatism shown by the Pearson chi-square and the Read and Cressie statistics, the likelihood ratio chi-square showed a marked tendency to be excessively liberal in Type I error control. Using Bradley's (1975) liberal criterion of statistical robustness (i.e., estimated Type I error rate within the range of $\alpha_{\text{nominal}} \pm .5 \alpha_{\text{nominal}}$), the likelihood ratio chi-square exceeded these limits in 38 of the 90 conditions examined with a nominal alpha
Contingency Tables

level of .01. The most extreme Type I error rate was seen in the 5 X 7 tables with equal
marginals and samples of size 65, in which the estimate of the Type I error rate for the test was
.041, more than four times the nominal alpha level.

The results were consistent with nominal alpha levels of .05 and .10, although as should be
expected the Type I error rates were closer to nominal levels for these more liberal alpha levels.
The Pearson chi-square and the Read and Cressie statistic remained conservative with small
samples and asymmetric marginals at both of these nominal alpha levels. The likelihood ratio
chi-square remained liberal in many of the conditions examined, exceeding Bradley's liberal
limits of robustness in 44 of the 90 conditions examined at a nominal alpha level of .05 and in
47 of the 90 conditions at a nominal alpha level of .10.

Of interest to note in Table 1, the relative conservatism of the Pearson chi-square and the
Read and Cressie statistic depended upon the table size, with the Read and Cressie statistic being
less conservative than the Pearson chi-square in the smaller tables, but more conservative in the
larger tables. For example, at the .05 nominal alpha level in the 2 X 7 tables with equal
marginals and samples of size 16, the estimated Type I error rate for the Read and Cressie
statistic was .027 while that of the Pearson chi-square was .017. With the smallest sample size in
the 3 X 5 and 3 X 7 tables with equal marginals, however, these tests were nearly equally
conservative (.032 vs. .033 for the 3 X 5 tables, and .021 vs. .022 for the 3 X 7 tables). In
contrast, for the 5 X 5 and 5 X 7 tables with the smallest samples and equal marginals, the
Pearson chi-square was less conservative than the Read and Cressie statistic (.027 vs. .014 for the 5 X 5 tables, and .021 vs. .006 for the 5 X 7 tables). Regardless of table dimensions, however, with larger sample sizes the estimated Type I error rates of these two statistics converged (see Figures 1 and 2).

Estimates of statistical power are provided in Tables 2 and 3. Because the likelihood ratio chi-square did not adequately control Type I error rates for the sparse tables examined in this study, power estimates for this test are not included. Table 2 provides power estimates averaged across effect sizes, while detailed results are provided in Table 3.

An examination of Table 2 shows that the average power differences between the Read and Cressie statistic and the Pearson chi-square were small. However, the Read and Cressie statistic was consistently more powerful, on average, than the Pearson chi-square for the smaller tables examined, while the Pearson chi-square evidenced power advantages for the larger tables. Such power differences, however, were found only with the smaller sample sizes. With the larger sample sizes examined, the two tests showed similar power across conditions examined. This pattern of power differences is consistent with the pattern of Type I error rate estimates presented.
in Table 1, with each test having lower power for those table sizes in which the test is the more conservative in terms of Type I error control.

For example, at the nominal alpha level of .01, the greatest difference favoring the Read and Cressie statistic was obtained in the 2 X 7 tables with samples of size 64 and extreme skewness in the marginals. In this condition, the power of the Read and Cressie statistic was .230, while that of the Pearson chi-square was .179, a difference of .051. The greatest difference favoring the Pearson chi-square was obtained in the 5 X 7 tables with samples of size 130 and extreme marginal skewness. In this condition, the estimated power of the Pearson chi-square was .285, while that of the Read and Cressie statistic was .242, a difference of .043. The Read and Cressie statistic provided slight but consistent power advantages relative to the Pearson chi-square for the 2 X 5 and 2 X 7 tables.

Similarly, for the nominal alpha level of .05, the greatest difference in power estimates favoring the Read and Cressie statistic was seen in the 2 X 7 tables with samples of size 32 and extreme marginal skewness, where the power of the Read and Cressie statistic was .124 while that of the Pearson chi-square was .087, a difference of .037. The greatest difference favoring the Pearson chi-square was obtained in the 5 X 7 tables with samples of size 65 and extremely skewed marginals. In this condition, the estimated power of the Pearson chi-square was .163, while that of the Read and Cressie statistic was .099, a power difference of .064.

