15 *Ordinal longitudinal data analysis*

Jeroen K. Vermunt and Jacques A. Hagenaars
Tilburg University

Introduction

Growth data and longitudinal data in general are often of an ordinal nature. For example, developmental stages may be classified into ordinal categories and behavioral variables repeatedly measured by discrete ordinal scales. Consider the data set presented in Table 15.1. This table contains information on marijuana use taken from five annual waves (1976–80) of the National Youth Survey (Elliot *et al.*, 1989; Lang *et al.*, 1999). The 237 respondents were 13 years old in 1976. The variable of interest is a trichotomous ordinal variable 'Marijuana use in the past year' measured during five consecutive years. There is also information on the gender of the respondents.

Ordinal data like this is often analysed as if it were continuous interval level data, that is, by means of methods that imply linear relationships and normally distributed errors. However, the data in Table 15.1 is essentially categorical and measured at ordinal, and not at interval level. Consequently, a much better way to deal with such an ordinal response variable is to treat it as a categorical variable coming from a multinomial distribution; the ordinal nature of the categories is then taken into account by imposing particular constraints on the odds of responding, i.e. of choosing one category rather than another. As will be further explained below, an ordinal analysis can be based on cumulative, adjacent-categories, or continuation-ratio odds (Agresti, 2002). The constraints are in the form of equality or inequality constraints on one of these types of odds.

In this chapter, we will discuss the three main approaches to the analysis of longitudinal data: transition models, random-effects or growth models and marginal models (Diggle *et al.*, 1994; Fahrmeir and Tutz, 1994). Roughly speaking, transition models like Markov chain models concentrate on overall gross changes or transitions between consecutive time points, marginal models investigate net changes at the aggregated level and random-effects or growth models

Methods in Human Growth Research, eds. R. C. Hauspie, N. Cameron and L. Molinari.
Published by Cambridge University Press.

Table 15.1 *Data on marijuana use in the past year and gender, taken from five yearly waves of the National Youth Survey*

Gender[a]	1976[b]	1977	1978	1979	1980	Frequency	Gender	1976	1977	1978	1979	1980	Frequency
0	1	1	1	1	1	63	1	1	1	1	3	1	1
0	1	1	1	1	2	10	1	1	1	1	3	3	1
0	1	1	1	1	3	3	1	1	1	2	1	3	1
0	1	1	1	2	1	4	1	1	1	2	2	1	2
0	1	1	1	2	2	2	1	1	1	2	2	2	2
0	1	1	1	3	1	1	1	1	1	2	2	3	1
0	1	1	1	3	2	1	1	1	1	2	3	3	5
0	1	1	1	3	3	3	1	1	1	3	1	2	1
0	1	1	2	1	1	2	1	1	1	3	2	2	1
0	1	1	2	1	2	2	1	1	1	3	3	3	3
0	1	1	2	2	1	3	1	1	2	1	1	2	1
0	1	1	2	2	2	7	1	1	2	1	2	1	1
0	1	1	2	2	3	1	1	1	2	2	1	1	2
0	1	1	2	3	3	1	1	1	2	2	2	1	1
0	1	2	1	1	1	1	1	1	2	2	1	3	1
0	1	2	1	1	2	2	1	1	2	2	3	3	1
0	1	2	2	1	2	1	1	1	2	3	2	2	1
0	1	2	2	2	1	1	1	1	2	3	2	3	1
0	1	2	2	3	3	2	1	1	2	3	3	2	1
0	1	2	3	1	2	1	1	1	2	3	3	3	4
0	1	2	3	3	2	1	1	1	3	1	3	3	1
0	1	2	3	3	3	1	1	1	3	2	2	2	1
0	1	3	3	2	2	1	1	1	3	3	3	3	2
0	2	1	1	3	3	1	1	1	3	3	2	2	1
0	2	1	2	2	2	1	1	2	1	1	1	1	3
0	2	1	3	3	3	1	1	2	2	2	2	2	1
0	2	3	3	3	3	1	1	2	2	3	3	3	1
0	2	3	3	3	2	1	1	2	3	2	1	1	1
0	3	3	3	2	3	1	1	2	3	2	3	3	1
1	1	1	1	1	1	48	1	2	3	3	3	3	2
1	1	1	1	1	2	8	1	3	1	1	1	1	1
1	1	1	1	1	3	4	1	3	2	3	3	3	1
1	1	1	1	2	1	2	1	3	3	3	3	1	1
1	1	1	1	2	2	4	1	3	3	3	3	3	1
1	1	1	1	2	3	1							

[a] 0, female; 1, male.

[b] 1, never; 2, not more than once a month; 3, more than once a month.

study developments at the individual level. Variants of each of these have been developed for ordinal categorical variables (Agresti, 2002).

There are various complicating issues that have to be dealt with when analysing longitudinal data in general and ordinal longitudinal data in particular. The first is the issue of misclassification or measurement error (Hagenaars, 1990, 2002; Bassi *et al.*, 2000). Measurement of developmental stages is almost never perfect. For example, even in situations in which this is theoretically

impossible, backward transitions will be observed and usually different indicators of the same phenomenon will provide partly inconsistent information. An important way to cope with measurement error is to introduce latent variables into the analysis. When dealing with categorical ordinal data that are assumed to measure ordinal true states, it is most natural to use categorical latent variables and latent class methods to this purpose.

Another important complicating issue, especially potentially harmful in transitional and growth models is the problem of unobserved heterogeneity (Vermunt, 1997, 2002). Random-effects approaches may be used to solve this problem. With ordinal variables, because of their non-linear nature, these models are somewhat more complicated and their estimation is more time consuming than with continuous outcome variables. A possible way out is to use a non-parametric random-effects approach based on latent class or finite mixture modelling.

