LOGIT MODELS FOR REPEATED ORDERED CATEGORICAL RESPONSE DATA

Alan Agresti, University of Florida

ABSTRACT

Logit models are presented for repeated observations on an ordinal categorical response variable. The models describe how the marginal distribution of the response depends on values of explanatory variables. Primary attention is given to a cumulative logit model that generalizes models proposed by McCullagh (1977, 1980). Also discussed are an adjacent-categories logit model that is equivalent to a loglinear model for marginal response probabilities, and a mean response model. The models are illustrated with an example in which a drug and placebo are compared with respect to treatment of insomnia problems. When the covariates are categorical, a weighted least squares fit for each type of model can be obtained using PROC CATMOD.

1. INTRODUCTION

In many applications a response variable is observed for each subject under several conditions -- for instance, at several time points, or before and after receiving some treatment. In other applications there are matched sets of subjects, with one subject from each set assigned to each condition. In either case, the matching results in statistically dependent observations. Repeated measurement on a categorical response variable occurs commonly in health-related applications, such as when a clinician makes a subjective evaluation of patients at weekly intervals about whether a new drug treatment has been successful. This paper proposes models for the case of an ordered categorical response variable, such as evaluation of success measured on the scale (excellent, good, fair, poor).

When a response is obtained for each subject under only two conditions, the data can be described by a square contingency table, the same categories occurring in each dimension. The cell in row i and column j gives the number of subjects classified in the ith level for the first condition and the jth level for the second condition. There is a large body of research literature on the analysis of square tables and their multidimensional analogs; for a review of much of it, see Bishop et al. (1975, Chap. 8). In practice, it is often desirable to incorporate a vector $\underline{x}$ of explanatory variables in the model. The literature for this case is considerably more sparse. The article by Koch et al. (1977) is a good source both for describing types of repeated measures data and for suggesting possible models.

Table 1, taken from Francom et al. (1987), is an example of repeated ordered categorical response data. The table gives results of a randomized, double-blind clinical trial comparing an active hypnotic drug with placebo in patients having insomnia problems. The response variable is patient response to the question, "How quickly did you fall asleep after going to bed?", measured using the ordered categories (<20 minutes, 20-30 minutes, 30-60 minutes, and >60 minutes). Patients were asked this question before and at the conclusion of a two-week treatment period. The two treatments, Active and Placebo, can be regarded as levels of a binary explanatory variable. In Table 1, the subjects receiving the two treatments are independent samples. In summary, the repeated measurement over time produces an ordinal response variable (patient's response) that is bivariate, measured under conditions (initial, follow-up). Also, there is a single explanatory variable, corresponding to the two types of treatment. There is repeated measurement over condition, but not over treatment. We shall use models formulated in this article to analyze these data. More generally, these models can be used to analyze ordinal variables measured under several conditions, for several independent groups (or levels of explanatory variables).

To analyze repeated categorical data, we could formulate a model for the cell expected frequencies. Such a model would describe the dependence structure of the entire cross-classification. An alternative approach is to formulate a model for the marginal distributions of the response within the various conditions and levels of the explanatory variables. In Table 1, there are four such marginal distributions: for each treatment, there is a patient's response distribution before treatment and after treatment.

For descriptive and inferential purposes, the full multivariate dependence among the repeated responses is often of less interest than the behavior of the marginal distributions of the response. Modeling the marginal distribution permits investigation of questions such as, "For a particular treatment, does the response improve with time?" or "At a particular time, are there differences among the response distributions for the various treatments?" or "Is the difference in response for any two treatments the same at all times?" For further discussion of this point, see Koch et al. (1977), Liang and Zeger (1985), and Stram et al. (1988).

We consider two types of logit models that describe how marginal distributions of a repeated ordinal categorical response depend on conditions and covariates. One type uses cumulative probabilities and the other uses adjacent pairs of response probabilities. We also briefly discuss a model suggested by Koch et al. (1977), based on describing variation in the means of the marginal distributions, for some fixed choice of response scores.

Section 2 surveys different types of logit transformations for ordinal response variables. Section 3 gives notation for modeling marginal distributions of a repeated response, and the models are presented in Sections 4-6. Section 7 discusses model-fitting, Section 8 uses the models to analyze the data in Table 1, and Section 9 shows how to fit the models using PROC CATMOD. Section 10 briefly discusses an alternative logit model that is much like an analysis of covariance model, describing response differences between groups while controlling for a baseline response. The final section discusses relative merits of the various models and special problems provided by repeated categorical measurement data.

2. TYPES OF LOGITS

If π denotes the probability of a particular response, the logit for that probability is defined to be

$$\text{logit}(\pi) = \log[\pi/(1-\pi)],$$

the log of the odds of making that response. Logits are defined for binary responses, but can be generalized for $r > 2$ response categories, there being r-1 non-redundant logits.

For instance, suppose the response variable Y is ordinal. Let

$$\pi_j = P(Y = j), \quad j = 1,...,r.$$

We shall take the order of categories into account by constructing logits for cumulative probabilities,

$$\text{logit}[P(Y \le j)] = \log[(\pi_1 + \cdots + \pi_j)/(\pi_{j+1} + \cdots + \pi_r)]$$

$j = 1,...,r-1$, called cumulative logits, and for adjacent-response

probabilities,

$$\log\left[\pi_j/\pi_{j+1}\right] = \text{logit}\left[P(Y=j|Y = j \text{ or } j+1)\right],$$

$j = 1,...,r-1$, called adjacent-categories logits. A third type is the set of logits

$$\log\left[\pi_j/(\pi_{j+1}+\cdots+\pi_r)\right] = \text{logit}\left[P(Y=j|Y\geq j)\right]$$

or, alternatively, the set

$$\log\left[\pi_j/(\pi_1+\cdots+\pi_{j-1})\right] = \text{logit}\left[P(Y=j|Y\leq j)\right]$$

$j = 2,...,r$. These are called continuation-ratio logits.

