PSYCHOMETRIKA—VOL. 42, NO. 1
MARCH, 1977

SOME EXACT CONDITIONAL TESTS OF INDEPENDENCE FOR $R \times C$ CROSS-CLASSIFICATION TABLES

ALAN AGRESTI

AND

DENNIS WACKERLY

UNIVERSITY OF FLORIDA

Exact conditional tests of independence in cross-classification tables are formulated based on the χ^2 statistic and statistics with stronger operational interpretations, such as some nominal and ordinal measures of association. Guidelines for the table dimensions and sample sizes for which the tests are economically implemented on a computer are given. Some selected sample sizes and marginal distributions are used in a numerical comparison between the significance levels of the approximate and exact conditional tests based on the χ^2 statistic.

Key words: exact test, independence, contingency tables, ordinal and nominal measures of association, chi-square test, computer algorithm.

1. Introduction

Over the years, much has been written about the adequacy of the chi-square distribution as an approximation for the sampling distribution of the statistic used to compare observed frequencies in a cross-classification table to the frequencies "expected" under the null hypothesis of independence. Most textbooks of statistical methodology contain a warning that the expected frequencies should fulfill some requirement (e.g., all exceed five), but there is considerable variability in these suggestions. Some papers in the literature emphasize the robustness of the chi-square test for small expected frequencies, but a simulation study by Roscoe and Byars [1971] showed that average expected frequencies as low as two or as high as ten may be needed, depending on the underlying structure.

Constraints in the research problem often do not permit sample sizes to be large enough so that even very lax constraints of this type can be met. For these cases, some researchers have suggested modifications to the chi-square test to yield "better" approximations to the sampling distributions

The authors are grateful for the suggestions of the referees and for computer funding provided by the Northeast Regional Data Center at the University of Florida.

Requests for reprints should be sent to D. D. Wackerly, Department of Statistics, Nuclear Sciences Center, University of Florida, Gainesville, Florida 32611.

of test statistics which are similar to the usual chi-square test statistic [see, e.g., Haldane, 1943; Nass, 1959]. However, probably the most common approaches in practice are to combine or eliminate categories so that suggested regularity conditions are met on a reduced table, to use Fisher's exact test (for 2×2 tables), or simply to ignore any possible difficulties with the chi-square approximation and implement the usual chi-square test. No matter how much research is done on the adequacy of applying asymptotic (large sample) techniques to tables with small cell frequencies, there will always be tables for which the observed cell frequencies are too small, so that the researcher will have to combine or eliminate categories in order comfortably to implement an approximation procedure. Many times this combination of categories cannot be done meaningfully. Even when it can be done, there remains the question of how much information is lost due to this (sometimes arbitrary) combination or elimination. An example of this is given in Section 2. The main purpose of this paper is to provide useable (with computer facilities) and conceptually simple *exact* conditional methods of analyzing such tables.

In the next section we demonstrate how the chi-square statistic or other statistics may be used in exact tests of independence against the general alternative of dependence. In Section 3, ways to further "strengthen" such exact conditional tests of independence by considering more specific alternative hypotheses are discussed. In particular, Kendall's tau *b* is used to illustrate how alternative hypotheses involving commonly used measures of association (whose asymptotic sampling distributions provide inadequate approximations for small samples) can be used to give small sample exact significance probabilities for tests of different aspects of cross-tabulated data.

The particularly attractive aspect of these tests is that they are *exact* tests of preordained level and that no worries about the adequacy of large sample approximations are encountered. The only constraint on the technique is the amount of time necessary to implement the procedure if the cell entries or table dimensions are large.

In Section 4 we discuss the extent to which these exact conditional tests can be implemented using the computer. In particular, it is shown that the exact tests of independence using the χ^2 formula or a measure of association as the test statistic are manageable (i.e., require less than a minute of CPU time on the IBM 370/165 computer) for 2×3, 2×4, 2×5, 2×6, 2×7, 3×3, 3×4, 3×5, and 4×4 tables for many of the sample sizes and marginal distributions for which one would question the use of large sample approximations.

