Unordered multinomial response models

Multilevel multinomial models

We have seen how logistic (and other) models handle the situation where we have a binary response (two response categories e.g. alive/dead or pass/fail).

Where we have a response variable with more than two categories we use multinomial models.

Two types of multinomial response:

- Unordered
  - e.g. voting preference (Labour, Tory, Lib Dem, Other)
  - cause of death
- Ordered
  - e.g. attitude scales (strongly disagree,. . . ,strongly agree)
  - exam grades

First we deal wtih unordered multinomial responses

Example: British Election Study

We will be using data from the British Election Study to illuminate our explanation and provide some examples. First we provide some brief information on UK politics to put this in context.

The UK is divided into areas called constituencies, with each constituency electing one MP to a seat.

There are 3 main parties, Conservative (also known as Tory), Labour and Liberal Democrat (often shortened to Lib Dem). In recent years it has always been the Conservatives or Labour who have won the greatest number of seats, with the Liberal Democrats coming third.

There are other parties but to simplify the data for this example we have discarded all individuals who did not vote for one of these three parties.

Extending a binary to a multinomial model

Take a binary variable (y_i) which is 1 if an individual votes Tory and 0 otherwise.

The underlying probability of individual i voting Tory is π_i.

We model the log odds of voting Tory as a function of explanatory variables

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \ldots \tag{1}$$

Let's change notation:
the probability of individual i voting Tory, π_i, becomes π_{1i} and
the probability of individual i not voting Tory, $1-\pi_i$, becomes π_{2i}.

We can now write (1) as

$$\log\left(\frac{\pi_{1i}}{\pi_{2i}}\right) = \beta_0 + \beta_1 x_{1i} + \ldots$$

Moving to more than two response categories

Suppose now that y_i can take three values, $\{1 = \text{vote Tory}, 2 = \text{vote Labour}, 3 = \text{vote Lib Dem}\}$. Now

π_{1i} is the probability that individual i votes Tory
π_{2i} is the probability that individual i votes Labour
π_{3i} is the probability that individual i votes Lib Dem

Now we must choose a reference category, say voting Lib Dem, and model the log odds of all remaining categories against the reference category. Therefore with m categories we need $m - 1$ equations to model this set of log odds ratios. In our case

$$\log\left(\frac{\pi_{1i}}{\pi_{3i}}\right) = \beta_0 + \beta_1 x_{1i} + \ldots$$
$$\log\left(\frac{\pi_{2i}}{\pi_{3i}}\right) = \beta_2 + \beta_3 x_{1i} + \ldots$$

Note that in the 2 category example on the previous slide our reference category was not voting Tory.

Notation

The MLwiN software uses the notation

$$\begin{aligned}\log\left(\pi_{1i}/\pi_{3i}\right) =& \beta_0 + \beta_1 x_{1i} + \ldots \\ \log\left(\pi_{2i}/\pi_{3i}\right) =& \beta_2 + \beta_3 x_{1i} + \ldots \\ &\vdots\end{aligned}$$

Often in papers you will see the more succinct notational form

$$\log\left(\frac{\pi_i^{(s)}}{\pi_i^{(t)}}\right) = \beta_0^{(s)} + \beta_1^{(s)} x_{1i} + \ldots, \quad s = 1, \ldots, t-1, t+1, \ldots, m$$

Which becomes

For $s = 1$

$$\log\left(\pi_i^{(1)}/\pi_i^{(3)}\right) = \beta_0^{(1)} + \beta_1^{(1)} x_{1i} + \ldots$$

For $s = 2$

$$\log\left(\pi_i^{(2)}/\pi_i^{(3)}\right) = \beta_0^{(2)} + \beta_1^{(2)} x_{1i} + \ldots$$

Interpretation (odds ratios)

We can interpret as with logistic regression. In the political example, $\{1 = \text{vote Tory}, 2 = \text{vote Labour}, 3 = \text{vote Lib Dem}\}$.

$$\log\left(\pi_{1i}/\pi_{3i}\right) = \beta_0 + \beta_1 x_{1i} + \dots$$
$$\log\left(\pi_{2i}/\pi_{3i}\right) = \beta_2 + \beta_3 x_{1i} + \dots$$

β_1 is the change in the log odds of voting Tory as opposed to Lib Dem for a 1 unit increase in x_1.