Depending on the application, one type of logit may lead to more natural interpretations than the others. Cumulative logits are useful for extending interpretations to an underlying continuous distribution for the response variable; see McCullagh (1980). For categorical explanatory variables, models for adjacent-categories logits are equivalent to loglinear models; see Goodman (1983). Continuation-ratio logits have the appealing feature that results of fitting models separately for different j are statistically independent; see Fienberg (1980, pp. 114-116). For cumulative logits and adjacent-categories logits, it is often possible to use a simple model in which effects of explanatory variables are identical for the various cutpoints for forming the logits.

3. MODELING MARGINAL DISTRIBUTIONS

Denote by d the number of conditions under which each subject in the sample is observed. Since there are r possible responses at each condition, the frequencies of the possible multivariate response profiles can be summarized in a contingency table with r^d cells. Let

$$\pi_{\underset{\sim}{j}} \text{ with } \underset{\sim}{j} = (j_1, \ldots, j_d)$$

denote the probability that the response for a randomly selected subject is j_g under condition g, $1\leq j_g\leq r$, $g=1,...,d$. The case $d=2$ corresponds to a square contingency table with r categories in each direction.

When there are covariates $\underset{\sim}{x}$, we allow a separate distribution

$$\left\{\pi_{\underset{\sim}{j}}(\underset{\sim}{x}),\ 1 \leq j_g \leq r, \quad g = 1, \ldots, d\right\}$$

at each level of $\underset{\sim}{x}$. When $\underset{\sim}{x}$ is categorical, its levels usually correspond to subpopulations, or treatments, whose distribution on the response we would like to compare.

Let the + subscript denote summation over that index. Then $\{\pi_{+\cdots+k+\cdots+}(\underset{\sim}{x}),\ k=1,...,r\}$, where k is in position g, is the marginal distribution of the response under the g^{th} condition, at level $\underset{\sim}{x}$ of the explanatory variables. Denote these marginal probabilities by $\{\phi_{gk}(\underset{\sim}{x}),\ k=1,...,r\}$. That is, $\phi_{gk}(\underset{\sim}{x})$ is the probability of making response k, for the g^{th} condition at level $\underset{\sim}{x}$ of explanatory variables, so $\Sigma_k\phi_{gk}(\underset{\sim}{x}) = 1$. The primary concern may be to analyze how the marginal distribution changes across the conditions $g=1,...,d$, for fixed $\underset{\sim}{x}$, or it may be to analyze how it depends on $\underset{\sim}{x}$, for fixed g. In either case, interpretations depend on whether the difference (on some scale) between the marginal distributions for two covariate values is the same for all conditions; that is, on whether there is condition × covariate interaction.

4. CUMULATIVE LOGIT MODELS

In the single-response ($d=1$) case, models based on the cumulative logit transformation of the response probabilities have been utilized by Williams and Grizzle (1972) and McCullagh (1980), among others. For such models it is unnecessary to assign scores to response categories, and model parameters can be simply interpreted using odds ratios. Models apply simultaneously to the cumulative logits formed for each of the r-1 possible cutpoints for collapsing the response into two categories. At condition g and covariate value $\underset{\sim}{x}$, the cumulative logits for marginal response probabilities are, for $k=1,...,r-1$,

$$L_{gk}(\underset{\sim}{x}) = \log\left[\frac{\phi_{g1}(\underset{\sim}{x})+\cdots+\phi_{gk}(\underset{\sim}{x})}{\phi_{g,k+1}(\underset{\sim}{x})+\cdots+\phi_{gr}(\underset{\sim}{x})}\right]$$

At a particular setting $\underset{\sim}{x}$, the d marginal distributions of the contingency table can be represented by a d×r table of $\{\phi_{gk}\}$ values. In terms of cumulative logits, the saturated model for this table has form

$$L_{gk} = \alpha_{gk}, \quad g = 1, \ldots, d,\ k = 1, \ldots, r\text{-}1.$$

The "cutpoint" parameters $\{\alpha_{gk}\}$ increase in k for fixed g (since the cumulative probabilities increase in k), and the saturated model permits them to vary by condition g in an unstructured manner.

A cumulative logit model that incorporates effects of covariates is

$$L_{gk}(\underset{\sim}{x}) = \alpha_{gk} + \underset{\sim}{\beta}'_g\underset{\sim}{x}, \tag{4.1}$$

for $g=1,...,d$, $k=1,...,r-1$. In this and subsequent formulas, the vectors (such as $\underset{\sim}{\beta}$ and $\underset{\sim}{x}$) are column vectors. This model allows condition×covariate interaction -- the effect $\underset{\sim}{\beta}_g$ of the covariates on the response is permitted to depend on the condition g.

As in McCullagh's (1980) model for a single response, the covariate effect in (4.1) is assumed to be identical for all cutpoints k. Under this assumption, it is simple to describe the covariate effect at each condition, since

$$L_{gk}(\underset{\sim}{x}_1) - L_{gk}(\underset{\sim}{x}_2) = \underset{\sim}{\beta}'_g(\underset{\sim}{x}_1 - \underset{\sim}{x}_2)$$

is identical for all k. The antilog of this measure is simply an odds ratio -- the odds of making response $\leq$ k (rather than $>$ k) when $\underset{\sim}{x} = \underset{\sim}{x}_1$, divided by the odds of making response $\leq$ k when $\underset{\sim}{x} = \underset{\sim}{x}_2$. For fixed g, model (4.1) induces an ordering of cumulative probabilities among levels of $\underset{\sim}{x}$ according to the values of $\underset{\sim}{\beta}'_g\underset{\sim}{x}$, the ordering being the same for each k; hence, for each condition, marginal distributions at different covariate values are stochastically ordered on the response variable.

