Ordered multinomial response models

Ordered categorical data

Where there is an underlying ordering to the categories a convenient parameterisation is to work with cumulative probabilities, i.e. the probabilities that an individual crosses each threshold. For example, with exam grades

Grade	Probability	Threshold		Cumulative probability
D	π_{1i}	$\leqslant$ D	(D)	$\gamma_{1i} = \pi_{1i}$
C	π_{2i}	$\leqslant$ C	(C, D)	$\gamma_{2i} = \pi_{1i} + \pi_{2i}$
B	π_{3i}	$\leqslant$ B	(B, C, D)	$\gamma_{3i} = \pi_{1i} + \pi_{2i} + \pi_{3i}$
A	π_{4i}	$\leqslant$ A	(A, B, C, D)	$\gamma_{4i} = \pi_{1i} + \pi_{2i} + \pi_{3i} + \pi_{4i} = 1$

With an ordered multinomial we work with the set of cumulative probabilities γ_{ki}. As before, with t categories, we put $t - 1$ categories in the model. The remaining cumulative probability, which is the sum of the probabilities for all the categories, must have the value 1 by definition

A model with no explanatory variables

$\log(\gamma_{1i}/(1-\gamma_{1i}))=\beta_0$ log odds of $\leqslant$ D
$\log(\gamma_{2i}/(1-\gamma_{2i}))=\beta_1$ log odds of $\leqslant$ C
$\log(\gamma_{3i}/(1-\gamma_{3i}))=\beta_2$ log odds of $\leqslant$ B

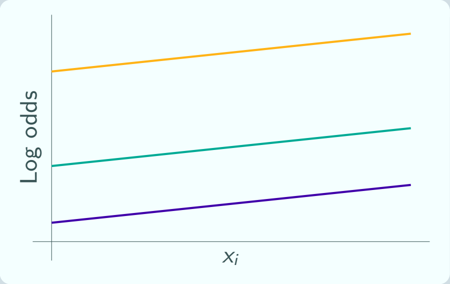
The threshold probabilities γ_{ki} are given by antilogit(β_k)

Because $\gamma_{1i} \leqslant \gamma_{2i} \leqslant \gamma_{3i}$ it follows that $\beta_0 \leqslant \beta_1 \leqslant \beta 2$

Adding covariates to the model

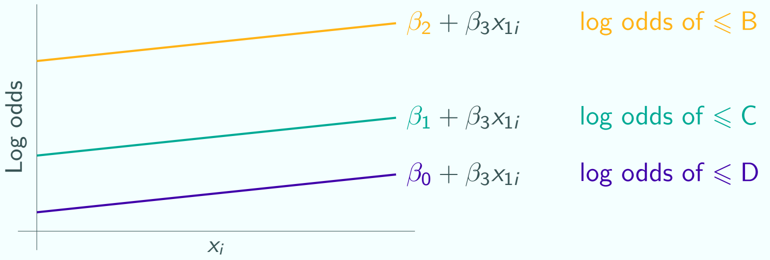
$$\log(\gamma_{1i}/(1-\gamma_{1i})) = \beta_0 + h_i \qquad \text{log odds of} \leqslant \text{D}$$
$$\log(\gamma_{2i}/(1-\gamma_{2i})) = \beta_1 + h_i \qquad \text{log odds of} \leqslant \text{C}$$
$$\log(\gamma_{3i}/(1-\gamma_{3i})) = \beta_2 + h_i \qquad \text{log odds of} \leqslant \text{B}$$
$$h_i = \beta_3 x_{1i} + \ldots$$

Note that the covariates and their coefficients, which comprise the term h_i, are the same for each of the response threshold categories



This means that the log odds ratios and odds ratios for threshold category membership are independent of the predictor variables. That is...

