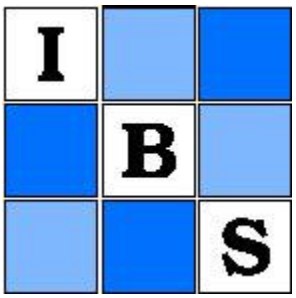

BIOMETRICS 52, 1103–1111
September 1996

THE CONSULTANT'S FORUM

Order-Restricted Tests for Stratified Comparisons of Binomial Proportions

Alan Agresti and Brent A. Coull

Department of Statistics, University of Florida,
Gainesville, Florida 32611-8545, U.S.A.

SUMMARY

The data set presented relates a binomial response to ordered levels of an explanatory variable, representing doses of a drug, with data collected at several centers. A study goal is to test independence of the response and the ordinal factor, assuming under the alternative only that the binomial parameter is a monotonically increasing function of the ordinal predictor. We present two likelihood-ratio tests that are sensitive to order-restricted alternatives. Simulating the exact distributions of the test statistics yields nearly exact P-values. We also discuss related analyses for comparing two groups on an ordinal response, and we propose a test that is sensitive to a stochastic ordering alternative.

1. Introduction

Table 1 was shown to the first author in a recent consultation with a pharmaceutical company. At each of 13 centers, subjects were randomly assigned to three dose levels of a drug for treating a certain medical condition. For each center, Table 1 shows the number of observations and the number of 'success' responses at each dose level. Some centers did not have observations at all doses. One study objective was to test the hypothesis of conditional independence of response and dose, given center. The study's investigators expected the probability of success to increase with dosage level. However, they preferred not to assume a particular functional form, such as logistic, for the relationship. They wanted to test conditional independence using a test statistic designed for the alternative that the probability of success is monotonically increasing in dose.

This article proposes order-restricted analyses for a binary response variable that has an explanatory variable with I ordered levels for data that are stratified by K levels of a control variable. For concreteness, we refer to the explanatory levels as 'treatments' and the levels of the control variable as 'centers.' Let n_{ik} denote the number of observations for treatment i at center k, and let Y_{ik} denote the number of 'successes.' We treat Y_{ik} as a binomial variate with success probability π_{ik}. We construct tests of the hypothesis of no treatment effect, H_0: $\pi_{1k} = \cdots = \pi_{Ik}$, for $k = 1, \ldots, K$. This is the hypothesis of conditional independence of treatment and response, given center. The tests refer to the order-restricted alternative, H_a: $\pi_{1k} \leq \cdots \leq \pi_{Ik}$, for $k = 1, \ldots, K$.

For data from a single center ($K = 1$), large-sample solutions exist for this ordered alternative. Bartholomew (1959) and Shi (1991) constructed test statistics based on the approximate normality of the sample proportions, and Robertson, Wright, and Dykstra (1988, p. 167) presented the likelihood-ratio statistic. These test statistics have large-sample chi-bar-squared distributions.

This article presents two extensions of the likelihood-ratio test statistic to three-way tables. These two approaches differ in terms of whether they permit patterns of association to vary among centers. Because Table 1 has small counts, rather than relying on large-sample theory, we simulate an exact distribution of the likelihood-ratio statistics to generate 'nearly exact' P-values.

Key words: Conditional independence; Dose-response curve; Isotonic regression; Logistic regression; Loglinear models; Stochastic ordering.

Table 1
A binomial response classified by dose and center, with two order-restricted fits

Center	Dose	Number of trials	Number of successes	Proportion successes	Homogeneous fit	Heterogeneous fit
1	1	1	0	0.000	0.486	0.000
	2	1	1	1.000	0.486	0.800
	3	4	3	0.750	0.757	0.800
2	1	1	1	1.000	0.355	0.500
	3	1	0	0.000	0.645	0.500
3	1	7	4	0.571	0.361	0.385
	2	6	1	0.167	0.361	0.385
	3	2	1	0.500	0.651	0.500
4	1	1	0	0.000	0.000	0.000
	2	2	0	0.000	0.000	0.000
	3	2	0	0.000	0.000	0.000
5	1	2	2	1.000	0.775	0.857
	2	1	1	1.000	0.775	0.857
	3	4	3	0.750	0.919	0.857
6	1	12	9	0.750	0.801	0.750
	2	10	8	0.800	0.801	0.800
	3	9	9	1.000	0.930	1.000
7	1	6	5	0.833	0.847	0.833
	2	5	5	1.000	0.847	0.909
	3	6	5	0.833	0.948	0.909
8	2	1	0	0.000	0.176	0.000
	3	2	1	0.500	0.412	0.500
9	1	2	0	0.000	0.182	0.000
	2	2	0	0.000	0.182	0.000
	3	3	2	0.667	0.423	0.667
10	1	2	0	0.000	0.000	0.000
11	1	2	1	0.500	0.495	0.400
	2	3	1	0.333	0.495	0.400
	3	2	2	1.000	0.763	1.000
12	1	4	3	0.750	0.814	0.750
	2	5	4	0.800	0.814	0.800
	3	5	5	1.000	0.935	1.000
13	1	1	0	0.000	0.000	0.000
	2	1	0	0.000	0.000	0.000
	3	1	0	0.000	0.000	0.000

