

Categorical Response Data

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University

1 Introduction

Earlier chapters in this volume discussed linear mixed models for continuous responses and generalized linear mixed models for dichotomous variables and counts with binomial and Poisson errors at the lower level, respectively. This chapter deals with multilevel models for discrete response variables with more than two categories; that is, with the situation where these errors can be assumed to come from a multinomial distribution, which can be seen as either a multivariate extension of the binomial distribution or a restricted version of the multivariate Poisson distribution. The dependent variables of interest can have ordered response categories modeled using an ordinal regression model or unordered response categories modeled using a multinomial logit, a probit, or a discrete choice model for first choices, rankings, or other types of choice formats.

While multilevel ordinal regression models are rather similar to models for dichotomous responses, models for nominal responses are somewhat more complex because they typically contain category-specific random effects. The latter makes them computationally much more demanding. Possible ways out of this problem are the use of factor analytic structures for the random effects yielding models similar to item response theory (IRT) models, the use of discrete approximations of the random effects using finite mixture models, or combinations of these. These options will be discussed in more detail below.

An issue that is getting more attention in the analysis of categorical responses is the distinction between heterogeneity in location or preference versus heterogeneity in scale. The latter concerns the error variance in an underlying latent variable specification of the ordinal or nominal regression model concerned. Taking into account heterogeneity in scale will usually yield much simpler explanations for group differences. Below, I will pay attention to this issue.

The remainder of this chapter is organized as follows. The theory is presented by first providing a generic model formulation, and subsequently

discussing models for ordinal responses and models for nominal responses. Then attention is paid to factor analytic and to discrete specifications for the random effects. After presenting the theory, I will illustrate multilevel analysis with categorical responses with a rather simple three-level random intercept model for an ordinal response variable and with a more complex model for a combination of ordinal and nominal responses (ratings and best-worst choices).

2 Generic model formulation

2.1 GLMMs for categorical response variables

Using the notation introduced in the previous chapters, I denote the response of subject i by y_i and the vector of predictors (which includes the constant) by \mathbf{x}_i . The total number of response categories is denoted by M , implying $1 \leq y_i \leq M$. The probability that person i 's response equals m , $P(y_i = m|\mathbf{x}_i)$, is denoted by π_{im} , and the vector of response probabilities by $\boldsymbol{\pi}_i$. Note that $\sum_{m=1}^M \pi_{im} = 1$.

Generalized Linear Models (GLMs) for categorical response variables have the following form (McCullagh and Nelder, 1989):

$$h_m(\boldsymbol{\pi}_i) = \mathbf{x}'_i \boldsymbol{\beta}_m, \quad (1)$$

for $2 \leq m \leq M$. That is, $M - 1$ transformations $h_m(\cdot)$ are defined for the response probabilities $\boldsymbol{\pi}_i$ and each of these is modelled via a linear predictor. Each linear predictor may have its own parameter vector $\boldsymbol{\beta}_m$. The main difference compared to a univariate GLM is thus that we have $M - 1$ such linear equations instead of one. As is shown in more detail below, in GLMs for ordinal response variables, predictor effects may be assumed to be equal across equations, eliminating the index m . Moreover, in discrete choice models for nominal responses, the predictor values that vary across responses instead of the regression coefficients; that is, $h_m(\boldsymbol{\pi}_i) = \mathbf{x}'_{im} \boldsymbol{\beta}$.

A two-level variant of the model defined in equation (1) is obtained as follows (Agresti, Booth, Hobert, and Caffo, 2000):

$$h_m(\boldsymbol{\pi}_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\beta}_m + \mathbf{z}'_{ij} \mathbf{u}_{jm}, \quad (2)$$

where the index i refers to a lower (or level-1) unit and j to a higher (or level-2) unit, and where \mathbf{u}_{jm} is the vector of random effects and \mathbf{z}_{ij} the associated

design vector. Note that different from univariate Generalized Linear Mixed Models (GLMMs; Verbeke and Molenberghs, this volume), \mathbf{u}_{jm} contains an index m , which indicates that the random effects may vary across response categories. Extension to a third level is rather straightforward,

$$h_m(\boldsymbol{\pi}_{ijk}) = \mathbf{x}'_{ijk}\boldsymbol{\beta}_m + \mathbf{z}'_{ijk}\mathbf{u}_{jkm}$$

where k refers a unit of a third level, and where $\mathbf{u}_{jkm} = (\mathbf{u}_{jkm}^{(2)'}, \mathbf{u}_{jkm}^{(3)'})'$ or $\mathbf{u}_{jkm} = (\mathbf{u}_{jm}^{(a)'}, \mathbf{u}_{km}^{(b)'})'$ for nested and crossed random effects, respectively.¹

The above equations, show the form of the $M - 1$ linear predictors. More details about the link functions are provided below when I describe the models for the different types of categorical responses. Besides the link function and the linear predictor, a GLMM requires specifying the distribution of the random effects and the distribution of the responses conditional on the random effects. The latter is modeled using a multinomial distribution. The $M - 1$ sets of random effects are typically assumed to come from a multivariate normal distribution. Using these components, we can define the likelihood for the GLMM interest, which contrary to the one for Linear Mixed Models contains intractable integrals. Similar to other types of GLMMs, approximations are typically used, such as linear approximations yielding analytically solvable integrals, numerical integration using quadrature, or Monte Carlo integration methods (McCulloch and Neuhaus, this volume; Skrondal and Rabe-Hesketh, 2004). Another option is to solve the parameter estimation problem within a Bayesian MCMC framework (Train, 2003).