Proportional odds



The log odds ratio

$$(\beta_2 + \beta_3 x_{1i}) - (\beta_1 + \beta_3 x_{1i}) = \log\left(\frac{\text{odds of} \leqslant \text{B}}{\text{odds of} \leqslant \text{C}}\right)$$

is constant for all x_{1i}. Similarly, the log odds ratios

$$\log\left(\frac{\text{odds of} \leqslant \text{B}}{\text{odds of} \leqslant \text{D}}\right) \quad \text{and} \quad \log\left(\frac{\text{odds of} \leqslant \text{C}}{\text{odds of} \leqslant \text{D}}\right)$$

are also constant with respect to x_{1i}

Testing the assumption of proportional odds

We can test the assumption that the odds ratios for each pair of response categories are constant across all values of the predictor variables by fitting the model

$$\begin{aligned}
\log(\gamma_{1i}/(1-\gamma_{1i})) &= \beta_0 + \beta_3 x_{1i} && \text{log odds of} \leqslant \text{D} \\
\log(\gamma_{2i}/(1-\gamma_{2i})) &= \beta_1 + \beta_3 x_{1i} && \text{log odds of} \leqslant \text{C} \\
\log(\gamma_{3i}/(1-\gamma_{3i})) &= \beta_2 + \beta_3 x_{1i} && \text{log odds of} \leqslant \text{B}
\end{aligned}$$

which allows each response category to have a different slope.

Now if our assumptions are correct, β_3, β_4 and β_5 will be very similar. We can formally test the hypothesis that $\beta_3 = \beta_4 = \beta_5$ using a Wald test (in the Intervals and tests window of MLwiN)

If the proportional odds assumption is valid we have a more parsimonious analysis because we fit a single common coefficient instead of t-1 coefficients.

Understanding the model

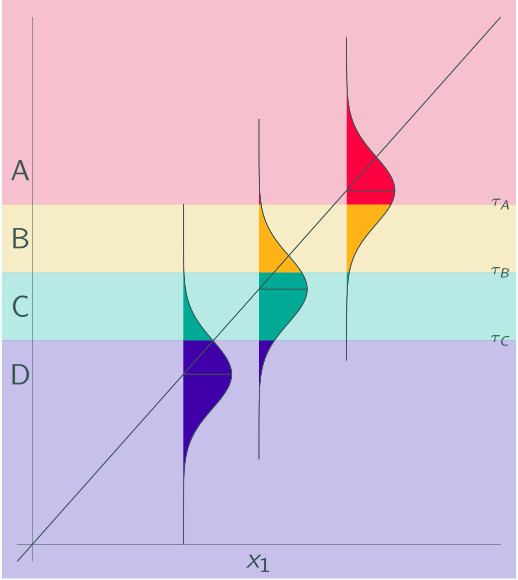
Our model has two features which distinguish it from our model for unordered categorical data:

- we have used cumulative probabilities (the γ_{ki}) instead of the probability for each category (the π_{ki})
- we have constrained the coefficients of the explanatory variables to be the same across response categories (our proportional odds assumption)

We have incorporated into the model the information that our categorical variable is ordered by using these two features together.

It is easiest to understand this by considering the latent variable representation. Our model with a separate intercept for each response category and a common slope across response categories corresponds to a single latent variable with $t - 1$ thresholds or cut points.

Latent variable representation



$$y_i^* = \beta_3^* x_{1i} + e_i^*$$

$$e_i^* \sim \text{logistic with variance } 3.29$$

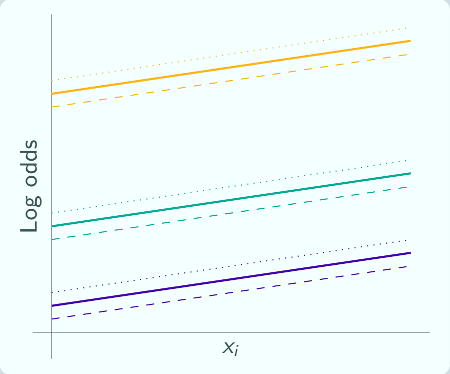
$$y_i = \begin{cases} \text{A} & y_i^* \geqslant \tau_A \\ \text{B} & y_i^* \geqslant \tau_B \\ \text{C} & y_i^* \geqslant \tau_C \\ \text{D} & y_i^* < \tau_C \end{cases}$$

Diagram adapted from notes by Anders Skrondal

Multilevel ordered multinomial models

$$\log(\gamma_{1i}/(1-\gamma_{1i})) = \beta_0 + h_i \qquad \text{log odds of} \leqslant \text{D}$$
$$\log(\gamma_{2i}/(1-\gamma_{2i})) = \beta_1 + h_i \qquad \text{log odds of} \leqslant \text{C}$$
$$\log(\gamma_{3i}/(1-\gamma_{3i})) = \beta_2 + h_i \qquad \text{log odds of} \leqslant \text{B}$$

$$h_i = \beta_3 x_{1i} + u_{0j}$$



u_{0j} is a random effect for school j, which shifts all the log odds lines equally for all students in school j.