We can simplify (4.1) by modeling how $\underset{\sim}{\beta}_g$ varies across conditions. When estimated effects in (4.1) are fairly stable across the conditions, the model

$$L_{gk}(\underset{\sim}{x}) = \alpha_{gk} + \underset{\sim}{\beta}'\underset{\sim}{x} \tag{4.2}$$

may be adequate. This model assumes an absence of condition covariate interaction as well as a covariate effect that is independent of the cutpoint. Hence, $L_{gk}(\underset{\sim}{x}_1) - L_{gk}(\underset{\sim}{x}_2)$ is the same for all conditions and for all cutpoints.

Condition effects are simpler to interpret when the model provides some structure for the marginal inhomogeneity over the conditions. For instance, consider the special case of (4.2)

$$L_{gk}(\underset{\sim}{x}) = \alpha_k + \mu_g + \underset{\sim}{\beta}'\underset{\sim}{x} \tag{4.3}$$

The cutpoint parameters $\{\alpha_k\}$ are usually nuisance parameters;

the condition effects are described by the $\{\mu_g\}$ and the covariate effects by $\underset{\sim}{\beta}$. For this model,

$$L_{bk}(\underset{\sim}{x}) - L_{ak}(\underset{\sim}{x}) = \mu_b - \mu_a.$$

The difference in cumulative logits is the same for all cutpoints; that is, the marginal distributions for different conditions are location shifts on a cumulative logit scale. The odds under condition b of making response below category k are $\exp(\mu_b-\mu_a)$ times greater than the corresponding odds under condition a, for all k and for all $\underset{\sim}{x}$.

Also for model (4.3),

$$L_{gk}(\underset{\sim}{x}_1) - L_{gk}(\underset{\sim}{x}_2) = \underset{\sim}{\beta}'(\underset{\sim}{x}_1 - \underset{\sim}{x}_2),$$

so simple interpretations apply to covariate effects, which are the same for all conditions and for all cutpoints. More generally,

$$L_{bk}(\underset{\sim}{x}_1) - L_{bk}(\underset{\sim}{x}_2) = (\mu_b - \mu_a) + \underset{\sim}{\beta}'(\underset{\sim}{x}_1-\underset{\sim}{x}_2)$$

for all k, which implies a stochastic ordering of the response marginal distributions at all combinations of condition and covariate values.

McCullagh (1977) proposed model (4.3) for d=2, s=1 (two conditions, no covariate). In that case, $\mu_1 = \mu_2$ gives the much-studied model of marginal homogeneity for a two-way table. More generally, when all μ_g are equal, model (4.3) simplifies to marginal homogeneity across d conditions. On the other hand, if $\underset{\sim}{\beta} = \underset{\sim}{0}$ in (4.3), there are no covariate effects, and condition effects are location shifts on a cumulative logit scale. When all μ_g are equal and $\underset{\sim}{\beta} = \underset{\sim}{0}$ in (4.3), the response is independent of condition and of the covariates.

Model (4.3) can be generalized, for instance by replacing $\underset{\sim}{\beta}$ by $\underset{\sim}{\beta}_g$, thus permitting condition×covariate interaction but maintaining a simple structure for the marginal inhomogeneity. Table 2 lists a variety of models that have condition and covariate effects independent of cutpoint.

Suppose $\underset{\sim}{x}$ is fully categorical, and denote by s the number of settings of $\underset{\sim}{x}$ at which observations occur. Then the data consist of cell counts in a $s\times r^d$ contingency table; that is, there is a separate r^d table for each of the s settings of $\underset{\sim}{x}$. In this case, Table 3 lists the residual degrees of freedom (df) for testing goodness of fit of the models in Table 2. These formulas are based on ds(r-1) marginal cumnlative logits [(r-1) logits for each of the ds marginal distributions], d(r-1) $\{\alpha_{gk}\}$ parameters, and an identifiability constraint (such as $\mu_d \equiv 0$) placed on the $\{\mu_g\}$.

Fignre 1 shows the hierarchical relationships of the models listed in Table 2. The figure becomes considerably more complicated when we allow for models in which $\underset{\sim}{\beta}$ is replaced by $\underset{\sim}{\beta}_k$; that is, models that have covariate×cutpoint interaction in the effect of $\underset{\sim}{X}$ on the response. For the models in Table 2, we can test the assumption of a lack of interaction of this type by fitting the corresponding model that allows it. For instance, model (4.3) would be compared to

$$L_{gk}(\underset{\sim}{x}) = \alpha_k + \mu_g + \underset{\sim}{\beta}_k'\underset{\sim}{x},$$

and the difference in goodness-of-fit statistics would give a test, based on df = (r-2)dim($\underset{\sim}{\beta}$), of $H_0 : \underset{\sim}{\beta}_1=\cdots=\underset{\sim}{\beta}_{r-1}$. When this more complex model gives a better fit, we can still investigate marginal homogeneity across the d conditions by testing H_0 : $\mu_1=\cdots=\mu_d$ in it.

5. ADJACENT-CATEGORIES LOGIT MODELS

We can generalize models introduced in the previous section by using, instead of the logit, an alternative link function that transforms monotonically the (0,1) cumulative probability scale onto the real line. Such models include cumulative probit models, proportional hazards models, and others of the sort studied by McCullagh (1980). There are also alternative types of logit transformations that can be used, such as continuation-ratio logits and adjacent-category logits. In this section we discuss the latter transformation.

