EXACT PROBABILITIES FOR FIRST-ORDER AND SECOND-ORDER INTERACTIONS IN $2 \times 2 \times 2$ CONTINGENCY TABLES

PAUL W. MIELKE, JR. AND KENNETH J. BERRY

Colorado State University

Reprint requests and correspondence should be addressed to Paul W. Mielke, Jr., Department of Statistics, Colorado State University, Fort Collins, CO 80523 or to e-mail address: mielke@lamar.colostate.edu.

EXACT PROBABILITIES FOR FIRST-ORDER AND SECOND-ORDER INTERACTIONS IN $2 \times 2 \times 2$ CONTINGENCY TABLES

[Abstract]

A FORTRAN program to calculate exact probabilities for first- and second-order interactions in $2 \times 2 \times 2$ contingency tables with fixed marginals is presented. Computational speed and accuracy are assured with the use of an arbitrary constant for the initial table and recursively defined values for all subsequent tables.

EXACT PROBABILITIES FOR FIRST-ORDER AND SECOND-ORDER INTERACTIONS IN $2 \times 2 \times 2$ CONTINGENCY TABLES

It is occasionally necessary to test the independence among three classification variables, each of which consists of two mutually exclusive classes, i.e., a $2 \times 2 \times 2$ contingency table. Mielke, Berry, and Zelterman (1994) provide an algorithm for the exact global probability value obtained from an examination of all possible permutations of the eight cell frequencies, conditioned on the realized marginals, of a $2 \times 2 \times 2$ table. An alternative approach which is not as computationally intensive and, quite possibly more fruitful, is to examine the first- and second-order interactions in a $2 \times 2 \times 2$ table, when the realized marginals are fixed. This approach was first proposed by Bartlett (1935) and has been discussed by Darroch (1962, 1974), Haber (1983, 1984), Odoroff (1970), Plackett (1962), Pomar (1984), Simpson (1951), and Zachs and Solomon (1976). In this paper an algorithm and associated FORTRAN program (EI222) are described which compute the exact probabilities of the three first-order (two-variable) interactions and the one second-order (three-variable) interaction.

Program

The logic on which program EI222 is based was first developed by Quetelet (1849) to calculate binomial probability values. Beginning with a small arbitrary initial value, a simple recursion procedure generates relative frequency values for all possible $2 \times 2 \times 2$ contingency tables, given the realized marginal frequency totals. The desired exact probability value is obtained by summing the relative frequency values equal to or less than the observed relative frequency value and dividing the resultant sum by the unrestricted relative frequency total.

Consider a sample of n independent observations arranged in a $2 \times 2 \times 2$ contingency table. Let n_{ijk} denote the observed cell frequency of the ith row, jth column, and kth slice, and let p_{ijk} denote the corresponding cell probability ($i = 1,2$; $j = 1,2$; $k = 1,2$). Also let $n_{.jk}$, $n_{i.k}$, $n_{ij.}$, $n_{i..}$, $n_{.j.}$, $n_{..k}$, and $n_{...}$ indicate the observed marginal frequency totals of the $2 \times 2 \times 2$ contingency table and let the corresponding marginals over p_{ijk} be indicated by $p_{.jk}$, $p_{i.k}$, $p_{ij.}$, $p_{i..}$, $p_{.j.}$, $p_{..k}$, and $p_{...}$, respectively ($i = 1,2$; $j = 1,2$; $k = 1,2$). Since the categories are mutually exclusive and exhaustive, $n_{...} = n$ and $p_{...} = 1$.

The null hypotheses for the three first-order interactions are

$$H_{01}: p_{.11}p_{.22} = p_{.12}p_{.21},$$

$$H_{02}: p_{1.1}p_{2.2} = p_{1.2}p_{2.1},$$

and

$$H_{03}: p_{11.}p_{22.} = p_{12.}p_{21.}$$

(Bartlett, 1935). Program EI222 utilizes the exact probability routine presented in Berry and Mielke (1985) and also in Mielke and Berry (1992) for the three first-order interaction probabilities, since each is only an analysis of a 2×2 contingency table collapsed over rows, columns, and slices, respectively, from the original $2 \times 2 \times 2$ table.

The null hypothesis for the second-order interaction is

$$H_{04}: p_{111}p_{122}p_{212}p_{221} = p_{112}p_{121}p_{211}p_{222}$$

(Bartlett, 1935; Haber, 1984, O'Neill, 1982). For simplicity, set $x = n_{111}$, $a = n_{.11}$, $b = n_{1.1}$, $c = n_{11.}$, $A = n_{1..}$, $B = n_{.1.}$, $C = n_{..1}$, and $n = n_{...}$. The point probability (P) of any x is given by

$$P(x|a,b,c,A,B,C,n) =$$
$$A!(n-A)!B!(n-B)!C!(n-C)!/$$
$$[(n!)^2 x!(a-x)!(b-x)!(c-x)!(A-b-c+x)!$$
$$(B-a-c+x)!(C-a-b+x)!(n-A-B-C+a+b+c-x)!].$$