Odds ratios for category membership are unaffected by the value of u_{0j}

—— $u_{0j} = 0$ ····· a positive u_{0j}

– – a negative u_{0j}

Higher level variances

$$u_{0j} \sim \mathrm{N}(0, \sigma^2_{u0})$$

The greater σ^2_{u0} the greater the variability in the school level departures for the response threshold probabilities

An example: psychiatric data

The data from this example is taken from Don Hedeker's web site

http://tigger.uic.edu/∼hedeker/ml.html

Data is from a psychiatric clinical trial.
Data on 437 schizophrenia patients (108 in placebo group, 329 in drug treatment group)
Longitudinal design, with measurements at weeks 0, 1, 3 and 6

Response is severity of illness scored as

1	normal or borderline mentally ill
2	mildly or moderately mentally ill
3	markedly ill
4	severely or among the most extremely ill

For more details of the study see Hedeker & Gibbons (1997)

Single level model

Model

$\text{logit}(\gamma_{1i}) = \beta_0 + h_i$	log odds of $\leqslant$ normal or borderline
$\text{logit}(\gamma_{2i}) = \beta_1 + h_i$	log odds of $\leqslant$ mild or moderate
$\text{logit}(\gamma_{3i}) = \beta_2 + h_i$	log odds of $\leqslant$ marked
$h_i = \beta_3 \text{week}_i$	change in log odds per week

Results

$\beta_0 = -3.296$ (0.114)	log odds of $\leqslant$ normal or borderline (week 0)
$\beta_1 = -1.327$ (0.077)	log odds of $\leqslant$ mild or moderate (week 0)
$\beta_2 = -0.076$ (0.069)	log odds of $\leqslant$ marked (week 0)
$\beta_3 = 0.423$ (0.023)	change in log odds per week

To interpret the results it helps to look at graphs

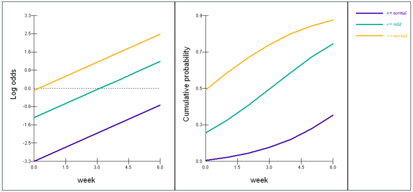
Graphs of results

$\beta_0 = -3.296$ (0.114) log odds of $\leqslant$ normal or borderline (week 0)
$\beta_1 = -1.327$ (0.077) log odds of $\leqslant$ mild or moderate (week 0)
$\beta_2 = -0.076$ (0.069) log odds of $\leqslant$ marked (week 0)

$\beta_3 = 0.423$ (0.023) change in log odds per week

At week 0, log odds of $\leqslant$ marked are -0.076, odds of 0.92, P($\leqslant$ marked) $= 0.48$, P(extreme) $= 1 - 0.48 = 0.52$

At week 6, log odds of $\leqslant$ marked are $-0.076 + 0.423 \times 6 = 2.495$, odds of 12, P($\leqslant$ marked) $= 0.92$, P(extreme) $= 1 - 0.92 = 0.08$



No matter which threshold we choose (normal, mild or marked), as the trial progresses fewer people are falling on the higher (more ill) side of the threshold. Later we will assess whether this improvement is stronger in the treatment than placebo group.

Testing the proportional odds assumption

Proportional odds

$\text{logit}(\gamma_{1i}) = \beta_0 + h_i$
$\text{logit}(\gamma_{2i}) = \beta_1 + h_i$
$\text{logit}(\gamma_{3i}) = \beta_2 + h_i$

$h_i = \beta_3 \text{week}_i$

$\beta_0 = -3.296$ (0.114)
$\beta_1 = -1.327$ (0.077)
$\beta_2 = -0.076$ (0.069)

$\beta_3 = 0.423$ (0.023)

Non-proportional odds

$\text{logit}(\gamma_{1i}) = \beta_0 + \beta_3 \text{week}_i$
$\text{logit}(\gamma_{2i}) = \beta_1 + \beta_4 \text{week}_i$
$\text{logit}(\gamma_{3i}) = \beta_2 + \beta_5 \text{week}_i$

