A Logistic Model for Paired Comparisons with Ordered Categorical Data

P. McCullagh

Biometrika, Volume 64, Issue 3 (Dec., 1977), 449-453.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197712%2964%3A3%3C449%3AALMFPC%3E2.0.CO%3B2-1

Biometrika (1977), **64**, 3, *pp.* 449–53
Printed in Great Britain

A logistic model for paired comparisons with ordered categorical data

BY P. McCULLAGH*
Department of Mathematics, Imperial College, London

SUMMARY

Clayton (1974) has proposed some simple odds-ratio statistics for comparing two independent samples of ordered categorical data. In this paper we extend these ideas to a related model for paired data. Two types of estimator are proposed and their efficiency and consistency investigated. In the special case where there are only two categories the model reduces to Cox's (1958) model for binary paired comparisons.

Some key words: Binary paired comparisons; Multivariate binomial distribution; Ordered categorical data.

1. INTRODUCTION

Paired data frequently arise in 'before and after' experiments where a given number of individuals are measured before and after treatment to determine the treatment effect. The recorded data are often on an ordered categorical scale. We might, for example, measure degree of injury on a four category scale: uninjured, slight, serious, fatal.

Such paired data are conveniently summarized in a square $k \times k$ table of counts $\{m_{ij}\}$, where k is the number of categories and m_{ij} is the number of pairs for which the first recording is category i and the second category j. In this paper we develop statistics which, for some purposes, can be used as a convenient summary of the whole table. These statistics measure the lack of marginal homogeneity when one margin differs only in location relative to the other.

2. THE MODEL

The recorded categorical variable X is assumed to arise from some underlying continuous variable Y which is not directly observable or simply is unavailable because of censoring or grouping. For k category data it is assumed that there exist $k-1$ unknown fixed boundary points $\theta = (\theta_1, \ldots, \theta_{k-1})$ which define the categories, and that

$$X = j \quad \text{if} \quad \theta_{j-1} \leqslant Y < \theta_j \quad (j = 1, \ldots, k). \tag{2·1}$$

For convenience of notation, let $\theta_0 = -\infty$, $\theta_k = \infty$.

Any model for the Y's together with (2·1) implies a corresponding model for the recorded X's. We consider a logistic model for pairs (Y_{i1}, Y_{i2}) such that Y_{i1} has mean $\lambda_i - \frac{1}{2}\Delta$ and Y_{i2} has mean $\lambda_i + \frac{1}{2}\Delta$. Then Δ is the parameter of interest, being the common difference of the means for each pair (Y_{i1}, Y_{i2}). The block parameters $\{\lambda_i\}$ are considered to be nuisance parameters and may be fixed or random. Algebraically the model is for $i = 1, \ldots, n$

$$\begin{aligned} \text{pr}(Y_{i1} \leqslant y | \lambda_i, \Delta) &= \exp(y - \lambda_i + \tfrac{1}{2}\Delta)/\{1 + \exp(y - \lambda_i + \tfrac{1}{2}\Delta)\}, \\ \text{pr}(Y_{i2} \leqslant y | \lambda_i, \Delta) &= \exp(y - \lambda_i - \tfrac{1}{2}\Delta)/\{1 + \exp(y - \lambda_i - \tfrac{1}{2}\Delta)\}, \end{aligned} \tag{2·2}$$

and, conditional on λ_i, Y_{i1} and Y_{i2} are independent.

* Present address: Department of Statistics, University of Chicago.

When $k = 2$, (2·2) reduces to the usual logistic model for binary paired comparisons. The number of parameters $n+k$ increases linearly with the number of observations $2n$, so that likelihood techniques are not applicable. Indeed it is difficult to find a convenient notation for writing down the likelihood when the block parameters are considered fixed. We use an unusual conditionality argument to make inference about Δ independently of both the block parameters $\{\lambda_i\}$ and the boundary parameters θ.

Consider the 2×2 table obtained from the original square $k\times k$ table by amalgamating categories $1,\ldots,j$ as, say, 'failure' and categories $j+1,\ldots,k$ as 'success'. There are $k-1$ such tables and for each, the model (2·2) reduces to the usual logistic form for paired binary data. Let the off-diagonal elements in these tables be r_j and $n_j - r_j$, where r_j is the number of pairs for which the first observation is category j or less, and the second observation, category $j+1$ or above. The standard procedure for paired binary data is to assume n_j fixed and look at the distribution of r_j conditional on n_j. This is binomial with index n_j and parameter $e^{\Delta}/(1+e^{\Delta})$. To extend this analysis to the multicategory case we examine the joint distribution of $r = (r_1,\ldots,r_{k-1})$ conditional on $n = (n_1,\ldots,n_{k-1})$. Unfortunately the joint distribution of r given n does depend on the nuisance parameters but this dependence is in the second and higher order properties. It is convenient to consider separately the null, $\Delta = 0$, distribution of r conditional on n and the nonnull distribution.