3 Models for ordinal responses

3.1 Underlying latent variable approach

A popular approach to the regression analysis of ordinal response variables is to treat the ordinal responses as discretized continuous responses (Agresti, 2002). The continuous responses y_{ij}^* are unobserved or latent variables for which a linear regression model is defined. That is,

$$y_{ij}^* = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_j + \varepsilon_{ij}.$$

¹The superscript denotes the level of the random effect to remove any ambiguity and provide direct reference in this vector notation.

The relationship between y_{ij}^* and the observed discrete response y_{ij} is that $y_{ij} = m$ if $\delta_m < y_{ij}^* < \delta_{m+1}$. The δ_m are the so-called thresholds, which are parameters to be estimated for $2 \leq m \leq M$, and with $\delta_1 = -\infty$ and $\delta_{M+1} = \infty$. Depending on the distributional assumptions about the error term ε_{ij} one obtains a specific type of ordinal regression model. The most common specifications for ε_{ij} are standard normal, logistic, and extreme value, yielding a probit, logit, and complementary log-log model, respectively.

3.2 Cumulative link functions

The same models as can be defined using an underlying latent variable approach, can also be defined using a cumulative link function combined with the assumption that predictor effects are equal across the $M - 1$ linear equations. Let $\pi_{ijm}^+ = \sum_{k=m}^M \pi_{ijk}$, or the probability that person i in group j gives response m or higher. Note that this implies that $\pi_{ijm} = \pi_{ijm}^+ - \pi_{ij,m-1}^+$.

The random-effects cumulative logit or proportional odds model is obtained using a logit transformation for π_{ijm}^+ (Hedeker and Gibbons, 1994):

$$\log \frac{\pi_{ijm}^+}{1 - \pi_{ijm}^+} = \alpha_m + \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_j,$$

or

$$\pi_{ijm}^+ = \frac{\exp(\alpha_m + \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_j)}{1 + \exp(\alpha_m + \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_j)}.$$

Here α_m are fixed intercept terms which depend on the response category. These are in fact the same as the thresholds defined above, but with a reversed sign: $\alpha_m = -\delta_m$. The vector $\boldsymbol{\beta}$ contains the fixed slopes and \mathbf{u}_j the random intercept and slopes. Note that these effects are constant across response categories, a restriction that is usually referred to as the proportional odds assumption when using a model with a logit link. Sometimes the proportionality assumption is relaxed, yielding regression coefficients $\boldsymbol{\beta}_m$ and \mathbf{u}_{jm} which depend on the response category.

Alternatives to the logit transformation are among others inverse standard normal and inverse extreme value transformations, yielding a probit and a complementary log-log model, respectively.

3.3 Scale heterogeneity

The latent variable approach described above can be extended by allowing the error terms ε_{ij} to be heteroscedastic (Hedeker, Berbaum, and Mermelstein, 2006). This is usually achieved by defining a log-linear model for the standard deviation of ε_{ij} ; that is,

$$\sigma_{\varepsilon_{ij}} \propto \exp(-\mathbf{w}'_{ij}\boldsymbol{\gamma}).$$

This amounts to expanding the cumulative logit model as follows:

$$h_m(\boldsymbol{\pi}_{ij}) = \log \frac{\pi_{ijm}^+}{1 - \pi_{ijm}^+} = (\alpha_m + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_j) \exp(\mathbf{w}'_{ij}\boldsymbol{\gamma}).$$

The term $\exp(\mathbf{w}'_{ij}\boldsymbol{\gamma})$ is sometimes referred to as the scale factor. Though not shown explicitly here, it is also possible to include random terms in the log-linear model for $\sigma_{\varepsilon_{ij}}$.

3.4 Other link functions

Instead of a cumulative link function, it is also possible to use other types of link functions, the most popular of which are the adjacent-category and the continuation-ratio logit models (Agresti, 2002). As in cumulative logit models, the linear term equals $\alpha_m + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_j$. The adjacent-category model is a logit model for $P(y_{ij} = m | m - 1 \leq y_{ij} \leq m)$, the probability of selecting category m out of adjacent categories $m - 1$ and m . The two possible variants of the continuation-ratio logit model concern $P(y_{ij} \geq m | y_{ij} \geq m - 1)$ and $P(y_{ij} \leq m - 1 | y_{ij} \leq m)$, respectively.