$\beta_0 = -3.296$ (0.114) $\beta_3 = -0.481$ (0.038)
$\beta_1 = -1.327$ (0.077) $\beta_4 = 0.418$ (0.026)
$\beta_2 = -0.076$ (0.069) $\beta_5 = 0.384$ (0.031)

The proportional odds assumption that $\beta_3 = \beta_4 = \beta_5$ is reasonable

Multilevel random intercept vs. single level

Single level

$$\text{logit}(\gamma_{1i}) = \beta_0^{(SL)} + h_i \qquad \beta_0^{(SL)} = -3.296 \ (0.114)$$

$$\text{logit}(\gamma_{2i}) = \beta_1^{(SL)} + h_i \qquad \beta_1^{(SL)} = -1.327 \ (0.077)$$

$$\text{logit}(\gamma_{3i}) = \beta_2^{(SL)} + h_i \qquad \beta_2^{(SL)} = -0.076 \ (0.069)$$

$$h_i = \beta_3^{(SL)} \text{week}_i \qquad \beta_3^{(SL)} = 0.423 \ (0.023)$$

Multilevel

$$\text{logit}(\gamma_{1ij}) = \beta_0^{(RI)} + h_{ij} \qquad \beta_0^{(RI)} = -5.004 \ (0.185)$$

$$\text{logit}(\gamma_{2ij}) = \beta_1^{(RI)} + h_{ij} \qquad \beta_1^{(RI)} = -2.067 \ (0.133)$$

$$\text{logit}(\gamma_{3ij}) = \beta_2^{(RI)} + h_{ij} \qquad \beta_2^{(RI)} = -0.021 \ (0.123)$$

$$h_{ij} = \beta_3^{(RI)} \text{week}_{ij} + u_j \qquad \beta_3^{(RI)} = 0.623 \ (0.028)$$

$$u_j \sim \text{N}(0, \sigma_u^2) \qquad \sigma_u^2 = 3.625 \ (0.325)$$

We have substantial between individual variation: $\sigma_u^2 = 3.625$; this corresponds to an ICC of

$$\frac{3.625}{3.29 + 3.625} = 52\%$$

Comparing the coefficients

Recall that

$$\frac{\beta^{(\text{RI})}}{\beta^{(\text{SL})}} \approx \sqrt{\frac{3.29 + \sigma_u^2}{\sigma_u^2}}$$

which in this case is

$$\sqrt{\frac{3.29 + 3.625}{3.625}} = 1.450$$

In our example we have

$$\beta_0^{(\text{RI})}/\beta_0^{(\text{SL})} = 1.56$$
$$\beta_1^{(\text{RI})}/\beta_1^{(\text{SL})} = 1.51$$
$$\beta_2^{(\text{RI})}/\beta_2^{(\text{SL})} = 0.27$$
$$\beta_3^{(\text{RI})}/\beta_3^{(\text{SL})} = 1.48$$

Thus with the exception of $\beta_2^{(\text{RI})}/\beta_2^{(\text{SL})}$ the pattern is as expected. We note that compared to their standard errors $\beta_2^{(\text{RI})}$ and $\beta_2^{(\text{SL})}$ are small and are thus indistinguishable from 0.

As in the binomial case the RI coefficients are cluster specific estimates and the SL coefficients are population average estimates.

Deriving population average predictions

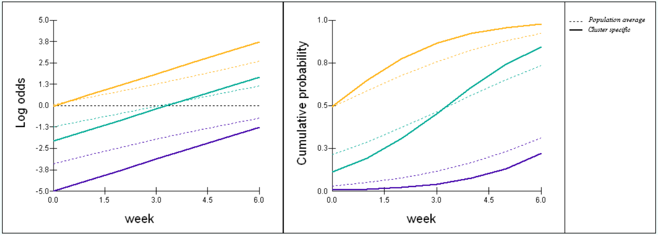
Although a single level model returns population average (PA) estimates, it still retains disadvantages of a single level model in that it ignores clustering and so gives misestimated precisions (SEs are too small). We can derive PA predictions from a cluster specific model by averaging over simulated values of u_j. This is method 3 described in the binary response handouts.