Source: Allan Pallay, Wyeth-Ayerst Research.

We also discuss a related problem in which the explanatory variable is binary and the response categories are ordered. We present an order-restricted test that is sensitive to the alternative of stochastically-ordered response distributions.

2. Order-Restricted Treatment Effects in a Logit Model

In many applications, one expects the nature of the change in the probability of success as a function of the treatment level to be similar in each center. This suggests a model with a lack of interaction, on some scale, in the manner in which π_{ik} depends on treatment and center effects. For the logit scale, such a model is

$$\text{logit}(\pi_{ik}) = \alpha_k + \beta_i. \tag{1}$$

For this model, conditional independence of response and treatment is equivalent to $\beta_1 = \cdots = \beta_I$. The model does not assume a structural form for how π_{ik} varies over treatment levels, but it does assume the same type of variation for each center. We construct a test of conditional independence with power directed toward the order-restricted alternative, $\beta_1 \leq \cdots \leq \beta_I$.

When the ordinary maximum likelihood (ML) fit of this model has monotonically increasing $\{\hat{\beta}_i\}$, the ML order-restricted fit is identical to it. Otherwise, suppose the ordinary ML estimates satisfy $\hat{\beta}_i > \hat{\beta}_{i+1}$ for at least one pair of treatments i and $i+1$. Since the log likelihood for model (1) is concave, it follows from McDonald and Diamond (1983, 1990) that the ML order-restricted fit of the

model lies on the boundary of the order-restricted parameter space. The solution equates estimates of $\{\beta_i\}$ that violate the ordering for the ordinary ML fit. One refits the model by constraining $\beta_i = \beta_{i+1}$. The algorithm can utilize the fact that fitting a model while equating certain treatment parameters corresponds to fitting simpler unconstrained models to collapsed tables that combine those treatment levels. The same remarks apply to alternative potential link functions for model (1) for which the log likelihood is concave, such as the probit and complementary log-log.

Let $G^2(M)$ denote the likelihood-ratio statistic for testing the goodness of fit of model (1) to the observed $I \times 2 \times K$ contingency table. Let $G^2(M^*)$ denote the likelihood-ratio statistic for testing the fit of the order-restricted version of the model, constraining $\beta_1 \leq \cdots \leq \beta_I$. Let $G^2(CI)$ denote the likelihood-ratio statistic for testing the fit of the model of conditional independence of treatment and response; that is, model (1) with $\beta_1 = \cdots = \beta_I$. Then $G^2(CI \mid M^*) = G^2(CI) - G^2(M^*)$ is the likelihood-ratio statistic for testing conditional independence against the order-restricted alternative. It contrasts with $G^2(CI \mid M) = G^2(CI) - G^2(M)$, which tests conditional independence against a no-interaction alternative but ignores the ordering of treatment levels.

For Table 1, $G^2(CI) = 25.55$ with d.f. $= 22$. The conditional independence model apparently fits decently, though this is difficult to determine using the chi-squared reference distribution since the data are sparse. The full logit model (1) has $G^2(M) = 21.02$ with d.f. $= 20$. The test of conditional independence against that alternative has $G^2(CI \mid M) = 4.53$, based on d.f. $= 2$, providing weak evidence of conditional association ($P = 0.10$). The unrestricted ML estimates of the treatment parameters in model M (setting $\beta_1 = 0$) are $\hat{\beta}_1 = 0, \hat{\beta}_2 = -0.252$, and $\hat{\beta}_3 = 1.067$.

Since $\hat{\beta}_1$ and $\hat{\beta}_2$ for model (1) violate the order restriction, the order-restricted solution refits the model by constraining these two estimates to be identical. This fit yields $\hat{\beta}_1 = 0, \hat{\beta}_2 = 0, \hat{\beta}_3 = 1.190$, and $G^2(M^*) = 21.21$. Table 1 displays the fitted proportions of success for this order-restricted fit. Section 4 shows how to use the statistic $G^2(CI \mid M^*) = 4.34$ in an order-restricted test of conditional independence.