4 Models for nominal responses

4.1 First choice

Also for the case of unordered or nominal responses it is possible to define the regression model based on an underlying latent variable approach (Grilli and Rampachini, 2007; Skrondal and Rabe-Hesketh, 2004; Swait and Louviere, 1993; Train, 2003). The utilities y_{ijm}^* of the M categories (or alternatives) are unobserved or latent variables for which linear regression models are defined. That is,

$$y_{ijm}^* = \mathbf{x}'_{ij}\boldsymbol{\beta}_m + \mathbf{z}'_{ij}\mathbf{u}_{jm} + \varepsilon_{ijm}.$$

The relationship between y_{ijm}^* and the observed discrete response y_{ij} is as follows: $y_{ij} = m$ if $y_{ijm}^* > y_{ijr}^*$ for all $r \neq m$. In other words, the response corresponds with the category for which the utility is largest. Depending on the assumed distribution of the error terms ε_{ijm} , one obtains another type of regression model. Assuming ε_{ijm} to be extreme value distributed and independent across alternatives yield the multinomial logit model. The multinomial probit model follows from the assumption that the ε_{ijm} come from a multivariate normal distribution. Here, I will use only the more popular multinomial logit specification in which

$$\pi_{ijm} = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_m + \mathbf{z}'_{ij}\mathbf{u}_{jm})}{\sum_{r=1}^M \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_r + \mathbf{z}'_{ij}\mathbf{u}_{jr})}.$$

As in models for ordinal responses, it may be useful to include a scale factor model, which is a model for capturing heterogeneity in the standard deviation of ε_{ijm} (Swait and Louviere, 1993); that is,

$$\sigma_{\varepsilon_{ijm}} \propto \exp(-\mathbf{w}'_{ij}\boldsymbol{\gamma}).$$

This modifies the multinomial logit model into

$$\pi_{ijm} = \frac{\exp\left[(\mathbf{x}'_{ij}\boldsymbol{\beta}_m + \mathbf{z}'_{ij}\mathbf{u}_{jm}) \exp(\mathbf{w}'_{ij}\boldsymbol{\gamma})\right]}{\sum_{r=1}^M \left[\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_r + \mathbf{z}'_{ij}\mathbf{u}_{jr}) \exp(\mathbf{w}'_{ij}\boldsymbol{\gamma})\right]}.$$

Below, I will provide an example in which the scale factor varies randomly across (latent classes) of respondents.

The nominal response models presented above can be used when predictor values vary across level-1 or level-2 units. Another related class of models has been developed for the situation in which predictor values vary across response categories or choice alternatives. These models are usually referred to as discrete choice models or conditional logit models (McFadden, 1974; Swait and Louviere, 1993). An example application is the analysis of transportation mode choices, which may depend on characteristics of the mode, such as the required time to reach the destination and the price. A model with alternative-specific predictor values for both fixed and random effects has the following form:

$$\pi_{ijm} = \frac{\exp(\mathbf{x}'_{ijm}\boldsymbol{\beta} + \mathbf{z}'_{ijm}\mathbf{u}_j)}{\sum_{r=1}^M \exp(\mathbf{x}'_{ijr}\boldsymbol{\beta} + \mathbf{z}'_{ijr}\mathbf{u}_j)}.$$

Also in this model one can incorporate a regression model for scale heterogeneity. Note also that it is straightforward to defining hybrids of the multinomial and conditional logit model. This can be achieved by including alternative-specific constants and/or interactions with individual-level predictors in the vectors \mathbf{x}_{ijm} and \mathbf{z}_{ijm} .

4.2 Other choice formats

When confronting individuals with choice tasks, it is often useful to obtain more information than on just the first choice. For example, one may ask for the best and second best alternative, the best and the worst alternative, the complete ranking of the alternatives, or the best two alternatives (without making a distinction between best and second best). These are all variants of ranking tasks, which can be modelled using logistic regression models similar to those for first choices (Skrondal and Rabe-Hesketh, 2003).

The simplest type of ranking model is obtained when it is assumed that the rankings are obtained sequentially, that is, one first selects the best alternative and subsequently selects the second best (or the worst) from the remaining $M - 1$ alternatives. The probability of selecting alternative m_1 as the best and alternative m_2 as the second best (worst) can be written as $\pi_{ij,m_1,m_2} = \pi_{ij,m_1}\pi_{ij,m_2|m_1}$, where the conditioning on m_1 implies that this alternative can no longer be selected. Generalization to three choices yields $\pi_{ij,m_1,m_2,m_3} = \pi_{ij,m_1}\pi_{ij,m_2|m_1}\pi_{ij,m_3|m_1,m_2}$, etc. The logit model for the second best alternative has the following form:

$$\pi_{ijm_2|m_1} = \frac{\exp(\mathbf{x}'_{ijm_2}\boldsymbol{\beta} + \mathbf{z}'_{ijm_2}\mathbf{u}_j)}{\sum_{r \neq m_1} \exp(\mathbf{x}'_{ijr}\boldsymbol{\beta} + \mathbf{z}'_{ijr}\mathbf{u}_j)}, \quad (3)$$

which is the same as a first choice model except for the sum in the denominator which excludes the previous choice. In a best-worst choice design, one first selects the most preferred alternative and subsequently the worst alternative. The model for the worst alternative has the same form as the model described in equation (3), but with the difference that the sign of the linear term changes since the alternative with the lowest instead of the highest utility is selected (Magidson, Thomas, and Vermunt, 2009).

