Module 7: Multilevel Models for Binary Responses

Practical

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-6

Contents

Most of the sections within this module have online quizzes for you to test your understanding. To find the quizzes:

From within the LEMMA learning environment

- Go down to the section for **Module 7: Multilevel Models for Binary Responses**
- Click " 7.1 Two-Level Random Intercept Model" to open Lesson 7.1
- Click Q 1 to open the first question

Introduction to the Bangladesh Demographic and Health Survey 2004 Dataset

You will be analysing data from the Ba[1a]ngladesh Demographic and Health Survey (BDHS), a nationally representative cross-sectional survey of women of reproductive age (13-49 years).

Our response variable is a binary indicator of whether a woman received antenatal care from a medically-trained provider (a doctor, nurse or midwife) at least once before her most recent live birth. To minimise recall errors, the question was asked only about children born within five years of the survey. For this reason, our analysis sample is restricted to women who had a live birth in the five-year period before the survey. Note that if a woman had more than one live birth during the reference period, we consider only the most recent.

These data were analysed in Module 6 using single-level models. In this module, we consider multilevel models to allow for and to explore between-community variance in antenatal care. The data have a two-level hierarchical structure with 5366 women at level 1, nested within 361 communities at level 2. In rural areas a community corresponds to a village, while an urban community is a neighbourhood based on census definitions.

We consider a range of predictors. At level 1, we consider variables such as a woman's age at the time of the birth and education. Level 2 variables include an indicator of whether the region of residence is classified as urban or rural. We will also derive community-level measures by aggregating woman-level variables, for example the proportion of respondents in the community who are in the top quintile of a wealth index.

[1a]We thank MEASURE DHS for their permission to make these data available for training purposes. Additional information about the 2004 BDHS and other Demographic and Health Surveys, including details of how to register for a DHS Download Account, is available from www.measuredhs.com

The file contains the following variables:

Variable name	Description and codes
comm	Community identifier
womid	Woman identifier
antemed	Received antenatal care at least once from a medically-trained provider, e.g. doctor, nurse or midwife (1=yes, 0=no)
bord	Birth order of child (ranges from 1 to 13)
mage	Mother's age at the child's birth (in years)
urban	Type of region of residence at survey (1=urban, 0=rural)
meduc	Mother's level of education at survey (1=none, 2=primary, 3=secondary or higher)
islam	Mother's religion (1=Islam, 0=other)
wealth	Household wealth index in quintiles (1=poorest to 5=richest)
cons	A column of ones. This variable will be included as an explanatory variable in all models and its coefficient will be the intercept

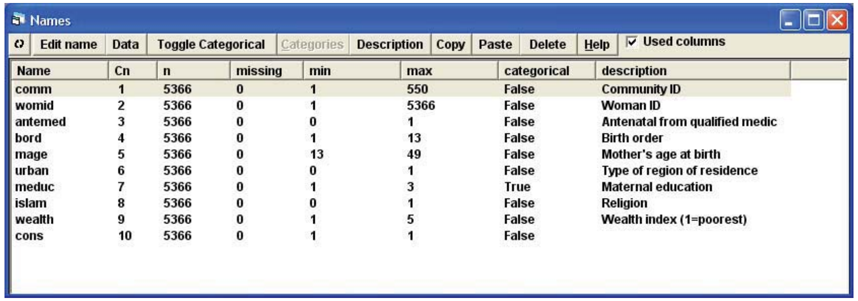
To open the worksheet:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***MLwiN* Datafiles**
- If you do not already have MLwiN to open the datafile with, click (**get *MLwiN***).
- Click "7.1.wsz"

The **Names** window will appear.

- Click the check box next to **Used columns** to view only those columns that contain data



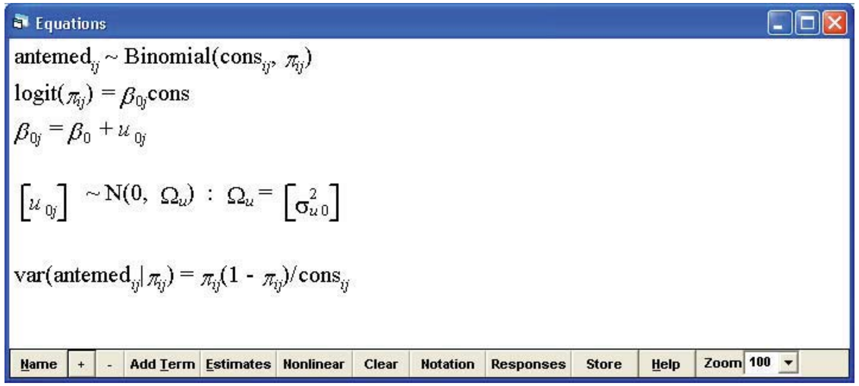
Name	Cn	n	missing	min	max	categorical	description
comm	1	5366	0	1	550	False	Community ID
womid	2	5366	0	1	5366	False	Woman ID
antemed	3	5366	0	0	1	False	Antenatal from qualified medic
bord	4	5366	0	1	13	False	Birth order
mage	5	5366	0	13	49	False	Mother's age at birth
urban	6	5366	0	0	1	False	Type of region of residence
meduc	7	5366	0	1	3	True	Maternal education
islam	8	5366	0	0	1	False	Religion
wealth	9	5366	0	1	5	False	Wealth index (1=poorest)
cons	10	5366	0	1	1	False	