The third issue is the presence of partially observed data. Methods for longitudinal data analysis are less useful if they can only deal with complete records. Fortunately, transition and growth models for ordinal variables can easily be adapted to deal with missing data (Hagenaars, 1990; Vermunt, 1997).

These complicating issues will be further discussed at the end of this chapter. First, logit models for ordinal response variables will be introduced (pp. 00–00). They form the basic building blocks for models for categorical longitudinal data. The main ways of analysing longitudinal data, viz. marginal, transitional and random-effects models for ordinal categorical variables, are also discussed (pp. 00–00).

Ordinal logit models

Given a dichotomous or polytomous ordinal outcome variable, the most popular model is the logit model. The logit model is a regression model in which the odds of choosing a particular category (or categories) of the response variable rather than another category (other categories) are assumed to depend on the values of certain independent variables (Agresti, 2002). Note that an odds is simply a ratio of two probabilities. The term logit comes from log odds, which refers to the fact that a logit model is a linear model for log odds.

The measurement level of the independent variables is also often ordinal. This ordinal character of the variables involved frequently leads to the assumption that the log odds are a linear function of predictors, which implies certain *equality* constraints on the odds ratios. As shown below, another important way of dealing with the ordinal nature of the variables of interest is to specify

Table 15.2 *Cross-tabulation of year and marijuana use in the past year based on data in Table 15.1*

	Time (X)				
Marijuana use (Y)	1. 1976	2. 1977	3. 1978	4. 1979	5. 1980
1. Never	218	195	167	156	138
2. No more than once a month	14	27	41	41	52
3. More than once a month	5	15	29	40	47

inequality restrictions on the odds ratios instead of *equality* restrictions. Various types of ordinal logit models can be defined depending on the type of odds that are being used (as mentioned above). It is also possible to use another link function than the logit link (using the generalized linear modeling jargon), and formulate probit, log-log, or complementary log-log models; Agresti provides an excellent overview of these possibilities (Agresti, 2002).[1] Here, we will only deal with logit models.

In order to make the discussion more concrete, assume that we have a two-way cross tabulation of X – time (or age) and Y – marijuana use, as shown in Table 15.2 (derived from Table 15.1). Variable X with category index i has $I = 5$ levels or categories and variable Y with category index j has $J = 3$ levels. It is assumed that X (time/age) serves as independent or predictor variable, that Y (marijuana use) is the dependent or response variable, and that we are interested in the conditional distribution of Y given X. The probability of giving response j at time i is denoted by $P(Y = j|X = i)$.

The substantive research question of interest is whether there is an increase of marijuana use with age; that is, whether the proportion of respondents in the highest categories increases over time. Note that this hypothesis does not imply any specific parametric form for the relationship between X and Y: we only assume that if X increases, Y will increase as well.

The most common way of modelling relationships between ordinal categorical variables is by means of a (linear) logit model that imposes equality constraints on certain odds ratios. Four types of odds can be used for this purpose (Agresti, 2002): cumulative odds denoted here as $O(\text{cum})_{i,j}$, adjacent-category (or local) odds $O(\text{adj})_{i,j}$, or one of two types of continuation-ratio odds

[1] The term link refers to the transformation of the dependent variable yielding the linear model. In a logit model, the response probabilities are transformed to log odds. It is, however, possible to work with other types of transformations. A probit link, for example, involves transforming the response probabilities to z values using the cumulative normal distribution.

$O(\text{conI})_{i,j}$, or $O(\text{conII})_{i,j}$. These are defined as

$$
\begin{aligned}
O(\text{cum})_{i,j} &= P(Y \leq j|X = i)/P(Y \geq j+1|X = i) \\
O(\text{adj})_{i,j} &= P(Y = j|X = i)/P(Y = j+1|X = i) \\
O(\text{conI})_{i,j} &= P(Y = j|X = i)/P(Y \geq j+1|X = i) \\
O(\text{conII})_{i,j} &= P(Y \leq j|X = i)/P(Y = j+1|X = i)
\end{aligned}
$$

for $1 \leq j \leq J-1$ and $1 \leq i \leq I$. Below, the symbol $O_{i,j}$ will be used as a generic symbol referring to any of these odds. As follows from these formal definitions, for an ordinal variable with four categories, the 'first' cumulative odds will be the odds of choosing category 1 rather than one of the other categories 2 or 3 or 4: $(1)/(2+3+4)$ and the other cumulative odds denoted in a similar way are $(1+2)/(3+4)$ and $(1+2+3)/(4)$. Using the same shorthand notation, the adjacent odds are $(1)/(2)$, $(2)/(3)$, and $(3)/(4)$ and the first type of continuation odds are $(1)/(2+3+4)$, $(2)/(3+4)$, and $(3)/(4)$.

In practical research situations, one has to make a choice between these four types of odds, that is, one has to specify a model for the type of odds that fits best to the process assumed to underlie the individual responses. The cumulative odds are the most natural choice if the discrete ordinal outcome variable is considered as resulting from a discretization of an underlying continuous variable. The adjacent category odds specification fits best if one is interested in each of the individual categories; that is, if one perceives the response variable as truly categorical. The continuation-ratio odds correspond to a sequential decision-making process in which alternatives are evaluated from low to high (type I) or from high to low (type II).

If X and Y are both treated as nominal level variables, no restrictions will be imposed on the odds. The logit model for this *nominal–nominal* case can be expressed as

$$\log O_{i,j} = \alpha_j - \beta_{ij}$$

which is just a decomposition of the log odds. Here, α_j is an intercept parameter and β_{ij} is a slope parameter.[2] As is the case in most models for categorical dependent variables, the intercept is category specific (for the categories of the dependent variable). Typical for the nominal–nominal case is that the slope depends on both the category of X and the category of Y.