β_3 is the change in the log odds of voting Labour as opposed to Lib Dem for a 1 unit increase in x_1.

and $\exp\left(\beta_k\right)$ gives the change in odds ratios for a 1 unit increase in x_1 $(k = 1, 3)$

Example: the British Election Study

Response is **voted** (1=Tory, 2=Labour, 3=Lib Dem) in 1997.
Sample of 1559 voters from 186 constituencies

Let's look at how the log odds of voting Tory vs. Lib Dem and Labour vs. Lib Dem change as functions of age (centred around 49 years).

Model

$$\log(\pi_{1i}/\pi_{3i}) = \beta_0 + \beta_1(\text{age} - 49)_i$$

$$\log(\pi_{2i}/\pi_{3i}) = \beta_2 + \beta_3(\text{age} - 49)_i$$

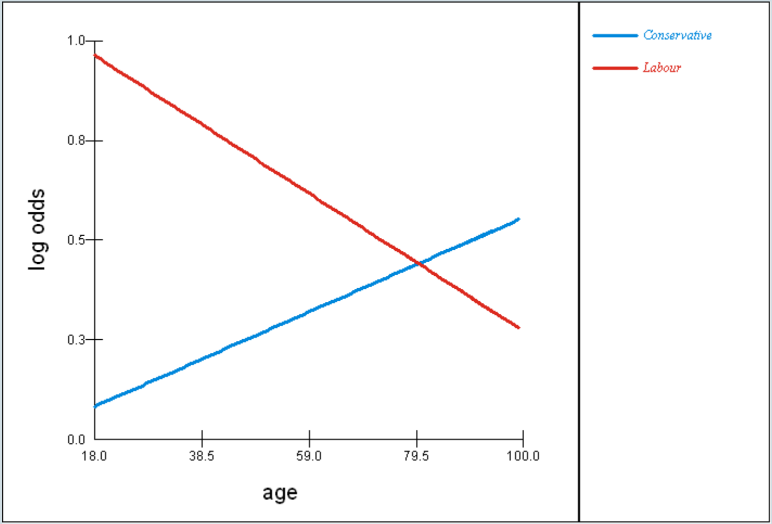
Results

$\beta_0 =$	0.262	(0.056)
$\beta_1 =$	0.006	(0.003)
$\beta_2 =$	0.704	(0.051)
$\beta_3 =$	-0.008	(0.003)

So log(odds) of voting Tory vs. Lib Dem increase by 0.006 for every additional year

So log(odds) of voting Labour vs. Lib Dem decrease by 0.008 for every additional year

Graph of log odds as a function of age



Interpretation (probabilities)

Probability of voting Tory for individual i

$$\pi_{1i} = \frac{e^{(\beta_0+\beta_1 x_{1i})}}{1+\left(e^{(\beta_0+\beta_1 x_{1i})}+e^{(\beta_2+\beta_3 x_{1i})}\right)}$$

Probability of voting Labour for individual i

$$\pi_{1i} = \frac{e^{(\beta_2+\beta_3 x_{1i})}}{1+\left(e^{(\beta_0+\beta_1 x_{1i})}+e^{(\beta_2+\beta_3 x_{1i})}\right)}$$

Probability of voting Liberal Democrat for individual i

$$\pi_{3i} = 1 = \pi_{2i} - \pi_{1i}$$

$$\log\left(\pi_{1i}/\pi_{3i}\right) = \beta_0 + \beta_1 x_{1i} + \ldots$$
$$\log\left(\pi_{2i}/\pi_{3i}\right) = \beta_2 + \beta_3 x_{1i} + \ldots$$

Or in general notation:

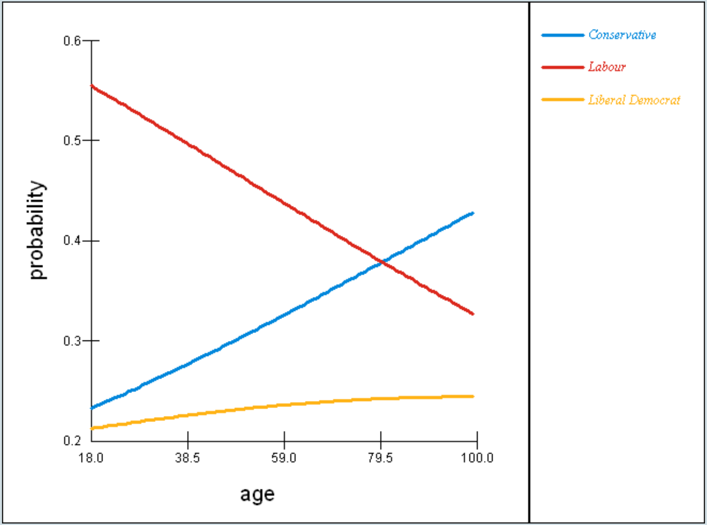
$$\pi_i^{(s)} = \frac{e^{\left(\beta_0^{(s)}+\beta_1^{(s)} x_{1i}\right)}}{1+\sum_{k=1}^{t-1} e^{\left(\beta_0^{(k)}+\beta_1^{(k)} x_{1i}\right)}}$$