P7.1 Two-Level Random Intercept Model

P7.1.1 Specifying and estimating a two-level model

We will begin by fitting a null or empty two-level model, that is a model with only an intercept and community effects. To specify a two-level logit model in MLwiN:

- From the **Model** menu, select **Equations**
- Click the red ***y*** in the **Equations** window
- From the drop-down list labelled **y:** select **antemed**
- From the drop-down list labelled **N levels:** select **2-ij**
- From the drop-down list labelled **level 2(j):** select **comm**
- From the drop-down list labelled **level 1(i):** select **womid**
- Click **done**
- Click on **N(*XB*, Ω)** and check **Binomial**. A list of link functions will appear. We will retain the default of **logit** so click **Done**
- We now need to specify the denominator which for a binary response is always equal to 1, so click on the red n_{ij} and select **cons** from the drop-down list. Check **Done**. If you look in the **Names** window you will see that a new variable called **denom** has been added to the worksheet. This has been created from **cons**
- Click on x_0 and select **cons** from the drop-down list. Check **j(comm)** to add a *j* subscript to the intercept β_0, then click **Done**
- Click the + button twice to see the full model specification



The intercept β_{0j} consists of two components: a fixed effect β_0, shared by all communities, and a random effect u_{0j}, specific to community *j*. The random effect

is assumed to follow a normal distribution with covariance matrix Ω_u which in this simple model contains just one element, the between-community variance σ^2_{u0}.

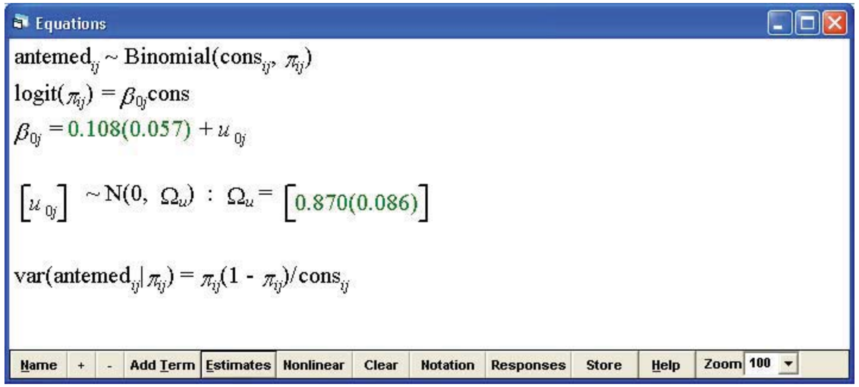
Before fitting the model, we have to specify details about the estimation procedure to be used. As will be described in C7.7 (and in more detail in the Technical Appendix), there are several estimation procedures available for binary and other categorical response models. In MLwiN, there are two options: quasi-likelihood methods and Markov chain Monte Carlo (MCMC) methods. There are four varieties of quasi-likelihood methods in MLwiN: 1st and 2nd order marginal quasi-likelihood (MQL1 and MQL2) and 1st and 2nd order penalised or predictive quasi-likelihood (PQL1 and PQL2). For most of these exercises we will use quasi-likelihood procedures, starting with the default procedure MQL1. However, MQL1 can produce estimates that are biased downwards so, after reaching convergence with MQL1, we will extend to the best available quasi-likelihood method, PQL2. We build the model in this way because convergence problems can sometimes be experienced using PQL, and these can be minimised by using MQL estimates as starting values. MCMC methods will be considered in P7.7.