[2] Note that the index j goes from 1 to $J-1$ and the index i from 1 to I. This means that no further restrictions need to be imposed on the $J-1$ α_j parameters. On the other hand, there are and $(J-1) \times (I)$ free β_{ij} parameters, but we can identify only $(J-1) \times (I-1)$ of them. For identification, one can, for instance, assume that $\beta_{1j} = 0$ for each j, which amounts to treating the first category of X as reference category.

The most restricted specification is obtained if both the dependent and the independent variable are treated as ordered. This yields the *ordinal–ordinal* model

$$\log O_{i,j} = \alpha_j - \beta x_i$$

where x_i denotes the fixed score assigned to category i of X, and α_j and β are the intercept and the slope of the logit model. In most cases, x_i will be equal-interval scores (for example, 1, 2, 3, etc., or 1976, 1977, 1978, etc.), but it is also possible to use other scoring schemes for the X variable. As can be seen, the slope is assumed to be independent of the categories of X and Y.

Because of the restrictions involved, a better, although not common, name for the ordinal–ordinal would be the *interval–interval* model. For the adjacent-categories odds, this model is also known as the linear-by-linear association model (Goodman, 1979; Clogg and Shihadeh, 1994; Hagenaars, 2002). In fact, we do not only specify scores for the categories of X, but implicitly also assume that the categories of Y are equally spaced with a mutual distance of 1; for example, 1, 2 and 3, or 0, 1 and 2.

The other two, intermediate, cases are the *nominal–ordinal* specification (with Y nominal and X ordinal) and the *ordinal–nominal* case (with Y ordinal and X nominal). These are defined as follows:

$$\log O_{i,j} = \alpha_j - \beta_j x_i$$
$$\log O_{i,j} = \alpha_j - \beta_i$$

As can be seen, in the nominal–ordinal case, the slope is category specific for Y, but does not depend on X. In the ordinal–nominal case, the slope does not depend on Y.

These models are also known, for the adjacent categories odds, as row- or column-association models (Goodman, 1979; Clogg and Shihadeh, 1994; Hagenaars, 2002). In our case, the nominal–ordinal model is a row-association model because the nominal variable (marijuana use) serves as the row variable in Table 15.1. In a row-association model, the scores of the categories of the column variable are fixed and the scores of the categories of the row variable are unknown parameters to be estimated. In our parameterization, β_j can be interpreted as the distance between the unknown scores of categories $j+1$ and j. For equivalent reasons, the ordinal–nominal model is a column-association model, where β_i represents an unknown column score. The row scores are treated as fixed and assumed to have a mutual distance of 1.

Except for the nominal–nominal specification, each of these specifications implies certain equality constraints on the odd ratios $O_{i,j}/O_{i+1,j}$. When we use equal-interval x_i, the ordinal–ordinal model implies that the log-odds increase

or decrease linearly with X and that, in other words, the log odds ratios between adjacent levels of X are assumed to be constant, to be equal to β:

$$\log(O_{i,j}/O_{i+1,j}) = \log O_{i,j} - \log O_{i+1,j} = \beta$$

for $1 \leq j \leq J-1$ and $1 \leq i \leq I-1$. The fact that these differences between log odds do not depend on the values of X and Y can also be expressed by the following two sets of equality constraints:

$$(\log O_{i,j} - \log O_{i+1,j}) - (\log O_{i,j'} - \log O_{i+1,j'}) = 0, \tag{15.1}$$

$$(\log O_{i,j} - \log O_{i+1,j}) - (\log O_{i',j} - \log O_{i'+1,j}) = 0, \tag{15.2}$$

where $i' \neq i$ and $j' \neq j$.

In the nominal–nominal specification of the logit model, the log odds ratio between adjacent categories of X equals $\beta_{i+1,j} - \beta_{i,j}$ and none of the two equality constraints (15.1) and (15.2) are valid. In the nominal–ordinal case, the log odds ratio equals β_j and equality constraints (15.2) apply: the odds ratios vary with Y but do not depend on X. In the ordinal–nominal case, the log odds ratio equals $\beta_{i+1} - \beta_i$ and equality constraints (15.1) are applicable: the odds ratios vary with X and do not depend on Y.

What can be observed is that the ordinal nature of the predictor variable is dealt with by assuming that the odds ratio is the same for each pair of adjacent categories. This amounts to assuming that the distances between all adjacent categories are equal, which is a much stronger assumption than ordinal, and essentially the interval-level assumption. The constraints related to the outcome variable imply that the effect of the predictor variable is assumed to be the same (constant) for each of the category-specific odds, which is also a more restrictive assumption than ordinal. An advantage of imposing the ordinal logit constraints is, however, that they force the solution to be in agreement with an ordinal relationship. Moreover, they provide a parsimonious and easy-to-interpret representation of the data: the effect of an ordinal predictor variable on an ordinal outcome variable is described by a (very) small number of parameters.

However, in many situations, the ordinal–ordinal specification is too restrictive even if the relationship between the variables of interest is truly ordinal. In such cases, one may consider staying closer to the definition of ordinal measurement. In terms of odds, the purest definition of a positive, (weakly) monotonically increasing relationship between two ordinal variables involves the following set of inequality constraints:

$$\log O_{i,j} - \log O_{i+1,j} \geq 0, \tag{15.3}$$

for $1 \leq j \leq J-1$ and $1 \leq i \leq I-1$ (Vermunt, 1999). As can be seen, we are assuming that all log odds ratios are at least zero, or, equivalently that all odds

ratios are larger than or equal to 1. Such a set of constraints is often referred to as simple stochastic ordering, likelihood ratio ordering, or uniform stochastic ordering for cumulative, adjacent category, and continuation odds, respectively (Dardanoni and Forcina, 1998).