Finally, for the nominal alpha level of .10, the greatest power difference favoring the Read and Cressie statistic was seen in the 2 X 7 tables with samples of size 32 and extreme marginal skewness. In this condition the estimated power of the Read and Cressie statistic was .250 while
that of the Pearson chi-square was .203, and a difference of .047. Conversely, the greatest estimated power difference favoring the Pearson chi-square was obtained in the 5 X 7 tables with samples of size 65 and extreme skewness in the marginal distributions. In this condition the power of the Pearson chi-square was .256 while that of the Read and Cressie statistic was .175, a difference of .081.

In an examination of the more detailed results on the statistical power estimates presented in Table 3, it should be noted that for those conditions that showed the greatest overall power differences, the differences increase with the effect size of the condition. For example, at a nominal alpha level of .05, with the 2 X 7 tables with extreme skewness and sample size 32, both tests showed almost no power at the small effect size (.004 for the Pearson chi-square and .006 for the Read and Cressie statistic). For the medium effect size (0.3) the estimated power of the Pearson chi-square was .043 while that of the Read and Cressie was .061. Finally, for the large effect size (0.5) the power estimates were .213 and .304, for the two tests respectively (see Figure 3). Similarly, for the 5 X 7 tables with extreme skewness and samples of size 65, under a nominal alpha of .05, the Pearson chi-square and the Read and Cressie statistic showed power estimates for the small effect size of .022 and .007, respectively. With a medium effect size the power estimates were .099 for the Pearson chi-square and .049 for the Read and Cressie statistic, while for the large effect size the power estimates were .369 and .240 for the two tests (see Figure 4).
A summary of the power comparisons is presented in Table 4. This table presents the number and percent of conditions in which the Pearson chi-square was nominally more powerful than the Read and Cressie statistic, the number and percent of conditions in which the Read and Cressie statistic was more powerful, and the number and percent of conditions in which the power estimates were identical. For example, at a nominal alpha level of .01, for the 2 X 5 and 2 X 7 tables, the Read and Cressie statistic was more powerful than the Pearson chi-square in 73% of the conditions examined, and this test was not less powerful than the chi-square in any of the conditions examined. For the 3 X 5 and 3 X 7 tables, the tests showed identical power in 40% of the conditions, the chi-square test was more powerful in 33% of the conditions and the Read and Cressie statistic was more powerful in 27% of the conditions. For the 5 X 5 and 5 X 7 tables, the chi-square test was more powerful than the Read and Cressie statistic in 78% of the conditions. As may be seen in Table 4, this pattern was evident across nominal alpha levels. However, for the 3 X 5 and 3 X 7 tables, the power estimates for the Read and Cressie statistic exceeded those of the Pearson chi-square at nominal alpha levels of .05 and .10 more frequently than it did for the nominal alpha level of .01.

Discussion

The likelihood ratio test demonstrated a poor ability to maintain the nominal Type I error rate, disqualifying it for further consideration. (A related, earlier investigation of 2x2 and 2x3
contingency table analyses found the likelihood ratio test to display relatively poor power along with liberal Type I error rates compared to other statistics investigated. See Parshall & Kromrey, 1994.) The performance of the remaining, competing statistics in this study converged as sample size increased, regardless of the size of the table. On a practical basis, this means that for the conditions studied, once a sample is large enough it may be a matter of indifference to the practitioner whether Pearson’s chi-square or the Read and Cressie statistic is used.

With the smaller samples however, differences in the performance of the two statistics were found. The Read and Cressie statistic displayed greater power than Pearson’s statistic in the smaller tables, while Pearson’s chi-square test out-performed the Read and Cressie statistic in the larger tables. While both tests demonstrated increased power as the effect size increased, this was not a general improvement. Rather, the Read and Cressie statistic displayed a greater increase in power associated with increased effect size in just those smaller tables where it already displayed a performance advantage over Pearson’s test, while Pearson’s test demonstrated a greater power increase associated with increased effect size in the larger tables where it displayed better performance. Both statistics are more powerful under equal marginals conditions. Both also demonstrate a marked conservativeness in the tables with highly skewed marginals, although the general pattern of better performance of the Read and Cressie for smaller tables and better performance for the Pearson for larger test hold here as well.