Consider the symmetric matrix N of order $k-1$ with elements n_{ij} given by

$$n_{ij} = \sum_{\alpha\leqslant i}\sum_{\beta>j}(m_{\alpha\beta}+m_{\beta\alpha}) \quad (i\leqslant j),$$

so that the diagonal elements of N are equal to the elements of n. Consider the null distribution of r given N which is equivalent to the null distribution of r given all $\frac{1}{2}k(k-1)$ quantities $\{m_{ij}+m_{ji}\}$ for $i<j$. The distribution of r given N can be found from the fact that the elements of r are simply sums of binomial random variables each with parameter $\frac{1}{2}$. Hence the null distribution of r conditional on N is multivariate binomial with overlapping trials (Patil & Joshi, 1968, p. 60; Lancaster, 1974). The mean vector and variance matrix of this distribution are $\frac{1}{2}n$ and $\frac{1}{4}N$ respectively.

The null distribution of r conditional only on n can be considered as a mixture of multivariate binomial distributions. The mixing distribution on N is such that its diagonal elements are held constant and equal to n. We use only the first two moments which are

$$E_0(r|n) = E_0\{E_0(r|N)|n\} = \tfrac{1}{2}n, \tag{2·3}$$

$$V_0(r|n) = E_0\{V_0(r|N)\,|\,n\}+V_0\{E_0(r|N)|n\} = \tfrac{1}{4}E_0(N|n), \tag{2·4}$$

respectively, where the zero subscript refers to the null distribution with $\Delta = 0$. Regardless of any nuisance parameters, $\frac{1}{4}N$ is an unbiased estimator of the null variance matrix of r conditional on n.

The only property of the nonnull distribution which is required is that $r_j/(n_j-r_j)$ tends almost surely to e^{Δ} for $j = 1,\ldots,k-1$. For small Δ, the nonnull variance matrix can be approximated by $V_0(r|n)$. A better approximation is

$$V(r|n) = V_0(r|n)\,(1-\tfrac{1}{4}\Delta^2). \tag{2·5}$$

We now examine two estimators of Δ based on r. The first, $\tilde{\Delta}$, is a weighted sum of consistent and asymptotically normal estimators of Δ, so that the combined estimator, $\tilde{\Delta}$, is also consistent and asymptotically normal. The second, Δ^*, includes the weights in both numerator and denominator. In this respect it is analogous to the Mantel–Haenszel procedure for pooling odds-ratios from several 2×2 tables (Mantel & Haenszel, 1959). The weights $\tilde{w}$ and w^* are based on the null covariance matrix $\frac{1}{4}N$, so that the efficiency of both estimators will be largest when Δ is small. The two estimators are

$$\tilde{\Delta} = \sum_{j=1}^{k-1} \tilde{w}_j \log\{(r_j+\tfrac{1}{2})/(n_j-r_j+\tfrac{1}{2})\}, \tag{2·6}$$

$$\Delta^* = \log\left[\left\{\frac{1}{2}+\sum_{j=1}^{k-1} w_j^* r_j\right\}\Big/\left\{\frac{1}{2}+\sum_{j=1}^{k-1} w_j^*(n_j-r_j)\right\}\right], \tag{2·7}$$

where

$$\tilde{w}_1+\ldots+\tilde{w}_{k-1} = 1.$$

The weights $\tilde{w} = (\tilde{w}_1, \ldots, \tilde{w}_{k-1})^{\mathrm{T}}$ and $w^* = (w_1^*, \ldots, w_{k-1}^*)^{\mathrm{T}}$ are chosen so that the estimators $\tilde{\Delta}$ and Δ^* have minimum variance. In general N is nonsingular and the weights are given by

$$\tilde{w} = (DN^{-1}n)/(n^{\mathrm{T}}N^{-1}n), \quad w^* = N^{-1}n,$$

where $D = \operatorname{diag}(n_1, \ldots, n_{k-1})$. The addition of $\frac{1}{2}$ to both numerator and denominator in (2·6) and (2·7) ensures that the estimates remain finite but does not affect their asymptotic properties. It is not necessary that the weights w^* should sum to unity since they occur in both the numerator and denominator of (2·7). Both estimators have the same asymptotic variance which can be estimated by

$$\operatorname{var}(\tilde{\Delta}) = \operatorname{var}(\Delta^*) \simeq 4(1+\tfrac{1}{4}\Delta^2)/(n^{\mathrm{T}}N^{-1}n). \tag{2·8}$$