For a variety of sample sizes and marginal distributions, numerical comparisons are made under the null hypothesis of independence between the exact conditional distribution of the χ^2 statistic and the corresponding chi-square distribution which serves as the large sample approximate conditional distribution of that statistic. The discreteness of the exact distribution for small sample sizes makes it difficult to describe when the approximate test is robust.

2. Exact Conditional Tests of Independence

Suppose that a random sample of n observations is categorized jointly according to two classification schemes, one with r categories (rows) and the other with c categories (columns). Suppose further that the cross-classification frequencies $\{f_{ij}, 1 \leq i \leq r, 1 \leq j \leq c\}$ are represented by an $r \times c$ table. In the tests considered in this paper, we regard the set S of all cross-classification tables with nonnegative integer entries which have the same row marginal frequencies $\{f_{i.}, 1 \leq i \leq r\}$ and the same column marginal frequencies $\{f_{.j}, 1 \leq j \leq c\}$ as the observed table. Conditional upon these marginal entries and under the null hypothesis of independence, the probability of the table with entries $\{f_{ij}'\}$ satisfying $\sum_{i=1}^{r} f_{ij}' = f_{.j}$ and $\sum_{j=1}^{c} f_{ij}' = f_{i.}$ for $1 \leq i \leq r$ and $1 \leq j \leq c$ can be shown [Halton, 1969] to equal

$$P(\{f_{ij}'\} \mid \{f_{i.}, f_{.j}\}) = \frac{\prod_{i=1}^{r} f_{i.}! \prod_{j=1}^{c} f_{.j}!}{n! \prod_{i=1}^{r} \prod_{j=1}^{c} f_{ij}'!}.$$

In the remainder of this paper, we shall denote the probability

$$P(\{f_{ij}\} \mid \{f_{i.}, f_{.j}\})$$

of the observed table by P and the probability $P(\{f_{ij}'\} \mid \{f_{i.}, f_{.j}\})$ of any other table with the given marginal frequencies by P'.

Freeman and Halton [1951] formulated an exact conditional test of the null hypothesis of independence versus the general alternative by taking as the attained significance level

$$p = \sum_{S} I(P' \leq P)P',$$

the sum of the probabilities of those tables which occur with no higher probability than the observed table. ($I(A)$ denotes the indicator function of the set A.) Procedures which order sample points solely on the basis of the probability of occurrence have received strong criticism [see, e.g., Radlow & Alf, Jr., 1975]. The rationale behind this criticism is that some configurations of cell frequencies may be less likely than the observed table under the null hypothesis, but yet, in some sense, exhibit less discrepancy from the null hypothesis than the observed table.

On the other hand, useful alternative exact conditional tests can be simply formulated by using other criteria for ranking the tables according to the deviation each exhibits from independence. For example, the χ^2 statistic or the likelihood ratio statistic (which also is asymptotically distributed chi-square) for testing independence could be computed for every set $\{f_{ij}'\}$ with the given marginal frequencies. The attained significance level is then defined to be the sum of the probabilities of all tables in S for which the value of the statistic is at least as large as the value of the statistic for the observed table. That is,

$$p_1 = \sum_S I(\chi^2 \geq \chi_0^2)P',$$

is the attained level for the exact conditional test based on this criterion, where χ_0^2 denotes the value of the χ^2 test statistic for the observed table. If a significance level, α, has previously been specified, then one would reject the null hypothesis if $p_1 \leq \alpha$. Notice that the test is implemented conditional upon the marginal frequencies, which are sufficient statistics for the unknown marginal proportions. The overall unconditional probability of a Type I error using this strategy is

$$P_{H_0} \text{ (reject } H_0) = E[P_{H_0} \text{ (reject } H_0 \mid \text{marginal frequencies})] \leq \alpha,$$

where the expectation is taken with respect to the distribution of all sets of marginal frequencies with the same total sample size. For 2×2 tables, this procedure is equivalent to the well-known "Fisher's exact test", for which tables and computer program packages are widely available.