The inequality constraints (15.3) are equivalent to the following constraints on the slope parameters of the nominal–nominal model

$$\beta_{i+1,j} - \beta_{i,j} \geq 0$$

This shows that this 'non-parametric' ordinal approach can be seen as a nominal–nominal approach with an additional set of constraints. As long as these constraints are not violated, the order-restricted solution will be the same as the nominal–nominal solution. Similar types order constraints can be defined for the nominal–ordinal and ordinal–nominal models to guarantee that the solution is ordered in the sense of a monotonic relationship. The order constraints corresponding with a positive association are $\beta_j \geq 0$ in the nominal–ordinal case and $\beta_{i+1} - \beta_i \geq 0$ in the ordinal–nominal case.

Another type of model for odds is a class of logit models with bi-linear terms. When working with adjacent category odds, these are called row–column association models (Goodman, 1979; Clogg and Shihadeh, 1994). The model of interest has the form

$$\log O_{i,j} = \alpha_j - \beta_j \gamma_i$$

where the γ_i are unknown parameters to be estimated. These can be seen as free scores for the categories of X. This model is, in fact, a restricted version of the nominal–nominal model. It can also be seen as a less restricted variant of the nominal–ordinal model (free instead of fixed scores for X) or of the ordinal–nominal model (non-constant slope). If $\beta_j \geq 0$ and $\gamma_{i+1} - \gamma_i \geq 0$ for all i and j, the solution is in agreement with an ordinal relationship.

In order to illustrate the equality and inequality constraints implied under the various specifications, we applied the models to the data in Table 15.2 with an adjacent-category odds formulation. Table 15.3 reports the estimated odds ratios obtained with the estimated models. Note that an odds ratio larger than one is in agreement with the postulated positive relationship between time and marijuana use. As can be seen from the outcomes of the nominal–nominal model, the data contains only one violation of an ordinal relationship. The test results indicate that this can be attributed to sampling fluctuation. In the estimation and testing, we treated the observations at different time points as independent samples, which is not correct. Results should therefore be treated with some caution. The next section discusses methods that take the dependence between observations into account. Moreover, the testing of models with inequality constraints is not straightforward. Because the number

Table 15.3 *Adjacent-category odds ratios under various models for the data in Table 15.3*

			Time/Age (X)			
Model[a]	L-sq (df)[b]	Marijuana use (Y)	1/2	2/3	3/4	4/5
Nominal–nominal	0.00 (0)	1/2	2.16	1.77	1.07	1.43
		2/3	1.56	1.27	1.38	0.93
Ordinal–ordinal	11.37 (7)	1/2	1.36	1.36	1.36	1.36
		2/3	1.36	1.36	1.36	1.36
Nominal–ordinal	8.74 (6)	1/2	1.48	1.48	1.48	1.48
		2/3	1.19	1.19	1.19	1.19
Ordinal–nominal	1.88 (4)	1/2	2.00	1.57	1.19	1.19
		2/3	2.00	1.57	1.19	1.19
Row–column	0.72 (3)	1/2	2.05	1.65	1.27	1.27
		2/3	1.75	1.41	1.08	1.09
Order-restricted	0.06 (1)	1/2	2.16	1.77	1.09	1.38
		2/3	1.56	1.27	1.32	1.00

[a] The models were estimated with the LEM program (Vermunt, 1997).
[b] L-sq is the likelihood-ratio statistic, which is defined as twice the difference between the log-likelihood of the data and the log-likelihood of the model concerned. The number of degrees of freedom is denoted by df. In the order-restricted model, df refers the number of odds ratios that are equated to 1.

of degrees of freedom is a random variable, the asymptotic distribution of the test statistics is a mixture of χ^2 distributions, which is usually denoted as $\bar{\chi}^2$ distribution (Vermunt, 1999; Galindo-Garre *et al.*, 2002). Especially the order-restricted model using only inequality restrictions fits almost perfectly, but also the ordinal–ordinal model, using equality restrictions, fits well and provides an excellent very parsimonious description of the data. However, comparison of the ordinal–ordinal with the ordinal–nominal model (L-sq $= 9.5$ with df $= 3$) indicates that the ordinal constraint is somewhat too restrictive for the column variable time. (Nested models can be compared by a likelihood-ratio test. For this purpose we subtract their L-sq and df values, which yields a new asymptotic χ^2 test.)

Three approaches to longitudinal data

There are three main approaches to the analysis of longitudinal data: transitional models, random-effects models and marginal models (Diggle *et al.*, 1994; Fahrmeir and Tutz, 1994). Transitional models such as Markov-type models

concentrate on changes between consecutive time points. Marginal models can be used to investigate changes in univariate distributions, and random-effects or growth models study development of individuals over time. (Here, we concentrate on situations in which there is a single response variable. With multiple response variables, one may wish to study change in multivariate distributions (e.g. Croon *et al.*, 2000)). These three approaches do not only differ in the questions they address, but also in the way they deal with the dependencies between the observations. Because of their structure, transitional models take the bivariate dependencies between observations at consecutive occasions into account. Growth models capture the dependence by introducing one or more latent variables. In marginal models, the dependency is often not modelled, but dealt with as found in the data and in general is taken into account in a more ad hoc way in the estimation procedure. In this section, we present marginal, transition, and random-effect models for ordinal outcome variables.

Before describing the ordinal data variants of the three approaches, we first extend our notation to deal with the longitudinal character of the data. The total number of time points is denoted by T, and a particular time point by t, where $1 \leq t \leq T$. Moreover, we denote the response variable at time point t by Y_t, a particular value of Y_t by j_t, and the number of levels of Y_t by J. Notice that number of levels of the response variable is assumed to be the same for each time point. Predictor k is denoted by X_{kt}, where the index t refers to the fact that a predictor may change its value over time. When referring to a vector of random variables, we use boldface characters. For example, the conditional distribution of the time-specific responses given a particular covariate pattern is denoted by $P(\mathbf{Y} = \mathbf{j}|\mathbf{x})$.