Rather than a sequential ranking mechanism, it is also possible to define ranking models based on a simultaneous ranking mechanism: best and second best, best and worst, or the two best are selected simultaneous after

evaluation all possible simultaneous choices. The model will be one for a single choice out of an expanded set of joint choices and where the design matrices are weighted to account for the mutual ranking (first choice get weight of 2 and second choice of 1), if known. For example

$$\pi_{ij,m_1,m_2} = \frac{\exp[2(\mathbf{x}'_{ijm_1}\boldsymbol{\beta} + \mathbf{z}'_{ijm_1}\mathbf{u}_j) + (\mathbf{x}'_{ijm_2}\boldsymbol{\beta} + \mathbf{z}'_{ijm_2}\mathbf{u}_j)]}{\sum_{r_1=1}^M \sum_{r_2 \neq r_1} \exp[2(\mathbf{x}'_{ijr_1}\boldsymbol{\beta} + \mathbf{z}'_{ijr_1}\mathbf{u}_j) + (\mathbf{x}'_{ijr_2}\boldsymbol{\beta} + \mathbf{z}'_{ijr_2}\mathbf{u}_j)]}$$

5 Factor analytic restrictions on random effects

Random effects models for categorical response responses are strongly related to item response theory (IRT) models, which are factor analytic models for multivariate response data (Van der Linden and Hamilton, 1997). The latter can, however, also be seen as two-level data with item responses nested within persons. A simple unidimensional IRT model for nominal responses has the following form:

$$h_m(\boldsymbol{\pi}_{ij}) = \lambda_{im}(\theta_j - \delta_{im}).$$

where θ_j represents the latent trait of person j , δ_{im} the difficulty parameter for category m of item i , and λ_{im} the discrimination parameter (or factor loading) for category m of item i .²

The link function is typically logit. Models for ordinal responses impose the restriction that the discrimination parameter equals across response categories. Another possible restriction used in IRT is that the discrimination parameter equals across items.

Using the multilevel notation introduced above, the same model can also formulated as follows:

$$h_m(\boldsymbol{\pi}_{ij}) = \beta_{im} + \lambda_{im}u_j,$$

where $u_j = \theta_j$ and $\beta_{im} = -\delta_{im}\lambda_{im}$. What can be observed is that a single random term u_j is used for modeling the random intercepts of all items and all response categories, were a separate scaling term λ_{im} is estimated for

²The term difficulty comes from the application of IRT models with dichotomous responses from educational tests: the higher δ the more difficult to answer the question concerned correctly. The term discrimination refers to the fact that the higher λ the better the item discriminates (distinguishes) persons with different θ values.

each item category. In fact, the random intercept for item i and category m is specified to have a variance which is proportional to $(\lambda_{im})^2$ and the correlation between the random intercepts equals either 1 or -1 depending on whether the two λ parameters concerned have the same sign or not. A multidimensional extension of an IRT model for nominal data can be defined as follows:

$$h_m(\boldsymbol{\pi}_{ij}) = \beta_{im} + \boldsymbol{\lambda}'_{im} \mathbf{u}_j,$$

where $\boldsymbol{\lambda}_{im}$ is a vector for discrimination parameters or loadings, and \mathbf{u}_j a vector of latent variables.

Hedeker (2003) proposed using a lower-dimensional approximation of the random effects distribution similar to IRT models for nominal data. For this purpose, the general model defined in equation (2) is reformulated as follows:

$$h_m(\boldsymbol{\pi}_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\beta}_m + \mathbf{z}'_{ij} \boldsymbol{\Lambda}_m \mathbf{u}_j,$$

where $\boldsymbol{\Lambda}_m$ is a kind of factor loadings matrix scaling the random effects variances (Skrondal and Rabe-Hesketh, 2004). The main advantage of this approach is that the dimensionality of the integrals in the likelihood function is reduced by a factor $M - 1$. Factor analytic structures can not only be used to impose simple structures for the random effects across response categories, but also to impose simple structures across the random slopes of different predictors, as is in fact done in the above IRT models.

6 Finite mixture specification of random effects

Rather than using a continuous random effects distribution, it is also possible to use a discrete distribution for the random effects, either as an approximation of a continuous distribution or because one wishes to cluster level-2 units based on differences in the regression coefficients (Aitkin, 1999; Browne and McNicholas, this volume; Vermunt and Van Dijk, 2001; Wedel and DeSarbo, 1994). Let C denote the number of latent classes and c a particular latent class; thus, $1 \leq c \leq C$. The resulting two-level mixture regression model can be formulated as follows:

$$h_m(\boldsymbol{\pi}_{ij|c}) = \mathbf{x}'_{ij} \boldsymbol{\beta}_m + \mathbf{z}'_{ij} \mathbf{u}_{cm},$$

where \mathbf{u}_{cm} are free parameters to be estimated. These parameters indicate the location of the effects of latent class c relative to the average effects β_m . The class membership is assumed to follow a multinomial distribution with probabilities π_c . The \mathbf{u}_{cm} and π_c are sometimes referred to as mass point locations and mass point weights. It should be noted that the computational burden for such a finite mixture two-level regression model is much lower than with continuous random effects since the likelihood contains a sum over C classes (or mass points) instead of a multidimensional integral.