If $H(k)$, given a, b, c, A, B, C, and n, is a recursively defined positive function, then solving the recursive relation $H(k+1) = H(k) \cdot g(k)$ yields

$$g(k) = \frac{(a-k)(b-k)(c-k)(n-A-B-C+a+b+c-k)}{(k+1)(A-b-c+k+1)(B-a-c+k+1)(C-a-b+k+1)},$$

which may be employed to enumerate the complete distribution of $P(k|a,b,c,A,B,C,n)$, $v \le k \le w$, where

$$v = \max(0,\ b+c-A,\ a+c-B,\ a+b-C),$$

$$w = \min(a,\ b,\ c,\ n-A-B-C+a+b+c),$$

and where $H(v)$ is set initially to some small value, e.g., 10^{-200}. The total (T) over the completely enumerated distribution may be found by

$$T = \sum_{k=v}^{w} H(k) .$$

The exact second-order interaction probability (P) value is found by

$$P = \sum_{k=v}^{w} I_k H(k) / T$$

where

$$I_k = \begin{cases} 1, \text{ if } H(k) \le H(x), \\ 0, \text{ otherwise.} \end{cases}$$

Example

Table 1 contains a $2 \times 2 \times 2$ contingency table, based on $n = 76$ Responses to a question (Yes, No) classified by Gender (Female, Male) in two elementary school Grades (First, Fourth).

Insert Table 1 about here

The first-order interaction probabilities associated with the data in Table 1 are 0.8134 (Grade by Gender over Response), 0.2496 (Gender by Response over Grade), and 0.4830 (Grade by Response over Gender). The second-order interaction probability is 0.9036×10^{-3}. The global probability of a table this extreme or more extreme than the observed table is 0.4453×10^{-2} (Mielke, Berry, & Zelterman, 1994).

Program Language

Program EI222 is written in FORTRAN-77 in double precision. Comment lines provide for input/output specification and documentation. Input into program EI222 consists of the eight observed cell frequencies in the $2 \times 2 \times 2$ contingency table. Output consists of the observed cell frequencies, the three exact first-order interaction probability values with their associated 2×2 contingency tables, and the exact second-order interaction probability value.

Availability

A listing of program EI222 is available from Kenneth J. Berry, Department of Sociology, Colorado State University, Fort Collins, CO 80523, or by e-mail from Internet address: berry@lamar.colostate.edu.

References

Bartlett, M. S. (1935). Contingency table interactions. *Journal of the Royal Statistical Society Supplement, 2*, 248-252.

Berry, K. J., & Mielke, P. W. (1985). Subroutines for computing exact chi-square and Fisher's exact probability tests. *Educational and Psychological Measurement, 45*, 153-159.

Darroch, J. N. (1962). Interactions in multi-factor contingency tables. *Journal of the Royal Statistical Society, B, 24*, 251-263.

Darroch, J. N. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika, 61*, 207-214.

Haber, M. (1983). Sample sizes for the exact test of 'no interaction' in $2 \times 2 \times 2$ tables. *Biometrics, 39*, 493-498.

Haber, M. (1984). A comparison of tests for the hypothesis of no three-factor interaction in $2 \times 2 \times 2$ contingency tables. *Journal of Statistical Computing and Simulation, 20*, 205-215.

Mielke, P. W., & Berry, K. J. (1992). Fisher's exact probability test for cross-classification tables. *Educational and Psychological Measurement, 52*, 97-101.

Mielke, P. W., Berry, K. J., & Zelterman, D. (1994). Fisher's exact test of mutual independence for $2 \times 2 \times 2$ cross-classification tables. *Educational and Psychological Measurement, 54*, 110-114.

Odoroff, C. L. (1970). A comparison of minimum logit chi-square estimation and maximum likelihood estimation in $2 \times 2 \times 2$ and $3 \times 2 \times 2$ contingency tables: tests for interaction. *Journal of the American Statistical Association, 65*, 1617-1631.

O'Neill, M. E. (1982). A comparison of the additive and multiplicative definitions of second-order interaction in $2 \times 2 \times 2$ contingency tables. *Journal of Statistical Computing and Simulation, 15*, 33-50.

Plackett, R. L. (1962). A note on interactions in contingency tables. *Journal of the Royal Statistical Society, B, 24*, 162-166.

Pomar, M. I. (1984). Demystifying loglinear analysis: four ways to assess interaction. *Sociological Perspectives, 27*, 111-135.

Quetelet, M. A. (1849). *Letters addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the theory of probabilities as applied to the moral and political sciences* (O. G. Downes, Trans.). London: Charles & Edwin Layton.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, B, 13*, 238-241.

Zachs, S., & Solomon, H. (1976). On testing and estimating the interaction between treatments and environmental conditions in binomial experiments: the case of two stations. *Communications in Statistics: Theory and Methods, A, 5*, 197-223.

Table 1

Cross-Classification of Responses, Categorized by Gender and Elementary School Grade

	Gender			
	Females		Males	
Grade	Yes	No	Yes	No
First	10	4	2	16
Fourth	6	11	15	12