3. A General Order-Restricted Alternative

Model (1) assumes that the odds ratio between the response and a pair of treatment levels is identical in each center. More generally one might permit interaction, allowing the odds ratio to vary in an unrestricted manner across centers. This corresponds to the general model

$$\text{logit}(\pi_{ik}) = \beta_{ik}, \tag{2}$$

which is saturated. This model represents the most general alternative hypothesis for a test of conditional independence, for which the likelihood-ratio test statistic is $G^2(CI)$. The related order-restricted version of this model assumes that $\beta_{1k} \leq \cdots \leq \beta_{Ik}$ for $k = 1, \ldots, K$. This model poses an order restriction in each center, but makes no further assumption about how patterns of association vary across centers. The use of the logit link in model formula (2) or its order-restricted version is irrelevant, and the model is equivalent if we use any link function.

One can fit the order-restricted version of model (2) by constructing such a fit separately for each center. This can be done using the pool–adjacent–violators algorithm with the treatment sample proportions of success at each center (Ayer et al., 1955). If, for some center, sample proportions are out-of-order for a particular adjacent pair of treatments, one combines the levels, recomputes the sample proportion, and continues the comparisons. The ultimate solution does not depend on the order in which one compares treatments in applying the algorithm.

Let $G^2(I_k)$ denote the likelihood-ratio statistic for testing independence for center k alone, and let $G^2(M_k^*)$ denote the likelihood-ratio statistic for testing the fit of the order-restricted model for that center. The likelihood-ratio statistic for testing conditional independence against the order-restricted version of (2) is then $\Sigma_k[G^2(I_k) - G^2(M_k^*)] = G^2(CI) - \Sigma_k G^2(M_k^*)$.

Table 1 also shows the fitted proportions of success for the order-restricted fit of model (2). For instance, in Center 1, the sample proportions for dose levels 2 and 3 are out of order. After pooling, the common estimate for these two levels equals $(1 + 3)/(1 + 4) = 0.800$. These have the proper order, compared to dose level 1, so this is the order-restricted solution for this center. The order-restricted fit has $\Sigma_k G^2(M_k^*) = 8.31$. The likelihood-ratio statistic for testing conditional independence against this general order-restricted alternative equals $G^2(CI) - \Sigma_k G^2(M_k^*) = 17.24$.

4. Exact Tests for the Ordered Alternatives

We now discuss the distributions of the test statistics presented in the previous two sections. We first consider the order-restricted fit of the no-interaction model (1). Consider the collapsing of the $I \times 2 \times K$ table in which rows are combined that have identical treatment estimates in that

fit. The fit of model (1) to this collapsed table provides the same estimates of $\{\beta_i\}$ as does the order-restricted fit for the complete table. Let $G^2(CI')$ and $G^2(M')$ denote the likelihood-ratio goodness-of-fit statistics for conditional independence and for model (1) fitted to the collapsed table. From standard arguments on partitioning chi-squared, such as given by Agresti, Chuang, and Kezouh (1987) for the order-restricted fit of a loglinear model, it follows that

$$G^2(CI) - G^2(M^*) = G^2(CI') - G^2(M').$$

Suppose the collapsed table has $a \leq I$ treatment levels. For that particular fixed collapsing, the asymptotic distribution of $G^2(CI') - G^2(M')$ is chi-squared with $d.f. = a - 1$. The set of possible order-restricted solutions, corresponding to the various possible groupings of adjacent parameters that are equated for such a solution, relates to the set of possible collapsings of the table. Those solutions can be grouped in terms of the number of rows for the collapsing. It follows that the asymptotic null distribution of $G^2(CI) - G^2(M^*)$ is the same as that of the chi-bar-squared random variable $\Sigma_{a=1}^{I} p_a \chi^2_{a-1}$, where χ^2_{a-1} denotes a chi-squared random variable with $d.f. = a - 1$ and where p_a denotes the probability that the order-restricted solution has a distinct ordered treatment estimates.

For the order-restricted test for the more general alternative (2), the component statistic $G^2(I_k \mid M_k^*)$ for each center has a null asymptotic chi-bar-squared distribution. The weights vary among centers. The overall likelihood-ratio statistic for the general order-restricted alternative is asymptotically a sum of chi-bar-squared random variables.