The models of interest will be illustrated with the data set reported in Table 15.1. This means that we have a trichotomous response variable measured at five occasions; that is, $J = 3$ and $T = 5$. There is a single time-constant predictor gender, whose value is denoted by x, where $x = 1$ for males and $x = 0$ for females.

Marginal models

The analysis presented in the previous section is an example of a marginal analysis. We studied the trend in the age- or time-specific marginal distributions of the response variable of interest. However, when estimating the model parameters, we assumed that observations at different time points are independent, which is, of course, unrealistic with longitudinal data. The purpose of the marginal modelling framework is to test hypotheses like the one discussed in the previous section, while taking the dependencies between the observations into

account. Parameters can be estimated by maximum likelihood (ML), but also by other methods, such as generalized estimating equations (GEE) or weighted least squares (WLS).

The ML approach takes the full multidimensional distribution of the response variables, $P(\mathbf{Y} = \mathbf{j}|\mathbf{x})$, as the starting point. In addition, to the marginal model of interest, a model has to be specified for the joint distribution. This is not necessarily a restricted model and often the saturated model is simply used. The ML estimates of the probabilities in the joint distribution should be in agreement with both the model for the joint distribution and the model for the marginal distribution. A disadvantage of the ML approach is that it is not practical with more than a few time points. Although theoretically inferior to ML, the GEE approach has the advantage that it can also be applied with larger numbers of time points.

Using the ordinal logit formulation introduced in the previous section, a marginal model for the data displayed in Table 15.1 could, for instance, be of the form

$$\log O_{jt} = \alpha_j - \beta_1 t - \beta_2 x$$

Here, α_j is the intercept for the log odds corresponding to category j, β_1 is the time effect and β_2 is the gender effect. (In the example $J = 3$, which means that there are two sets of odds since $1 \leq j \leq J - 1$). The fact that the time and gender effects do not depend on the category of the outcome variable shows that we are using an ordinal logit model specification for the outcome variable. Moreover, the log odds are assumed to change linearly with time or age, which amounts to using an 'ordinal' specification for the time effects. Since gender is a dichotomous variable, for this variable there is no difference between a nominal or ordinal specification.

In the previous section, we showed how to relax the 'ordinality' assumptions. For example, the assumption that the time trend is linear can be relaxed by replacing the term $\beta_1 t$ by β_{1t}; that is, by introducing a separate parameter for each time point. Moreover, inequality constraints can be used to transform such a nominal specification for the time effect into ordinal. The ordinal-logit assumption for the dependent variable can be relaxed by having a separate set of effect parameters for $j = 1$ and $j = 2$.

Another possible modification of the above model is the inclusion of an interaction effect between time and gender; that is,

$$\log O_{jt} = \alpha_j - \beta_1 t - \beta_2 x - \beta_3 t x$$

This model relaxes the assumption that the (linear) time trend is the same for males and females.

What should be clear from this example is that marginal models are very much similar to the logit models for ordinal response variables presented in the

previous section. The only fundamental difference appears in the estimation procedure in which the dependencies between the time-specific observations have to be taken into account. As was already mentioned, ML estimation involves estimating the cell probabilities in the joint distribution of the time-specific response variables given gender. These cell probabilities should be in agreement with the marginal model of interest and the model that is specific for the joint distribution. As shown by Lang and Agresti (1994) and Bergsma (1997), this estimation problem can be defined as a restricted ML estimation problem.

Transitional models

Typical for transitional models is that a regression model is specified for the conditional distribution of the response variable Y_t given the responses at previous time points (Y_{t-1}, Y_{t-2}, Y_{t-3}, etc.) and predictor values. The fact that Y_t is regressed on a person's state at previous occasions distinguishes transitional from marginal and growth models. Further, a transitional model implies a model for the joint distribution of the time-specific responses. The most popular transitional model is the first-order Markov model in which Y_t is assumed to depend on the state at $t-1$, but not on responses at earlier occasions. For our example, a first-order Markov model implies the following structure for the joint distribution $P(\mathbf{Y}=\mathbf{j}|x)$:

$$\begin{aligned} P(\mathbf{Y}=\mathbf{j}|x) = {} & P(Y_1=j_1|x)P(Y_2=j_2|Y_1=j_1,x)P(Y_3=j_3|Y_2=j_2,x) \\ & \times P(Y_4=j_4|Y_3=j_3,x)P(Y_5=j_5|Y_4=j_4,x) \end{aligned}$$

Further restrictions may be imposed on the initial and transition probabilities. For example, one might assume that the transition probabilities are time homogenous, yielding what is called a stationary first-order Markov model. The ordinal nature of the response variable can be exploited by restricting the model probabilities by means of an ordinal logit model. An example of a restricted ordinal logit model for the transition probabilities $P(Y_t=j_t|Y_{t-1}=j_{t-1},x)$ is

$$\log O_{jt} = \alpha_j - \beta_1 y_{t-1} - \beta_2 t - \beta_3 x$$

Except for the presence of Y_{t-1} as a predictor, this transitional logit model is similar to the marginal logit model presented above. Note that y_{t-1} denotes the fixed score corresponding to category of the response at time point $t-1$. As in a marginal model, ordinal specifications can be changed into a nominal specification, inequality constraints can be imposed, and interaction terms can be included.

ML estimation of transitional models is straightforward. One makes use of a log-likelihood based on a multinomial density with probabilities $P(\mathbf{Y}=\mathbf{j}|x)$. By

including states at previous occasions as predictors, the dependence between the time-specific observations is automatically taken into account. More precisely, observations are assumed to be independent given these previous states.

