Module 9: Single-level and Multilevel Models for Ordinal Reponses Concepts

Fiona Steele[1]
Centre for Multilevel Modelling

Pre-requisites

- Modules 5, 6 and 7

Contents

[1] With thanks to Antony Fielding for helpful comments on an earlier draft.

Introduction

In Module 6 we saw how multiple regression models for continuous responses can be generalised to handle binary responses, and in Module 7 these models were further extended for the analysis of binary data with a two-level hierarchical structure. In this module and the next, we consider standard (single-level) and multilevel models for categorical response variables with more than two categories. We begin in this module with models for ordinal variables, where the numeric codes assigned to categories imply some ordering (see C1.3.8 for a classification scheme for variables).

Ordinal responses are especially common in the social sciences. Examples include Likert scale items where respondents are asked to indicate their strength of agreement with a statement from ‘strongly agree’ to ‘strongly disagree’, and educational tests where marks are available as grades rather than percentage scores. While many researchers simply treat ordinal variables as if they were continuous and analyse them using ordinary least squares regression, this is not recommended for two main reasons: (i) the differences between numeric codes assigned to categories have no meaning (only their relative values can be interpreted), and (ii) ordinal variables very often have skewed distributions. Another approach might be to collapse categories of an ordinal variable to obtain a binary variable, and to apply the methods of Modules 6 and 7, but this clearly wastes information. Fortunately, methods have been developed specifically for ordinal variables and these have been extended to handle multilevel data structures.

In this module, we begin by describing two models for single-level ordinal responses: the cumulative logit model and the continuation ratio model. We then consider multilevel cumulative logit models for two-level structures. We shall see that models for ordinal 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 ordinal responses, and the same issues in interpretation arise.

Introduction to the Example Dataset

Our main example dataset for this module comes from the 2008 National Travel Survey (NTS)[2]. 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 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 frequency of walking, which is recorded on a seven-point scale:

[2] 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/

Code	Label
1	Less than once a year or never
2	1-2 times a year
3	More than twice a year but less than monthly
4	1-2 times a month
5	More than twice a month but less than weekly
6	1-2 times a week
7	3+ times a week

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

- *Gender*
- *Age* (16-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+ years)
- *Employment status* (employed full-time, employed part-time, unemployed, retired or permanently sick, student, looking after family/home, or other)

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.[3] 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 16,393 individuals nested within 683 PSUs.

C9.1 Cumulative Logit Model for Single-Level Data

Two widely used approaches for analysing ordinal responses are the *cumulative logit* model and the *continuation ratio* model. The cumulative logit model is described in this section. We focus on logit models in this module, but each model

[3] 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/

has a probit counterpart – called the *ordered probit* model and *sequential probit* model respectively.

C9.1.1 Preliminaries

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

We define *response probabilities* for each category as

$$\Pr(y = k) = \pi_k \qquad (9.1)$$

where $\pi_1 + \pi_2 + \ldots + \pi_C = 1$

In addition we can define *cumulative response probabilities* which reflect the ordering of the values of y. We define by γ_k the cumulative probability of being in category k or lower:

$$\gamma_k = \Pr(y \le k) = \pi_1 + \pi_2 + \cdots \pi_k \qquad (9.2)$$

where $\gamma_1 = \pi_1$ and $\gamma_C = 1$.

Equation (9.2) shows how cumulative probabilities are based on the response probabilities. We can also work backwards to derive response probabilities from cumulative probabilities using

$$\pi_k = \Pr(y = k) = \Pr(y \le k) - \Pr(y \le k-1) = \gamma_k - \gamma_{k-1}$$

Note that in the binary case, cumulative probabilities are redundant because $\Pr(y \le 0) = 1 - \pi$ and $\Pr(y \le 1) = 1$.

C9.1.2 The cumulative logit model

We begin be considering models for a single-level ordinal response. The cumulative logit model, sometimes called the ordered logit model, is based on the cumulative response probabilities defined in (9.2) above. Suppose we have one continuous or binary explanatory variable x, then the model for the cumulative response probability for individual i (i = 1, . . ., n) can be written

$$\log\left(\frac{\Pr(y_i \le k)}{\Pr(y_i > k)}\right) = \text{logit}(\gamma_{ki}) = \alpha_k + \beta x_i, \quad k = 1, \cdots C-1 \qquad (9.3)$$

where α_k are referred to as threshold parameters (analogous to the intercept in a binary response model, and explained below) and β is the coefficient of x.

