Module 6: Regression Models for Binary Responses MLwiN Practicals

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-3

Contents

Introduction to the Bangladesh Demographic and Health Survey 2004 Dataset .......... 1
P6.1 Preliminaries: Mean and Variance of Binary Data ........................................... 3
  P6.1.1 Mean and standard deviation of the response variable ......................... 3
  P6.1.2 Bivariate relationships between the response and explanatory variables..... 4
P6.2 Moving Towards a Regression Model: The Linear Probability Model .......... 9
P6.3 Generalised Linear Models ........................................................................... 15
P6.4 Latent Variable Representation of a Generalised Linear Model ................. 16
P6.5 Application of Logit and Probit Models in Analyses of Antenatal Care Uptake .. 17
  P6.5.1 Probabilities, odds and odds ratios .......................................................... 17
  P6.5.2 Interpretation of a logit model ................................................................. 19
  P6.5.3 Comparison of probit and logit coefficients ........................................... 23
  P6.5.4 Interpretation of a probit model ............................................................... 24
  P6.5.5 Significance testing and confidence intervals ......................................... 25
P6.6 Adding Further Predictors in the Analysis of Antenatal Care .................... 32
  P6.6.1 Extending the logit model ...................................................................... 32
  P6.6.2 Model interpretation ............................................................................... 35
P6.7 Interaction Effects......................................................................................... 39
P6.8 Modelling Proportions.................................................................................. 43
  P6.8.1 Creating a community-level dataset ...................................................... 43
  P6.8.2 Fitting a binomial logit model ............................................................... 46
  P6.8.3 Extrabinomial variation ......................................................................... 47
All of the sections within this module have online quizzes for you to test your understanding. To find the quizzes:

EXAMPLE

From within the LEMMA learning environment
  - Go down to the section for Module 6: Regression Models for Binary Responses
  - Click "6.1 Preliminaries: Mean and Variance of Binary Data" to open Lesson 6.1
  - Click Q1 to open the first question

Introduction to the Bangladesh Demographic and Health Survey 2004 Dataset

You will be analysing data from the Bangladesh 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.

We consider a range of predictors, including the woman’s age at the time of the birth, her level of education, and an indicator of whether she was living in an urban or rural area at the time of the survey. The file contains the following variables:

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description and codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>comm</td>
<td>Community identifier (not used until P6.8)</td>
</tr>
<tr>
<td>womid</td>
<td>Woman identifier</td>
</tr>
<tr>
<td>antemed</td>
<td>Received antenatal care at least once from a medically-trained provider, e.g. doctor, nurse or midwife (1=yes, 0=no)</td>
</tr>
<tr>
<td>bord</td>
<td>Birth order of child (ranges from 1 to 13)</td>
</tr>
</tbody>
</table>

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
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

There are 5366 women in the data file.

To open the worksheet:

From within the LEMMA Learning Environment
- Go to Module 6: Regression 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 “6.1.wsz”

The Names window will appear.

- Click the check box next to Used columns to view only those columns that contain data
P6.1 Preliminaries: Mean and Variance of Binary Data

P6.1.1 Mean and standard deviation of the response variable

We will begin by tabulating our response variable, antemed.

- From the Basic Statistics menu, select Tabulate
- Check Percentages of row totals
- From the drop-down list next to Columns, select antemed
- Click Tabulate

The following table will appear in the Output window:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>2613</td>
<td>2753</td>
<td>5366</td>
</tr>
<tr>
<td>%</td>
<td>48.7</td>
<td>51.3</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The sample estimate of the proportion of women receiving antenatal care\(^\text{10}\) is \(\hat{\pi} = 0.513\).

Next, we will calculate the mean and standard deviation of antemed.

- From the Basic Statistics menu, select Averages and Correlations
- Select antemed from the variable list
- Click Calculate

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Missing</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>antemed</td>
<td>5366</td>
<td>0</td>
<td>0.51305</td>
<td>0.49988</td>
</tr>
</tbody>
</table>

Notice that the mean of 0.513 is equal to the proportion receiving antenatal care that we obtained from the tabulation.

Using the formula for the standard deviation of a binary variable given in C6.1, we obtain

\[
s = \sqrt{\hat{\pi}(1 - \hat{\pi})} = \sqrt{0.513(1 - 0.513)} = 0.4998, \text{ which agrees with the s.d. value in the output.}
\]

\(^{10}\)Throughout the practical we will frequently refer to antenatal care from a medically-trained provider simply as antenatal care.
P6.1.2  Bivariate relationships between the response and explanatory variables

Before fitting any models to the relationship between \( y \) (antemed) and explanatory variables, we will first examine the bivariate relationship between \( y \) and three potential predictors: maternal age (mage), type of region of residence (urban) and maternal education (meduc).

We begin with \( \text{mage} \), a continuous variable. Let’s start with a scatterplot of \( \text{antemed} \) versus \( \text{mage} \).

- From the Graphs menu, select Customised Graph(s)
- From the drop-down list labelled \( \text{plot type} \), select point
- From the drop-down list labelled \( y \), select \( \text{antemed} \)
- From the drop-down list labelled \( x \), select \( \text{mage} \)
- Click Apply
- You can add titles by clicking anywhere on the graph and selecting the Titles tab

![Scatterplot of antemed versus mage](image)

Clearly the scatterplot is not very informative because our response takes only two values. Instead we will plot the proportion receiving antenatal care (i.e. the mean
of \text{antemed} against \text{mage}. To do this, we calculate the mean of \text{antemed} for each distinct value of \text{mage}, but first we need to sort the values of \text{antemed} by \text{mage}. We will store the sorted values of \text{antemed} and \text{mage} in columns c11 and c12, which we will call \text{ante-sort} and \text{mage-sort}.

- From the Data Manipulation menu, select Sort
- Under Key code columns, select \text{mage}
- Under Input columns, highlight \text{antemed} and \text{mage} (using Ctrl-click)
- Under Output columns, click Free columns (so that the next empty columns, c11 and c12 will be used)
- Click Add to action list followed by Execute
- Go to the Names window and type in the names \text{ante-sort} and \text{mage-sort} for the new variables in c11 and c12

We can now calculate the mean of \text{ante-sort} for each value of \text{mage-sort}.

- From the Data Manipulation menu, select Multilevel data manipulations
- Under Operation, retain the default of Average
- Under On blocks defined by, select \text{mage-sort}
- Under Input columns, highlight \text{ante-sort} and \text{mage-sort}
- Under Output columns, click Same as input
- Click Add to action list followed by Execute

If you look at \text{ante-sort} and \text{mage-sort} (using Data Manipulation \rightarrow View or edit data) you will see that values of \text{ante-sort} are the means for each value of \text{mage-sort}.

We are now in a position to plot the mean of \text{antemed} (proportion receiving antenatal care) versus maternal age.

- From the Graphs menu, select Customised Graph(s)
- From the drop-down list labelled plot type, select point
- From the drop-down list labelled y, select \text{ante-sort}
- From the drop-down list labelled x, select \text{mage-sort}
- Click Apply
- Click anywhere on the plot and then on the Titles tab. Change the y title to mean(antemed).
This document is only the first few pages of the full version.
To see the complete document please go to learning materials and register:
http://www.cmm.bris.ac.uk/lemma

The course is completely free. We ask for a few details about yourself for our research purposes only. We will not give any details to any other organisation unless it is with your express permission.