Using these asymptotic results in practice to construct P-values for the order-restricted tests is problematic. First, determining weights $\{p_a\}$ for the asymptotic chi-bar-squared distribution for $G^2(CI) - G^2(M^*)$ is difficult. Even if one knew $\{\pi_{ik}\}$ and $\{n_{ik}\}$, one could not calculate the weights; the best one could do is estimate their values through simulation. For sparse data such as Table 1, estimates of these weights are poor. The situation is even worse for the test for the more general model (2), in which one would need to estimate weights $\{p_a\}$ separately for each center and then determine P-values based on a sum of chi-bar-squared variates. In any case, for either model the asymptotic distribution might well be inadequate for small samples or highly sparse data such as Table 1.

An alternative approach, involving much less approximation, estimates a P-value for the exact conditional distribution of the test statistic. In testing conditional independence, one eliminates unknown parameters by conditioning on the row and column marginal totals in each center. Given these totals, the null distribution of the cell counts in each table is multivariate hypergeometric, and counts from separate centers are independent (see, e.g., Agresti, 1992). Exact conditional distributions of $G^2(CI \mid M^*)$ and $\Sigma_k G^2(I_k \mid M_k^*)$ for the order-restricted versions of alternatives (1) and (2) are generated using these distributions.

Even when the data set contains only a few centers, it is a computationally intensive task to generate the exact conditional distribution. When the data are not too extensive, the software StatXact (Cytel Software, 1991) can generate such distributions for linear rank statistics for $I \times 2 \times K$ contingency tables, but it does not handle the order-restricted statistics considered in this article.

On the other hand, it is computationally relatively simple and inexpensive to simulate the exact conditional distribution. To do this, one randomly samples tables from the relevant multivariate hypergeometric distributions having the required fixed margins. For each generated three-way table, one computes $G^2(CI \mid M^*)$ and/or $\Sigma_k G^2(I_k \mid M_k^*)$. The estimate of the exact P-value is the proportion of sampled tables for which the statistic is at least as large as the observed value. Computing the test statistic for 50,000 generated tables ensures that the estimated P-value falls within 0.004 of the true value with probability at least 0.95; the estimate is considerably more accurate for P-values far from 0.5.

Agresti, Wackerly, and Boyett (1979), Patefield (1981, 1982), and Kreiner (1989) provided details about Monte Carlo simulation of exact distributions for contingency tables. An advantage of the simulation approach, compared to exact enumeration, is that the computation time does not grow dramatically in the sample size or the number of centers. One can easily generate a sufficiently large number of tables to estimate the exact P-value to within a desired accuracy with some fixed probability. Thus, this approach works well both for small and large sample sizes.

For Table 1, we used this approach to estimate exact P-values for various tests. Assuming independence between response and dosage for each center, we randomly generated 50,000 $3 \times 2 \times 13$ tables such that the row and column totals for each center were identical to those in Table 1. We applied Patefield's (1981) algorithm separately to each center for generation of each partial table. The 95% confidence interval for the exact P-value is based on inverting the large-sample test for a

proportion; that is, it consists of all null hypothesis proportion values not rejected in a two-sided 0.05-level test using the normal test statistic.

First, we consider the ordinary likelihood-ratio tests of conditional independence. The data are highly sparse, so estimated exact P-values are preferred to asymptotic ones. The statistic $G^2(CI) = 25.55$, based on d.f. $= 22$, has an asymptotic chi-squared P-value of 0.272. A 95% confidence interval for the exact P-value of this statistic equals (0.551, 0.560). A comparison of these reflects the tendency of the asymptotic G^2-based test to be relatively liberal for mildly sparse data. For the more focused alternative (1) of no interaction, the statistic $G^2(CI \mid M) = 4.53$, based on d.f. $= 2$, has an asymptotic chi-squared P-value of 0.104. A 95% confidence interval for the exact P-value equals (0.143, 0.150).

Next we consider the two tests of conditional independence based on order-restricted alternatives. First, using the order-restricted version of the no-interaction model (1) as the alternative yields $G^2(CI \mid M^*) = 4.34$; a 95% confidence interval for the exact P-value equals (0.068, 0.072). Next, using the order-restricted version of the saturated model (2) as the alternative yields $\Sigma_k G^2(I_k \mid M_k^*) = 17.24$; a 95% confidence interval for the exact P-value equals (0.125, 0.131).