The exact conditional distribution of a statistic such as χ^2 may be highly discrete if the sample size is small or if the marginal frequencies are such that the set S is small in size. In order to describe how much the discreteness at the observed value affects the attained significance level, one could also report the value of

$$p_1^* = \sum_S I(\chi^2 > \chi_0^2)P'.$$

In Section 4, some numerical comparisons are made between results using this exact test for the χ^2 statistic and the test using the chi-square distribution as an approximation. The poorness of the approximation for many tables suggests that this exact test would often be of practical use.

Since there is no need to restrict the exact conditional tests to the traditional chi-square type of statistic, the statistic used to rank the tables could alternatively be a measure with a stronger operational interpretation. For example, Goodman and Kruskal [1954] introduced an asymmetric measure of association for nominal scale variables called tau, which measures the proportional reduction in error obtained when an independent variable is used for proportional prediction of a dependent variable. Tau ranges in value between zero and one, where a value of zero is equivalent to independence (all $f_{ij} = e_{ij}$), and a value of one occurs when, for each category of the independent variable, all observations fall into only one category of the dependent variable. Thus, one could base the exact test of independence vs. the alternative of dependence on

$$p_2 = \sum_S I(\tau \geq \tau_0)P' \quad \text{or} \quad p_2{}^* = \sum_S I(\tau > \tau_0)P',$$

where τ_0 is the value of tau for the observed table. The sampling distribution of tau also is only known asymptotically [Goodman & Kruskal, 1963, 1972], so this exact test is most useful in the small sample case, as with other alternatives to the traditional chi-square test.

Some examples of the computer time required for conducting the above tests are given in Section 4. Given the computing feasibility, the exact conditional attained levels of statistics such as χ^2 and τ should be evaluated in many practical situations. Information is lost or obscured whenever categories are omitted or combined for the purpose of improving the chi-square approximation. Dependence existing between the more complete classifications may no longer be exhibited. For example, $p_1 = .004$ ($p_1{}^* = .004$) for Table 1, but if Categories two and three of both the row and column classifications are combined to fulfill typical requirements for goodness of the chi-square approximation, a non-significant value of $\chi^2 = .009$ results. In addition, many tables cannot be collapsed meaningfully. The data† in Table 2, for example, represent twenty students classified according to the choice of whether to have each question on an exam graded immediately upon completion (I) or after the entire examination had been completed (C), and according to the proctor administering the exam. Unless the proctors themselves were classified with respect to some variable, it would not be reasonable to combine proctors in investigating whether their attitudes influenced the students' choices. Thus the chi-square test would be inappropriate. The table can be analyzed using an exact conditional test, however; it shows significance at the $p_1 = .043$ ($p_1{}^* = .030$) level using the χ^2 criterion.

The small sample effect of using the chi-square distribution for the unadjusted χ^2 statistic will in most cases be unknown. The examples described in Section 4 (see Table 7) contain good approximations as well as poor approximations in both underestimating and overestimating the exact significance level. This is not surprising, since typically the exact conditional distributions are highly discrete. A modified statistic such as the one suggested by Nass [1959] would probably improve the approximation, but still in some situations a researcher should worry about the unknown size of error and may thus prefer to use a more exact approach. In fact, in some

† Obtained from C. R. Lea, Department of Psychology, University of Florida, and K. A. Lockhart, Department of Psychology, Western Michigan University.

TABLE 1

10	1	6
3	5	0
5	0	1

TABLE 2

Choice of Grading Method

		I	C
	1	2	1
	2	1	3
	3	4	0
Proctor	4	0	2
	5	0	3
	6	1	1
	7	0	2

situations it is natural to consider the marginal frequencies to be fixed, in which case these exact procedures are certainly to be preferred.