At condition g and covariate value $\underset{\sim}{x}$, the adjacent-category logits are

$$L'_{gk}(\underset{\sim}{x}) = \log\left[\phi_{gk}(\underset{\sim}{x})/\phi_{g,k+1}(\underset{\sim}{x})\right],$$

k=1,...,r-1. The models formulated in Table 2 for cumulative logits also make sense for adjacent-category logits. The formulas for residual df are identical to those given in Table 3, but the models for adjacent-category logits are not equivalent to those for cumulative logits when r > 2. The interpretations refer to identical odds ratios for all pairs of adjacent response categories.

Goodman (1983) discussed the use of adjacent-category logits in models for a single response. Those models are equivalent to loglinear models presented in Goodman (1979). Similarly, adjacent-category logit models can be expressed as loglinear models for the marginal probabilities $\{\phi_{gk}(\underset{\sim}{x})\}$.

For instance, consider the case in which there is a single covariate, say a nominal variable corresponding to s subpopulations. Let ϕ_{gik} denote the probability of response k under condition g, for subpopulation i. A model for adjacent-category logits that is analogous to model (4.3) for cumulative logits is

$$\begin{aligned} L'_{gik} &= \log(\phi_{gik}/\phi_{gi,\,k+1}) \\ &= \alpha_k + \mu_g + \beta_i \end{aligned}$$

For each pair of adjacent response categories, the odds of making the lower (rather than higher) response are $\exp(\mu_b - \mu_a)$ times greater under condition b than under condition a (for each subpopulation), and they are $\exp(\beta_b - \beta_a)$ times greater for subpopulation b than for subpopulation a (under each condition). The response marginal distributions for the ds different condition-subpopulation combinations are stochastically ordered according to the values of $\{\mu_g + \beta_i\}$. This model is equivalent to the loglinear model

$$\log(\phi_{gik}) = \mu + \lambda_g^C + \lambda_i^S + \lambda_k^R + \lambda_{gi}^{CS} - a_k\beta_i - a_k\mu_g$$

with $\{a_k=k\}$ and $\alpha_k = \lambda_k^R - \lambda_{k+1}^R$, where C = condition, R = response, S = subpopulation. This loglinear model applies to d×s×r marginal probabilities of the original $s\times r^d$ contingency table, and, because of the dependence in the samples across conditions, it cannot be fitted using standard loglinear methods for three-way tables.

Adjacent-category logits and cumulative logits coincide when there are only r=2 response categories. For that case, logit models for marginal distributions of repeated measures data were discnssed by Koch et al. (1977).

6. MEAN RESPONSE MODELS

The models discussed in this section are simpler to interpret than the logit models of the previous two sections, but are structurally more controversial. Unlike the logit models, they require assignment of scores $\{a_1,\ldots,a_r\}$ to response categories. This choice is often straightforward for grouped

Table 1. Frequency Distribution of Time to Falling Asleep, by Treatment and Condition.

Time to Falling Asleep

		Initial	Follow-Up		
Treatment		<20	20-30	30-60	>60
Active	<20	7	4	1	0
	20-30	11	5	2	2
	30-60	13	23	3	1
	>60	9	17	13	8
Placebo	<20	7	4	2	1
	20-30	14	5	1	0
	30-60	6	9	18	2
	>60	4	11	14	22

Source: Francom et al. (1987)

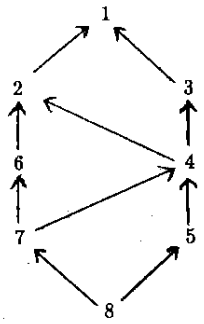
Table 2. Summary of Logit Models and Interpretations.

Model $L_{gk}(\underset{\sim}{x}) = :$	Description of Model
1. $\alpha_{gk}+\underset{\sim}{\beta}'_g\underset{\sim}{x}$	condition × covariate interaction
2. $\alpha_{gk}+\underset{\sim}{\beta}'\underset{\sim}{x}$	no condition × covariate interaction
3. $\alpha_k+\mu_g+\underset{\sim}{\beta}'_g\underset{\sim}{x}$	condition × covariate interaction, location condition effects for each $\underset{\sim}{x}$.
4. $\alpha_k+\mu_g+\underset{\sim}{\beta}'\underset{\sim}{x}$	covariate effects, location condition effects
5. $\alpha_k+\underset{\sim}{\beta}'\underset{\sim}{x}$	covariate effects, no condition effects
6. α_{gk}	condition effects, no covariate effects
7. $\alpha_k+\mu_g$	location condition effects, no covariate effects
8. α_k	no condition effects, no covariate effects

Table 3. Residual Degrees of Freedom for Models in Table 2, for Categorical Covariates.

Model	Residual df
1.	$d(r-1)(s-1) - \Sigma_g \dim(\underset{\sim}{\beta}_g)$
2.	$d(r-1)(s-1) - \dim(\underset{\sim}{\beta})$
3.	$(r-1)(ds-1) - (d-1) - \Sigma_g \dim(\underset{\sim}{\beta}_g)$
4.	$(r-1)(ds-1) - (d-1) - \dim(\underset{\sim}{\beta})$
5.	$(r-1)(ds-1) - \dim(\underset{\sim}{\beta})$
6.	$d(r-1)(s-1)$
7.	$(r-1)(ds-1) - (d-1)$
8.	$(r-1)(ds-1)$

Figure 1. Nesting of Cumulative Logit Models (Arrow Points to More General Model).



continuous variables, but there may be several reasonable assignments for subjective rating scales. The mean response under condition g, for covariate $\underset{\sim}{x}$, is

$$M_g(\underset{\sim}{x}) = \Sigma_k a_k \phi_{gk}(\underset{\sim}{x}).$$

Unlike the logit models, mean response models do not characterize the marginal distribntions in their entirety, but only through a measure of location.