The computational intensity for these analyses was not too horrendous. For instance, using a FORTRAN program to estimate the P-value for the statistic $\Sigma_k G^2(I_k \mid M_k^*)$ using 50,000 randomly generated tables took about 3 minutes on a Sun Sparc Station 10. The analogous test for the statistic $G^2(CI \mid M^*)$ took longer (about 50 minutes) because we used iterative proportional fitting to fit the no-interaction model for each randomly generated table.

5. Power Considerations

Either order-restricted test provides stronger evidence of an association than the corresponding test that ignores the ordering, the estimated exact P-values being 0.070 vs. 0.147 for the no-interaction alternative and 0.128 vs. 0.554 for the general alternative. Currently, power computations for the tests using estimated exact P-values would be extremely time consuming to perform. However, when the binomial parameters are truly ordered, one expects a power advantage from a procedure that utilizes the ordering compared to one that ignores it.

Asymptotically, the reasons for this are clear. We illustrate for the statistics $G^2(CI \mid M^*)$ and $G^2(CI \mid M)$ relating to logit model (1). Suppose that model holds and the $\{\beta_i\}$ are truly strictly monotone. The probability that the order-restricted fit is identical to the ordinary fit, and hence that $G^2(CI \mid M^*) = G^2(CI \mid M)$, converges to 1 as n increases. Thus, both test statistics then have the same asymptotic noncentral chi-squared distribution. The null distribution of $G^2(CI \mid M)$ is chi-squared with $d.f. = I - 1$, whereas the null distribution of $G^2(CI \mid M^*)$ is asymptotically a mixture of chi-squared distributions having d.f. values between 0 and $I - 1$. Thus, in the null case, $G^2(CI \mid M)$ is stochastically larger than $G^2(CI \mid M^*)$. Under the strictly ordered alternative, since the two statistics will tend to take similar value, it follows that $G^2(CI \mid M^*)$ will tend to be farther out in the right-hand tail of its null distribution than $G^2(CI \mid M)$ falls in the tail of its null distribution, leading to smaller P-values and a more powerful test.

An analogous argument holds for comparing the statistics $G^2(CI)$ and $\Sigma_k G^2(I_k \mid M_k^*)$ when the binomial parameters are strictly ordered in each stratum. The statistic $G^2(CI)$ is a sum of chi-squared statistics for different strata, each of which has d.f. $= I-1$; $\Sigma_k G^2(I_k \mid M_k^*)$ is stochastically smaller, being a sum of statistics each of which is a null mixture of chi-squared variates having d.f. between 0 and $I - 1$. For a rigorous argument comparing chi-squared tests having common noncentrality but differing d.f. values, see Das Gupta and Perlman (1974).

By the same reasoning, when one expects the association between treatment and response to be similar in each center, it makes sense to use the statistic $G^2(CI \mid M^*)$ rather than $\Sigma_k G^2(I_k \mid M_k^*)$. The former statistic provides greater power than the more general statistic, unless the nature of the association varies dramatically across centers. This strategy parallels the usual one for ordinary tests of conditional independence, in which it is common to use $G^2(CI \mid M)$ (or the corresponding efficient score test, the Cochran–Mantel–Haenszel test) rather than $G^2(CI)$ in order to direct the power toward a more focused alternative based on fewer degrees of freedom.

6. Order-Restricted Comparison of Two Groups on an Ordinal Response

The two order-restricted tests also apply to a rows-and-columns-reversed situation in which the explanatory variable is binary and the response variable has ordered categories. Let π_{ijk} denote the probability of response outcome j, for $j = 1, \ldots, J$, with treatment i ($i = 1, 2$) in center k; thus, $\Sigma_j \pi_{ijk} = 1$. Denote the conditional local odds ratios within centers by

$$\theta_{j(k)} = (\pi_{1,j+1,k}\pi_{2jk})/(\pi_{1jk}\pi_{2,j+1,k}), \qquad j = 1, \ldots, J-1, \quad k = 1, \ldots, K.$$

One can reverse the roles of response and explanatory variables and apply the test statistics of Sections 2 and 3. The statistic based on the approach of Section 3 permits the pattern of association to vary among centers. The order restriction used there is equivalent to $\log(\theta_{j(k)}) \geq 0$ for all j and k. The structure of uniformly nonnegative or uniformly nonpositive local log odds ratios for each table is sometimes called *likelihood-ratio dependence* (Lehmann, 1966).

The test statistic of Section 2 is based on the additional no-interaction structure for these odds ratios,

$$\theta_{j(1)} = \cdots = \theta_{j(K)}, \qquad j = 1, \ldots, J-1.$$

Denote the common value of the jth odds ratio across the K partial tables in this case by θ_j. This order-restricted no-interaction structure for an ordinal response corresponds to an adjacent-categories logit model for that response of form

$$\log(\pi_{i,j+1,k}/\pi_{ijk}) = \alpha_{jk} + \beta_j I(i = 1),$$

where $\beta_j = \log(\theta_j) \geq 0$ for all j.