When using the model for clustering purposes, one will usually interpret the terms $\beta_{pm} + u_{pcm}$, which are the class-specific regression parameters for the p th predictor. When using the model as an approximation for a continuous random effects model, one will typically be interested in the fixed effects and the random effects (co)variances. The β_{pm} have the usual interpretation if the u_{pcm} terms are centered across classes; that is, if $\sum_c u_{pcm} \pi_c = 0$. The covariance between two random effects can be obtained as follows: $\sum_c u_{pcm} u_{p'cm'} \pi_c$.

As is shown in the second application below, it sometimes makes sense to use a hybrid model combining discrete and continuous random effects. Such an approach is rather popular in the field of longitudinal data analysis, where these models are referred to as mixture growth models (Muthén, 2004; Vermunt, 2007). The aim is to identify latent classes with different change trajectories, while taking into account within class residual heterogeneity in these trajectories.

7 Applications

7.1 A three-level random intercept regression model for an ordinal response variable

The first application concerns the analysis of a three-level data set using an ordinal regression model. The data come from the Television School and Family Smoking Preventions and Cessation Project (TVSFP) and were analyzed using a linear mixed model by Hedeker, Gibbons, and Flay (1994). The multilevel data structure arises from the fact that the 1600 children are nested within 135 classes which are themselves nested within 28 schools. Consistent with the notation introduced above, the indexes i , j , and k are used to refer to a child, a class, and a school, respectively.

Schools were randomized into four conditions obtained by crossing two

experimental factors: media intervention present or absent (x_{1k}) and social-resistance classroom curriculum present or absent (x_{2k}). The outcome is the child’s tobacco health knowledge (y_{ijk}) measured on an ordinal scale with 8 categories, $1 \leq m \leq 8$. Tobacco health knowledge was also measured before the intervention (x_{3ijk}). Taking into account the nesting of children within classes and classes within schools, we define the following three-level ordinal regression model:

$$\log \frac{\pi_{ijkm}^+}{1 - \pi_{ijkm}^+} = \alpha_m + \beta_1 x_{1k} + \beta_2 x_{2k} + \beta_3 x_{1k} x_{2k} + \beta_4 x_{3ijk} + u_{jk}^{(2)} + u_k^{(3)}.$$

The question of interest is whether the treatment effects and their interaction are significant. This was investigated using different specifications for the random intercepts; that is, the full model containing both school and class random effects, models with either class effects or school effects, and a model without random effects. Table 1 reports the fit measures and parameters estimates for these four models. I obtained the parameters with the Latent GOLD software (Vermunt and Magidson, 2008) using its default settings for numerical integration. Other software can also be used, such as the GLLAMM routine written for Stata (Skrondal and Rabe-Hesketh, 2003, 2004).

As can be seen from Table 1, the model with only a random intercept at the class level should be preferred according to the BIC and AIC statistics. In this model, the main effect of social-resistance classroom curriculum (β_2) is significant, but the main effect of mediation intervention (β_1) and the interaction (β_3) are not. This is clearly different from the results of the model that ignores the multilevel structure, in which all fixed effects are significant.

7.2 A two-level model for a choice experiment combining best-worst choices with ratings

The second application concerns the analysis of data from an experimental study done for Roche Diagnostics Corporation, a supplier of lab equipment, with the aim to construct a “needs based segmentation” of the company’s clients. The segmentation is based on the importance 305 lab managers attach to 36 different features of lab equipment. The results of this study were used in the training of sales people; that is, when selling to a lab manager,

Table 1: Fit measures and parameters estimates (z values between braces) for the tobacco health knowledge intervention study.

	Full model		Model with only class effect		Model with only school effect		Model without random effects	
α_2	2.03	(10.33)	2.01	(11.21)	2.00	(10.42)	1.93	(12.24)
α_3	0.05	(0.29)	0.03	(0.17)	0.05	(0.31)	-0.01	(-0.06)
α_4	-1.23	(-7.32)	-1.25	(-8.51)	-1.19	(-7.37)	-1.23	(-10.16)
α_5	-2.43	(-13.85)	-2.46	(-15.72)	-2.37	(-13.98)	-2.39	(-18.18)
α_6	-3.87	(-19.95)	-3.89	(-22.04)	-3.78	(-20.20)	-3.79	(-24.65)
α_7	-5.57	(-22.00)	-5.59	(-23.29)	-5.47	(-22.13)	-5.47	(-24.57)
α_8	-7.69	(-14.44)	-7.71	(-14.65)	-7.58	(-14.33)	-7.59	(-14.64)
β_1	0.23	(1.14)	0.20	(1.21)	0.27	(1.38)	0.26	(2.09)
β_2	0.89	(4.35)	0.88	(5.13)	0.91	(4.63)	0.88	(6.92)
β_3	-0.38	(-1.30)	-0.32	(-1.32)	-0.46	(-1.68)	-0.39	(-2.21)
β_4	0.43	(11.20)	0.44	(11.41)	0.43	(11.23)	0.45	(11.90)
σ_{u3}	0.20	(2.07)			0.26	(3.86)		
σ_{u2}	0.39	(4.80)	0.44	(6.01)				
loglik	-2652		-2653		-2658		-2663	
# of par.	13		12		12		11	
BIC(N=28)	5348		5346		5356		5363	
AIC	5331		5330		5340		5349	
pseudo R^2	0.19		0.19		0.14		0.12	