Of course, the attained significance levels using different statistics in an exact conditional test need not be identical. In our experience, though, the attained significance levels for the criteria discussed above are usually similar, if not equal. For example, in Table 2, $p_1 = .043$ ($p_1^* = .030$) and $p_2 = .046$ ($p_2^* = .030$, $\tau = .601$, treating choice of grading method as the dependent variable).

The specific tests discussed thus far are exact α-level conditional procedures for the null hypothesis of independence versus the *general* alternative of dependence. If one is especially interested in some *particular* alternative and would like the protection of higher power for that alternative (at the same time losing power for some less interesting alternatives), an overall type of test should not be used. In the next section, an example is provided of a similar procedure designed for testing against a particular type of alternative.

3. Other Alternative Hypotheses for Exact Tests of Independence

The exact test procedure can clearly be generalized to provide exact conditional tests of the null hypothesis of independence against alternative hypotheses more specific and of greater interest in the study of cross-classification tables than that of dependence (non-independence). For example, several measures have been formulated to describe the degree of various types of association between two variables.

As an example, Kendall's tau measures the difference between the proportions of concordant and discordant pairs of observations for variables with ordered observations. A generalized version of this measure for cross-classification tables (called Kendall's τ_b) corrects for pairs of observations tied with respect to at least one of the categorizations. Letting

$$C = \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij} \left(\sum_{i'>i} \sum_{j'>j} f_{i'j'} \right)$$

and

$$D = \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij} \left(\sum_{i'>i} \sum_{j'<j} f_{i'j'} \right)$$

denote the numbers of concordant and discordant pairs, respectively,

$$\tau_b = \frac{(C - D)}{\left\{ \left[\frac{n(n-1)}{2} - \frac{1}{2} \sum_{i=1}^{r} f_{i.}(f_{i.} - 1) \right] \left[\frac{n(n-1)}{2} - \frac{1}{2} \sum_{j=1}^{c} f_{.j}(f_{.j} - 1) \right] \right\}^{1/2}}.$$

The value of τ_b ranges between -1 and $+1$; τ_b is zero under the condition of independence.

The random sample version t_b of τ_b is asymptotically normally distributed about τ_b [see Agresti, 1976]. Our investigations, however, have provided evidence that the normal approximation may be quite poor for small sample sizes, especially if σ is replaced by its maximum likelihood estimate $\hat{\sigma}$. Table 3 summarizes the results of a simulation study in which the samples are generated randomly according to a bivariate normal distribution (with correlation $\rho = 0, .2, .5, .8$). They are then categorized into a 4×4 table according to the quartile of each marginal distribution into which the observation falls. The corresponding τ_b values are 0, .146, .368, and .641 [Agresti, 1976, Table 2]. Even for this "nice" (equal marginal proportions) case, the proportion of times that $|t_b - \tau_b|/\hat{\sigma}$ exceeds the normal percentage point $Z_{\alpha/2}$ is consistently larger than α. The approximation is especially poor for small

TABLE 3

The number of times in 1000 samples from a bivariate normal distribution with correlation ρ that $|t_b - \tau_b|/\hat{\sigma} > z_{\alpha/2}$, for 4x4 cross-classifications with equal marginal proportions (Underlined values are more than two standard deviations $(2\sqrt{1000\alpha(1-\alpha)})$ from expected value 1000α).

Sample Size	α	ρ = 0	.2	.5	.8
10	.10	205	202	208	218
	.05	153	142	161	179
	.01	82	77	93	127
20	.10	160	171	152	151
	.05	107	116	107	106
	.01	48	50	55	62
30	.10	129	124	127	136
	.05	70	73	86	93
	.01	23	26	26	36
40	.10	127	120	118	118
	.05	76	65	70	79
	.01	23	24	33	33
50	.10	116	114	116	123
	.05	64	56	68	66
	.01	15	21	25	24
70	.10	108	106	105	130
	.05	50	58	58	77
	.01	14	19	16	33

values of n and small values of α. In these cases, a test of $H_0 : \tau_b = 0$ using the large sample statistic $Z = t_b/\hat{\sigma}$ would yield an attained significance level much less than the true (unknown) level. A more extensive study by Rosenthal [1966] for a related measure (gamma) showed that such approximations tend to further deteriorate when the marginal proportions are very different. Both investigations reveal a tendency for the maximum likelihood value $\hat{\sigma}$ to underestimate σ. In the extreme case in which $t_b = 1$, $\hat{\sigma} = 0$. The bias of $\hat{\sigma}$ is likely to be at least partially responsible for the poorness of the approximations.