In some applications with an ordinal response, one might prefer to use a weaker order-restricted condition. For instance, one could test conditional independence against the alternative of a stochastic ordering of the two response distributions in each center. For each treatment and center, let $\gamma_{ijk} = \pi_{i1k} + \cdots + \pi_{ijk}, j = 1, \ldots, J$. The stochastic ordering alternative is

$$\gamma_{1jk} \leq \gamma_{2jk}, \qquad j = 1, \ldots, J-1, \ k = 1, \ldots, K.$$

Letting

$$\lambda_{j(k)} = \gamma_{2jk}(1 - \gamma_{1jk})/\gamma_{1jk}(1 - \gamma_{2jk}),$$

this alternative has the form $\log(\lambda_{j(k)}) \geq 0$ for all j and k.

Grove (1980) and Robertson and Wright (1981) presented an order-restricted test for this alternative with $K = 1$, based on results for ML estimation with stochastic orderings by Brunk et al. (1966). For partial table k, let $r_{jk} = (n_{11k} + \cdots + n_{1jk})/(n_{21k} + \cdots + n_{2jk}), j = 1, \ldots, J$. The J columns are divided into subsets as follows: The first subset ends at the column v_1 for which $r_{v_1 k}$ is the maximum of $\{r_{1k}, \ldots, r_{Jk}\}$. If this does not include all columns, then the next subset consists of columns $\{v_1 + 1, \ldots, v_2\}$ such that $r_{v_2 k}$ is the maximum of $\{r_{v_1+1,k}, \ldots, r_{Jk}\}$. One continues in this manner, forming the collection of subsets. For instance, suppose that a particular partial table has counts (1, 4, 1, 3, 1) in row 1 and (2, 2, 2, 2, 2) in row 2. Then the ratios of partial sums are (1/2, 1.25, 1, 9/8, 1); the first subset consists of columns 1 and 2, the second subset consists of columns 3 and 4, and the final subset is column 5. Grove (1980) provided a geometric representation of this construction.

In this construction of subsets, suppose a particular subset consists of columns $a, a + 1, \ldots, b$. Then the ML fitted value under the stochastic ordering restriction for cell (i, j, k) in those columns equals

$$n_{ijk} \frac{(n_{+ak} + \cdots + n_{+bk})(n_{i+k})}{(n_{iak} + \cdots + n_{ibk})(n_{++k})}.$$

Note that when $a = b$ (i.e., a subset contains a single column), the fitted value in that cell equals the fitted value for the conditional independence model, namely $n_{+jk}n_{i+k}/n_{++k}$. The construction yields a single subset when the maximum ratio of cumulative proportions for row 1 to row 2 equals 1, the value for the final column. In that case, the sample counts themselves satisfy a stochastic ordering and the ML fitted values are simply those sample counts.

Let $G^2(S_k)$ denote the likelihood-ratio statistic for testing the fit of the stochastic-ordering restriction to the data for partial table k. Note that necessarily,

$$\sum_k G^2(S_k) \leq \sum_k G^2(M_k^*) \leq G^2(M^*),$$

since model M^* implies that model M_k^* holds for each partial table, which itself implies that S_k holds in each partial table. In partial table k, the likelihood-ratio statistic $G^2(I_k \mid S_k) = G^2(I_k) - G^2(S_k)$ compares the fitted values under independence to the fitted values under the stochastic ordering alternative. The asymptotic distribution of this statistic for testing independence is chi-bar-squared (Robertson and Wright, 1981).

The likelihood-ratio statistic for testing conditional independence against a simultaneous stochastic ordering alternative for each partial table is $\Sigma_k[G^2(I_k \mid S_k)]$. The order-restricted statistics for testing conditional independence satisfy

$$\sum_k G^2(I_k \mid S_k) \geq \sum_k G^2(I_k \mid M_k^*) \geq G^2(CI \mid M^*).$$

In the same manner as discussed in Section 4 for the other statistics, one can simulate the exact null conditional distribution of $\Sigma_k[G^2(I_k \mid S_k)]$, given the margins in each table, and estimate precisely an exact P-value for this more general ordered alternative. We illustrate with Table 1, although this analysis is somewhat unnatural for these data since the ordinal variable is the explanatory variable rather than the response. The simultaneous stochastic ordering fit has $\Sigma_k G^2(S_k) = 5.85$. Comparing this to $G^2(CI) = 25.55$ gives a likelihood-ratio statistic for an order-restricted test of conditional independence equal to 19.69. Simulating the exact distribution yields a 95% interval estimate of the exact P-value of (0.196, 0.203).