A general mean response model for repeated ordinal categorical data is

$$M_g(\underset{\sim}{x}) = \alpha + \mu_g + \underset{\sim}{\beta}_g' \underset{\sim}{x},$$

g=1,..,d. This model allows conditionxcovariate interaction, through a different regression of the response on the explanatory variables at each condition. The simpler model

$$M_g(\underset{\sim}{x}) = \alpha + \mu_g + \underset{\sim}{\beta}' \underset{\sim}{x}, \qquad (6.1)$$

g=1,..,d, assumes an absence of such interaction. When this model fits adequately, variation in the marginal means is simple to describe, with condition effects given by the $\{\mu_g\}$ and the influence of the explanatory variables at each condition given by $\underset{\sim}{\beta}$. There are d marginal response means at each of the s settings of $\underset{\sim}{x}$, so when the covariates are categorical, the residual degrees of freedom for testing the fit of model (6.1) equals df = ds - [d + dim($\underset{\sim}{\beta}$)].

As with the cumulative logit model (4.3), further special cases correspond to lack of condition effects or lack of covariate effects. For instance, the special case in which the $\{\mu_g\}$ are equal corresponds to marginal homogeneity of the means across the d conditions, for each $\underset{\sim}{x}$.

When the repeated response is ordinal, modeling the mean response was suggested by Koch et al. (1977). For r=2 responses, this approach assumes a linear influence of $\underset{\sim}{X}$ on the probability of making a particular response. Such a model has been criticized for being structurally unsound, since it can yield predicted probabilities outside the [0,1] range. Similarly, for the r response case, mean response models cannot hold over an unbounded range of covariate values. However, when r is moderately large and there is reasonable dispersion of responses among the r categories at each setting of $\underset{\sim}{x}$, it is unusual to obtain predicted means below a_1 or above a_r.

The modeling of only a measure of location of the response does not allow simple comparisons of entire response distributions (such as whether they are stochastically ordered). Also, when r >2, special cases of the model do not correspond exactly to conditions such as marginal homogeneity or statistical independence of response and covariates. On the other hand, positive features of the mean response model include (1) the interface with standard regression modeling that occurs as r increases, so that the response is more nearly continuous, and (2) the simplicity of description obtained with it -- condition effects and covariate effects being given in terms of differences of means. Hence, it is a useful model when the technical level of one's clients necessitates an approach having simple interpretations.

7. MODEL FITTING

A commonly assumed sampling model for repeated measures on a categorical response is multinomial sampling over the r^d possible response profiles, with independent samples at each of the s levels of $\underset{\sim}{x}$. In other words, if $\underset{\sim}{\pi}$ denote the cell probabilities in the full table, tben $\underset{\sim}{\pi}$ consists of s independent sets of multinomial probabilities.

Haber (1985a) gave an iterative Newton-Raphson routine for obtaining ML estimates for models of the form

$$A\underset{\sim}{\pi} = X\underset{\sim}{\beta} \quad \text{or} \quad A \log \underset{\sim}{\pi} = X\underset{\sim}{\beta}.$$

The models discussed in Section 6 are of the first form, so Haber's routine can be applied to fit those models. The logit models of Sections 4 and 5 have the form

$$A \log B\underset{\sim}{\pi} = X\underset{\sim}{\beta},$$

Haber (1985b) showed how to obtain ML fits for this class of models. Haber's routines are an application of theory described by Aitchison and Silvey (1958) for maximizing a likelihood subject to constraints. These routines are impractical when there is a large number of cells in the table (such as happens when there are several conditions, several gronps, and a several-point scale), because of the size of the matrix that must be inverted.

Stram et al. (1988) have proposed a somewhat different approach to cumulative logit modeling of repeated measures data, one that more easily permits ML solutions. They proposed fitting a cumulative logit model separately to each of the sxr marginal tables obtained for the d different conditions. They also showed how to construct the joint asymptotic covariance matrix of the model parameter estimates obtained under the various conditions. This approach makes it simple to allow for time-dependent covariates and for missing data. However, since their approach involves fitting separate models under the various conditions, it does not give as special cases models for the full table that assume either marginal homogeneity or a constant covariate effect $\underset{\sim}{\beta}_g = \underset{\sim}{\beta}$, g=1,..,d. Their approach is related to recent work by Liang and Zeger (1985), who modeled repeated measures data using a generalized linear model for each condition.

When the covariates are fully categorical, the weighted least squares (WLS) approach suggested by Koch et al. (1977) for repeated measures data can be applied to all models given in this paper. That approach requires a nonsingular estimated covariance matrix for the ds(r-1) sample marginal logits for the models in Sections 4 and 5 and for the ds sample means for the models in Section 6. A WLS solution can be difficult to obtain for tables with small sample sizes or for large, sparse tables, because of the possibility of ill-defined sample logits (for instance, because of a zero marginal count) or a singular sample covariance matrix for the sample response functions used in the model. However, WLS methods are more likely to be manageable for the models discussed here than for models for the joint cell probabilities in the full sxr^d table. Even if the full table is quite sparse, the sxrxd table of marginal counts may not be.

An advantage of the WLS approach is that it can be implemented using PROC CATMOD. An important disadvantage is its inefficiency in handling continuous covariates, which must be collapsed into categorical variahles in order to use WLS.

8. EXAMPLE

We now use Table 1 to illustrate several models for repeated ordinal categorical data. Tahle 4 contains the marginal distributions of Table 1 for the four combinations of treatment and condition. From the initial to follow-up observation, the distribution of time to falling asleep seems to shift downwards both for the active and placebo treatments. The degree of shift seems to be greater for the active treatment, though, indicating that an interaction term may be needed in the model. Since there is only a single covariate (treatment), we will replace the notation $L_{gk}(\underset{\sim}{x})$ by the simpler L_{gik}, where i indexes treatment.