A logical alternative procedure to a z-test for small samples is to perform an exact conditional test of the null hypothesis of independence against the alternative that the population value of Kendall's τ_b is non-zero (i.e., that the proportions of concordant and discordant pairs of observations are unequal in the real or conceptual population from which the data were sampled). That is, one could calculate

$$p_3 = \sum_S I(|\tau_b| \geq |\tau_{b,0}|)P' \quad \text{or} \quad \underline{p_3^* =} \sum_S I(|\tau_b| > |\tau_{b,0}|)P',$$

where $\tau_{b,0}$ is the value of τ_b for the observed table. The null hypothesis would be rejected at (preordained) level α if $p_3 \leq \alpha$.

Since the exact test procedures using the individual probabilities, the χ^2 statistic, the likelihood ratio statistic, or Goodman and Kruskal's tau all ignore any natural ordering among the categories of the two classifications, the exact test procedure using Kendall's τ_b as the test statistic is more powerful in rejecting a false null hypothesis of independence for many underlying bivariate distributions. This is also true, of course, in the asymptotically formulated tests for ordered categorical data [Proctor, 1973]. For example, the attained significance level for the positive trend displayed in Table 4 is .053 using τ_b as the criterion, but is .514 using the exact χ^2 test.

In practical applications, $2 \times c$ tables often arise in comparing two groups with respect to some variable with c ordered levels. Independence here corresponds to homogeneity, or equality of the two discrete distributions. It can be shown that the exact conditional test of identical discrete populations using the Mann–Whitney statistic is a special case of the exact conditional test of independence using Kendall's τ_b as the criterion. Klotz [1966] has considered the problem of enumeration of matrices for this test.

There are several other summary statistics which are often worthwhile to consider using the exact conditional test framework. For example, Goodman

TABLE 4

	High	Medium	Low
High	6	4	2
Medium	4	4	4
Low	2	4	6

and Kruskal [1954] have introduced other measures of association for cross-classification tables representing variables measured on an ordinal scale (e.g., gamma) or a nominal scale (e.g., lambda), which describe specific types of dependence. The principle behind these tests extends naturally to the multi-dimensional situation. Thus exact conditional tests can be used for alternatives phrased in terms of measures of interaction when the sample sizes are too small to apply asymptotic approximations.

4. Computation of Exact Tests, with Numerical Comparisons

The exact conditional tests of independence described in the previous two sections are seldom if ever used for tables larger than 2 $\times$ 2, mainly because of the burden of identifying all of the tables with the same marginal distributions as the observed table and computing the conditional probabilities and values of the test statistics for those tables. In this section, the extent to which these procedures can be applied using modern computing facilities is considered.

TABLE 5

IBM 370/165 CPU time in seconds for conducting simultaneously three exact tests, for tables with uniform marginal frequencies (size of conditional set S given in parentheses; time exceeds one minute for omitted entries).