If N is singular, as can happen when the original data matrix is sparse, it is sufficient to find $\tilde{w}$ and w^* which satisfy $ND^{-1}w = n$ with $\tilde{w}_1+\ldots+\tilde{w}_{k-1} = 1$ and $Nw^* = n$. In this case the asymptotic variance of both estimators is $4(1+\frac{1}{4}\Delta^2)/(n^{\mathrm{T}}w^*)$.

3. Example

The example in Table 1 from Wise & Oldham (1963), concerns the degree of pneumoconiosis in coalface workers as measured radiologically. The scale $1, \ldots, 4$ indicates increasing severity of the disease. Each pair of readings was taken on the same individual, the interval between readings being 2·5 years.

We first compute the quantities N, n, r which are

$$N = \begin{bmatrix} 14 & 4 & 0 \\ 4 & 12 & 3 \\ 0 & 3 & 6 \end{bmatrix}, \quad n = \begin{bmatrix} 14 \\ 12 \\ 6 \end{bmatrix}, \quad r = \begin{bmatrix} 11 \\ 11 \\ 5 \end{bmatrix}.$$

The weights $\tilde{w}$ and $\tilde{w}^*$ are (0·523, 0·283, 0·194) and (0·847, 0·534, 0·733), respectively. The estimates $\tilde{\Delta}$ and Δ^* are 1·450 and 1·503 with standard deviation 0·53.

Table 2 gives the estimates $\tilde{\Delta}$ and Δ^* together with their standard deviations for the other mines $A, \ldots, H$ from the data of Wise & Oldham (1963). The values of Δ are to be interpreted as progressions on a scale such that the observer variation has a standard logistic distribution. It is reasonable, in this example, to assume that observer variation is constant over all the groups, so that the different values of Δ are on a comparable scale.

There is strong evidence of positive progression in all mines except A and B. In addition $\tilde{\Delta}$ or Δ^* could be used to rank the mines in order of increasing progression, or to group them in clusters according to the similarity of the progression of disease.

It is interesting to compare the present analysis with others which use only the information in the marginals. The data have been analyzed in this way in an unpublished University College report by T. P. Hutchinson using an exponential model and by Wise & Oldham (1963) using a normal distribution for the marginals. Table 3 shows the results of these analyses for mine G together with the estimate obtained by using the method of Clayton (1974). All three methods show some evidence of progression although none is conclusive. Contrast the standard normal test value of 2·8 obtained from Δ^* or $\tilde{\Delta}$ divided by their standard deviation, with those in Table 3. Clearly, considerable information is lost when the pairings are ignored.

Armitage (1975) has investigated the efficiencies of the paired and unpaired model for binary data when pairing is random. In this case both models are applicable but the unpaired model gives a more efficient estimate except when $\Delta = 0$ when both models are equally efficient asymptotically.

Table 1. *Paired readings for 82 coalface workers at mine G from Table 1 of Wise & Oldham (1963)*

	Category	Second reading 1	2	3	4	Total
First reading	1	43	8	3	0	54
	2	2	2	5	3	12
	3	1	0	7	2	10
	4	0	0	1	5	6
	Total	46	10	16	10	82

Table 2. *Progression estimates,* $\tilde{\Delta}$ *and* Δ^**, for* 8 *mines A, ..., H*

	A	B	C	D	E	F	G	H
$\tilde{\Delta}$	0·84	0·37	2·38	2·88	3·16	3·20	1·45	1·90
Δ^*	0·92	0·51	2·22	3·22	3·70	3·26	1·50	1·62
St. error	0·49	1·18	0·66	0·74	1·11	0·60	0·53	0·58
n	90	33	87	83	82	148	82	84

Table 3. *Progression estimates for mine G from marginals only*

	Model	Estimate	St. error	Ratio
Hutchinson	Exponential	0·27	0·17	1·59
Wise & Oldham	Normal	0·29	0·18	1·61
Clayton	Logistic	0·48	0·31	1·55

4. Further comments

A referee suggested an alternative approach based on models for quasisymmetry (Bishop, Fienberg & Holland, 1975, p. 286). However, a random effects version of (2·2) is not, in general, quasisymmetric, but it is possible that a distribution could be found for the $\{\lambda_i\}$ which leads to a joint quasisymmetric distribution. With a normal distribution for the errors and a normal distribution for the $\{\lambda_i\}$ the joint distribution of (Y_{i1}, Y_{i2}) is bivariate normal which is quasisymmetric. Unfortunately the parameters in the model for quasisymmetry are related in a complicated way to those of the bivariate normal distribution. This is a severe limitation of the model for quasisymmetry.