The model (4.3) of location shifts for conditions and treatments is

$$L_{gik} = \alpha_k + \mu_g + \beta_i, \qquad (8.1)$$

g=1,2, i=1,2, k=1,2,3. For identifiability, we impose constraints $\mu_2 = 0$ and $\beta_2 = 0$. The WLS fit of this model gives $\hat{\mu}_1 = -1.29$ (a.s.e. = 0.13) and $\hat{\beta}_1 = 0.37$ (a.s.e. = 0.20). According to this model, for each treatment, the odds that time to falling asleep is below any fixed level is estimated to be exp(1.29) = 3.6 times as high at the follow-up observation as at the initial observation. Similarly, for each condition, the odds that time to falling asleep is below any fixed level is estimated to be exp(0.37) = 1.4 times as high for the active treatment as for the placebo.

These interpretations assume the no-interaction model is reasonable. However, the goodness-of-fit statistic for testing (8.1) equals 14.5. The residual degrees of freedom equal 7, since there are 12 cumulative logits (3 for each marginal distribution) and 5 parameters (α_1, α_2, α_3, μ_1, β_1) in the model. The lack of fit is not surprising, since the treatments have similar sample marginal distributions at the initial observation, but not at the follow-up.

Next consider the model

$$L_{gik} = \alpha_k + \mu_g + \beta_i + \eta_{gi} \qquad (8.2)$$

where $\mu_2 = \beta_2 = \eta_{12} = \eta_{21} = \eta_{22} = 0$. The WLS fit gives $\hat{\alpha}_1$ = -1.16, $\hat{\alpha}_2 = 0.10$, $\hat{\alpha}_3 = 1.37$, $\hat{\mu}_1 = -1.05$ (a.s.e. = 0.16), $\hat{\beta}_1 = 0.69$ (a.s.e. = 0.23), and $\hat{\eta}_{11} = -0.65$ (a.s.e. = 0.25). There is substantial evidence of interaction. The model fits reasonably well, with a residual chi-squared of 7.4 based on df = 6.

For the active treatment, the odds that time to falling asleep is below any fixed level is estimated to be exp(1.05+0.65) = 5.5 times as high at the follow-up observation as at the initial; for the placebo group, the corresponding effect is exp(1.05) = 2.9. At the initial observation, the odds that falling asleep is below any fixed level is estimated to be exp(0.69-0.65) = 1.04 times as high for the active group as for the placebo; at the follow-up observation, the corresponding effect is exp(0.69) = 2.0. In other words, the placebo and active groups had about the same distributions of time to falling asleep at the initial observation, but at the follow-up the active group tended to fall asleep more quickly than the placebo. The interaction model gives a simple and economical description of the variation among the four marginal distributions.

Table 5 summarizes results for the two cumulative logit models just described. The table also gives results for the corresponding models using the adjacent-categories logit transform, for which similar results hold. The parameter estimates are somewhat smaller for the adjacent-categories logit transform, which is expected since the effects for that transform refer to a restricted range of the response.

The interaction model for adjacent-categories logits fits adequately, and can be interpreted as follows: For the active group, the odds that time to falling asleep is <20 minutes instead of 20-30 minutes (or 20-30 minutes instead of 30-60 minutes, or 30-60 minutes instead of >60 minutes) is estimated to be exp(0.55+0.36) = 2.5 times as high at the follow-up observation as at the initial observation; for the placebo group, the corresponding effect is exp(0.55) = 1.7. Similarly, at the initial observation, the odds that time to falling sleep is in category k instead of k+1 (for k = 1,2,3) is estimated to be exp(0.38-0.36) = 1.01 times as high for the active group as the placebo; at the follow-up observation, the corresponding effect is exp(0.38) = 1.46. We obtain the same substantive conclusions as we did using cumulative logits.

Table 4. Observed and Fitted Marginal Proportions (in Parentheses) for Cumulative Logit Model (8.2)

		Response			
Treatment	Condition	<20	20-30	30-60	>60
Active	Initial	.101	.168	.336	.395
		(.102)	(.184)	(.303)	(.411)
	Follow-Up	.336	.412	.160	.092
		(.385)	(.303)	(.200)	(.111)
Placebo	Initial	.117	.167	.292	.425
		(.098)	(.179)	(.301)	(.421)
	Follow-Up	.258	.242	.292	.208
		(.239)	(.286)	(.273)	(.202)

Table 5. Summary of Results for Logit Models Fitted to Table 1 (a.s.e. Values in Parentheses).

Model	Effect	Cumulative Logit	Adjacent Categories Logit
No Interation	Treatment	.37(.20)	.20(.11)
	Condition	-1.29(.13)	-.67(.07)
	Residual χ^2	14.5,df=7	15.4,df=7
Interaction	Treatment	.69(.23)	.39(.138)
	Condition	-1.05(.16)	-.55(.09)
	Treat.×Cond.	-.65(.25)	-.36(.14)
	Residual χ^2	7.4,df=6	8.3,df=6

Fitted models can be used to generate fitted marginal logits, and hence fitted marginal proportions. To illustrate, Table 4 also contains the fitted marginal proportions for the cumulative logit model (8.2). The only poor fit occurs for response 20-30 for the active treatment under the follow-up observation, in which case the observed proportion is considerably larger than the fitted one.

Let M_{gi} denote the mean response for the marginal distribution for condition g and treatment i, g = 1,2, i = 1,2. Model (6.1) corresponds to

$$M_{gi} = \alpha + \mu_g + \beta_i \qquad (8.3)$$

We fitted this model using WLS, with response scores $\{a_1 = 10, a_2 = 25, a_3 = 45, a_4 = 75\}$ for time to falling asleep. For the constraints $\mu_2 = \beta_2 = 0$, the estimates are $\hat{\alpha} = 35.3$, $\hat{\mu}_1 = 17.2$ (a.s.e. = 1.5), $\hat{\beta}_1 = -5.7$ (a.s.e. = 2.4). There are four response means and three parameters, so the residual chi-squared of 9.3 is based on df = 1. There is strong evidence of interaction on this scale, as well.