Table Size	d.f.	Sample Size								
		5	10	15	20	30	40	50	70	100
2x3	2	.01 (5)	.01 (14)	.02 (27)	.02 (44)	.04 (91)	.07 (154)	.10 (234)	.19 (444)	.36 (884)
2x4	3	.01 (7)	.03 (28)	.04 (70)	.06 (146)	.18 (408)	.40 (891)	.72 (1,638)	1.81 (4,218)	4.92 (11,726)
2x5	4	.01 (10)	.03 (51)	.08 (155)	.20 (381)	.80 (1,451)	1.89 (3,951)	4.05 (8,801)	13.89 (30,381)	52.88 (116,601)
3x3	4	.01 (11)	.04 (65)	.12 (231)	.27 (546)	1.10 (2,211)	3.02 (6,020)	6.86 (13,566)	23.35 (47,450)	
2x6	5	.01 (10)	.04 (70)	.13 (273)	.39 (826)	1.99 (4,332)	6.41 (14,476)	16.65 (38,802)		
2x7	6	.01 (10)	.06 (96)	.26 (483)	.85 (1,672)	5.16 (11,008)	21.31 (46,398)			
3x4	6	.02 (18)	.10 (180)	.57 (993)	1.97 (3,600)	13.42 (25,191)	57.05 (110,328)			
3x5	8	.03 (30)	.31 (440)	2.19 (3,391)	9.66 (16,250)					
4x4	9	.04 (33)	.44 (626)	4.53 (6,241)	29.02 (40,176)					

March [1972] has developed a subroutine for calculating the conditional probabilities needed for these tests. Boulton [1974] improved March's routine and thus decreased the amount of computer time needed to identify the tables in S. We have developed a FORTRAN subroutine (incorporating Boulton's routine for identifying the tables in S and calculating their probabilities) for conducting all of the exact conditional tests for $r \times c$ tables described in this paper, and we have measured the computing time for various values of $r \leq c$ and various sample sizes. For a given table dimension and sample size, the table entries were designed to keep the row marginal frequencies equal or within at most one of each other, and similarly for the column marginal frequencies. In general, for a given sample size, the number of tables in the set S is much larger when the marginal frequencies are uniform than when the row or column marginal frequencies are highly unequal. Thus if an exact test of independence is manageable for an $r \times c$ table with sample size n_0 and uniform marginal distributions, it should also at least be manageable for *any* $r \times c$ table with sample size $n \leq n_0$ and any marginal frequencies such that

$$\sum_{i=1}^{r} f_{i.} = \sum_{j=1}^{c} f_{.j} = n.$$

For various table dimensions and sample sizes with uniform marginal distributions, Table 5 lists the number of tables in S and the IBM 370/165 computer (CPU) time spent in the FORTRAN subroutine for calculating p, p_1, and p_2. The times listed in Table 5 should be interpreted as an approximate gauge, since CPU time will vary from one computer model to another. The memory requirement for the entire program is about 40,000 bytes, of which approximately 15,000 are allotted for the subroutine.

Table 5 indicates the practical limits of the exact conditional tests. Since the computer time increases very rapidly for tables with sample sizes larger than those given in Table 5 (at least when the marginal frequencies are uniform), the potential user should exercise caution when the table dimensions or sample sizes exceed those listed. However, notice that when

TABLE 6

60	4	1	0
14	5	4	1
3	3	3	2

the degrees of freedom for a table dimension do not exceed approximately six, these exact tests can be simultaneously conducted economically for most of the sample sizes for which one might doubt the goodness of the chi-square approximation for the distribution of the χ^2 statistic. When the marginal frequencies are highly unequal, the exact tests are much more economical. For example, Table 6 is a 3×4 table with a sample size of 100, yet S contains 33,675 distinct tables and the computer time for simultaneously conducting these exact tests was only 18.77 seconds. The computing times using Kendall's τ_b as the test statistic (when both sets of marginal categories are naturally ordered) are very similar. A copy of the FORTRAN subroutine which we have used is available upon request.