Since the results about Δ^* and $\tilde{\Delta}$ are asymptotic only, we present briefly the results of some simulations for realistic sample sizes. For four-category data and 100 pairs per sample with corr (Y_1, Y_2) equal to 0·84 to mimic the correlation of pairs of observations for mine G, neither estimator appeared to be biased for Δ in the range $(-1·5, 1·5)$. However, the correction factor of $1 + \frac{1}{4}\Delta^2$ for the variance as given in (2·8) was necessary for the larger values of Δ.

It has been suggested that there might be considerable loss of information from using the above analysis. Consequently a random effects approach was tried where the distribution on $\{\lambda_i\}$ was chosen to be a logit transformation of the beta distribution. This density is

$$f(\lambda) = e^{-\lambda\alpha}/\{(1+e^{-\lambda})^{\alpha+\beta}B(\alpha,\beta)\} \quad (-\infty < \lambda < \infty;\ \alpha,\beta > 0),$$

where $B(\alpha,\beta)$ is the beta function. This density was chosen because it is the only one which could be found for which the joint marginal density of pairs (Y_1, Y_2) could be expressed in closed form. Even so, this closed form involved the ${}_2F_1$ hypergeometric function.

When this model was fitted to the data of Table 1 the value of Δ obtained was 1·69 with estimated standard deviation 0·56. One possible reason for the difference from the previous estimates is the addition of $\frac{1}{2}$ to both numerator and denominator in (2·6) and (2·7). This $\frac{1}{2}$ has the effect of shrinking the estimate towards the origin and hence produces the discrepancy. In larger samples the discrepancy would be expected to disappear. From this example it appears that there is little loss of information from using Δ^* or $\tilde{\Delta}$.

An interesting problem arises concerning the consistency of Δ^* when the complete ranking of all $2n$ observations is available. In this case it can be shown that Δ^* reduces to

$$e^{\Delta^*} = \frac{\frac{1}{2} + \text{no. of pairs where } Y_2 > Y_1}{\frac{1}{2} + \text{no. of pairs where } Y_1 > Y_2}.$$

In this limit Δ^* is inconsistent. In fact for small Δ

$$\text{plim}\,(\Delta^*) = \tfrac{2}{3}\Delta + O(\Delta^3).$$

It is probable that $\tilde{\Delta}$ is also inconsistent in the same limit but it does not have such a simple limiting form.

This research was supported by the Health & Safety Executive.

References

ARMITAGE, P. (1975). The use of the cross-ratio in aetiological surveys. In *Perspectives in Probability and Statistics. Papers in Honour of M. S. Bartlett*, Ed. J. Gani, pp. 349–55. New York: Academic Press.

BISHOP, Y., FIENBERG, S. E. & HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: M.I.T. Press.

CLAYTON, D. G. (1974). Some odds ratio statistics for the analysis of ordered categorical data. *Biometrika* **61**, 525–31.

COX, D. R. (1958). Two further applications of a model for binary regression. *Biometrika* **45**, 562–5.

LANCASTER, H. O. (1974). Multivariate binomial distributions. In *Studies in Probability & Statistics: Papers in Honour of Edwin J. G. Pitman*, Ed. E. J. Williams, pp. 13–9. Jerusalem: Academic Press.

MANTEL, N. & HAENSZEL, W. (1959). Statistical aspects of the analysis of data from restrospective studies. *J. Nat. Cancer Inst.* **22**, 719–48.

PATIL, G. G. & JOSHI, S. W. (1968). *A Dictionary and Bibliography of Discrete Distributions*. Edinburgh: Oliver and Boyd.

WISE, M. E. & OLDHAM, P. D. (1963). Estimating progression of coalworkers' simple pneumoconiosis from readings of radiological categories. *Br. J. Indust. Med.* **20**, 124–44.

[*Received October* 1976. *Revised May* 1977]