Read and Cressie (1988) point out a number of the variables and assumptions which may need to be considered, including “the sample size n, the number of cells k, the form of the null model (loglinear, etc.), and the ‘direction of departure’ of the alternative from the null” (pg. 80). Their
model for analysis suggests using the family of power-divergence statistics, selecting the particular value of lambda which is optimal under given conditions. For example, they indicate that Pearson's chi-square (lambda=1) may be optimal when expected cell frequencies are near equal across a table, and when the number of table cells is at least 20, and $G^2$ is optimal for certain nonlocal alternatives with a limited number of near-zero probabilities. When the expected cell frequencies are unequal, they recommend $\lambda = 2/3$, specifying that Pearson's chi-square statistic can display serious bias for sparse tables with unequal cell probabilities.

For the conditions examined in this study, it would appear that a decision regarding the best statistic to apply in a given situation needs to be based on table size as well as sample size. (Skewness also impacted the results, by lessening power and increasing conservativeness, but this effect was consistent across statistical test.) In general, as Read and Cressie suggest, $\lambda = 2/3$ may be a good compromise solution when a researcher has little knowledge about possible alternative hypotheses. Conversely, a researcher may opt to use Pearson's test under large table, small sample conditions, the Read and Cressie power-divergence statistic under small table, small sample conditions, and either statistic with large samples.

Future research should include a more detailed investigation into sparseness. One aspect of this study which may limit the generalization of results concerns the manner in which data were generated. The data were simulated according to expected marginal proportions (e.g., equal marginals, slightly skewed, and highly skewed marginals), while the total table sample size was held near constant. For example, a 5 x 7 table with sample size of 256 will be sparser than a 2 x 7
table with the same total sample size. Thus, as table size increased, simultaneously sparseness also increased. A follow-up study could alter this data-generation design to provide more information about the interaction of sample size with table size. Additionally, other approaches for modeling the alternative hypothesis could be simulated. While one reasonable alternative was modeled in the current study, a given contingency table can differ from the null in many ways, potentially affecting the selection of the optimal statistical test.
References


<table>
<thead>
<tr>
<th>Table 1: Empirical Estimates of Type I Error Rates by Table Size, Sample Size, Marginal Distribution and Nominal Alpha Level.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal Alpha Level</strong></td>
</tr>
<tr>
<td>.01</td>
</tr>
<tr>
<td><strong>Marginal Distribution</strong></td>
</tr>
<tr>
<td><strong>Equal</strong></td>
</tr>
<tr>
<td><strong>Slight Skew</strong></td>
</tr>
<tr>
<td><strong>Extreme Skew</strong></td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
</tr>
<tr>
<td><strong>Table Size</strong></td>
</tr>
<tr>
<td>2 x 5</td>
</tr>
<tr>
<td>Sample Size: 16</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 5</td>
</tr>
<tr>
<td>Sample Size: 32</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 5</td>
</tr>
<tr>
<td>Sample Size: 64</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 5</td>
</tr>
<tr>
<td>Sample Size: 128</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 5</td>
</tr>
<tr>
<td>Sample Size: 256</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td><strong>Equal</strong></td>
</tr>
<tr>
<td><strong>Slight Skew</strong></td>
</tr>
<tr>
<td><strong>Extreme Skew</strong></td>
</tr>
<tr>
<td><strong>Marginal Distribution</strong></td>
</tr>
<tr>
<td><strong>Sample Size</strong></td>
</tr>
<tr>
<td><strong>Table Size</strong></td>
</tr>
<tr>
<td>2 x 7</td>
</tr>
<tr>
<td>Sample Size: 16</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 7</td>
</tr>
<tr>
<td>Sample Size: 32</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 7</td>
</tr>
<tr>
<td>Sample Size: 64</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 7</td>
</tr>
<tr>
<td>Sample Size: 128</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>2 x 7</td>
</tr>
<tr>
<td>Sample Size: 256</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
<tr>
<td>Ch1 / Ch2 / IR / RC</td>
</tr>
</tbody>
</table>

Note: Estimates are based on 5000 samples of each condition. Type I error estimates have been multiplied by 1000 and rounded.
Table 2
Average Empirical Estimates of Statistical Power by Table Size, Sample Size, Marginal Distribution and Nominal Alpha Level.

