Distribution-Free Fitting of Logit Models with Random Effects for Repeated Categorical Responses

Alan Agresti
Department of Statistics
University of Florida
Gainesville, Florida, USA 32611
phone (904) 392-1941

SUMMARY

This article discusses random effects models for within-subject comparisons of repeated responses on the same categorical scale. The models account for the correlation that normally occurs between repeated responses. The standard way of fitting such models maximizes the marginal likelihood after integrating with respect to a distribution for the random effect. An alternative nonparametric approach does not assume a distributional form for the random effects. Recent literature shows that for certain simple logit models, this approach yields essentially the same model parameter estimates as conditional maximum likelihood. Moreover, these estimates also result from fitting corresponding quasi-symmetric loglinear models. For simple data sets in which primary interest relates to subject-specific comparisons of the repeated responses, one can easily obtain the estimates with standard software for loglinear models. Examples include data from cross-over designs and from comparisons of treatment and control groups regarding the change between baseline and follow-up observations.

Key words: adjacent-category logit; cross-over designs; cumulative logit; item response models; marginal homogeneity; matched pairs; ordinal data; overdispersion; proportional odds; quasi symmetry; Rasch model; square contingency tables; subject-specific effects.

1 INTRODUCTION

Suppose n subjects make T repeated responses on the same categorical scale, with r categories. For instance, each subject may be classified on a scale (mild, moderate, severe) relating to severity of some chronic condition, before and after receiving treatment for it. Or, each subject might be classified by T different observers regarding whether a particular disease is present. Normally there are also explanatory variables. For instance, one might compare the change in severity of disease for two treatments, or for treatment and control groups. To simplify, we suppress explanatory variables in our notation until we consider specific examples.

For subject i and response measurement $j, j = 1, ..., T,$ let ϕ_{hij} denote the probability of response outcome in category $h, h = 1, ..., r,$ so $\sum_h \phi_{hij} = 1$. The notation reflects allowances for within-subject and between-subject heterogeneity. Let $\phi_{h.j}$ denote the marginal probability of outcome h for response j, for a subject randomly selected from the population of interest (The '.' notation represents averaging over the population of subjects).

Table 1, from Matthews and Farewell[1] (p. 64), is an example of the simplest setting for such data, with $T = r = 2$. Here, $T = 2$ pathologists examined the same $n = 100$ tumors and classified each tumor in one of $r = 2$ categories, malignant or benign. How does one describe the difference between the pathologists' evaluations? For instance, consider estimation of θ in the conclusion, "The odds that pathologist B judges a malignancy are θ times the odds for pathologist A." One estimate is $(18/82)/(10/90) = 1.98$, based on the ratio of sample odds in the marginal distributions. This estimates $\theta = \exp(\beta_2 - \beta_1)$ in the *population-averaged* model

$$logit(\phi_{1.j}) = \alpha + \beta_j \ , \ j = 1, 2. \tag{1}$$

The effect θ refers to the odds of pathologist B judging a malignancy for a tumor from a randomly selected member in the population of interest, relative to the odds of pathologist A judging a malignancy for a tumor from another randomly selected member.

Alternatively, one might prefer an interpretation that relates to the *same* tumor evaluated by

each pathologist. Such effects refer to estimating $\theta = \exp(\beta_2 - \beta_1)$ in the *subject-specific* model

$$logit(\phi_{1ij}) = \alpha_i + \beta_j \ , \ j = 1, 2. \tag{2}$$

For this model the usual sample estimate of θ is the ratio of the "discordant" counts[2], $9/1 = 9$ in Table 1, and the related test of the hypothesis of no effect is McNemar's test.

Parameters in population-averaged models describe differences in marginal distributions of the repeated response purely in terms of the marginal probabilities. By contrast, parameters in subject-specific models describe differences in a way that directly incorporates the dependence of repeated responses in the joint distribution. If a subject-specific model holds of logistic form (2), the implied population-averaged model for the marginal distributions is not necessarily of logistic form. This is because the marginal probability $\phi_{1.j}$ is the expectation with respect to α_i of $\phi_{1ij} = \{1 + exp[-(\alpha_i + \beta_j)]\}^{-1}$, which need not have form $\{1 + exp[-(\alpha + \beta_j)]\}^{-1}$. An exception is when the subject parameters are identical, in which case the population-averaged model holds with the same effects. Ordinarily, when we fit both models, between-subject heterogeneity results in population-averaged effect estimates that are closer to zero than subject-specific effect estimates, with greater heterogeneity producing more of a discrepancy.

This distinction does not occur for the usual linear models for normally distributed responses. For matched pairs ($T = 2$), for instance, the mean of the paired differences (a subject-specific effect) is the same as the difference of the two means (a population-averaged effect). Neuhaus[3] *et al.* and Zeger[4] *et al.* gave interesting discussions of relationships between subject-specific and population-averaged effects. In this article, we focus on simple ways of estimating subject-specific effects.

Subject-specific models are often used to account for overdispersion that may result from uncontrolled factors or dependence or clustering in a sample. For instance, mice from different litters subjected to the same experimental conditions might have different "success" probabilities for a binary response, and results for different mice from the same litter might have dependent responses. Either condition results in counts of successes displaying greater variation than expected with the binomial sampling model for independent, identical trials. Models with random effects terms for subjects or clusters can reflect such departures from the usual simple assumptions, and they can also represent effects of omitted explanatory variables or measurement error in the explanatory variables. Chap. 6 of Collett[5] gives a good introduction to random effects models for the binary-response case.

The psychometrics literature reports an enormous body of research relating to random effects models for subject-specific comparisons of repeated responses, and referred to as *item-response models*. Suppose n subjects respond to T questions (items), each having a correct response (1) and an incorrect response (2). A basic item-response model, the Rasch model[6,7], states that the probability of correct response by subject i on item j depends on subject abilities $\{\alpha_i\}$ and item difficulties $\{\beta_j\}$ through the equation

$$logit(\phi_{1ij}) = \alpha_i - \beta_j. \tag{3}$$

In item response models, interest may focus on estimating the subject abilities, the item difficulties, or both. A complication in fitting the models is that the number of parameters increases as the sample size increases. Regularity conditions needed to ensure consistency of maximum likelihood (ML) estimators do not apply[8]. The degree of bias can be substantial. For model (2), for instance, the ordinary ML estimator of β converges in probability to 2β (Andersen[9], p.244).

For purposes of estimating item parameters, there are two common ways of dealing with this complication. The first, a fixed-effects solution, treats the subject effects as nuisance parameters and conditions on sufficient statistics for them. One maximizes the resulting log likelihood, which depends only on the parameters of interest (*e.g.*, $\{\beta_j\}$ in model (3)). For $T = 2$ this *conditional ML* approach leads to the estimate quoted above based solely on discordant pairs. Andersen's[9] text is a good reference for the conditional approach.

The second approach treats the subject terms as random effects, assuming some distribution for them, such as normal with unknown standard deviation. One integrates the likelihood with respect to that distribution to obtain a marginal likelihood. One then maximizes the marginal likelihood, which depends also on the parameters of the random effects distribution. For discussion of this *marginal ML* approach, see, Anderson and Hinde[10], Bock and Aitkin[11], Conaway[12], Laird[13], Stiratelli[14] *et al.*, Thissen[15], and also many articles listed in the annotated bibliography of Ashby[16] *et al.*. An advantage of marginal ML is that it is applicable also for models with non-canonical links such as the probit or log-log, for which sufficient statistics do not exist and conditional ML cannot be used. Another disadvantage often quoted for the conditional ML approach is its computational complexity. However, the marginal ML approach is also somewhat complex, because of the necessity of integrating out the random effect. This normally requires numerical integration in coordination with an algorithm such as the EM algorithm[11], or at least Gibbs sampling[17] or some type of Monte

Carlo method[18]. One can obtain such estimates using specialized software or subroutines, such as in NAG and IMSL.