they could get a feel as to which segment a manager belongs and therefore emphasize which lab equipment best filled these needs.³

This study used a combination of best-worst choices and ratings. For the choice task, 64 different choice sets were created, each consisting of 5 alternatives ($M = 5$). More specifically, each set contained 5 randomly selected features out of the list of 36 lab equipment features. Each of the 305 lab managers participating in this study was confronted with 16 of the 64 choice sets and requested to pick the best (most important) and the worst (least important) feature. The question asked was “Which of the following criteria is most important and which is least important to you in choosing a supplier of lab equipment?”. In addition, each lab manager provided an ‘importance’ rating on a 5-point scale for each of the 36 features (1=not important, ..., 5=extremely important). The hybrid choice-rating model described below was proposed by Magidson, Thomas, and Vermunt (2009).

Note that the data set has a two-level structure: the 16 best-worst choices

³For another multilevel market research application using preference data, see Sagan (this volume).

and the 36 ratings are nested within persons. The question of main interest concerns the importance lab manager managers attach to the features and, more specifically, whether segments can identified with different “needs”. This means that it is most natural to model variation in feature effects using a mixture regression specification with C latent “needs” classes, which we index by c . The class-specific preference parameters are denoted by β_{pc} , for $1 \leq p \leq 36$ and $1 \leq c \leq C$.

Various other issues need to be taken into account in the model specification:

- Different types of regression models are needed for the 16 best-worst choices and for the 36 ratings. For the former ($1 \leq i \leq 16$), I use a sequential ranking model and for the latter ($17 \leq i \leq 52$) a proportional odds model.
- The preference parameters can only be assumed to be proportional to one another in the ranking and rating models. This is achieved by using a different log-linear scale parameter in the ranking and the rating model (γ_{1d} and γ_{2d}).
- Respondents may not only differ in preference but also in how certain they are about their preferences. This is same as saying that the level-1 variances of the underlying latent variables are heteroscedastic. This heterogeneity is modeled by allowing the log-linear scale parameters to be vary across D latent scale classes index by d (γ_{1d} and γ_{2d}). An alternative would be to model scale heterogeneity using continuous random effects.
- Ratings are affected by the fact that some persons tend to give higher ratings than others irrespective of the specific content of the question (the feature). We correct for such individual differences in scale use which is unrelated to the important attached to the features by a random intercept (u_j). These differences in response style affecting the ratings could also be modeled using a latent class specification (Moors, 2003; Morren, Gelissen, Vermunt, 2011).

The model used is formulated as follows:

$$\pi_{ijm_1|cd} = \frac{\exp \left[\left\{ \sum_{p=1}^{36} (\beta_p + u_{pc}) x_{pim_1} \right\} \exp(\gamma_{1d}) \right]}{\sum_{r=1}^M \exp \left[\left\{ \sum_{p=1}^{36} (\beta_p + u_{pc}) x_{pir} \right\} \exp(\gamma_{1d}) \right]}$$

$$\pi_{ijm_2|m_1cd} = \frac{\exp \left[\left\{ - \sum_{p=1}^{36} (\beta_p + u_{pc}) x_{pim_2} \right\} \exp(\gamma_{1d}) \right]}{\sum_{r \neq m_1} \exp \left[\left\{ - \sum_{p=1}^{36} (\beta_p + u_{pc}) x_{pir} \right\} \exp(\gamma_{1d}) \right]},$$

for $1 \leq i \leq 16$, and

$$\pi_{ijm|cd}^+ = \frac{\exp \left[\left\{ \alpha_{0m} + \alpha_{1c} + \sum_{p=1}^{36} (\beta_p + u_{pc}) x_{pi} + u_j \right\} \exp(\gamma_{2d}) \right]}{1 + \exp \left[\left\{ \alpha_{0m} + \alpha_{1c} + \sum_{p=1}^{36} (\beta_p + u_{pc}) x_{pi} + u_j \right\} \exp(\gamma_{2d}) \right]}.$$

for $17 \leq i \leq 52$. Note that x_{pim} equals 1 if alternative m of set i of person j corresponds to attribute p and otherwise 0, and x_{pi} equals 1 if rating i concern attribute p (if $p = i - 16$) and otherwise 0. The parameters of main interest, the class-specific utilities u_{pc} , appear in both the best-worst and the rating model. Note that these indicate how much class c deviates from the average utility β_p . The class-specific utilities β_{pc} , which are the actual parameters estimated in the mixture regression analysis, equal $\beta_p + u_{pc}$.