The exact test procedures are used in studies concerning how small the sample size may be for various approximation techniques to work well. That is, conditional approximate procedures can be compared to the conditional exact tests under the null hypothesis of independence so that guidelines can be developed as to when the exact tests must be used. For example, for several table dimensions and sample sizes, we have compared the percentage points of the chi-square distribution (which is the asymptotic distribution of the χ^2 statistic even for fixed marginal frequencies) to the corresponding percentage points of the exact conditional distribution of the χ^2 statistic, under the null hypothesis of independence. Table 7 compares the nominal significance level α (for $\alpha = .01, .05, .10$) to the exact probability of exceeding the $100(1 - \alpha)$ percentage point of the chi-square distribution. When the frequencies within each marginal distribution are equal, the chi-square distribution gives a good approximation for a sample size of about 30, for the table dimensions considered. Otherwise, though, Table 7 reaffirms the difficulty of making general statements about the robustness of the chi-square test for small samples. In some cases the approximation is quite good, whereas in others it is clearly inadequate for the percentage points chosen. Moreover, the poorness of the approximation is not always in the conservative direction.

One reason for the inconsistency in the goodness of the chi-square approximation is that the exact conditional distribution may be highly discrete for small sample sizes. Thus, the precision of the approximation described in Table 7 at a given percentage point of the chi-square distribution is affected by whether there is a substantial jump in the exact distribution function just below or just above that point. As an example, Table 8 displays the exact conditional distributions of the χ^2 statistic in the upper 10% of the chi-square distribution with 2 d.f., for six of the 2×3 tables used in Table 7. Notice that even when $n = 30$ and the expected frequencies all equal five, $\chi^2 = 5.600$ has a probability of occurrence of .088. When n is increased to 50, there are 52 distinct χ^2 values in the upper tail (above 4.605) for the 2×3 table with marginal frequencies (25, 25) and (16, 17, 17), and the maximum probability is only .021. When the marginal frequencies

TABLE 7

Comparison of exact conditional distribution of χ^2 to approximate conditional (chi-square) distribution of χ^2 under H_o: independence. (Tabled are the exact conditional probabilities that χ^2 exceeds the (.10, .05, .01) upper percentage points of the chi-square distribution.)

		Table dimension								
		2x3			2x5			3x3		
row marg. dist.		(.5,.5)	(.8,.2)	(.8,.2)	(.5,.5)	(.8,.2)	(.8,.2)	(.33,.33,.33)	(.6,.3,.1)	(.6,.3,.1)
col. marg. dist.		(.33, .33, .33)	(.33, .33, .33)	(.6, .3, .1)	(.2, .2, .2, .2, .2)	(.2, .2, .2, .2, .2)	(.4, .2, .2, .1, .1)	(.33, .33, .33)	(.33, .33, .33)	(.6, .3, .1)
sample size	percentage points									
10	.10	.095	.133	.267	.238	.111	.067	.117	.050	.111
	.05	.095	.0	.0	.0	.111	.067	.049	.020	.111
	.01	.0	.0	.0	.0	.0	.0	.004	.0	.001
20	.10	.106	.075	.092	.147	.067	.090	.107	.081	.064
	.05	.058	.018	.043	.069	.067	.090	.049	.036	.022
	.01	.011	.018	.020	.006	.001	.006	.007	.003	.007
30	.10	.122	.122	.173	.102	.043	.058	.130	.101	.092
	.05	.034	.027	.045	.038	.043	.039	.067	.036	.061
	.01	.014	.001	.010	.008	.009	.012	.005	.004	.015
50	.10	.084	.095	.126	.107	.096	.108	.101	.096	.099
	.05	.056	.050	.031	.056	.044	.036	.050	.041	.041
	.01	.009	.005	.010	.009	.012	.013	.009	.007	.011

TABLE 8

The conditional distribution of the χ^2 statistic in the upper 10% of the chi-square distribution with 2 d.f., for some 2x3 tables (The .10, .05, and .01 upper percentage points of the chi-square distribution are 4.605, 5.991, and 9.210).