The main purpose of this article is to show that for logit models that describe subject-specific effects with repeated categorical responses, it is actually easy to obtain conditional ML estimates for some standard types of data sets. We review recent literature showing that conditional ML estimates are also ordinary ML estimates for certain loglinear models. Hence, one can use existing software such as SAS and GLIM to obtain the conditional estimates. In addition, it is possible to check model adequacy using residuals and goodness-of-fit tests for the corresponding loglinear model.

An advantage of the conditional approach is that it is unnecessary to assume a distribution for the subject random effects. The analysis is distribution-free, and hence one fewer basic assumption needs to be satisfied for the resulting estimators to be consistent. In addition, the conditional ML estimates are identical to estimates obtained for an extended form of marginal maximum likelihood that makes no assumption about the random effects distribution. Thus, the estimates are also nonparametric marginal ML estimates. We illustrate how to obtain the estimates for three distinct models for three data sets analyzed recently in this journal. The first is a logit model for a repeated binary response obtained from a cross-over experiment with three treatments. The second is an adjacent-categories logit model for an ordinal response observed at a baseline and at follow-up, for control and treatment groups. The third is a cumulative logit model for an ordinal response observed in a cross-over experiment. There is a need for additional research to extend this approach to more complex types of repeated measurement data, in order to handle complications such as missing data and time-varying covariates.

2 THE BINARY RASCH MODEL AND QUASI SYMMETRY

We first discuss the Rasch model (3), its connection to the loglinear model of quasi symmetry, and its application to modeling repeated binary responses. The term α_i common to all responses by subject i represents the effects of characteristics of that subject, net of any explanatory variables in the model, that could affect the response. This term represents a latent variable, and for a given subject i with fixed $\{\alpha_i\}$, the model's "local independence" assumption treats separate responses as independent. We also assume that responses by different subjects are independent. The relative similarity of

different responses by the same subject, reflected by the $\{\alpha_i\}$, implies that joint distributions of pairs of responses (averaged over subjects) have positive associations.

Section 1 expressed subject-specific models in terms of an $r \times n \times T$ array of probabilities $\{\phi_{hij}\}$. Upon elimination of the subject terms, marginal and conditional ML analyses refer to a r^T cross-classification of the T responses by each subject. Let $(h_1, ..., h_T)$ denote a potential response pattern for the T responses, where the outcome h_j for the j^{th} response is an integer between 1 and r, for $j = 1, ..., T$. Let $n(h_1, ..., h_T)$ denote the number of subjects in the sample having this response pattern, and let $m(h_1, ..., h_T)$ denote its expected frequency. We treat $\{n(h_1, ..., h_T)\}$ as a multinomial sample of size n, with cell probabilities proportional to $\{m(h_1, ..., h_T)\}$.

For binary responses ($r = 2$), Tjur[19] showed a connection between conditional ML estimates for the Rasch model (3) and standard ML estimates for a loglinear model for $\{m(h_1, ..., h_T)\}$ in the 2^T cross classification of the subjects' T responses. We next outline his argument, omitting the technical details. Tjur first described a nonparametric marginal ML approach. One considers the probability of a particular response pattern $(h_1, ..., h_T)$, for a given α. Integrating this probability with respect to an unspecified distribution $F(\alpha)$ for that subject effect leads to the likelihood used in the nonparametric marginal ML solution. This likelihood involves complicated functions of the parameters. Tjur noted, however, that an extended version of this likelihood that is slightly more general (in terms of allowing additional values for parameters) has simple form, and is equivalent to the likelihood for a certain loglinear model. Tjur referred to this loglinear model for the "extended" likelihood as the "extended random model." He also noted that one can decompose the likelihood for that loglinear model as the product of a function of $\{\beta_j\}$ and a function of the remaining parameters, in such a way that the function of $\{\beta_j\}$ is the conditional likelihood function for the Rasch model; that is, it is the likelihood obtained by conditioning on sufficient statistics for $\{\alpha_i\}$. Thus, Tjur noted that ML estimates of $\{\beta_j\}$ and their estimated standard errors are identical for the loglinear model and the conditional Rasch model.

It follows that one can obtain conditional ML estimates of $\{\beta_j\}$ in the Rasch model by fitting a loglinear model. Blackwood[20] and Darroch and McCloud[21] provided alternative motivation for the connection between the models. Following Tjur's work, others (*e.g.*, Fienberg[17]) noted that his loglinear model is simply the quasi-symmetry model, and the conditional ML estimates of $\{\beta_j\}$ relate to ordinary ML estimates of main effects for that model. Also, de Leeuw and Verhelst[22] showed that

when $F(\alpha)$ is unspecified but has an infinite number of points of increase, the ordinary nonparametric marginal ML estimators of $\{\beta_j\}$ are identical to Tjur's extended marginal ML estimates (and hence the conditional ML estimates) with probability converging to 1 as $n \to \infty$. Thus, one can also interpret the conditional ML estimates as nonparametric marginal ML estimates, being free of any assumption about the form of the distribution for the random effects. See Cressie and Holland[23] and Lindsay[24] *et al.* for related remarks, and Conaway[25,26], Darroch and McCloud[21], and Kenward and Jones[27] for extensions of Tjur's result to generalized logit models for r unordered categories.

The quasi-symmetry model has form

$$\log m(h_1, ..., h_T) = \sum_{j=1}^{T} \beta_j(h_j) + \lambda(h_1, ..., h_T) \tag{4}$$

where the term $\lambda(.)$ is identical for all permutations of its argument; for instance, for $T = 3$ and $r = 2$, $\lambda(1,1,2) = \lambda(1,2,1) = \lambda(2,1,1)$ and $\lambda(1,2,2) = \lambda(2,1,2) = \lambda(2,2,1)$. This model is a generalization of the complete symmetry model that permits different main effect parameters for each response[28,29]. The conditional ML estimates of $\{\beta_j\}$ in the Rasch model (3) are related to the ordinary ML parameter estimates in this model by

$$\hat{\beta}_b - \hat{\beta}_a = (\hat{\beta}_b(2) - \hat{\beta}_b(1)) - (\hat{\beta}_a(2) - \hat{\beta}_a(1))$$

The conditioning argument that results in elimination of the subject parameters in the Rasch model yields the symmetry parameters in (4), whereby subjects who have the same total 'score' on the T responses (*i.e.*, the same number of $h_j = 1$) have the same interaction term.