Running a series of models for different values of C and D showed that a model with five preference classes ($C = 5$) and two scale classes ($D = 2$) performed best according to the BIC. To illustrate the impact of ignoring differences in the use of the rating scale, in error variances, and in preference, I also show the fit of restricted models in which either $u_j = 0$, $D = 1$, or $C = 1$, and the model in which all three restrictions are imposed simultaneously (see Table 2). For parameter estimation, I used the Latent GOLD software, including its Choice module (Vermunt and Magidson, 2005, 2008). The model can also be estimated using LEM (Vermunt, 1997).

Table 2: Fit measures for the segmentation of lab managers study.

Model	Loglik.	# of par.	BIC	AIC	R^2 choice	R^2 rating
Unrestricted	-23157	192	47412	46698	0.23	0.46
$\sigma_u = 0$	-23618	191	48329	47618	0.22	0.39
$D = 1$	-23289	189	47659	46956	0.23	0.47
$C = 1$	-24170	44	48591	48427	0.17	0.39
Restricted	-25080	40	50389	50240	0.17	0.24

As can be seen from the BIC and AIC values, ignoring any of the three multilevel components deteriorates the fit significantly. However, the pseudo R^2 values for the choice and ranking variables show that ignoring the heteroscedasticity in the error variances ($D = 1$) does not deteriorate within sample prediction.

Table 3 reports the β_{pc} parameters, the class-specific utilities (importances) of the 36 features. As can be seen, there are large differences in importance across features and across latent classes. Some of the features are considered to be important in all 5 segments, such as features 11 and 17, whereas others are specific for one particular class. For example, features 24, 7, 12, and 16 are important only for sales managers belonging to classes 1, 3, 4, and 5, respectively. Both results are of interest; that is, which features are important in all segments and which are important in specific segments only.

8 Discussion

This chapter discussed multilevel regression models for ordinal and nominal response variables. Special attention was paid to issues such as the distinction between location and scale heterogeneity, the specification of restricted random effects using factor analytic structures, and the use of discrete specifications for the random effects, possibly combined with continuous random effects.

In this chapter, I did not discuss models for multivariate categorical data. However, recently various interesting multilevel models for multivariate categorical responses have been proposed. These models fit within the generalized latent variable modeling framework described by Skrondal and Rabe-Hesketh (2004). The most important special cases are multilevel latent class and mixture models (Vermunt, 2003, 2008), multilevel IRT and factor analytic models for categorical responses (Fox and Glas, 2001; Varriale and Vermunt, 2012), and multilevel mixture growth models (Palardy and Vermunt, 2010).

References

- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 117-128.
- Agresti, A. (2002). *Categorical Data Analysis (Second Edition)*. Hoboken, NJ: John Wiley & Sons.
- Agresti, A., Booth, J.G., Hobert, J.P., and Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*,

30, 27-80.

- Browne, R.P., and McNicholas, P.D. (this volume). Mixture and latent class models in the multilevel setting. M.A. Scott, J.S. Simonoff, and B.D. Marx (eds.), *The SAGE handbook of multilevel modeling*. Sage.
- Fox, J. P., and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 269-286.
- Grilli, L., and Rampichini, C. (2007). A multilevel multinomial logit model for the analysis of graduates' skills. *Statistical Methods and Applications*, 16, 381-393.
- Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22, 1433-1446.
- Hedeker, D., Berbaum, M., and Mermelstein, R. (2006). Location-scale models for multilevel ordinal data: Between- and within-subjects variance modeling. *Journal of Probability and Statistical Science*, 4, 1-20.
- Hedeker, D., and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-944.
- Hedeker, D., Gibbons, R. D., and Flay, B. R. (1994). Random-effects regression models for clustered data: with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62, 757-765.
- Magidson, J., Thomas, D., and Vermunt, J.K. (2009). A new model for the fusion of MaxDiff scaling and ratings data. *2009 Sawtooth Software Conference Proceedings*, 83-103.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall: London.
- McCulloch, C.E., and Neuhaus, J.M. (this volume). Generalized linear mixed models: Estimation and inference. M.A. Scott, J.S. Simonoff, and B.D. Marx (eds.), *The SAGE handbook of multilevel modeling*. Sage.

- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. P. Zarembka (ed.), *Frontiers in Econometrics*, 105-142, New York: Academic Press.
- Moors, G. (2003). Diagnosing response style Behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality and Quantity*, 37, 277-302.
- Morren, M., Gelissen, J.P.T.M., and Vermunt, J.K. (2011). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. *Sociological Methodology*, 41, 13-47.
- Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, 345-368. Thousand Oakes: Sage Publications.
- Palardy, G., and Vermunt, J.K.. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics*, 35, 532-565.
- Sagan, A. (this volume). Market research and preference data. M.A. Scott, J.S. Simonoff, and B.D. Marx (eds.), *The SAGE handbook of multilevel modeling*. Sage.
- Skrondal, A., and Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, 68, 267-287.
- Skrondal, A., and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. London: Chapman & Hall/CRC.
- Swait, J., and Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30, 305-314.
- Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

- Van der Linden, W. J., Ronald, K., and Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York, NY: Springer.
- Varriale, R., and Vermunt, J.K. (2012). Multilevel mixture factor models. *Multivariate Behavioral Research*, 47, 247-275.
- Verbeke, G, and Molenberghs, G. (this volume). Generalized linear mixed models Overview. M.A. Scott, J.S. Simonoff, and B.D. Marx (eds.), *The SAGE handbook of multilevel modeling*. Sage.
- Vermunt, J.K (1997). *LEM 1.0: A general program for the analysis of categorical data*. Tilburg, NL: Tilburg University.
- Vermunt, J. K. (2003). *Multilevel latent class models*. *Sociological Methodology*, 33, 213-239.
- Vermunt, J.K. (2007). Growth models for categorical response variables: standard, latent-class, and hybrid approaches. In: K. van Montfort, H. Oud, and A. Satorra (eds.), *Longitudinal Models in the Behavioral and Related Sciences*, 139-158. Mahwah, NJ: Erlbaum.
- Vermunt, J.K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 17, 33-51.
- Vermunt, J. K. and Magidson, J. (2005). *Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced*, Belmont, MA: Statistical Innovations.
- Vermunt, J. K. and Magidson, J. (2008). *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*, Belmont, MA: Statistical Innovations.
- Vermunt, J. K., and Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.
- Wedel, M., and DeSarbo, W.S. (1994). A review of recent developments in latent class regression models. R.P. Bagozzi (ed.), *Advanced Methods Marketing Research*, 352-388. Cambridge: Blackwell Publishers.

Table 3: Estimates of the importance of the 36 features for the 5 latent classes of lab managers (β_{pc} parameters), their average (β_p), and their standard deviation. Rows are sorted by decreasing average importance.

Feature number	Class 1	Class 2	Class 3	Class 4	Class 5	Average	Std. Dev.
11	2.80	1.94	3.05	2.45	2.35	2.42	0.41
17	2.60	1.81	2.44	3.24	2.16	2.34	0.50
26	2.76	1.96	2.31	2.28	1.75	2.15	0.31
13	2.36	1.13	0.75	1.22	2.43	1.47	0.64
10	1.65	0.75	3.25	1.54	0.63	1.44	0.94
15	1.87	1.11	1.25	1.51	1.63	1.40	0.27
36	0.35	1.28	0.17	1.41	1.62	1.04	0.55
28	1.52	0.13	1.85	0.71	1.30	0.93	0.67
21	0.82	0.62	0.73	1.69	0.47	0.83	0.41
9	0.67	1.04	-0.50	1.00	0.55	0.62	0.56
3	0.60	0.86	-0.02	1.00	0.11	0.56	0.40
33	0.57	0.84	0.16	0.55	0.17	0.51	0.28
24	1.05	0.15	0.59	0.26	-0.08	0.33	0.35
22	0.37	-0.30	-0.65	0.80	1.58	0.24	0.78
32	-0.23	0.53	-0.34	0.13	-0.09	0.09	0.34
30	-0.84	0.50	-0.65	0.16	0.52	0.06	0.55
1	-0.83	-0.15	0.89	-0.45	-0.24	-0.12	0.53
8	0.31	-0.39	0.45	0.67	-1.58	-0.17	0.77
19	0.25	-0.60	0.38	-0.83	0.05	-0.24	0.47
7	-1.34	-0.28	1.06	-0.33	-0.71	-0.26	0.71
16	0.44	-1.16	-0.95	-0.38	1.50	-0.31	0.98
2	-0.78	0.00	-1.03	-0.34	-0.25	-0.39	0.39
23	-0.44	-0.12	-1.06	-0.32	-0.58	-0.45	0.33
12	-0.97	-1.27	0.34	1.67	-1.51	-0.47	1.16
27	-0.88	-0.56	-0.53	-0.84	0.35	-0.48	0.41
25	-1.61	-0.19	-0.40	-0.70	-0.63	-0.58	0.45
6	-1.48	-0.19	-0.15	-0.50	-1.34	-0.61	0.54
18	0.40	-0.54	-1.67	-2.41	-0.19	-0.89	0.92
4	-1.78	-0.81	-0.04	-2.04	-0.23	-0.91	0.74
5	-0.20	-0.37	-1.30	-1.99	-1.36	-0.97	0.65
29	-1.99	-0.59	-1.13	-1.34	-0.91	-1.06	0.45
20	0.35	-1.32	-0.65	-1.65	-1.63	-1.09	0.65
14	-1.20	-1.43	-2.34	-1.20	-2.14	-1.65	0.45
31	-2.20	-1.13	-1.95	-2.01	-1.58	-1.65	0.42
34	-2.17	-1.39	-1.72	-2.41	-1.91	-1.82	0.37
35	-2.81	-1.85	-2.60	-2.58	-2.20	-2.30	0.37
Class proportion	0.13	0.34	0.18	0.17	0.18		