Random-effects growth models

The model structure of a growth model is similar to that of a marginal model. The probability of being in a certain state at occasion t is assumed to be a function of time and predictors. A difference is, however, that the dependence between observations is dealt with in another way. More specifically, the dependence between observations is attributed to systematic differences between individuals. This unobserved heterogeneity is captured by the introduction of random effects in the regression model. A random effect is a parameter that takes on a different value for each individual and that is assumed to come from a particular distribution. Random-effect terms are, in fact, latent variables, which means that a random-effects model is a latent variable model.

An ordinal logit model with a linear growth structure, a gender effect, and a random intercept has the form

$$\log O_{ijt} = \alpha_j + u_i - \beta_1 t - \beta_2 x$$

Here, u_i is the random intercept for individual i. The most common specification is to assume that u_i comes from a normal distribution with a mean equal to zero and variance equal to σ^2; that is, $u_i \sim N(0,\sigma^2)$. Introducing such a random effect amounts to specifying that the intercept is person specific, where person i's intercept equals $\alpha_j + u_i$. It should be noted that apart from the random effect, this logit model is the same as the marginal logit model presented above.

Not only the intercept can be specified to be person specific, but also the time or predictor effects can be assumed to vary across individuals. For example, a model with a random time effect is obtained by

$$\log O_{ijt} = \alpha_j + u_{1i} - \beta_1 t - u_{2i} t - \beta_2 x$$

In this case, the joint distribution of the two random effects u_{1i} and u_{2i} has to be specified. A common choice is bivariate normal, which means that besides the variances also the covariance between the two random effects has to be estimated.

ML estimation is based on the log-likelihood function derived from the multinomial density with probabilities $P(\mathbf{Y} = \mathbf{j}|x)$. Similarly to a transitional model, a random-effects model implies a particular model for the joint distribution

$P(\mathbf{Y} = \mathbf{j}|x)$; that is,

$$P(\mathbf{Y} = \mathbf{j}|x) = \int_{\mathrm{u}} P(Y_1 = j_1|x, \mathbf{u})P(Y_2 = j_2|x, \mathbf{u})P(Y_3 = j_3|x, \mathbf{u}) \times P(Y_4 = j_4|x, \mathbf{u})P(Y_5 = j_5|x, \mathbf{u})f(\mathbf{u})\,\mathrm{d}\mathbf{u}$$

As can be seen, occasion-specific responses are assumed to be independent given the random effects $\mathbf{u}$. In order to obtain $P(\mathbf{Y} = \mathbf{j}|x)$, we have to integrate out the unobserved random effects. However, contrary to the case of a linear model with normal errors, in our case this integral cannot be solved analytically. Two possible ways to solve the integral are numerical integration by Gauss–Hermite quadrature or integration by simulation methods. Both methods can become quite time consuming with more than a few random-effects terms.

An alternative to the above random-effects approach in which a parametric form is specified for the mixing distribution $f(\mathbf{u})$ is to use a non-parametric specification for the mixing distribution (Laird, 1978; Vermunt and Van Dijk, 2001; Agresti; 2002). The distribution of the random effects is then approximated by a small number of mass points (or latent classes), whose locations and weights are unknown parameters to be estimated. This approach, which is usually referred to as latent class regression or finite mixture regression, has several advantages over using a multivariate normal mixing distribution. One advantage is that it not necessary to make non-testable assumptions about the form of the distribution of the random effects. Another advantage is the much smaller computation burden resulting from the fact that $P(\mathbf{Y} = \mathbf{j}|x)$ can be obtained by summing over a small number of latent classes instead of a large number of quadrature points.

Let the index c refer to a latent class or mixture component and let C be the number of latent classes. A non-parametric specification of the ordinal logit model with a random intercept and a random time effect is

$$\log O_{jt} = \alpha_j + u_{1jc} - \beta_1 t - u_{2c} t - \beta_2 x$$

Rather that assuming that each individual has its own intercept and time effect, we now say that each individual belongs to one of C latent classes, each of which has its own set of logit parameters. A more common way to express this is to index the regression coefficients by c; that is,

$$\log O_{jt} = \alpha_{jc} - \beta_{1c} t - \beta_2 x$$

where $\alpha_{jc} = \alpha_j + u_{1jc}$ and $\beta_{1c} = \beta_1 + u_{2c}$. The implied model for the joint distribution is now

$$P(\mathbf{Y} = \mathbf{j}|x) = \sum_{\mathrm{c}} P(Y_1 = j_1|x, c)P(Y_2 = j_2|x, c)P(Y_3 = j_3|x, c) \times P(Y_4 = j_4|x, c)P(Y_5 = j_5|x, c)P(c)$$

which is much simpler than in the case of the parametric random effects. Note that $P(c)$ is the probability that an individual belongs to latent class c.

Both in the parametric and the non-parametric model it is possible to compute individual-level effects. The most popular are expected a posteriori (EAP) estimates.

Combining the three approaches

Above, we described the three approaches for dealing with longitudinal data as if those provided three mutually exclusive options. However, in some situations, to answer certain questions, one may wish to combine approaches. As explained above, an ML approach to marginal modelling involves the specification of a model for the joint distribution. This could be a restricted model, for example, a first-order Markov model, a random-effects model, or a combination of the two. Vermunt *et al.* (2001), for instance, proposed a combination of the three approaches, in which the random-effects part of the model had the form of a log-linear Rasch model.