To illustrate this approach, we use Table 2, previously analyzed by Jones and Kenward[30]. The data refer to a three-period cross-over trial designed to compare placebo (treatment A) with a low-dose analgesic (treatment B) and high-dose analgesic (treatment C) for relief of primary dysmenorrhea. The subjects in the study were divided randomly into six groups, corresponding to the six possible sequences for administering the three treatments during the three periods. At the end of each period, each subject rated the treatment as giving either no relief (1) or some relief (2). Let $\phi_{hi(k)j}$ denote the probability of response h ($h = 1, 2$) for subject i using treatment j ($j = A, B, C$), where subject i is nested in treatment sequence k ($k = 1, ..., 6$). Let $m_k(a, b, c)$ denote the expected frequency of outcomes a for treatment A, b for treatment B, and c for treatment C, under treatment sequence k. The subject-specific logit model corresponding to a Rasch model with common

treatment effects for each sequence k is

$$logit(\phi_{1i(k)j}) = \alpha_{i(k)} - \beta_j. \tag{5}$$

One can estimate treatment parameters in this model by fitting the corresponding quasi-symmetry type of model

$$\log m_k(a,b,c) = \beta_A(a) + \beta_B(b) + \beta_C(c) + \lambda_k(a,b,c) \tag{6}$$

where, for instance, $\beta_A = \beta_A(2) - \beta_A(1)$, and where for each k, $\lambda_k(a,b,c) = \lambda_k(a,c,b) = \lambda_k(b,a,c) = \lambda_k(b,c,a) = \lambda_k(c,a,b) = \lambda_k(c,b,a)$. The more general subject-specific model in which the effects vary by treatment sequence has main effect parameters $\{\beta_{jk}\}$ that differ for each k, and corresponds to a separate quasi-symmetry fit for each treatment sequence. The difference in likelihood-ratio goodness-of-fit statistics between the two models is 12.77, based on $df = 10$. This suggests that the simpler model is adequate. One can also consider intermediate models that permit certain types of treatment-by-sequence interaction, such as period or carry-over effects, but these are insignificant for these data. To illustrate, suppose one included the period terms

$$-\pi_1 I(period\ 1) - \pi_2 I(period\ 2)$$

in the Rasch-type model (5). For instance, the first period indicator equals 1 when $\{(j,k) = (1,1),(1,2),(2,3),(2,4),(3,5),(3,6)\}$. It corresponds to including the term π_1 in model (6) for the cells for which $(a = 2, k = 1,2),(b = 2, k = 3,4),(c = 2, k = 5,6)$. The decrease in the likelihood-ratio statistic is then 0.66, compared to model (6), based on $df = 2$.

The treatment effect estimates obtained from the simple model (6) are

$$\hat{\beta}_B - \hat{\beta}_A = 1.641\ (ase = 0.338)$$

$$\hat{\beta}_C - \hat{\beta}_A = 2.230\ (ase = 0.388)$$

$$\hat{\beta}_C - \hat{\beta}_B = 0.589\ (ase = 0.393).$$

For instance, for a given subject, we estimate the odds of relief for the low-dose analgesic as $exp(1.641) = 5.2$ times the odds of relief for the placebo. Substantive results are the same as those given by Jones and Kenward using a different loglinear model, but their treatment parameters have an awkward conditional interpretation. In a later article[27], these authors showed how to fit a Rasch-type model for an expanded version of these data in which the response has three categories.

In the model (6) with common treatment effects for each sequence, suppose we sum expected frequencies over k, thus collapsing the table over the treatment-sequence factor. Then we see that the quasi-symmetry model with the same treatment parameters applies to the collapsed table. Thus, model (6) has the same treatment effect estimates as we would obtain for the usual quasi-symmetry model fitted to the single 2^3 table collapsed over treatment sequence. Representation (6) for the uncollapsed table is useful, however, for comparing results to more general models in which the treatment effect may vary according to period, carry-over, or other interaction effects. The lower margin of Table 2 shows the fit of the simple model to the collapsed table. It has a likelihood-ratio goodness-of-fit statistic of G^2 =3.27, based on $df = 2$. There is a reduction of 53.53 in the likelihood-ratio statistic (based on $df = 2$) compared to the complete symmetry model, giving strong evidence against the hypothesis of identical treatment effects.

3 AN ORDINAL RANDOM EFFECTS MODEL

Though the literature is extensive on random effects and item response models for binary data, this is not the case for ordinal responses. Relevant articles include Agresti[31], Ezzet and Whitehead[32], Harville and Mee[33], McCullagh[34] and Tutz[35]. We shall discuss two types of ordinal models, utilizing different types of logits. We need separate approaches to obtain distribution-free estimates for these models, but both approaches are related to that for binary responses whereby estimates result from the fitting of a corresponding quasi-symmetry model.

The first ordinal logit model is a special case of a model considered by Andersen[36]. It has the adjacent-categories logit representation

$$\log(\phi_{h+1,ij}/\phi_{hij}) = \alpha_{hi} + \beta_j. \tag{7}$$

Simpler models also provide structure for $\{\alpha_{hi}\}$, but are not relevant when main interest focuses on estimating the subject-specific effects $\{\beta_j\}$. In model (7), for each subject the odds of outcome $h+1$ instead of outcome h for response a are $exp(\beta_a - \beta_b)$ times the odds for response b. We assume the effects are identical for each pair of adjacent categories.

This model has simple sufficient statistics for the subject effects, so the conditional approach applies directly. These statistics relate to counts of subjects who have response sequences that, apart from a permutation, are identical. Generalizing Tjur's argument, Agresti[31] noted that the conditional ML estimates are also distribution-free marginal ML estimates. One can obtain them

by ordinary fitting of the loglinear model

$$\log m(h_1, ..., h_T) = \sum_{j=1}^{T} \beta_j v_{h_j} + \lambda(h_1, ..., h_T), \tag{8}$$

where $\lambda(.)$ is permutationally invariant and $\{v_h = h\}$. This is a special case of the quasi-symmetry model (4) that has the ordinal structure for the main effects,

$$\beta_j(h) = h\beta_j \ .$$

The complete symmetry model is the further special case $\beta_1 = ... = \beta_T$. Agresti[31] also discussed models in which $\{v_h\}$ are arbitrary fixed scores.

The sufficient statistics for $\{\beta_j\}$ in model (8) are the mean responses in the T one-way margins of the r^T table, when one assigns equally-spaced scores to the r response categories. The conditional ML estimates of $\{\beta_j\}$ have the same order as the sample marginal means, and an efficient score test for the hypothesis of T identical response distributions is based on the variation in the values of those means (see, *e.g.,* Koch[37] *et al.*, and Meeks and D'Agostino[38] for $T = 2$). Alternatively, one can use a likelihood-ratio test with $df = T - 1$, based on comparing the fit of this model to that of the complete symmetry model.

To illustrate this ordinal logit model, consider Table 3. These data, previously analyzed by Francom[39] *et al.* and Agresti[40], give results of a randomized, double-blind clinical trial comparing an active hypnotic drug with a placebo in patients with insomnia. The outcome is patient response to the question, "How quickly did you fall asleep after going to bed?" Patients responded at the start and conclusion of a two-week treatment period. Agresti analyzed the data using population-averaged models. That article noted that subject-specific ordinal models provided an alternative approach, but that likelihood methods had not been developed for them. We now conduct such an analysis.

Let $\phi_{hi(k)j}$ denote the probability of response h ($h = 1, 2, 3, 4$) at occasion j ($j = 1$ for initial, and $j = 2$ for follow-up), for subject i who is nested within treatment group k (where $k = 1$ for active drug and $k = 2$ for placebo). The model

$$\log(\phi_{h+1,i(k)j}/\phi_{hi(k)j}) = \alpha_{hi(k)} + \beta_j \tag{9}$$

assumes the same occasion effect $\beta = \beta_2 - \beta_1$ for each treatment group. For Table 3, let $m_k(a, b)$ denote the expected frequency for outcome a at the initial occasion and outcome b at follow-up, for

treatment group k. One can obtain the conditional ML estimate of the occasion effect by fitting the loglinear model

$$\log m_k(a,b) = a\beta_1 + b\beta_2 + \lambda_k(a,b) \tag{10}$$

where for $k = 1, 2$, $\lambda_k(a,b) = \lambda_k(b,a)$ for all a and b.