Marginal Frequencies	χ^2 value	probability	Marginal Frequencies	χ^2 value	probability
(5,5),(3,3,4)	6.000	.048	(2,8),(1,3,6)	4.792	.133
n=10	7.333	.048	n=10	5.833	.133
(10,10),(6,7,7)	5.524	.048	(4,16),(2,6,12)	4.896	.050
n=20	6.381	.032	n=20	7.500	.008
	7.143	.011		9.063	.015
	8.571	.005		9.583	.014
	9.714	.005		11.667	.006
	10.952	.003			
	11.238	.002			
	13.143	.001			
	14.000	<.001			
(15,15),(10,10,10)	5.600	.088	(6,24),(3,9,18)	4.653	.037
n=30	7.200	.020	n=30	5.000	.092
	9.600	.006		6.042	.011
	10.400	.007		6.389	.015
	12.800	.001		7.431	.008
	15.200	<.001		10.208	.004
	16.800	<.001		11.944	.001
	20.000	<.001		13.333	.002
				14.375	.002
				17.500	<.001

are (40, 10) and (30, 15, 5), there are 23 distinct values in the upper tail, and the maximum probability is .033.

5. *Conclusion*

The primary argument in this paper has been that exact conditional tests of the null hypothesis of independence should be applied to many cross-classification tables in which available approximations (such as the chi-square distribution for the χ^2 statistic) would be questionable and collapsing the table or eliminating categories would result in an intolerable loss of information. We have shown how such tests are especially useful for considering specific alternative hypotheses, such as one concerning the value of Kendall's τ_b when the categories are ordered. Further encouragement for the use of these exact conditional tests was provided by the fact that for many table dimensions, the procedures are easily manageable on the computer for nearly any set of frequencies for which application of the asymptotic approximating distributions would be dubious. The exact tests can also be used to gauge the accuracy of approximate techniques for small samples.

REFERENCES

Agresti, A. The effect of category choice on some ordinal measures of association. *Journal of the American Statistical Association*, 1976, *71*, 49–55.

Boulton, D. M. Remark on algorithm 434. *Communications of the ACM*, 1974, *17*, 326.

Freeman, G. H. & Halton, J. H. Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, 1951, *38*, 141–149.

Goodman, L. A. & Kruskal, W. H. Measures of association for cross classifications. *Journal of the American Statistical Association*, 1954, *49*, 732–764.

Goodman, L. A. & Kruskal, W. H. Measures of association for cross classifications III: approximate sampling theory. *Journal of the American Statistical Association*, 1963, *58*, 310–364.

Goodman, L. A. & Kruskal, W. H. Measures of association for cross classifications IV: simplification of asymptotic variances. *Journal of the American Statistical Association*, 1972, *67*, 415–421.

Haldane, J. B. S. The use of χ^2 as a test of homogeneity in a $(n \times 2)$-fold table when expectations are small. *Biometrika*, 1943, *33*, 234–238.

Halton, J. H. A rigorous derivation of the exact contingency formula. *Proceedings of the Cambridge Philosophical Society*, 1969, *65*, 527–530.

Klotz, J. The Wilcoxon, ties, and the computer. *Journal of the American Statistical Association*, 1966, *61*, 772–787.

March, D. L. Algorithm 434, exact probabilities for $r \times c$ contingency tables. *Communications of the ACM*, 1972, *15*, 991–992.

Nass, C. A. G. The chi-square test for small expectations in contingency tables, with special reference to accidents and absenteeism. *Biometrika*, 1959, *46*, 365–385.

Proctor, C. H. Relative efficiencies of some measures of association for ordered two-way contingency tables under varying intervalness of measurement error. *Proceedings of the Social Statistics Section, American Statistical Association*, 1973, 372–379.

Radlow, R. & Alf, E. Jr. An alternate multinomial assessment of the accuracy of the χ^2 test of goodness of fit. *Journal of the American Statistical Association*, 1975, *70*, 811–813.

Roscoe, J. T. & Byars, J. A. An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. *Journal of the American Statistical Association*, 1971, *66*, 755–759.

Rosenthal, I. Distribution of the sample version of the measure of association, gamma. *Journal of the American Statistical Association*, 1966, *61*, 440–453.

Manuscript received 4/14/75
Revised Manuscript received 10/6/75
Final version received 6/7/76