$$\pi_i^{(t)} = 1 - \sum_{k=1}^{t-1} \pi_i^{(k)}$$

Note that MLwiN will do this transformation for you

Graph of probabilities as a function of age

Looking at probabilities of voting for each party as a function of age gives



Multilevel multinomial models

Suppose the individuals in the voting example are clustered into constituencies and we wish to include constituency effects in our model. We include intercept level residuals for each log odds equation in our model

$$\log\left(\frac{\pi_{1ij}}{\pi_{3ij}}\right) = \beta_0 + \beta_1 x_{1ij} + u_{0j}$$
$$\log\left(\frac{\pi_{2ij}}{\pi_{3ij}}\right) = \beta_2 + \beta_3 x_{1ij} + u_{2j}$$

u_{0j} is the effect of the constituency j on the log odds of voting Tory as opposed to Lib Dem. So if u_{0j} is 1 the log odds of voting Tory as opposed to Lib Dem increase by 1 compared to a constituency where $u_{0j} = 0$ (the average constituency).

Likewise u_{2j} is the effect of constituency j on the log odds of voting Labour as opposed to Lib Dem.

Variance of level 2 random effects

$$\log(\pi_{1ij}/\pi_{3ij}) = \beta_0 + u_{0j} + \beta_1 x_{1ij}$$
$$\log(\pi_{2ij}/\pi_{3ij}) = \beta_2 + u_{2j} + \beta_3 x_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{2j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) \quad \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u02} & \sigma^2_{u2} \end{bmatrix}$$

σ^2_{u0} is the between constituency variance of the (vote Tory):(vote Lib Dem) log odds ratio

σ^2_{u2} is the between constituency variance of the (vote Labour):(vote Lib Dem) log odds ratio

σ_{u02} is the covariance between Tory and Labour constituency level effects. A negative covariance means that in constituencies where Labour do well compared to Lib Dems, Tories tend to do badly compared to Lib Dems and vice versa. A positive covariance means that in constituencies where Labour do well compared to Lib Dems, Tories also tend to do well.

Multilevel model of voting behaviour

Model

$$\log\left(\pi_{1ij}/\pi_{3ij}\right) = \beta_0 + u_{0j} + \beta_1 x_{1ij}$$
$$\log\left(\pi_{2ij}/\pi_{3ij}\right) = \beta_2 + u_{2j} + \beta_3 x_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{2j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) \quad \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u02} & \sigma^2_{u2} \end{bmatrix}$$

Results

$$\beta_0 = 0.580\ (0.093) \qquad \beta_1 = 0.005\ (0.003)$$
$$\beta_2 = 1.112\ (0.126) \qquad \beta_3 = \text{-}0.008\ (0.003)$$

$$\sigma^2_{u0} = 0.836\ (0.156)$$
$$\sigma_{u02} = 0.778\ (0.157) \qquad \sigma^2_{u2} = 1.755\ (0.288)$$

For $(\mathbf{age} - 49) = 0$, we have log(Tory:Lib Dem)=0.58, however between constituency variation is 0.836, so 95% range for this log odds is $0.58 \pm (\sqrt{0.836} \times 1.96) = (-1.22, 2.37)$ which corresponds to an odds range (by exponentiating) of $(0.3, 10.67)$

And for log(Labour:Lib Dem) at $(\mathbf{age} - 49) = 0$ we have $1.112 \pm (\sqrt{1.755} \times 1.96) = (-1.53, 3.8)$ which corresponds to an odds range of $(0.21, 44.7)$

So there is a great deal of between constituency variation

Between constituency variation in odds ratios

Based on shrunken constituency-level residuals