The model fits relatively well, the fit also being displayed in Table 3. It has a likelihood-ratio goodness-of-fit statistic of $G^2 = 15.16$, based on $df = 11$. The ML estimate of the common occasion effect equals $\hat{\beta} = -1.365$ ($ase = 0.183$). For each subject with either drug or placebo, the estimated odds that time to falling asleep at the initial observation is > 60 minutes instead of 30-60 minutes (or 30-60 minutes instead of 20-30 minutes, or 20-30 minutes instead of < 20 minutes) is $exp(1.365) = 3.9$ times the estimated odds at the follow-up observation. The model lacking an occasion effect is simply the model that displays a separate complete symmetry structure for each k (model (10) with $\beta_1 = \beta_2$). It fits much more poorly, having $G^2 = 117.19$ ($df = 12$).

More general models permit the occasion effect to vary by treatment group. For Table 3, the generalized model has $G^2 = 14.32$ ($df = 10$). The estimated occasion effects are then -1.519 for the active drug and -1.183 for placebo. The difference of 0.336 has an estimated standard error of 0.367. There is not much evidence of a stronger occasion effect for the active treatment, based on the Wald test ($(0.336/0.367)^2 = 0.84, df = 1$) or the likelihood-ratio test (change in G^2 of 0.84, $df = 1$). The estimated occasion effect for the simpler model (10) is necessarily identical to the one obtained by collapsing the table over the group factor.

The results for this model differ somewhat from results for population-averaged models, both in terms of sizes of estimated effects and whether there is much evidence of interaction. For instance, consider the model

$$\log(\phi_{h+1,jk}/\phi_{hjk}) = \alpha_{hk} + \beta_j$$

which applies to the treatment-by-occasion margins of Table 3. We fitted this model by maximizing a product multinomial likelihood for the 16 cells in each part of Table 3, obtaining $G^2 = 13.00$, based on $df = 5$. The model that lets the effect vary by treatment (*i.e.*, $logit = \alpha_{hk} + \beta_{jk}$) has $G^2 = 4.21$ ($df = 4$). For the latter model, the ML estimated occasion effect is -0.982 for the active drug and -0.505 for placebo, and the difference of 0.477 has an estimated standard error of 0.162.

For another example, we consider an ordinal-response version of the data analyzed in the previous section (Table 2), reported in the later paper by Kenward and Jones[27]. Relief was there measured as

(none, moderate, complete) instead of (none, some). We obtained estimates of treatment effects for a simple subject-specific model of form (9) with k referring to treatment sequence. Estimates are the same as those for the loglinear model (8) fitted to a 3^3 table that is a collapsing of their table over the six sequences. The model fits well, with $G^2 = 10.35$ ($df = 15$) compared to $G^2 = 69.00$ ($df = 17$) for the complete symmetry model. The estimated treatment effects are

$$\beta_B - \beta_A = 1.207 \quad (ase = 0.239)$$

$$\beta_C - \beta_A = 1.537 \quad (ase = 0.259)$$

$$\beta_C - \beta_B = 0.330 \quad (ase = 0.221).$$

For instance, for a given subject, the estimated odds that relief for the low-dose analgesic is moderate rather than none, or complete rather than moderate, are $exp(1.207) = 3.34$ times the estimated odds for the placebo. As in the previous analysis, both treatments B and C clearly differ from placebo, but there is only weak evidence that the high dose is better than the low dose. More complex models that contain period or carry-over effects do not fit appreciably better. Use of the ordinality results in greater parsimony than that obtained with the Kenward and Jones generalized Rasch model, but the substantive results are the same.

4 A PROPORTIONAL ODDS RANDOM EFFECTS MODEL

Currently, the most popular model form for ordinal responses uses cumulative logits. For subject i and response j, denote the cumulative probability at category h by $\gamma_{hij} = \phi_{1ij} + ... + \phi_{hij},\ h = 1, ..., r$. The cumulative logit alternative to model (7) is the model

$$\log[\gamma_{hij}/(1 - \gamma_{hij})] = \alpha_{hi} - \beta_j, \tag{11}$$

$h = 1, ..., r-1,\ i = 1, ..., n,\ j = 1, ..., T$. For each subject, the odds that the outcome for response a falls above any fixed level are $exp(\beta_a - \beta_b)$ times the odds for response b.

This cumulative logit model has the proportional odds property, for which the T response effects are identical at each h. The model holds[41] if for each pairing of subject i and response measurement j, there is an underlying continuous variable that has a logistic distribution with mean β_j, and the observed outcome falls in category h when the underlying continuous variable falls between $\alpha_{h-1,i}$ and α_{hi}. The model permits different subjects to use different cutpoints for determining the response

category. For instance, for a given value for the underlying continuous scale, one subject may regard it as part of category "good," while a second subject regards it as "very good." In the ordinal models using adjacent-categories logits or cumulative logits, the distributions for the T responses are stochastically ordered according to $\{\beta_j\}$. For relatively larger values of β_j, subjects tend to make higher responses on the ordinal scale for response j. For $r = 2$, the models simplify to the Rasch model. We refer to (11) as the *cumulative Rasch model*. McCullagh[34,42] and Ezzet and Whitehead[32] discussed related models for $T = 2$.

Unfortunately, the cumulative Rasch model does not have reduced sufficient statistics, so the standard conditional approach is unavailable. There are ways, however, of obtaining distribution-free consistent estimates of $\{\beta_j\}$. For instance, note that for fixed h, model (11) is the ordinary Rasch model for a collapsing of the response into binary outcomes ($\leq h$, $> h$). Thus, when the model holds, one can estimate $\{\beta_j\}$ by obtaining conditional maximum likelihood estimates for the Rasch model applied to the collapsed binary response scale, for any h (*e.g.*, one could do this by fitting the corresponding quasi-symmetry model). Such estimates are somewhat inefficient, more so as r increases.

To obtain more efficient estimates, we suggest the following approach. Consider $(r-1)$ separate 2^T contingency tables, in which the h^{th} 2^T table is the cross classification of responses of the n subjects on the h^{th} binary collapsing of the ordinal response scale, $h = 1, ..., r-1$. Each subject occurs in each 2^T table. In the collapsing for cutpoint h, let $n(h; h_1, ..., h_T)$ denote the number of subjects making collapsed response h_j to item j, where $1 \leq h \leq r-1,\ 1 \leq h_j \leq 2,\ j = 1, ..., T$. Let $m(h; h_1, ..., h_T) = En(h; h_1, ..., h_T)$. Agresti and Lang[43] showed that one can estimate $\{\beta_j\}$ by fitting the model

$$\log m(h; h_1, ..., h_T) = \sum_j \beta_j I(h_j = 2) + \lambda(h; h_1, ..., h_T), \tag{12}$$

simultaneously for all h, where $I(.)$ denotes the indicator function, and for each h, $\lambda(h; h_1, ..., h_T)$ is permutation invariant. For fixed h, model (12) is the ordinary quasi-symmetry model for a 2^T table. Model (12) assumes homogeneity of main effect parameters for each 2^T collapsing, reflecting the proportional odds assumption for the item effects in the cumulative Rasch model.