Another interesting (and popular) combination is that between a transitional and a random-effects model. Each of these models makes a very specific assumption about the dependence structure of the repeated measures. In a first-order Markov model, for example, it is assumed that dependencies between observations can be fully described by means of first-order autocorrelation terms. In a random effects model, on the other hand, it is assumed that after controlling for the random effects (or latent class memberships), there is no autocorrelation. It is very important when using transitional models to take unobserved heterogeneity into account since failure to do so may result in a strong negatively biased time dependence (Vermunt, 1997). A possible model that combines the two approaches is a first-order Markov model with a non-parametric specification of the random effect:

$$\begin{aligned} P(\mathbf{Y}=\mathbf{j}|x) = \sum_c P(c)P(Y_1=j_1|x,c)P(Y_2=j_2|Y_1=j_1,x,c) \\ \times P(Y_3=j_3|Y_2=j_2,x,c)P(Y_4=j_4|Y_3=j_3,x,c) \\ \times P(Y_5=j_5|Y_4=j_4,x,c) \end{aligned}$$

This transitional model with unobserved heterogeneity is usually referred to as mixed Markov model (Langeheine and Van de Pol, 1994; Vermunt, 1997).

And of course, also in these combined models, ordinal logit models can be specified to further restrict the model probabilities. A simple example is

$$\log O_{jt} = \alpha_{jc} - \beta_1 y_{t-1} - \beta_2 t - \beta_3 x$$

in which the intercept is assumed to vary across latent classes.

Table 15.4 *Test results for adjacent-category ordinal logit models estimated with the data in Table 15.1*

Longitudinal type	Specification for Y and X	L-sq[a]	Degrees of freedom
Marginal	M1. nominal–nominal	5.33	8
	M2. nominal–ordinal	19.46	14
	M3. nominal Y and no effect of X	96.09	16
	M4. ordinal–nominal	9.27	13
	M5. ordinal–ordinal	25.06	16
	M6. ordered-restricted	5.40	9[c]
Transitional	T1. nominal–nominal	208.47	466
	T2. nominal–ordinal	258.50	470
	T3. nominal Y and no effect of X	307.45	472
	T4. ordinal–nominal	237.39	473
	T5. ordinal–ordinal	274.84	475
	T6. order-restricted	209.37	468[c]
Random effects (parametric)	R1. nominal–nominal	216.65	470
	R2. nominal–ordinal	230.07	476
	R3. nominal Y and no effect of X	409.36	478
	R4. ordinal–nominal	221.67	476
	R5. ordinal–ordinal	235.36	479
	R6. order-restricted model	216.65[b]	470[c]
Random effects (three-class mixture)	L1. nominal–nominal	204.79	466
	L2. nominal–ordinal	211.57	472
	L3. nominal Y and no effect of X	388.01	474
	L4. ordinal–nominal	208.61	471
	L5. ordinal–ordinal	214.85	474
	L6. order-restricted model	204.79[b]	466[c]

[a] L-sq is the likelihood-ratio statistic, which is defined as twice the difference between the log-likelihood of the data and the log-likelihood of the model concerned. It is sometimes referred to as the deviance statistic.

[b] The fact that the L-sq of the order-restricted model has the same values as the one of the nominal–nominal model indicates that the latter was already in agreement with the order restrictions.

[c] The number of order-restricted parameters that is equated to zero is added to the degrees of freedom.

Application to data set on marijuana use

Table 15.4 reports the test results for a number of models that were estimated for the data in Table 15.1. As can be seen in Table 15.4, marginal, transitional and two types of random-effects models were estimated using different types of specifications for the response and time variables. As in the previous section, only adjacent-category logits have been applied. Many models contained a time–gender interaction term. But as in none of these models the interaction

effects were significant, they are not reported here. Obviously, the development of boys and girls over time with respect to marijuana use is the same according to this data set. In the transitional models, we conditioned on the state at the first occasion, which means that no logit restrictions were specified for the response at the first time point. The random-effects models are parametric and three-class mixture models with a random intercept: no more than three latent classes were needed in the non-parametric models and the random time effect was not significant.[3]

The test results show that in the marginal model and the two types of random-effects models, there is no problem if the response variable is treated as ordinal: the difference in L-sq between the ordinal–nominal and nominal–nominal specification is small given the difference in degrees of freedom. In the transitional model, this is somewhat more problematic. In each of the models, there is clear evidence for a time effect since the models without a time effect have much higher L-sq values than the models with a time effect, given the differences in degrees of freedom.[4] Although the ordinal (linear trend) specification captures the most important part of the time dependence, this specification is only satisfactory in the three-class mixture model.

In order to give an impression of the differences in parameter estimates between the four longitudinal data approaches, we present the parameters obtained for the ordinal–ordinal specification in Table 15.5. As can be seen, the signs of the time and gender effects are the same in each of approaches: marijuana use increases over time and males are more likely to use marijuana than females. The time effects are significant in each of the models. The estimates for the gender effects are on the borderline of significance in all models, except for the transitional model, in which it is clearly significant. A well-known phenomenon that can also be observed in this application is that effect sizes are generally larger in random-effects than in marginal models. The autocorrelation term in the transitional model shows that there is a strong dependence between responses at consecutive time points. The variance of random intercept shows that there are large differences in marijuana use among children.

[3] The non-parametric random-effects models were estimated with the Latent GOLD (Vermunt and Magidson, 2000; www.latentclass.com), which is a very user-friendly Windows program for latent class analysis. The other models were estimated with LEM (Vermunt, 1997; www.uvt.nl/mto), a general program for categorical data analysis. There are several software packages available for the estimation of parametric random-effects model for ordinal variables.

[4] The overall goodness of fit of most models is very good. The number of degrees of freedom is usually larger than the value of the likelihood-ratio statistic L-sq. The number of degrees of freedom equals the number of independent cells in the table that is analyzed, minus the number of parameters to be estimates plus the number of constraints that are imposed.