To specify the estimation procedure:

- Click on the **Nonlinear** button at the bottom of the **Equations** window
- In the **Nonlinear Estimation** window, click on **Use Defaults**, then **Done**
- Click once on **Estimates** so that the parameters to be estimated (β_0 and σ^2_{u0}) appear in blue

Now to fit the model:

- Click **Start**
- Click once on **Estimates** to see the estimated coefficients (and standard errors in brackets)

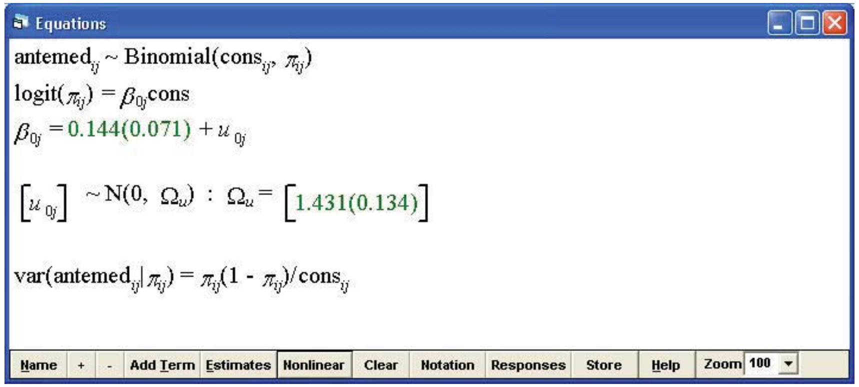


As the MQL1 procedure may lead to estimates that are biased downwards, PQL2 is preferred. To change the estimation procedure:

- Click on the **Nonlinear** button at the bottom of the **Equations** window. Under **Linearisation**, select **2nd Order**
- Under **Estimation type**, select **PQL**
- Click **Done**
- Now click **More** to fit the model

*Note that clicking **More** rather than **Start** means that the MQL1 estimates will be used as starting values in the PQL2 procedure. Because convergence problems may be encountered when using PQL it is advisable to use MQL first and then extend to PQL.*

You should obtain the following estimates:



In this case, the PQL2 estimates are much larger than the MQL1 estimates. From now on we will use PQL2, always using **More** rather than **Start** so that the estimates from the current model are used as starting values for the next model. If convergence problems are encountered when estimating a particular model, we can fit the model using MQL1 and then extend to PQL2 using **More**. In practice, and especially when there are large differences between the MQL1 and PQL2 estimates, we would use MCMC methods (or numerical quadrature – see C7.7). Here, we will wait until P7.7 before introducing MCMC. At this stage, the aim is to be able to specify and interpret multilevel models for binary responses, so we will ignore these technical issues for now.

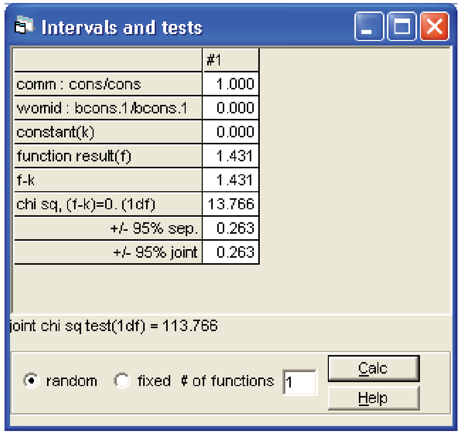
P7.1.2 Interpretation of the null two-level model

From the PQL2 estimates, we can say that the log-odds of receiving antenatal care from a medically-trained provider in an 'average' community (one with u_{0j} = 0) is

estimated as $\hat{\beta}_0 = 0.144$. The intercept for community *j* is 0.144 + u_{0j}, where the variance of u_{0j} is estimated as $\hat{\sigma}^2_{u0} = 1.431$.

To calculate the Wald statistic for testing the null hypothesis that $\sigma^2_{u0} = 0$:

- From the **Model** menu, select **Intervals and tests**
- We will retain the default settings of **random** and **# functions = 1**
- In the **#1** column, next to **comm: cons/cons**, type **1** and click **Calc**



The test statistic is 113.8 which is compared with a chi-squared distribution on 1 degree of freedom. Bearing in mind that the critical value for a test at the 5% level is 3.84, there is strong evidence that the between-community variance is non-zero.

We will now examine estimates of the community effects or residuals, $\hat{u}_{0j}$, obtained from the null model. To calculate the residuals and produce a 'caterpillar plot' with the community effects shown in rank order together with 95% confidence intervals:

- From the **Model** menu, select **Residuals**
- At the bottom of the **Residuals** window, change **level** from 1:womid to **2:comm**
- In the box to the left of **SD(comparative) of residual to**, edit 1.0 to **1.96**
- Click **Calc**
- Click on the **Plots** tab and check **residual +/- 1.96 sd x rank** (the 3rd option)
- Click **Apply**