We cannot treat counts $\{n(h; h_1, ..., h_T)\}$ from different 2^T tables as independent, since the same subjects occur in each table. To fit model (12), we can maximize a multinomial likelihood for the original r^T table subject to the constraint that (12) holds simultaneously for each collapsed 2^T table.

In terms of a vector of expected frequencies μ for the original r^T table, (12) has generalized loglinear form

$$\log A\mu = X\beta. \tag{13}$$

The $(r-1)2^T \times r^T$ matrix A, when applied to μ, forms the collapsed tables $\{m(h; h_1, ..., h_T)\}$ of expected frequencies. To obtain the ML fit of models of this form, we can utilize Lagrange's method of undetermined multipliers, such as presented by Aitchison and Silvey[44], Haber[45], and Agresti and Lang[43]. Or, we can use some other estimation method that may be simpler to use with existing software. For instance, SAS (PROC CATMOD) can fit models of form (13) using weighted least squares.

One can check the adequacy of the cumulative Rasch model by ordinary goodness-of-fit tests comparing observed and fitted counts in the original table for the generalized loglinear model. In this model and the one for adjacent-category logits, the fitted counts in the diagonal cells $(h, h, ..., h)$ in the r^T table are identical to the corresponding observed counts. The conditional ML estimates of $\{\beta_j\}$ do not depend on those values. The estimates do not exist when any diagonal counts is zero or when the counts are zero for all permutations of any cell index. When this happens, one can add a very small positive constant ($e.g., 10^{-8}$) to empty cells so the estimates exist. Then, one should conduct a sensitivity analysis to verify that the choice of constant does not affect substantive conclusions.

To illustrate the cumulative Rasch form of model, we analyze Table 4, from Ezzet and Whitehead[32]. The data refer to a cross-over study comparing the suitability of two inhalation devices for delivering salbutamol. The response refers to the patients' report of clarity of leaflet instructions accompanying the devices. The top part of Table 4 refers to patients who used device A for one week followed by device B for another week, whereas the bottom part refers to patients who used the devices in the reverse sequence. Let $\gamma_{hi(k)j}$ denote the cumulative probability at category h for subject i using device j $(A: j = 1; B: j = 2)$, when subject i is nested within sequence k. In the model

$$logit(\gamma_{hi(k)j}) = \alpha_{hi(k)} - \beta_j - \pi[I(j = 1, k = 1) + I(j = 2, k = 2)], \tag{14}$$

$\beta_2 - \beta_1$ denotes the treatment (device) effect and the period effect π denotes the potential effect of using a device first instead of second. For the response collapsed following category h, let $m_k(h; a, b)$ denote the expected frequency of collapsed outcome a with device A and collapsed outcome b with device B, for sequence k. We can obtain conditional estimates for model (14) by maximizing a

product multinomial likelihood for Table 4 subject to the constraint that the generalized quasi-symmetry model

$$\log m_k(h;a,b) = \beta_1 I(a=2) + \beta_2 I(b=2) + \pi[I(a=2,k=1) + I(b=2,k=2)] + \lambda_k(h;a,b)$$

holds, where $\lambda_k(h;a,b) = \lambda_k(h;b,a)$ for all $a, b,\ h = 1,2,3,\ k = 1,2$.

This cumulative Rasch model fits Table 4 reasonably well. The likelihood-ratio statistic of $G^2 = 8.16$ ($df = 4$) is not very reliable because of the table sparseness. To investigate more finely the lack of fit, we computed adjusted residuals based on differences between observed and fitted counts, divided by estimated standard errors. Such single-degree-of-freedom statistics can be somewhat informative even when more complex goodness-of-fit statistics are unreliable. When expected frequencies are very small, the distributions of the adjusted residuals are typically highly discrete and skewed to the right. Though we cannot interpret individual values too literally, absolute residuals exceeding about 2 or 3 can indicate potential problems with lack of fit. For Table 4, none of the 32 adjusted residuals exceeds 2.0. The conditional estimates for the cumulative Rasch model equal $\hat{\beta}_2 - \hat{\beta}_1 = 1.385$ ($ase = 0.224$) and $\hat{\pi} = 0.121$ ($ase = 0.224$). The estimated treatment effect is 1.385 - 0.121 = 1.264 when device A is used first, and 1.385 + 0.121 = 1.506 when device B is used first. The simpler model without the period effect has $G^2 = 8.45$ ($df = 5$) and an estimated treatment effect of 1.392 ($ase = 0.224$). Table 4 shows the fit of this model. For a given subject, the estimated odds that the response for device A falls below any particular category is $exp(1.392) = 4.02$ times the corresponding estimated odds for device B. The simpler model that also lacks a treatment effect fits poorly ($G^2 = 47.7, df = 6$), and the change in G^2 indicates extremely strong evidence of a treatment effect. For these data, Ezzet and Whitehead used a marginal ML approach with normally distributed random effect. They also indicated that the likelihood-ratio test showed a highly significant treatment effect. They reported estimates for a slightly more complex model, for which their estimated average treatment effect was 1.17.

For the insomnia data of Table 3, cumulative Rasch models provide similar conclusions as we obtained in the previous section using adjacent-categories logits. The model having the same change from baseline to follow-up for each treatment group fits well ($G^2 = 5.98, df = 5$), and has an estimated effect of -2.080 ($ase = 0.256$). The model with separate effects fits only slightly better ($G^2 = 5.14, df = 4$), with effects of -2.286 ($ase = 0.352$) for the active drug and - 1.813 ($ase = 0.377$) for placebo, for which the difference of 0.473 has $ase = 0.708$.

5 SOME SPECIALIZED RESULTS FOR ORDINAL MATCHED PAIRS

This section considers separately the special case $T = 2$ of a bivariate response, which occurs for matched-pairs data. In this case, quasi-symmetry models used to fit subject-specific models have logit representations, and computations are especially simple.

The quasi-symmetry model for $T = 2$ is

$$\log m(a, b) = \beta_1(a) + \beta_2(b) + \lambda(a, b) \ ,$$

where $\lambda(a, b) = \lambda(b, a)$. It has logit form

$$\log\{m(a, b)/m(b, a)\} = \tau_b - \tau_a$$

for all a and b, where $\tau_h = \beta_2(h) - \beta_1(h)$. In this logit form, the model is often called the Bradley-Terry model. We fit the logit random effects model (7) for adjacent categories using quasi-symmetry model (8), for which this logit representation is especially simple. Letting $\beta = \beta_2 - \beta_1$, the logit characterization is

$$\log\{m(a, b)/m(b, a)\} = \beta(b - a). \tag{15}$$

In fact, we can also fit the ordinal model using software for logistic regression models. For instance, to estimate the assumed common occasion effect in applying model (10) to Table 3, we can fit the logit model

$$\log\{m_k(a, b)/m_k(b, a)\} = \beta(b - a). \tag{16}$$

We then treat $\{n_k(a, b), a < b\}$ as binomial counts with sample sizes $\{n_k(a, b) + n_k(b, a)\}$.