We compare predictions for our three thresholds at week $= 0$, for single level (SL), cluster specific (CS), and PA derived from the multilevel model (PA):

	β_0	β_1	β_2
CS	-5.00 (0.19)	-2.07 (0.13)	-0.02 (0.12)
PA	-3.59 (0.16)	-1.35 (0.09)	-0.03 (0.08)
SL	-3.26 (0.11)	-1.33 (0.08)	-0.07 (0.07)

We see the SEs for SL are all lower than PA. In this case the differences are not great but in other cases they may be.

Graphs of PA and CS predictions



$\beta_3^{(CS)}$ is the effect of a 1 unit change in x (here week) on the log odds of being in each cumulative category holding constant all cluster (person) specific unobservables. The contrast is between two occasions in the same individual

$\beta_3^{(PA)}$ is the effect of a 1 unit change in x (week) on the log odds of being in each cumulative category in the study population, i.e. averaging over all cluster (person) specific unobservables

Drug vs. placebo effects

Model

$\text{logit}(\gamma_{1ij}) = \beta_0^{(RI)} + h_{ij}$

$\text{logit}(\gamma_{2ij}) = \beta_1^{(RI)} + h_{ij}$

$\text{logit}(\gamma_{3ij}) = \beta_2^{(RI)} + h_{ij}$

$h_{ij} = \beta_3^{(RI)}\text{week}_{ij} + \beta_4^{(RI)}\text{drug}_{ij} + \beta_5^{(RI)}\text{drug.week}_{ij} + u_j$

$u_j \sim \text{N}(0, \sigma_u^2)$

β_4 allows intercepts to be different for drug and placebo

β_5 allows week slopes to be different for drug and placebo

Results

$\beta_0^{(RI)} = -5.541\ (0.288)$

$\beta_1^{(RI)} = -2.477\ (0.256)$

$\beta_2^{(RI)} = -0.375\ (0.248)$

$\beta_3^{(RI)} = 0.296\ (0.051)$

$\beta_4^{(RI)} = 0.517\ (0.282)$

$\beta_5^{(RI)} = 0.435\ (0.059)$

$\sigma_u^2 = 3.522\ (0.319)$

So difference between placebo and drug as a function of time is

$$\log\left(\frac{\text{odds}(\leqslant \text{normal}^{(\text{drug})})}{\text{odds}(\leqslant \text{normal}^{(\text{placebo})})}\right) = \log\left(\frac{\text{odds}(\leqslant \text{mild}^{(\text{drug})})}{\text{odds}(\leqslant \text{mild}^{(\text{placebo})})}\right) = \log\left(\frac{\text{odds}(\leqslant \text{marked}^{(\text{drug})})}{\text{odds}(\leqslant \text{marked}^{(\text{placebo})})}\right)$$

$$= \beta_4\text{drug}_{ij} + \beta_5\text{week.drug}_{ij}$$

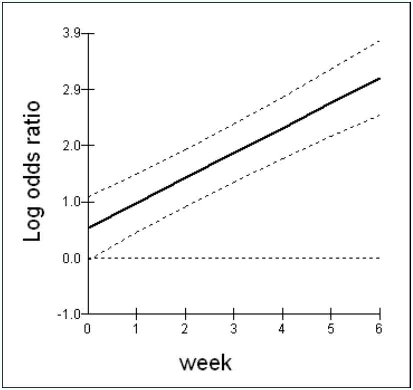
Since β_4 and β_5 are both positive, a positive drug effect is present at week 0 and becomes stronger over the trial period.

This means that for any threshold fewer people are falling on the higher (more ill) side in the drug than the placebo group.

Graph of drug vs. placebo log odds ratio

$$\log\left(\frac{\text{odds}(\leqslant \text{normal}^{(\text{drug})})}{\text{odds}(\leqslant \text{normal}^{(\text{placebo})})}\right) = \log\left(\frac{\text{odds}(\leqslant \text{mild}^{(\text{drug})})}{\text{odds}(\leqslant \text{mild}^{(\text{placebo})})}\right) = \log\left(\frac{\text{odds}(\leqslant \text{marked}^{(\text{drug})})}{\text{odds}(\leqslant \text{marked}^{(\text{placebo})})}\right)$$

$$= \beta_4 \text{drug}_{ij} + \beta_5 \text{week.drug}_{ij}$$



Graph plots the difference between the drug and the placebo groups (the log odds ratio) against week

Improvement is present at week 0 and increases with time

Graph of placebo vs. drug effects