Adding an interaction term to (8.3) gives a saturated model, whereby the fitted marginal means equal the observed ones. From this model, we find the initial means were 50.0 for the active group and 50.3 for the placebo, and the difference in response means between the initial observation and the follow-up was 22.2 for the active group and 13.0 for the placebo. The difference between these differences of means equals 9.2, with a.s.e. = 3.0, indicating that the change was significantly greater for the active group.

9. FITTING ORDINAL LOGIT MODELS USING SAS

All results quoted in the previous section were obtained using CATMOD. Table 6 shows code for fitting the cumulative logit model (8.2). The RESPONSE statement forms the three cumulative logits for each of the two conditions -- the first three for the initial observation margin and the final three for the follow-up margin. Since there are two treatments, the POPULATION GROUP statement results in the calculation of twelve cumulative logits, six for each treatment. The MODEL statement contains the design matrix. The first three columns refer to the cutpoint parameters, the fourth column refers to treatment (the first six rows refer to the active group, and the last six to the placebo), the fifth column refers to condition (rows 1-3 and 7-9 refer to initial, rows 4-6 and 10-12 to follow-up), and the sixth column refers to the interaction.

Table 7 shows code for fitting the corresponding adjacent-categories logit model to Table 1. The design matrix is the same as for the cumulative logit model, and the RESPONSE statement forms the six adjacent-categories logits for each treatment. Table 7 also contains code for the mean response model. The RESPONSE statement forms the means for the two condition margins, using the chosen scores. For the design matrix in the MODEL statement, the first column refers to the "y-intercept," the second column is the treatment effect, the third column is the condition effect, and the fourth column is the interaction.

10. COMPARING TREATMENT EFFECTS, CONTROLLING FOR INITIAL RESPONSE

Though this article has focused on model-building for the marginal distributions, modeling the interior of the table can also be informative, particularly when there are only d=2 responses. For data such as in Table 1, for instance, we might want to model the follow-up response in terms of effects of explanatory variables, controlling for a baseline (initial) observation. Let L_{ijk} denote the cumulative logit when the cutpoint for follow-up response is at category k, for group i with baseline observation j, and let $\{x_j\}$ be fixed scores for the baseline levels. For the model

$$L_{ijk} = \alpha_k + \beta_i + \beta x_j \qquad (10.1)$$

we use the $\{\beta_i\}$ to compare the groups in terms of distribution of follow-up, controlling for baseline observation. This model is an analog of an analysis of covariance model, in which the response and covariate are ordinal rather than quantitative.

Applying model (10.1) to Table 1 with scores {10, 25, 45, 75} for time to falling asleep, the constraint $\beta_2 = 0$, and 0.5 added to the two empty cells (so WLS estimates exist), we obtain $\hat{\beta}_1 = 0.81$ (a.s.e. = 0.25) and $\hat{\beta} = -0.037$ (a.s.e. = 0.006), with a residual chi-squared of 25.7 based on df = 19. Given initial observation, the odds that follow-up time to falling asleep is below any fixed level is estimated to be exp(0.81) = 2.25 times higher for the active group than for the placebo.

Inspection of Table 1 reveals that for the first two baseline levels, the two treatments have similar distributions of time to fall asleep at the follow-up, whereas the active treatment is relatively more successful at the higher baseline levels. The model with interaction of baseline observation and treatment fits slightly better, with residual chi-squared of 22.5, based on df = 18; the estimated interaction effect is 0.019 (a.s.e. = 0.011), whereas $\hat{\beta}_1 = -0.161$ and $\hat{\beta} = -0.046$. Across the four baseline levels, the treatment effect varies between -0.161 + 0.019(10) = 0.03 and -0.161 + 0.019(75) = 1.27. It is uniformly positive (i.e., time to falling asleep at follow-up is estimated to be smaller for the active group), but the effect increases from negligible to strong as baseline level increases.

11. DISCUSSION

Complications that often occur for repeated categorical measurement data include sparseness of cell counts, missing data, time-dependent covariates, and a sampling design more complex than independent multinomial. For WLS analyses, it is natural to handle the missing data problem as illustrated by Stanish et al. (1978) or Woolson and Clarke (1984). For cumulative logit models, Landis et al. (1987) showed how to incorporate sampling weights and design effects into test statistics, by using Taylor-series approximations to obtain weighted proportions and their corresponding covariance matrix. Repeated measurement of covariates (i.e., the covariates $\underset{\sim}{x} = \underset{\sim}{x}_g$ are time-dependent) can be handled without difficulty with the Stram et al. (1988) approach of fitting a model separately under each condition.

The advantages and disadvantages of the different types of models described in this article are similar to those for the corresponding models for a single response. See Agresti (1984, Chap. 11) for a discussion of these. Of the logit models, the cumulative logit has an important advantage of a certain invariance to response category choice. If a cumulative logit model holds for a particular set of response categories, it will also hold when some of the categories are combined, with the same value for the covariate effect parameter. This is not true for the adjacent-category logit models, for which the corresponding loglinear model reveals that it assumes an equal-interval scoring of response categories. When there is an arbitrary rather than fixed choice of response categories, the interpretation of the parameters may also be more natural for the cumulative logit models. The mean response model has the advantage of simple interpretation. However, it requires the assignment of response scores, and its structural form can be problematic.