Next, for the cumulative Rasch model (11) with $T = 2$, let (Y_{i1}, Y_{i2}) denote the responses for subject i. By independence,

$$\log[\frac{P(Y_{i1} \le h, Y_{i2} > h)}{P(Y_{i1} > h, Y_{i2} \le h)}] = logit(\gamma_{hi1}) - logit(\gamma_{hi2}) = (\beta_2 - \beta_1) \tag{17}$$

Collapsibility conditions imply that the same relationship holds for the joint distribution for the r^2 table averaged over subjects. For instance, for Table 4, let $m_k(a, b)$ denote the expected frequency under cross-over order k for outcome a with device A and outcome b with device B. The estimates for model (14) result from maximizing a product multinomial likelihood for Table 4, for the logit model

$$\log[\frac{\sum_{a \le h} \sum_{b > h} m_k(a, b)}{\sum_{a > h} \sum_{b \le h} m_k(a, b)}] = \beta + \pi[I(k = 2) - I(k = 1)] \tag{18}$$

for $h = 1, 2, 3$ and $k = 1, 2$, where $\beta = \beta_2 - \beta_1$. It is easy using SAS (CATMOD) to obtain a weighted least squares fit of this model applied simultaneously for the $(r-1)$ values of h.

For the cumulative Rasch model with $T = 2$, suppose we naively treated counts from different 2×2 tables with different cutpoint collapsings as independent, even though the same subjects occur in each table. When we use standard software to fit model (12), the estimate of $\beta = \beta_2 - \beta_1$ equals the estimate obtained for the 2×2 further collapsing of this table, collapsed over the cutpoint dimension. But this estimate is simply the log of the ratio of the two discordant counts in that table. In terms of cell counts $\{n_{ij}\}$ for the original $r \times r$ table, this estimate equals

$$\tilde{\beta} = log\{[\sum_{i<j}(j-i)n_{ij}]/[\sum_{i>j}(i-j)n_{ij}]\}. \tag{19}$$

Assuming the model holds, this is a consistent estimator of β, even though it is based on naive assumptions. This follows since we can express $exp(\tilde{\beta})$ such that the numerator is an average of the cell counts in row 1 and column 2 for the different collapsings, and the denominator is a corresponding average of the counts in row 2 and column 1. Each such collapsing has ratio of discordant counts that is itself consistent for $exp(\beta)$, being the conditional ML estimator for a particular binary response.

The estimate (19) is simple to compute and intuitively appealing. In using it, however, it would not be sensible to use a standard error that assumes the different 2×2 tables are independent. An estimated asymptotic variance for $\tilde{\beta}$ that is based on the more reasonable assumption that the original data $\{n_{ij}\}$ are a multinomial sample has form

$$\hat{V}(\tilde{\beta}) = \sum_{i<j}(j-i)^2 n_{ij}/[\sum_{i<j}(j-i)n_{ij}]^2 + \sum_{i>j}(i-j)^2 n_{ij}/[\sum_{i>j}(i-j)n_{ij}]^2. \tag{20}$$

We can base a simple test of marginal homogeneity for ordinal matched-pairs data on the ratio of $\tilde{\beta}$ to its estimated standard error.

The naive estimator $\tilde{\beta}$ seems to perform well, being nearly as efficient as much more complex ones. For instance, for the simple model without a period effect applied to Table 4, we can estimate the treatment effect by computing (19) for the 4×4 table collapsed over the cross-over sequences. This gives $\tilde{\beta} = 1.369$ and an estimated standard error of 0.237, compared to 1.392 and 0.224 for the ML approach of the previous section. For the model with common occasion effect applied to the insomnia data of Table 3, we obtain $\tilde{\beta} = 2.084$ and estimated standard error of 0.259, compared to 2.080 and 0.256 reported previously for the ML approach.

6 EXAMPLES OF SOFTWARE USE

A positive feature of the logit models discussed in this article is that they are easy to fit with existing software. One can obtain conditional ML estimates and equivalent distribution-free marginal ML estimates by fitting loglinear models that have the same likelihood in the response effects.

To illustrate, Table 5 illustrates the use of GLIM for fitting the quasi-symmetry model to obtain the results quoted for the Rasch model applied to Table 2. The factor called *'sym'* has levels corresponding to the cells that have the same terms for the complete symmetry model; specifically, cells having indices that are permutations of each other (such as (1,1,2), (1,2,1), and (2,1,1)) have the same level of this index. Quasi symmetry is the model that also has main effect terms. The calculate (*$calc*) directives for period effects add indicator variables that equal 1 for the cells in which the treatment for that period is at its second level. For instance, $p1$ has the values (0,0,0,0,1,1,1,1) for the 8 cells with $k = 1$ or 2 (when treatment A occurs in the first period), (0,0,1,1,0,0,1,1) when $k = 3$ or 4 (treatment B in the first period), and (0,1,0,1,0,1,0,1) when $k = 5$ or 6 (treatment C in the first period). A Poisson error assumption generates the same estimates and standard errors as our multinomial assumption. See Kenward and Jones[27] and Hatzinger[46] for related examples of using GLIM and SAS for these types of models, for nominal responses.

Table 6 uses GLIM to fit to Table 3 the quasi-symmetry model that corresponds to the adjacent-categories logit model described in Section 3. The symmetry factor (*sym*) generates a separate symmetry structure for each group. In some packages, it would be easier to fit the model using logit model software (such as PROC LOGISTIC or CATMOD in SAS) for the logit representation (16). Table 6 also uses GLIM to obtain estimates for the adjacent-categories logit model fitted to the ordinal-response version of Table 2 reported by Kenward and Jones. This fits a version of the model in which the treatment effects are identical for each treatment sequence, which has the same estimates as those for the 3^3 table collapsed over treatment sequence.

To obtain the fit reported for the cumulative logit model applied to the data in Table 4, one can use software that can fit generalized loglinear models of form (13) or can maximize a likelihood subject to constraints. Table 7 shows how to use SAS (CATMOD) to fit the model as expressed in form (18). The model is expressed as a generalized logit model of form $C \log A\mu = X\beta$, corresponding to generalized loglinear model (13). The model uses logits for each of the three collapsings of a 4×4 table to a 2×2 table. The *'response'* statement specifies the 6×12 matrix C and the 12×32

matrix A that are applied to the cell counts to construct the six logits, three for each cross-over order. The *'model'* statement contains the 6×2 model matrix X that is applied to the two parameters $(\beta_2 - \beta_1,\ \pi)$. To fit the simpler model with no period effect, one deletes the second column of X. One must add a small constant to all zero counts so that SAS treats them as sampling zeroes rather than structural zeroes. SAS provides a weighted least squares fit, which is more severely affected by small counts than the ML fit. For Table 4, substantive results are similar, however. The effect estimate is 1.275 ($ase = 0.224$), and the period estimate is 0.052 ($ase = 0.224$).

7 REMARKS

For any model discussed in this article, the unidimensionality of the subject effect suggests that the model is likely to fit decently only when the responses are different indicators of a common trait, rather than indicators of quite different traits. Also, the models assume that, conditional on the subject parameters, different responses are independent. In some applications involving measurements at different times or different locations, lack of fit may occur because of residual dependence between responses nearby in time or space, and one must include terms in the model to account for this[25,26].

Both forms of ordinal logit model tend to fit well in similar situations, since both imply stochastic orderings of responses. The cumulative logit model has a natural connection to a regression model for underlying continuous response variables. Moreover, if it holds for that underlying response, it holds with the same effects no matter how one collapses the response scale into discrete categories. The adjacant-categories logit link has the advantage of sufficient statistics and hence a direct conditional solution. Another factor that can influence the choice of model relates to whether one prefers odds ratio interpretations referring to the entire response scale (as in the cumulative logit model) or to pairs of categories (as in the adjacent-categories logit).

