Module 10: Single-level and Multilevel Models for Nominal Responses Concepts

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites

- Modules 5, 6 and 7

Contents

Introduction

In Module 6 we saw how multiple regression models can be generalised to handle binary responses, and in Module 7 these models were extended for the analysis of binary data with a two-level hierarchical structure. Module 9 considered single-level and multilevel models for categorical responses with more than two categories, where the numeric codes assigned to categories imply an ordering. Examples of ordinal variables include Likert scale items where respondents are asked to indicate their strength of agreement with a statement, and exam grades. In this module we look at models for nominal (or unordered) categorical responses, where the numeric codes assigned to categories are simply labels and serve only to distinguish between categories (see C1.3.8 for a classification scheme for variables).

Examples of nominal responses include political party preferences (e.g. Labour, Conservative, Liberal Democrat, other in the UK), mode of transport and brand preference. Aggregating such variables to a binary response not only wastes potentially important information, but may result in misleading conclusions if predictors have different effects for different categories. For example, the choice between driving to work or using public transport may depend on the availability of free car-parking, while the choice between driving and walking is likely to depend strongly on the distance between home and work. Fortunately, multinomial regression methods have been developed that allow such distinctions between categories of a nominal response, and these have been extended to handle multilevel data structures.

In this module, we begin by describing multinomial logit models for single-level nominal responses. As the coefficients of multinomial models can be difficult to interpret, we pay particular attention to calculating predicted response probabilities to aid interpretation. We then consider multilevel multinomial logit models for two-level structures. We shall see that models for nominal responses are direct extensions of the models for binary responses described in Modules 6 and 7. The same generalisations of the basic multilevel model – for example, random slopes and contextual effects – are possible for nominal responses. We end with a discussion of conditional logit models which are used when the effects of characteristics of the different response alternatives are of interest. For example, the choice between driving and using public transport may depend on their relative costs to an individual, according to where the individual lives and the travel time for each option.

Introduction to the Example Dataset

Our main example dataset for this module comes from the 2008 National Travel Survey (NTS)[1]. The 2008 NTS is one of a series of annual cross-sectional household surveys, designed to provide regular data on personal travel in Great Britain. We will

[1]Department for Transport, *National Travel Survey, 2002-2008* [computer file]. *5th edition*. Colchester, Essex: UK Data Archive [distributor], June 2010. SN: 5340. The data are free to download after registration from http://www.data-archive.ac.uk/

use data from personal face-to-face interviews (the survey also includes travel diaries), and restrict the sample to household members who were aged 16 or older.

The response variable for the analysis is the mode of transport used to travel to work, which has been grouped into three categories:

Code	Label
1	Car /motorcycle
2	Bicycle or walking
3	Public transport

We consider three individual-level characteristics as explanatory variables (all categorical):

- *Gender*
- *Age* (16-19, 20-29, 30-39, 40-49, 50-59 years)
- *Employed part-time* (versus full-time)

The survey is based on a stratified two-stage random probability sample of private households in Great Britain. The primary sampling units (PSUs) at the first stage of sampling are postcode sectors. At the second stage, a sample of households was drawn from the selected PSUs.[2] We will ignore the household level in this module, and treat the data as a two-level structure with individuals at level 1 and PSUs at level 2.

We consider one PSU-level explanatory variable:

- *Type of area* (London boroughs, metropolitan built-up areas, other urban areas over 250,000 population, urban 25,000-250,000 population, urban 10,000-25,000 population, urban 3000-10,000 population, rural)

After excluding a small number of individuals with missing data on at least one of the variables, the analysis file contains 8,512 individuals nested within 683 PSUs.

Note that the same dataset was analysed in Module 9 for an ordinal response (frequency of walking). In this module, the analysis sample has been restricted to employed respondents aged less than 60 because means of travel to work was only asked of this group.

[2] See Anderson, Christophersen, Pickering, Southwood and Tipping (2009) *National Travel Survey 2008 Technical Report*. Prepared for the Department of Transport. This report and other documentation can be downloaded with the dataset from http://www.data-archive.ac.uk/

C10.1 Multinomial Logit Model for Single-Level Data

In this module we focus on multinomial logit models, the most common approach for the analysis of nominal responses. Another model for nominal responses, the conditional logit model, is discussed in the final section C10.5.

C10.1.1 The multinomial logit model

Consider response variable y which takes values 1, 2, . . ., C.

We define *response probabilities* for each category k as

$$\Pr(y = k) = \pi_k$$

where $\pi_1 + \pi_2 + \dots \pi_C = 1$.

As for binary and ordered logit models, one of the response categories is chosen as the reference. We then model the log-odds of being in one of the remaining categories rather than the reference category. If we take the first category as the reference, for example, we model the log-odds of being in category k $(k = 2, \dots, C)$ rather than category 1.

We begin be considering models for a single-level nominal response. Suppose we have one continuous or binary explanatory variable x, then the model for the contrast between response category k and the reference category 1 for individual i $(i = 1, \dots, n)$ can be written

$$\log\left(\frac{\pi_{ki}}{\pi_{1i}}\right) = \beta_{0k} + \beta_{1k}x_i, \quad k = 2, \dots, C \qquad (10.1)$$

Equation (10.1) consists of $C - 1$ contrasts or sub-equations, one for each category apart from the reference, where β_{0k} is the intercept and β_{1k} the effect of x for the contrast of category k versus category 1.