Table 15.5 *Parameter estimates for the ordinal–ordinal models*

Model[a]	Auto correlation	Variance[b]	Time	Gender
Marginal			0.28	0.28
Transitional	1.27		0.40	0.33
Random effects (parametric)		4.17	0.69	0.60
Random effects (three-class mixture)		3.37/1.44	0.68	0.39

[a] Standard errors are obtained by computing the observed information matrix, the matrix of second-order derivates of the log-likelihood function towards all parameters. The square root of diagonal elements of the inverse of this matrix contains the estimated standard errors.
[b] The three-class mixture model contains two variance terms, one for the log odds between categories 1 and 2 and one for the log odds between categories 2 and 3.

Special issues

In Section 15.1, it was mentioned that longitudinal data analysis methods should be able to deal with three main problems: unobserved heterogeneity, measurement error and incomplete responses. The issue of unobserved heterogeneity was already addressed above within the context of random-effects modelling.

Longitudinal data is often incomplete. When using ML estimation it is always possible to use cases with missing data on some of the occasions in the analysis. The assumption generally made is that the missing data are missing at random (MAR). ML under missing at random is straightforward in transitional and random-effects models since only the observed time points contribute to the likelihood function. Although in marginal models things are bit more complicated, the missing data problem can be dealt with, for example, by ML estimation using an expectation maximization (EM) algorithm.

Another serious problem in longitudinal data analysis, especially in the transitional modelling approach, is measurement error in the response variable. As a result of measurement error in the response variable, the number of observed transitions will be much larger than the true number of transitions, a phenomenon that has the highest impact for the smallest response category (Hagenaars, 1990; Bassi *et al.*, 2000). Another effect of measurement error in the dependent or the independent variable is that covariate effects may be biased. Measurement error can easily be taken into account when using a transitional approach. The transition model is then specified for the true unobserved states Φ_t, which are connected to observed stated Y_t by means of probabilities $P(Y_t|\Phi_t)$. This yields a model that is usually referred to as hidden Markov,

latent class Markov, or latent transition model (Langeheine and Van de Pol, 1994; Vermunt, 1997).

The latent Markov model can also be used to deal with multiple response variables. Each of the response variables will then be linked to the latent states by a set of conditional probabilities $P(Y_{tp}|\Phi_t)$, where Y_{tp} denotes response variable p at time point t. With multiple response variables, it is also possible to specify random-effects models in which measurement error is taken into account (Vermunt, 2003). Also the marginal modelling approach could be extended to deal with measurement error in the response variable.

Essentially, the introduction of (partially) unobserved or latent variables is a powerful means to overcome many of the most important problems and complexities in longitudinal analysis. Incorporated into one of the basic approaches towards longitudinal analysis and in combination with flexible kinds of ordinal restrictions to take the ordered nature of the categories into account, it provides the developmental researcher with excellent tools for answering the relevant research questions. And maybe the best news for the researcher is: many easy-to-use computer programs are available to carry out the job.

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley.

Bassi, F., Hagenaars, J. A., Croon, M. A. and Vermunt, J. K. (2000). Estimating true changes when categorical panel data are affected by uncorrelated and correlated errors. *Sociological Methods and Research*, **29**, 230–68.

Bergsma, W. (1997). *Marginal Models for Categorical Data*. Tilburg, The Netherlands: Tilburg University Press.

Clogg, C. C. and Shihadeh, E. S. (1994). *Statistical Models for Ordinal Data*. Thousand Oaks, CA: Sage.

Croon, M., Bergsma, W. and Hagenaars, J. A. (2000). Analyzing change in discrete variables by generalized log-linear models. *Sociological Methods and Research*, **29**, 195–229.

Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference or stochastic orderings in a nonparametric context. *Journal of the American Statistical Association*, **93**, 1112–23.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford, UK: Clarendon Press.

Elliot, D. S., Huizinga, D. and Menard, S. (1989). *Multiple Problem Youth: Delinquence, Substance Use and Mental Health Problems*. New York: Springer-Verlag.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.

Galindo-Garre, F., Vermunt, J. K. and Croon, M. A. (2002). Likelihood-ratio tests for order-restricted log-linear models: a comparison of asymptotic and bootstrap methods. *Metodología de las Ciencias del Comportamiento*, **4**, 325–37.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications saving ordered categories. *Journal of the American Statistical Association*, **74**, 537–52.

Hagenaars, J. A. (1990). *Categorical Longitudinal Data: Log-Linear Analysis of Panel, Trend and Cohort Data*. London: Sage.

(2002). Directed loglinear modeling with latent variables: causal models for categorical data with nonsystematic and systematic measurement errors. In *Applied Latent Class Analysis*, eds. J. A. Hagenaars and A. L. McCutcheon, pp. 234–86. New York: Cambridge University Press.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, **73**, 805–11.

Lang, J. B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–32.

Lang, J. B., McDonald, J. W. and Smith, P. W. F. (1999). Association modeling of multivariate categorical responses: a maximum likelihood approach. *Journal of the American Statistical Association*, **94**, 1161–71.

Langeheine, R. and Van de Pol, F. (1994). Discrete time mixed Markov latent class models. In *Analyzing Social and Political Change: A Casebook of Methods*, eds. A. Dale and R. B. Davies, pp. 171–97. London: Sage.

Vermunt, J. K. (1997). *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage.

(1999). A general non-parametric approach to the analysis of ordinal categorical data. *Sociological Methodology*, **29**, 197–221.

(2002). A general latent class approach to unobserved heterogeneity in the analysis of event history data. In *Applied Latent Class Analysis*, eds. J. A. Hagenaars and A. L. McCutcheon, pp. 383–407. New York: Cambridge University Press.

(2003). Multilevel latent class models. *Sociological Methodology*, **33**.

Vermunt, J. K. and Magidson, J. (2000). *Latent GOLD User's guide*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K. and Van Dijk, L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, **13**, 6–13.

Vermunt, J. K., Rodrigo, M. F. and Ato-Garcia, M. (2001). Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods and Research*, **30**, 170–96.