We specified models permitting different subjects to use different response cutpoints. There was no sacrifice in doing so, since the analyses in the examples estimated within-subject effects. Different approaches are necessary if one wants to make between-subjects comparisons of independent samples at fixed occasions. In addition, it is of interest to investigate the degree to which the methods of this article can be generalized to handle more complex forms of repeated measurement data. Examples include cases in which there is missing data or not the same number of repeats per subject, time-varying covariates or more complex covariate structure. The generalized estimating equations approach[4] is well-suited for handling such complications or for more general regression-type model-

ing of repeated measurement data.

ACKNOWLEDGEMENTS

This research was partially supported by a grant from the National Institutes of Health. Computations for the ML fit of the cumulative Rasch model were conducted using a FORTRAN program for generalized loglinear models written by Dr. Joseph Lang.

REFERENCES

1. Matthews, D. E., and V. T. Farewell. *Using and Understanding Medical Statistics*, 2nd revised edition. Basel: Karger. (1988).

2. Cox, D. R. Two further applications of a model for binary regression. *Biometrika* **45**, 562-565 (1958).

3. Neuhaus, J. M., J. D. Kalbfleisch, and W. W. Hauck. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, **59** 25-35 (1991).

4. Zeger, S. L., K.-Y. Liang, and P. S. Albert. Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049-1060 (1988).

5. Collett, D. *Modelling Binary Data*. London: Chapman and Hall (1991).

6. Rasch, G. On general laws and the meaning of measurement in psychology. pp. 321-333 in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. 4, ed. J. Neyman. Berkeley: Univ. of California Press (1961).

7. Fienberg, S. E. The Rasch model. *Encyclopedia of Statistical Sciences* **7**, 627-632 (1986).

8. Neyman, J. and E. L. Scott. Consistent estimates based on partially consistent observations. *Econometrika*, **16**, 1-32. (1948).

9. Andersen, E. B. *Discrete Statistical Models with Social Science Application*. Amsterdam: North-Holland (1980).

10. Anderson, D. A., and J. P. Hinde. Random effects in generalized linear models and the EM algorithm. *Commun. Statist., Theor. Meth.* **17**, 3847-3856 (1988).

11. Bock, R. D., and M. Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46**, 443-459 (1981).

12. Conaway, M. A random effects model for binary data. *Biometrics*, **46**, 317-328 (1990).

13. Laird, N. Topics in likelihood-based methods for longitudinal data analysis. *Statistica Sinica* **1**, 33-50 (1991).

14. Stiratelli, R., N. Laird, and J. H. Ware. Random-effects models for serial observations with a binary response. *Biometrics* **40**, 1025-1035 (1984).

15. Thissen, D. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* **47**, 175-186 (1982).

16. Ashby, M., J. M. Neuhaus, W. W. Hauck, P. Bacchetti, D. C. Heilbron, N. P. Jewell, M. R. Segal, and R. E. Rusaro. An annotated bibliograpphy of methods for analysing correlated categorical data. *Statistics in Medicine* **11**, 67-99 (1992).

17. Zeger, S. L., and M. R. Karim. Generalized linear models with random effects; A Gibbs sampling approach. *J. Amer. Statist. Assoc.* **86**, 79-86, (1991).

18. Fahrmeir, L., W. Hennevogl, and Tutz, G. *Multivariate Cross-Sectional and Longitudinal Data Analysis Based on Generalized Linear Models*, Springer series (to appear, 1993).

19. Tjur, T. A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics* **9**, 23-30 (1982).

20. Blackwood, L. Latent variable models for the analysis of medical data with repeated measures of binary variables. *Statistics in Medicine* **7**, 975-981 (1988).

21. Darroch, J. N., and P. I. McCloud. Category distinguishability and observer agreement. *Austral. J. Statist.* **28**, 371-388 (1986).

22. De Leeuw, J., and N. Verhelst. Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics* **11**, 183-196 (1986).

23. Cressie, N. E., and P. W. Holland. Characterizing the manifest probabilities of latent trait models. *Psychometrika* **37**, 129-141 (1983).

24. Lindsay, B., C. C. Clogg, and J. Grego. Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86**, 96-107 (1991).

25. Conaway, M. Analysis of repeated categorical measurements with conditional likelihood methods. *J. Amer. Statist. Assoc.*, **84**, 53-62 (1989).

26. Conaway, M. The analysis of repeated categorical measurements subject to nonignorable nonresponse. *Journal of the Americal Statistical Association* **87**, 817-824 (1992).

27. Kenward, M. G., and Jones, B. The analysis of categorical data from cross-over trials using a latent variable model. *Statistics in Medicine* **10**, 1607-1619, (1991).

28. Agresti, A. *Categorical Data Analysis*. New York: Wiley (1990).

29. Bishop, Y., Fienberg, S., and Holland, P. *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press (1975).

30. Jones, B., and M. G. Kenward. Modeling binary data from a three-period cross-over trial. *Statistics in Medicine*, **6**, 555-564, (1987).

31. Agresti, A. Computing conditional maximum likelihood estimates for generalized Rasch models using simple loglinear models with diagonals parameters. *Scand. J. Stat.*, **19**, to appear (1992).

32. Ezzet, F., and J. Whitehead. A random effects model for ordinal responses from a crossover trial. *Statist. Medic.* **10**, 901-907. (1991).

33. Harville, D. A., and Mee, R. W. A mixed-model procedure for analyzing ordered categorical data. *Biometrics*, **40**, 393-408 (1984).

34. McCullagh, P. A logistic model for paired comparisons with ordered categorical data. *Biometrika* **64**, 449-453 (1977).

35. Tutz, G. Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology* **43**, 39-55 (1990).

36. Andersen, E. B. Conditional inference for multiple-choice questionnaires. *Brit. J. Math. Statist. Psychol.* **26**, 31-44 (1973).

37. Koch, G. G., J. R. Landis, J. L. Freeman, D. H. Freeman, and R. G. Lehnen. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* **33**, 133-158 (1977).

38. Meeks, S. L., and R. B. D'Agostino. A model for comparisons with ordered categorical data. *Communications in Statistics* **A 12**, 895-906 (1983).

39. Francom, S. F., C. Chuang-Stein, and J. R. Landis. A log-linear model for ordinal data to characterize differential change among treatments. *Statist. Medic.* **8**, 571-582 (1989).

40. Agresti, A. A survey of models for repeated ordered categorical response data. *Statist. Medic.* **8**, 1209-1224 (1989).

41. Anderson, J. A., and P. R. Philips. Regression, discrimination, and measurement models for ordered categorical variables. *Appl. Statist.* **30**, 22-31 (1981).

42. McCullagh, P. On the elimination of nuisance parameters in the proportional odds model. *Journal of the Royal Statistical Society* **B 46**, 250-256 (1984).

43. Agresti, A., and J. B. Lang. A proportional odds model with subject-specific effects for repeated ordered categorical responses. To appear, *Biometrika* (1993).

44. Aitchison, J., and S. D. Silvey. Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29**, 813-828 (1958).

45. Haber, M. Maximum likelihood methods for linear and log-linear models in categorical data. *Comput. Statist. Data Anal.* **3**, 1-10 (1985).