Before discussing interpretation of the multinomial logit model, we note that the binary logit described in Module 6 is a special case of (10.1). To see this, suppose that the response y_i is binary but coded 1 and 2 (rather than the usual 0 and 1). Taking the first category as the reference (now coded 1 rather than 0) equation (10.1) reduces to a single contrast:

$$\log\left(\frac{\pi_{2i}}{\pi_{1i}}\right) = \log\left(\frac{\pi_{2i}}{1-\pi_{2i}}\right) = \beta_{02} + \beta_{12}x_i,$$

where π_{2i} is the binary response probability.

Remarks

- The multinomial logit model given by (10.1) has the same predictor x in each equation. This restriction can be relaxed to allow a predictor to affect a subset of contrasts. In some software packages, it is possible to directly specify the contrast(s) for which a particular predictor should be included. In other packages, a predictor is removed from a contrast by constraining its coefficient to equal zero.

- The equations in (10.1) are estimated simultaneously, but an approximation to the multinomial logit model is obtained by estimating a series of binary logit models on subsets of the data. For example, the contrast of category 2 versus 1 may be approximated by selecting respondents with $y = 1$ or $y = 2$ and estimating a simple logit model for a new binary response distinguishing these two categories (coded 1 when $y = 2$ and 0 when $y = 1$).[3] However, this approach does not extend to the multilevel case where we will typically wish to allow for correlations between random effects for the different contrasts.

C10.1.2 Interpretation of coefficients and predicted probabilities

The intercept β_{0k} for contrast k is the log of the probability of being in category k relative to the probability of being in category 1 when $x = 0$, and its exponent $\exp(\beta_{0k})$ is the ratio of the probability of being in category k to the probability of being in category 1. The left-hand side of equation (10.1) is commonly referred to as the log-odds of being in category k rather than category 1, and we will refer to it as such as a shorthand even though we are really modelling the ratio of two probabilities.[4] However, it is incorrect to refer to $\log(\pi_{ki}/\pi_{1i})$ as simply the odds of being in category k (as we would for a binary response); if we do not explicitly refer to the reference category, the odds are $\log(\pi_{ki}/(1 - \pi_{ki}))$. This is an important difference between the binary logit model and the multinomial logit model for a multi-category response which has implications for the interpretation of coefficients from a multinomial model (as discussed below).

The coefficient of x for contrast k, β_{1k}, is the effect of a 1-unit increase in x on the log-odds of being in category k rather than category 1. As in the binary response case, we can interpret $\exp(\beta_{1k})$ as an odds ratio, comparing the odds of being in category k rather than category 1 for two randomly selected individuals whose x values differ by 1 unit.

As you can tell from the above, interpretation of the coefficients of a multinomial logit model (and the associated odds ratios) is rather awkward! In a binary logit model, the coefficients are the effects of predictors on being in one of the response categories rather than the other, but in the multinomial generalisation we could have many pairwise contrasts to consider. It would be much easier to interpret the effects of a predictor on each response category, rather than on a contrast between two categories. Fortunately, we can calculate predicted response probabilities from the estimated coefficients for whatever values of x we choose.

---

[3] This approximation was proposed by Begg, C.B. and Gray, R. (1984) "Calculation of polychotomous logistic regression parameters using individualized regressions". *Biometrika* **71**, 11-18.

[4] Exponentiated coefficients from a multinomial logit model are more accurately described as *relative risk ratios*, but this terminology is less commonly used than *odds ratio*.

Equation (10.1) can be rearranged to give the following expressions for the response probabilities:

$$\pi_{ki} = \frac{\exp(\beta_{0k} + \beta_{1k}x_i)}{1 + \sum_{l=2}^{C}\exp(\beta_{0l} + \beta_{1l}x_i)}, \quad k = 2, \dots, C \qquad (10.2)$$

with the probability for the reference category calculated by subtraction:

$$\pi_{1i} = 1 - \sum_{l=2}^{C}\pi_{li} = \frac{1}{1 + \sum_{l=2}^{C}\exp(\beta_{0l} + \beta_{1l}x_i)} \qquad (10.3)$$

Predicted response probabilities are calculated by 'plugging in' the estimates for β_{0k} and β_{1k} from the fitted model and applying (10.2) and (10.3) for selected values of x (some examples will be given in C10.2).

Retherford and Choe (1993, p.153)[5] note that coefficients (or odds ratios) are not only difficult to interpret, but may even be misleading because the sign of β_{1k} may not reflect the direction of the effect of x on either of the response probabilities being compared (π_k and π_1). To illustrate the problem, suppose we fit a multinomial logit model to a three-category response taking category 1 as the reference, and including a single binary predictor x. We consider two scenarios where the coefficient of x for the contrast of response categories 2 and 1, β_{12} in equation (10.2), does not reflect the effect of x on the response probabilities for these categories.

In Table 10.1 the probabilities for categories 1 and 2 (π_1 and π_2) are both lower for $x = 1$ than for $x = 0$, so we would say that there is a negative association between being in categories 1 or 2 of the response and x. However, the ratio of π_2 to π_1 is constant across values of x, so that $\exp(\beta_{12}) = 1$ which implies $\beta_{12} = 0$. Interpreting the coefficients of x, we might be tempted to incorrectly conclude that there is no relationship between x and being in response category 2. The correct interpretation of β_{12} is that the probability of being in category 2 *rather than category 1* does not depend on x.

Table 10.1. Scenario where the response probabilities depend on x but the regression coefficient for the contrast of category 2 versus 1 is zero

x	π_1	π_2	π_3	π_2/π_1	π_3/π_1
0	0.2	0.4	0.4	2	2
1	0.1	0.2	0.7	2	7
Ratio for $x = 1$ versus $x = 0$				1	3.5
β_{1k}				0	1.25

5 Retherford, R. D., & Choe, M. K. (1993). *Statistical Models for Causal Analysis*. New York: Wiley.