AN EXACT SOLUTION TO AN OCCUPANCY PROBLEM: A USEFUL ALTERNATIVE TO COCHRAN'S *Q* TEST

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

Colorado State University

Request reprints from Paul W. Mielke, Jr., Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877

AN EXACT SOLUTION TO AN OCCUPANCY PROBLEM: A USEFUL ALTERNATIVE TO COCHRAN'S *Q* TEST

Summary.—An exact solution to an occupancy problem is presented. Relationships between this occupancy problem, the committee problem, and Cochran's *Q* test are detailed. The exact solution of this occupancy problem may be more appropriate than Cochran's *Q* test when the number of subjects is small and the number of treatments is large.

AN EXACT SOLUTION TO AN OCCUPANCY PROBLEM: A USEFUL ALTERNATIVE TO COCHRAN'S *Q* TEST

In many research situations it is necessary to examine a sequence of observations on a group of subjects, where each observation is classified in one of two ways. For example, suppose a success (1) or failure (0) is recorded for each of $n \geq 2$ subjects on each of $g \geq 2$ motor skill tasks. The standard test in these cases is Cochran's Q test (Cochran, 1950) which is asymptotically distributed as the chi-squared distribution with $g - 1$ degrees of freedom. When the asymptotic distribution of Cochran's Q test is not appropriate, the nonasymptotic version of the Q test (Mielke & Berry, 19XX) provides improvement.

When n is small (e.g., $2 \leq n \leq 6$) and g is large (e.g., $20 \leq g \leq 400$), an exact test may be preferable to the asymptotic test of Cochran (1950) and the nonasymptotic version of the Q test (Mielke & Berry, 19XX). Such research conditions arise for a number of reasons. First, a long-term panel study is proposed but few subjects are willing to make a research commitment due to the extended time of the research and/or the treatment is either distasteful or time-intensive for the subjects. Second, a longitudinal study begins with an adequate number of subjects, but there is a high drop-out rate and survival analysis cannot be justified. Third, very few subjects satisfy the research protocol. Fourth, the cost of each observation/treatment is expensive for the researcher. Fifth, subjects are very expensive, as in primate studies. Sixth, a pilot study with a few number of subjects may be implemented to establish the validity of the research prior to applying for funding for a larger study.

METHOD

Consider an $n \times g$ occupancy matrix with n subjects (rows) and g treatments (columns). Let X_{ij} denote the observation of the ith subject ($i = 1,\ldots,n$) in the jth treatment ($j = 1,\ldots,g$) where a success is coded 1 and a failure is coded 0. For any subject, a success might result from the treatment administered or it might result from some other cause or a random response (i.e., a false positive). Therefore, a successful treatment response is counted only when all n subjects score a success, i.e., a full column of 1 values. Clearly, this approach does not generalize well to a great number of subjects since it is unrealistic for a large number of subjects to respond in concert. The asymptotic Q test of Cochran (1950) or the nonasymptotic version of the Q test (Mielke & Berry, 19XX) should be used when n is large.

Let

$$R_i = \sum_{j=1}^{g} X_{ij} \qquad [1]$$

for $i = 1,\ldots,n$ denote subject (row) totals, let

$$M = \prod_{i=1}^{n} \binom{g}{R_i} \qquad [2]$$

denote the total number of possible configurations of distinct matrices, and let $v = min(R_1,\ldots,R_n)$. The null hypothesis (H_0) stipulates that each of the M distinguishable configurations of 1s and 0s within each of the n rows, given that the $R_1,\ldots,R_n$ values are fixed, occurs with equal probability. If A_j is the number of distinct configurations where exactly j treatments (columns) are filled with successes (1s), then

$$A_v = \binom{g}{v}\left[\prod_{i=1}^{n} \binom{g-v}{R_i - v}\right], \qquad [3]$$

and

$$A_j = \binom{g}{j}\left[\binom{g-j}{R_i - j} - \sum_{k=j+1}^{v} \binom{g-j}{k-j} \frac{A_k}{\binom{g}{k}}\right] \qquad [4]$$

where $0 \le j \le v - 1$. Then, as defined,

$$M = \sum_{j=0}^{v} A_j \qquad [5]$$

and the exact probability (P) of observing s or more treatments (columns) completely filled with successes (1s) is given by

$$P = \frac{1}{M} \sum_{j=s}^{v} A_j \qquad [6]$$

where $0 \leq s \leq v$.