Table 6. SAS CATMOD Code for Fitting Cumulative Logit Model (8.2) to Table 1.

```
INPUT GROUP SLEEP COUNT @@ ;
CARDS;
1 1 7     1 2 4     1 3 1     1 4 0.000001
1 5 11    1 6 5     1 7 2     1 8 2
1 9 13    1 10 23   1 11 3    1 12 1
1 13 9    1 14 17   1 15 13   1 16 8
2 1 7     2 2 4     2 3 2     2 4 1
2 5 14    2 6 5     2 7 1     2 8 0.000001
2 9 6     2 10 9    2 11 18   2 12 2
2 13 4    2 14 11   2 15 14   2 16 22
;

PROC CATMOD ORDER=DATA;
WEIGHT COUNT;
POPULATION GROUP;
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 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
         1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0,
         0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1,
         1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0,
         0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1,
         1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0,
         0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1,
         1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0,
         0 1 1 1 0 1 1 1 0 1 1 1 0 1 1 1,
         1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0,
         0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1,
         1 1 1 0 1 1 1 0 1 1 1 0 1 1 1 0,
         0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1;
MODEL SLEEP = (1 0 0 1 1 1,
               0 1 0 1 1 1,
               0 0 1 1 1 1,
               1 0 0 1 0 0,
               0 1 0 1 0 0,
               0 0 1 1 0 0,
               1 0 0 0 1 0,
               0 1 0 0 1 0,
               0 0 1 0 1 0,

               1 0 0 0 0 0,
               0 1 0 0 0 0,
               0 0 1 0 0 0);
```

Table 7. SAS CATMOD Code for Fitting Adjacent Categories Logit Model and Mean Response Model to Table 1.

```
PROC CATMOD ORDER=DATA;
WEIGHT COUNT;
POPULATION GROUP;

RESPONSE 1 -1  0  0  0  0  0  0,
         0  1 -1  0  0  0  0  0,
         0  0  1 -1  0  0  0  0,
         0  0  0  0  1 -1  0  0,
         0  0  0  0  0  1 -1  0,
         0  0  0  0  0  0  1 -1  LOG
         1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0,
         0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0,
         0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0,
         0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1,
         1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0,
         0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0,
         0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0,
         0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1;
MODEL SLEEP = (1 0 0 1 1 1,
               0 1 0 1 1 1,
               0 0 1 1 1 1,
               1 0 0 1 0 0,
               0 1 0 1 0 0,
               0 0 1 1 0 0,
               1 0 0 0 1 0,
               0 1 0 0 1 0,

               0 0 1 0 1 0,
               1 0 0 0 0 0,
               0 1 0 0 0 0,
               0 0 1 0 0 0);

PROC CATMOD ORDER=DATA;
WEIGHT COUNT;
POPULATION GROUP;
RESPONSE 10 10 10 10 25 25 25 25 45 45 45 45 75 75 75 75,
         10 25 45 75 10 25 45 75 10 25 45 75 10 25 45 75;
MODEL SLEEP = (1 1 1 1,
               1 1 0 0,
               1 0 1 0,
               1 0 0 0);
```

REFERENCES

Agresti, A. 1984. Analysis of Ordinal Categorical Data. New York: Wiley.

Aitchison, J. and S.D. Silvey. 1958. "Maximum-Likelihood Estimation of Parameters Subject to Restraints." Ann. Math. Statist. 29: 813-828.

Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. 1975. Discrete Multivariate Analysis. Cambridge: MIT Press.

Fienberg, S. 1980. The Analysis of Cross-Classified Categorical Data. 2nd ed. Cambridge, MA: MIT Press.

Francom, S.F., C. Chuang, and J.R. Landis. 1987. "A Log-Linear Model to Analyze Change for Ordinal Data with Multiple Treatments." Unpublished manuscript.

Goodman, L.A. 1979. "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." J. Amer. Statist. Assoc. 74: 537-552.

Goodman, L.A. 1983. "The Analysis of Dependence in Cross-Classifications Having Ordered Categories, Using Log-Linear Models for Frequencies and Log-Linear Models for Odds." Biometrics 39: 149-160.

Haber, M. 1985a. "Maximum Likelihood Methods for Linear and Log-Linear Models in Categorical Data." Computational Statist. & Data Anal. 3: 1-10.

Haber, M. 1985b. "Log-Linear Models for Correlated Marginal Totals of a Contingency Table." Comm. Statist. - Theor. Meth. 14: 2845-2856.

Koch, G.G., J.R. Landis, J.L. Freeman, D.H. Freeman, Jr., R.G. Lehnen. 1977. "A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data." Biometrics 33: 133-158.

Landis, J.R., J.M. Lepkowski, C.S. Davis, and M.E. Miller. 1987. "Cumulative Logit Models for Weighted Data from Complex Sample Surveys." Unpublished Manuscript.

Liang, K. and S.L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." Biometrika 73: 13-22.

McCullagh, P. 1977. "A Logistic Model for Paired Comparisons with Ordered Categorical Data." Biometrika 64: 449-453.

McCullagh, P. 1980. "Regression Models for Ordinal Data" (with discussion). J. Roy. Statist. Soc. B 42: 109-142.

Stanish, W.M., D.B. Gillings, and G.G. Koch. 1978. "An Application of Multivariate Ratio Methods for the Analysis of a Longitudinal Clinical Trial with Missing Data." Biometrics 34: 305-317.

Stram, D.O., L.J. Wei, and J.H. Ware. 1988. "Analysis of Repeated Ordered Categorical Outcomes with Possibly Missing Observations and Time-Dependent Covariates." to appear in J. Amer. Statist. Assoc.

Williams, O.D., and J.E. Grizzle. 1972. "Analysis of Contingency Tables Having Ordered Response Categories." J. Amer. Statist. Assoc. 67: 55-63.

Woolson, R.F., and W.R. Clarke. 1984. "Analysis of Categorical Incomplete Longitudinal Data." J. Roy. Statist. Soc. A, 147: 87-99.

Alan Agresti
Department of Statistics
University of Florida
Gainesville, FL 32611