One could add further structure for the stochastic ordering alternative. For instance, one could constrain odds ratios of cumulative probabilities to be identical across strata, or one could constrain odds ratios within strata to be identical across cutpoints j for the cumulative probabilities. These structured alternatives correspond to cumulative logit models, the latter type having the *proportional odds* assumption. Such model-based alternatives make additional assumptions about structure but have the potential for increasing power, from building strength by focusing on a narrower alternative. More generally, one could also extend these tests to $I \times J \times K$ tables with ordinal rows, columns, or layers. For instance, one could extend a test of Patefield (1982) for $I \times J$ tables for which the alternative corresponds to the $(I-1)(J-1)$ local log odds ratios being uniformly nonnegative. For proportional odds models and related models with other links that assume the same treatment effects for each cutpoint and imply stochastic orderings of distributions, one can exploit results on the concavity of the log likelihood in determining order-restricted fits (Pratt, 1981).

7. Comments

It is perhaps surprising that the order-restricted literature, as extensive as it is, does not seem to contain the simple likelihood-ratio tests for three-way tables proposed here. There is, however, considerable literature on ways of handling monotonicity in bivariate analyses. Chassan (1960) gave an early attempt to construct order-restricted analyses for a single table (see also Chassan, 1962; Bennett, 1962). Morris (1988) provided confidence limits for a set of binomial parameters having a monotonicity assumption; Thomas (1983) provided nonparametric estimates for a monotone increasing hazard rate; and Schmoyer (1984) provided ML estimation when response probabilities are constrained to satisfy a sigmoidal shape. Model-based analyses for a binary response include Bacchetti (1989) using additive isotonic models, Ramsey (1972), Disch (1981), and Gelfand and Kuo (1991) using Bayesian approaches, and Geyer (1991) using logistic regression for convex response functions. Finally, Silvapulle (1994) has recently presented a class of one-sided tests that can be used for large-sample order-restricted inference about parameters in generalized linear models.

Many other ways exist of using ordering in a more structured manner to build power in testing the hypothesis of conditional independence. When the association is expected to be similar in each partial table, the most common approach is to use generalizations of the Cochran–Mantel–Haenszel test based on assigning scores (possibly rank-based) to the ordered categories (e.g., Landis, Heyman, and Koch, 1978). For $I \times 2 \times K$ tables such as discussed in Section 2, this is an efficient score test for model (1) when the $\{\beta_i\}$ satisfy the pattern $\{\beta_i = \beta u_i\}$ for the scores $\{u_i\}$ chosen for the treatment levels. The test statistics proposed in Sections 2 and 3 are designed for more general alternatives, in which $\{\beta_i\}$ in model (1) and $\{\beta_{ik}\}$ in model (2) are only assumed to be monotone across treatment levels. More generally, see Tarone and Gart (1980) for a discussion of score tests for various underlying models. For $2 \times J \times K$ tables comparing two groups on an ordinal response such as discussed in Section 6, the Landis et al. (1978) approach with midrank scores for the response categories provides an efficient score test for the special case of a proportional odds model having the same effects in each center.

A FORTRAN program for conducting the analyses described in this paper is available from the authors by e-mail or upon receipt of a formatted 3 1/2 inch floppy disk.

Acknowledgements

This research was partially supported by a grant from the National Institutes of Health. The authors thank Allan Pallay of Wyeth–Ayerst Research for raising the question that led to this solution.

They also acknowledge helpful discussions with Christy Chuang-Stein, computing assistance from Donguk Kim, and bibliographic suggestions from John W. McDonald and Zri Gilula.

Résumé

Le jeu de données présenté est une réponse binomiale ordonnée d'une variable exploratoire représentant les doses d'un produit. Les données proviennent de plusieurs centres. Un des buts de l'étude est de tester l'indépendance de la réponse et du facteur ordinal en supposant que sous l'hypothèse d'alternative le paramètre de la binomiale est une fonction monotone croissante du prédicteur ordinal. Nous présentons deux tests du rapport de vraisemblance qui sont sensibles à l'ordre des alternatives. Nous simulons les distributions exactes des tests statistiques qui donnent pratiquement les valeurs P. Puis nous discutons des analyses pour comparer deux groupes sur une réponse ordinale et nous proposons un test sensible à l'ordre aléatoire de l'alternative.