EXAMPLES

Two example data sets are considered. The first example is from Mantel (1974) and the second example is constructed to illustrate a typical long-term panel study.

Example 1. Consider an experiment with $n = 6$ subjects and $g = 8$ treatments (Mantel, 1974). The R_i (subject total) values are {4, 6, 5, 7, 4, 6} for $i = 1,\ldots,n$ and s (the number of observed treatment columns filled with 1s) is equal to 2. The exact probability of observing $s \geq 2$ is 0.0908. Treatment marginals are required to compute Cochran's Q test and the nonasymptotic version of the Q test (Mielke & Berry, 19XX). Let C_j be the observed total number of successes (1s) for the jth treatment

($j = 1,\dots,g$). The values of C_j for these data are {3, 6, 5, 4, 1, 3, 6, 4}. Cochran's Q test for this data set is $Q = 14.3590$ with a probability of 0.0452; the nonasymptotic version of the Q test yields a probability of 0.0309. Both Cochran's Q test and the nonasymptotic version of the Q test reject H_0 at the customary $\alpha = 0.05$, while the exact occupancy test yielded a probability of 0.0908.

Example 2. Consider a long-term longitudinal study where $n = 4$ subjects are observed daily for $g = 140$ days. The R_i (subject total) values are {23, 16, 12, 31} for $i = 1,\dots,n$ and s (the number of observed treatment columns filled with 1s) is equal to 3. The exact probability of $s \geq 3$ is 1.0731×10^{-5}. The values for C_j consisted of $C_j = 4$ for three treatments, $C_j = 1$ for 70 treatments, and $C_j = 0$ for 67 treatments. In contrast to the previous example, Cochran's Q test for these data is 141.9858 with a probability of 0.4138 and the nonasymptotic probability value is 0.4004, whereas the exact occupancy test yielded a probability of 1.0731×10^{-5}.

GENERAL DISCUSSION

This occupancy problem was first identified and solved by Mielke and Siddiqui (1965) and was recast and published as the committee problem by Mantel and Pasternack (1968) for committees of equal size. While this occupancy problem considers n subjects, g treatments, and scores a success as a 1 and a failure as a 0, the committee problem considers n committees, g individuals, and interchanges the 0 and 1 values. In the committee problem the question of interest is the probability that an individual will belong to no committees. Gittelsohn (1969) offered a moment generating function approach to the committee problem, Sprott (1969) generalized the problem to committees of unequal sizes, and White (1971) suggested alternative arithmetic solutions to the committee problem. Mantel (1974) acknowledged that the work of Mielke and Siddiqui (1965) on this occupancy problem correctly anticipated the committee problem and detailed the relationship between this occupancy problem and the committee problem with unequal committee sizes. Finally, while this occupancy problem focuses on a combinatorial solution which yields exact results, Cochran's Q test relies on a statistic which is governed by the same underlying distribution as this occupancy problem. It should be emphasized that the occupancy approach is appropriate for those cases involving a small number of subjects and a large number of treatments, whereas Cochran's Q test is designed for those cases involving larger numbers of subjects.

REFERENCES

COCHRAN, W. G. (1950) The comparison of percentages in matched samples. *Biometrika*, 37, 256-266.

GITTELSOHN, A. M. (1969) An occupancy problem. The *American Statistician*, 23, 11-12.

MANTEL, N. (1974) Approaches to a health research occupancy problem. *Biometrics*, 30, 355-362.

MANTEL, N., & PASTERNACK, B. S. (1968) A class of occupancy problems. *The American Statistician*, 22, 23-24.

MIELKE, P. W., JR., & SIDDIQUI, M. M. (1965) A combinatorial test for independence of dichotomous responses. *Journal of the American Statistical Association*, 60, 437-441.

MIELKE, P. W., JR., & BERRY, K. J. (19XX) Nonasymptotic inferences based on Cochran's Q test. *Perceptual and Motor Skills*, XX, XXX-XXX.

SPROTT, D. A. (1969) A note on a class of occupancy problems. *The American Statistician*, 23, 12-13.

WHITE, C. (1971) The committee problem. *The American Statistician*, 25, 25-26.