46. Hatzinger, R. The Rasch model, some extensions and their relation to the class of generalized linear models. *Statistical Modelling: Proceedings of GLIM 89 and the 4th International Workshop on Statistical Modelling*, Lecture Notes in Statistics, 57. Berlin: Springer-Verlag (1989).

Table 1. Example of repeated responses for a binary variable.

	Pathologist B		
Pathologist A	Malignant	Benign	Total
Malignant	9	1	10
Benign	9	81	90
Total	18	82	100

Table 2. Data for crossover study comparing treatments for relief of primary dysmenorrhea.

Treatment	Response pattern for (A, B, C)							
sequence	111	112	121	122	211	212	221	222
A B C	0	2	2	9	0	0	1	1
A C B	2	0	0	9	1	0	0	4
B A C	0	1	1	8	1	3	0	1
B C A	0	1	1	8	1	0	0	1
C A B	3	0	0	7	0	1	2	1
C B A	1	5	0	4	0	3	1	0
Total	6	9	4	45	3	7	4	8
Fit	6.00	9.63	5.34	43.04	1.04	8.34	4.63	8.00

Table 3. Distribution of time to fall asleep, by treatment and occasion.

Treatment	Initial occasion	Follow-up occasion < 20	20-30	30-60	> 60
Active	< 20	7	4	1	0
		7.0	3.1	0.9	0.1
	20-30	11	5	2	2
		11.9	5.0	5.1	1.2
	30-60	13	23	3	1
		13.1	19.9	3.0	2.8
	> 60	9	17	13	8
		8.9	17.8	11.2	8.0
Placebo	< 20	7	4	2	1
		7.0	3.7	0.5	0.1
	20-30	14	5	1	0
		14.3	5.0	2.0	0.7
	30-60	6	9	18	2
		7.5	8.0	18.0	3.3
	> 60	4	11	14	22
		4.9	10.3	12.7	22.0

Note: Fitted values refer to model (10).

Table 4. Data on clarity of instructions, for inhaler study using crossover design.

Order	Inhaler A	Inhaler B 1	2	3	4
A then B	1	59	35	3	2
		59.0	36.1	2.5	2.7
	2	11	27	2	1
		9.8	27.0	1.6	1.3
	3	0	0	0	0
		0.0	1.0	0.0	0.0
	4	1	1	0	0
		0.5	0.5	0.0	0.0
B then A	1	63	40	7	2
		63.0	40.9	5.7	1.6
	2	13	15	2	0
		12.0	15.0	1.6	0.0
	3	0	0	1	1
		0.0	1.6	1.0	1.0
	4	0	0	0	0
		0.0	0.7	0.0	0.0

Note: Category 1 = easy, 2 = only clear after rereading, 3 = not very clear, 4 = confusing. Fitted values refer to model (14).

Table 5. GLIM code for fitting Rasch-type models to Table 2.

```
$units 48
$data count    $read
0  2  2  9  0  0  1  1    2  0  0  9  1  0  0  4    0  1  1  8  1  3  0  1
0  1  1  8  1  0  0  1    3  0  0  7  0  1  2  1    1  5  0  4  0  3  1  0
$calc a = %gl(2,4):  b = %gl(2,2):  c = %gl(2,1) $  !  Generates levels of treatments
$calc seq = %gl(6,8) $  !  Generates levels for six treatment sequences
$ass sym = 1,2,2,3,2,3,3,4,5,6,6,7,6,7,7,8,9,10,10,11,10,11,11,12,13,14,14,15,14,15,15,16,
17,18,18,19,18,19,19,20,21,22,22,23,22,23,23,24 $  !  Generates levels of symmetry term
$fac  a  2    b  2    c  2    sym  24    seq  6   $
$yvar count    $err pois
$fit  sym + a + b + c $  !  Model with same main effects for each sequence
$fit  + a.seq + b.seq + c.seq $  !  Different main effects for each sequence
$calc  p1 = (%eq(seq,1) + %eq(seq,2))*%eq(a,2) + (%eq(seq,3) + %eq(seq,4))*%eq(b,2) + (%eq(seq,5)
+ %eq(seq,6))*%eq(c,2) $
$calc  p2 = (%eq(seq,3) + %eq(seq,5))*%eq(a,2) + (%eq(seq,1) + %eq(seq,6))*%eq(b,2) + %eq(seq,2)
+ %eq(seq,4))*%eq(c,2) $
$fit  sym + a + b + c + p1 + p2 $  !  period effects
$end
$units 8  !  Now consider table collapsed over sequence
$data a b c count
$read
1  1  1  6    1  1  2  9    1  2  1  4    1  2  2  45
2  1  1  3    2  1  2  7    2  2  1  4    2  2  2  8
$ass sym = 1,  2,  2,  3,  2,  3,  3,  4 $
$fac a  2    b  2    c  2    sym  4
$yvar count    $err pois
$fit sym : + a + b + c $   !   Fits complete symmetry, then quasi symmetry
$dis e  s  r   !   Displays estimates, std. errors of differences, fit and residuals
```

Table 6. GLIM code for fitting adjacent-categories logit models to Table 3 and expanded version of Table 2.

```
$units 32
$data count   $read   !   Data from Table 3
7  4  1  0    11  5  2  2    13  23  3  1    9  17  13  8
7  4  2  1    14  5  1  0    6  9  18  2    4  11  14  22
$calc i = %gl(4,4):  f = %gl(4,1) $   !   generates levels for initial and follow-up
$ass sym = 1, 2, 3, 4, 2, 5, 6, 7, 3, 6, 8, 9, 4, 7, 9, 10,
11, 12,  13,  14,  12,  15,  16,  17,  13,  16,  18,  19,  14,  17,  19,  20 $
$fac sym 20  $
$yvar count   $err pois
$fit sym : + i + f $  !  Fits complete symmetry and model (10)
$end
$units 27
$data count   $read   !   Data from Kenward-Jones expansion of Table 2
6  4  5    3  13  10    1  8  14
2  3  2    1  3  1    2  1  2
1  0  2    0  0  0    1 1 0
$calc a = %gl(3,9):  b = %gl(3,3):  c = %gl(3,1)  $
$ass sym = 1,2,3,2,4,5,3,5,6,2,4,5,4,7,8,5,8,9,3,5,6,5,8,9,6,9,10  $
$fac sym 10  $
$yvar count   $err pois
$fit sym:  + a + b + c  $  !  Fits complete symmetry and model (8)
$end
```

Table 7. SAS code for fitting cumulative Rasch model to Table 4.

```
data culogit;
input cell $ count @@;  if count=0 then count=1e-10;
cards;
1  59    2  35    3  3    4  2    5  11    6  27    7  2    8  1
9  0    10  0    11  0    12  0    13  1    14  1    15  0    16  0
17  63    18  40    19  7    20  2    21  13    22  15    23 2    24  0
25  0    26  0    27  1    28  1    29  0    30  0    31  0    32  0
;
proc catmod   order=data;   weight count;
response  1  -1  0  0  0  0  0  0  0  0  0  0,  0  0  1  -1  0  0  0  0  0  0  0  0,
0  0  0  0  1  -1  0  0  0  0  0  0,  0  0  0  0  0  0  1  -1  0  0  0  0,
0  0  0  0  0  0  0  0  1  -1  0  0,  0  0  0  0  0  0  0  0  0  0  1  -1
log
0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0;
model cell = (1 -1, 1 -1, 1 -1, 1 1, 1 1, 1 1)/pred;
```