<table>
<thead>
<tr>
<th>Table Size</th>
<th>Sample Size</th>
<th>Nominal Alpha Level</th>
<th>Marginal Distribution</th>
<th>Slight Skew</th>
<th>Extreme Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Equal CH12</td>
<td>LR</td>
<td>RC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CH12</td>
<td>LR</td>
<td>RC</td>
</tr>
<tr>
<td>2 x 5</td>
<td>16</td>
<td>.01</td>
<td>18</td>
<td>31</td>
<td>5</td>
</tr>
<tr>
<td>2 x 5</td>
<td>32</td>
<td>.05</td>
<td>12</td>
<td>31</td>
<td>6</td>
</tr>
<tr>
<td>7 x 5</td>
<td>64</td>
<td>.01</td>
<td>19</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>7 x 5</td>
<td>178</td>
<td>.01</td>
<td>19</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>7 x 5</td>
<td>256</td>
<td>.01</td>
<td>26</td>
<td>34</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. Estimates are based on 1000 samples of each condition. Power estimates have been multiplied by 1000 and rounded.
<table>
<thead>
<tr>
<th>Table 3</th>
<th>Nominal Alpha Level</th>
<th>Sample Size</th>
<th>Effect Size</th>
<th>Marginal Distribution</th>
<th>Chi^2</th>
<th>LR</th>
<th>RC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x 5</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equal</td>
<td>Slight Skew</td>
<td>Extreme Skew</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH12 LC RC</td>
<td>CH12 LC RC</td>
<td>CH12 LC RC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2 x 5</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equal</td>
<td>Slight Skew</td>
<td>Extreme Skew</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH12 LC RC</td>
<td>CH12 LC RC</td>
<td>CH12 LC RC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2 x 5</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equal</td>
<td>Slight Skew</td>
<td>Extreme Skew</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH12 LC RC</td>
<td>CH12 LC RC</td>
<td>CH12 LC RC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
### Table 3 (Continued)

Empirical Estimates of Statistical Power by Table Size, Sample Size, Effect Size, Marginal Distribution and Nominal Alpha Level.

<table>
<thead>
<tr>
<th>Table Size</th>
<th>Sample Size</th>
<th>Effect Size</th>
<th>Nominal Alpha Level</th>
<th>Marginal Distribution</th>
<th>Marginal Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>18</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>18</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>18</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>33</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>33</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>33</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>66</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>66</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>66</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>129</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>129</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 5</td>
<td>129</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>18</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>18</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>18</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>33</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>33</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>33</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>66</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>66</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>66</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>129</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>129</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>129</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>258</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>258</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>258</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>1099</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>1099</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 7</td>
<td>1099</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For columns 4 through 6, the values represent the empirical estimates of statistical power for different table sizes, sample sizes, effect sizes, and marginal distributions at nominal alpha levels of 0.01 and 0.05.
Table 3 (Continued)
Empirical Estimates of Statistical Power by Table Size, Sample Size, Effect Size, Marginal Distribution and Nominal Alpha Level.

<table>
<thead>
<tr>
<th>Table Size</th>
<th>Sample Size</th>
<th>Effect Size</th>
<th>Marginal Distribution</th>
<th>Nominal Alpha Level</th>
<th>.01</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x 5</td>
<td>15</td>
<td>0.1</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5 x 5</td>
<td>15</td>
<td>0.3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5 x 5</td>
<td>15</td>
<td>0.5</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5 x 5</td>
<td>30</td>
<td>0.1</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5 x 5</td>
<td>30</td>
<td>0.3</td>
<td>20</td>
<td>15</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>5 x 5</td>
<td>30</td>
<td>0.5</td>
<td>70</td>
<td>5</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>5 x 5</td>
<td>65</td>
<td>0.3</td>
<td>64</td>
<td>64</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td>5 x 5</td>
<td>65</td>
<td>0.5</td>
<td>347</td>
<td>345</td>
<td>264</td>
<td>197</td>
</tr>
<tr>
<td>5 x 5</td>
<td>150</td>
<td>0.1</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>5 x 5</td>
<td>150</td>
<td>0.3</td>
<td>235</td>
<td>220</td>
<td>197</td>
<td>172</td>
</tr>
<tr>
<td>5 x 5</td>
<td>150</td>
<td>0.5</td>
<td>834</td>
<td>656</td>
<td>702</td>
<td>723</td>
</tr>
<tr>
<td>5 x 5</td>
<td>255</td>
<td>0.1</td>
<td>33</td>
<td>33</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>5 x 5</td>
<td>255</td>
<td>0.3</td>
<td>614</td>
<td>608</td>
<td>564</td>
<td>513</td>
</tr>
<tr>
<td>5 x 7</td>
<td>15</td>
<td>0.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5 x 7</td>
<td>15</td>
<td>0.3</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5 x 7</td>
<td>15</td>
<td>0.5</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5 x 7</td>
<td>30</td>
<td>0.1</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5 x 7</td>
<td>30</td>
<td>0.3</td>
<td>16</td>
<td>8</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5 x 7</td>
<td>30</td>
<td>0.5</td>
<td>50</td>
<td>50</td>
<td>31</td>
<td>13</td>
</tr>
<tr>
<td>5 x 7</td>
<td>65</td>
<td>0.1</td>
<td>11</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5 x 7</td>
<td>65</td>
<td>0.3</td>
<td>59</td>
<td>55</td>
<td>43</td>
<td>27</td>
</tr>
<tr>
<td>5 x 7</td>
<td>65</td>
<td>0.5</td>
<td>310</td>
<td>297</td>
<td>252</td>
<td>190</td>
</tr>
<tr>
<td>5 x 7</td>
<td>150</td>
<td>0.1</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>5 x 7</td>
<td>150</td>
<td>0.3</td>
<td>101</td>
<td>189</td>
<td>176</td>
<td>160</td>
</tr>
</tbody>
</table>