References

Agresti, A. (1992). A survey of exact inference for contingency tables (with discussion). *Statistical Science* **7,** 131–177.

Agresti, A., Chuang, C., and Kezouh, A. (1987). Order-restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association* **82,** 619–623.

Agresti, A., Wackerly, D., and Boyett, J. (1979). Exact conditional tests for cross classifications: Approximation of attained significance levels. *Psychometrika* **44,** 75–83.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics* **26,** 641–647.

Bacchetti, P. (1989). Additive isotonic models. *Journal of the American Statistical Association* **84,** 289–293.

Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika* **46,** 36–48.

Bennett, B. M. (1962). On an exact test for trend in binomial trials and its power function. *Metrika* **5,** 49–53.

Brunk, H. D., Franck, W. E., Hanson, D. L., and Hogg, R. V. (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. *Journal of the American Statistical Association* **61,** 1067–1080.

Chassan, J. B. (1960). On a test for order. *Biometrics* **16,** 119–121.

Chassan, J. B. (1962). An extension of a test for order. *Biometrics* **18,** 245–247.

Cytel Software. (1991). *StatXact: Statistical Software for Exact Nonparametric Inference*, Version 2. Cambridge, Massachusetts: Cytel Software Corporation.

Das Gupta, S. and Perlman, M. D. (1974). Power of the noncentral F-test: Effect of additional variates on Hotelling's T^2 test. *Journal of the American Statistical Association* **69,** 174–180.

Disch, D. (1981). Bayesian nonparametric inference for effective doses in a quantal-response experiment. *Biometrics* **37,** 713–722.

Gelfand, A. E. and Kuo, L. (1991). Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika* **78,** 657–666.

Geyer, C. J. (1991). Constrained maximum likelihood exemplified by isotonic convex logistic regression. *Journal of the American Statistical Association* **86,** 717–724.

Grove, D. M. (1980). A test of independence against a class of ordered alternatives in a $2 \times c$ contingency table. *Journal of the American Statistical Association* **75,** 454–459.

Kreiner, S. (1987). Analysis of multidimensional contingency tables by exact conditional tests: Techniques and strategies. *Scandinavian Journal of Statistics* **14,** 97–112.

Landis, J. R., Heyman, E. R., and Koch, G. G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review* **46,** 237–254.

Lehmann, E. (1966). Some concepts of dependence. *Annals of Mathematical Statistics* **37,** 1137–1153.

McDonald, J. W. and Diamond, I. D. (1983). Fitting generalised linear models with linear inequality parameter constraints. *The GLIM Newsletter* **8,** 29–36.

McDonald, J. W. and Diamond, I. D. (1990). On the fitting of generalized linear models with nonnegativity parameter constraints. *Biometrics* **46,** 201–206.

Morris, M. D. (1988). Small-sample confidence limits for parameters under inequality constraints with application to quantal bioassay. *Biometrics* **44,** 1083–1092.

Patefield, W. M. (1981). An efficient method of generating random $R \times C$ tables with given row and column totals. *Journal of the Royal Statistical Society, Series C* **30,** 91–97.

Patefield, W. M. (1982). Exact tests for trends in ordered contingency tables. *Journal of the Royal Statistical Society, Series C* **31,** 32–43.

Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association* **76,** 103–106.

Ramsey, F. (1972). A Bayesian approach to bioassay. *Biometrics* **28,** 841–858.

Robertson, T. and Wright, F. T. (1981). Likelihood-ratio tests for and against a stochastic ordering between multinomial populations. *Annals of Statistics* **9,** 1248–1257.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* New York: Wiley.

Schmoyer, R. L. (1984). Sigmoidally constrained maximum likelihood estimation in quantal bioassay. *Journal of the American Statistical Association* **79,** 448–453.

Shi, N.-Z. (1991). A test of homogeneity of odds ratios against order restrictions. *Journal of the American Statistical Association* **86,** 154–158.

Silvapulle, M. J. (1994). On tests against one-sided hypotheses in some generalized linear models. *Biometrics* **50,** 853–858.

Tarone, R. E. and Gart, J. J. (1980). On the robustness of combined tests for trends in proportions. *Journal of the American Statistical Association* **75,** 110–116.

Thomas, D. C. (1983). Nonparametric estimation and tests of fit for dose-response relations. *Biometrics* **39,** 263–268.

Received January 1995; *revised July* 1995; *accepted August* 1995.