Before discussing interpretation of (9.3), it is useful to show that the model is a generalisation of the binary logit model described in Module 6. For a binary response y_i coded 0 and 1, with the response probability defined in the usual way as $\Pr(y = 1) = \pi$, (9.3) reduces to a single equation:

$$\log\left(\frac{\Pr(y_i \le 0)}{\Pr(y_i > 0)}\right) = \log\left(\frac{\Pr(y_i = 0)}{\Pr(y_i = 1)}\right) = \log\left(\frac{1-\pi_i}{\pi_i}\right) = -\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta x_i$$

Multiplying each side by -1 gives

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = -\alpha - \beta x_i$$

which is the familiar binary logit model with the signs of the coefficients reversed. (Recall that one of the properties of the logit (and probit) models for binary responses is that switching the coding of *y* simply reverses the signs of the coefficients. This property is a consequence of the symmetry of the logistic and probit transformations.)

Let's now return to the ordinal case (9.3). The coefficient β is interpreted as the effect of a 1-unit change in *x* on the log-odds of being in a lower category of *y* rather than a higher category. In this model, the effect of *x* is assumed to be constant wherever the lower category is fixed. (This is the proportional odds assumption which is discussed below.) Thus $\beta > 0$ implies that higher values of *x* are associated with lower values of *y*. This interpretation is counterintuitive because we are used to interpreting a positive regression coefficient as implying a positive relationship between *x* and *y*. For this reason, (9.3) is sometimes written with a negative sign in front of β so that a positive value for β then implies a positive relationship. In this module, however, we will continue to write down models with positive signs for all coefficients, as in (9.3), but note that software packages differ in whether they output β or $-\beta$.[4]

Model (9.3) also includes parameters α_k which are referred to as *cut-points* or *thresholds*, and can be interpreted as intercept terms. For example, α_2 is the log-odds of being in either category 1 or 2 (rather than 3 or above) for an individual with $x = 0$. While only one intercept is needed for a binary response (C=2), C-1 intercepts are required for an ordinal response with C categories. Furthermore, because we are modelling the logits of the cumulative response probabilities which must necessarily increase with *k*, the intercepts must also be ordered with $\alpha_1 < \alpha_2 < \ldots \alpha_{C-1}$

As in the binary case, we can take exponentials of the regression coefficients and interpret them as odds ratios. For a continuous *x*, $\exp(\beta)$ compares the odds of being in a lower category of *y* for two individuals with *x*-values spaced 1 unit apart (but with the same values on any other explanatory variables included in the model). Equivalently, we can interpret $\exp(\beta)$ as the multiplicative effect of a 1-unit increase in *x* on the odds of being in a lower category of *y*.

[4] If in any doubt about whether a software package outputs β or $-\beta$, it is a good idea to fit a simple model with one explanatory variable *x* and compare the results with a cross-tabulation of *y* and *x*.

For a binary x, $\exp(\beta)$ compares the odds of being in a lower category of *y* for an individual with $x = 1$ with the odds for an individual with $x = 0$. As for the binary logit model, $\exp(\beta) = 1$ implies that there is no relationship between *x* and *y*.

Examples of interpretation of β and $\exp(\beta)$ will be given in the next section. However, it is often easier to assess the magnitude of an effect by calculating predicted probabilities, either response probabilities π_k or cumulative response probabilities γ_k.

Proportional odds

Another important point about model (9.3) is that β does not have a *k* subscript. This is why we did not specify a particular category when interpreting β as the effect of a 1-unit change in *x* on the log-odds of being in a lower category rather than a higher category of *y*: the effect is the same whether we compare category 1 versus categories 2, . . ., C or categories 1 and 2 versus categories 3, . . ., C.

The above property is known as the proportional odds assumption, and model (9.3) is often referred to as a *proportional odds model*. This assumption is commonly made, but it can and should be tested. The proportional odds assumption will be explained in more detail in C9.2.3.

C9.1.3 Example: walking frequency

Correspondence between observed and predicted probabilities

Table shows the observed probabilities for categories of frequency of walking which was measured on a seven-point ordinal scale. Cumulative probabilities are also shown indicating, for example, that 30.3% of respondents walk less than once a month. In this section we will fit cumulative logit models to these data.

Table 9.1. Observed probabilities and cumulative probabilities of frequency of walking, National Travel Survey 2008

k	Frequency of walking	π_k	γ_k	n
1	Less than once a year or never	0.250	0.250	4094
2	1-2 times a year	0.020	0.270	333
3	More than twice a year but less than monthly	0.033	0.303	544
4	1-2 times a month	0.071	0.374	1167
5	More than twice a month but less than weekly	0.045	0.419	738
6	1-2 times a week	0.218	0.638	3580
7	3+ times a week	0.362	1.000	5937
	Total	1.000		16393