Note: Estimates are based on 1000 samples of each condition. Power estimates have been multiplied by 1000 and rounded.
Table 4

Number and Percent of Conditions with Differences in Empirical Power Estimates by Table Size.

<table>
<thead>
<tr>
<th>Table Size</th>
<th>Nominal Alpha = .01</th>
<th>Nominal Alpha = .05</th>
<th>Nominal Alpha = .10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RC&gt;Chi2  No Diff  Chi2&gt;RC</td>
<td>RC&gt;Chi2  No Diff  Chi2&gt;RC</td>
<td>RC&gt;Chi2  No Diff  Chi2&gt;RC</td>
</tr>
<tr>
<td>2 X 5</td>
<td>33 12 0</td>
<td>37 8 0</td>
<td>33 12 0</td>
</tr>
<tr>
<td></td>
<td>73% 27% 0%</td>
<td>82% 18% 0%</td>
<td>73% 27% 0%</td>
</tr>
<tr>
<td>2 X 7</td>
<td>33 12 0</td>
<td>38 7 0</td>
<td>36 9 0</td>
</tr>
<tr>
<td></td>
<td>73% 27% 0%</td>
<td>84% 16% 0%</td>
<td>80% 20% 0%</td>
</tr>
<tr>
<td>3 X 5</td>
<td>12 18 15</td>
<td>26 13 6</td>
<td>30 6 9</td>
</tr>
<tr>
<td></td>
<td>27% 40% 33%</td>
<td>58% 29% 13%</td>
<td>67% 13% 20%</td>
</tr>
<tr>
<td>3 X 7</td>
<td>12 18 15</td>
<td>25 8 12</td>
<td>20 7 18</td>
</tr>
<tr>
<td></td>
<td>27% 40% 33%</td>
<td>56% 18% 27%</td>
<td>44% 16% 40%</td>
</tr>
<tr>
<td>5 X 5</td>
<td>2 8 35</td>
<td>8 4 33</td>
<td>12 6 27</td>
</tr>
<tr>
<td></td>
<td>4% 18% 78%</td>
<td>18% 9% 73%</td>
<td>27% 13% 60%</td>
</tr>
<tr>
<td>5 X 7</td>
<td>1 9 35</td>
<td>5 4 36</td>
<td>10 3 32</td>
</tr>
<tr>
<td></td>
<td>2% 20% 78%</td>
<td>11% 9% 80%</td>
<td>22% 7% 71%</td>
</tr>
</tbody>
</table>
Figure 1

Type Error Rates of the Pearson Chi-Square and Read & Cressie Statistic

- Pearson Chi-square
- Read & Cressie
Figure 2

Type I Error Rates of the Pearson Chi-Square and Read & Cressie Statistic

- Pearson Chi-square
- Read & Cressie

Table: Equal Marginals, Nominal Alpha = .05
Figure 3

Power of the Pearson Chi-Square and Read & Cressie Statistic

- Pearson Chi-square
- Read & Cressie

Effect Size vs. Power Score
Figure 4: Power of the Pearson Chi-Square and Read & Cressie Statistic

Pearson Chi-square
Read & Cressie