

# Module 7: Multilevel Models for Binary Responses

## R Practical

*Camille Szmaragd and George Leckie<sup>1</sup>*  
Centre for Multilevel Modelling

### Pre-requisites

- Modules 1-6

## Contents

Introduction to the Bangladesh Demographic and Health Survey 2004 Dataset .....	1
P7.1 Two-Level Random Intercept Model .....	3
P7.1.1 Specifying and estimating a two-level model .....	3
P7.1.2 Interpretation of the null two-level model .....	5
P7.1.3 Adding an explanatory variable.....	7
P7.2 Latent Variable Representation of a Random Intercept Model .....	11
P7.2.1 Comparison of a single-level and multilevel threshold model .....	11
P7.2.2 Variance partition coefficient .....	14
P7.3 Population-Averaged and Cluster-Specific Effects .....	16
P7.4 Predicted Probabilities from a Multilevel Model.....	17
P7.5 Two-Level Random Slope Model .....	23
P7.5.1 Allowing the effect of wealth to vary across communities .....	23
P7.5.2 Interpretation of a random slope model .....	25
P7.5.3 Fitting random coefficients to categorical wealth.....	29
P7.6 Adding Level 2 Explanatory Variables: Contextual Effects .....	37
P7.6.1 Contextual effects.....	37
P7.6.2 Cross-level interactions .....	41
P7.7 Estimation of Binary Response Models: MCMC Methods .....	47
References.....	47

---

<sup>1</sup> This R practical is adapted from the corresponding MLwiN practical: Steele, F. (2008) Module 7: Multilevel Models for Binary Responses. LEMMA VLE, Centre for Multilevel Modelling. Accessed at <http://www.cmm.bris.ac.uk/lemma/course/view.php?id=13>.

**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 [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),<sup>2</sup> 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.

---

<sup>2</sup>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](http://www.measuredhs.com).

The file contains the following variables:

Variable name	Description and codes
<b>comm</b>	Community identifier
<b>womid</b>	Woman identifier
<b>antemed</b>	Received antenatal care at least once from a medically-trained provider, e.g. doctor, nurse or midwife (1=yes, 0=no)
<b>bord</b>	Birth order of child (ranges from 1 to 13)
<b>mage</b>	Mother's age at the child's birth (in years)
<b>urban</b>	Type of region of residence at survey (1=urban, 0=rural)
<b>meduc</b>	Mother's level of education at survey (1=none, 2=primary, 3=secondary or higher)
<b>islam</b>	Mother's religion (1=Islam, 0=other)
<b>wealth</b>	Household wealth index in quintiles (1=poorest to 5=richest)

The dataset also contains a number of extra variables derived from those above (see the practical for Module 6).

## P7.1 Two-Level Random Intercept Model

Download the Rdataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to **Rdatasets and Rfiles**
- Right click “7.1.txt” and select **Save Link As...** to save the dataset to your computer

Read the dataset into R and create a dataframe object named `mydata`<sup>3</sup>:

```
> mydata <- read.table("7.1.txt", header = TRUE, sep = ",")
```

and use the `str` command to produce a summary of the dataset:

```
> str(mydata)
'data.frame': 5366 obs. of 17 variables:
 $ comm : int 1 1 1 1 1 1 1 1 1 1 ...
 $ womid : int 1 2 3 4 5 6 7 8 9 10 ...
 $ antemed: int 0 1 1 0 0 1 0 0 0 1 ...
 $ bord : int 4 2 3 6 6 4 2 3 1 1 ...
 $ mage : int 33 21 26 28 37 29 20 29 19 19 ...
 $ urban : int 0 0 0 0 0 0 0 0 0 0 ...
 $ meduc : int 2 3 2 1 2 2 3 3 3 3 ...
 $ islam : int 1 1 1 1 1 1 1 1 1 1 ...
 $ wealth : int 3 4 2 2 4 4 2 3 3 4 ...
 $ magec : num 9.37 -2.63 2.37 4.37 13.37 ...
 $ magecsq: num 87.72 6.94 5.6 19.06 178.64 ...
 $ meduc2 : int 1 0 1 0 1 1 0 0 0 0 ...
 $ meduc3 : int 0 1 0 0 0 0 1 1 1 1 ...
 $ wealth2: int 0 0 1 1 0 0 1 0 0 0 ...
 $ wealth3: int 1 0 0 0 0 0 0 1 1 0 ...
 $ wealth4: int 0 1 0 0 1 1 0 0 0 1 ...
 $ wealth5: int 0 0 0 0 0 0 0 0 0 0 ...
```

### 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.

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + u_{0j}$$

The intercept  $\beta_0$  is shared by all communities while the random effect  $u_{0j}$  is specific to community  $j$ . The random effect is assumed to follow a normal distribution with variance  $\sigma_{u0}^2$ .

<sup>3</sup>At the beginning of your R session, you will need to set R's working directory to the file location where you saved the dataset. This can be done using the command line and the `setwd` function:

```
> setwd("C:\\userdirectory\\")
```

Or through selecting **Change Dir...** on the **File** menu.

R's main command for fitting multilevel models for binary and other discrete response variables is the `glmer` command which is part of an additional `lme4` library<sup>4</sup>, which we used already in Module 5. This library can be installed through the R Packages menu; select Install Package(s) and then select the correct Mirror and package from the scroll-down menus. As you will see, there is a variety of additional packages that can be installed with R. You only need to install a package once to your own computer. If you then want to use the package, you simply need to call it from within R prior to using the command for the first time in each R session<sup>5</sup>:

```
> library(lme4)
Loading required package: Matrix
```

The syntax for `glmer` is similar to that for the `lmer` command which we introduced in Module 5. To fit the above model using the `glmer` command and to create a model object `fit`, we type:

```
> fit <- glmer(antemed ~ (1|comm), family = binomial("logit"), data = mydata)
```

The binary response variable (**antemed**) follows the command which is then followed by a `~` and then by a list of fixed part explanatory variables (excluding the constant as this is included by default<sup>6</sup>). The above model contains only an intercept and so no fixed part explanatory variables are specified. The level 2 random part of the model is specified in brackets by the list of random part explanatory variables (the constant has to be explicitly specified by `1`, followed by a single vertical bar `|` and then by the level 2 identifier (**comm**). The `family` option is used to specify a binomial distribution for the response with a `logit` link function. The `data` option specifies the dataframe being used to fit the model.

We then display the results using the `summary` command, which gives the following output:

```
> summary(fit)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ (1 | comm)
Data: mydata

      AIC      BIC   logLik deviance df.resid
6639.5   6652.7  -3317.8   6635.5     5364

Scaled residuals:
    Min       1Q   Median       3Q      Max
-2.7779 -0.7458  0.3423  0.7118  2.6784

Random effects:
Groups Name      Variance Std.Dev.
```

---

<sup>4</sup>`lme4` is a package developed by Douglas Bates and Martin Maechler for fitting linear and generalized linear mixed-effect models. For more details about this library, see Module 5.

<sup>5</sup>You may get slightly different messages when calling `library(lme4)` as it will depend on your system and the version of R you have installed.

<sup>6</sup>Note, to omit the constant you need to add `-1` or `0` to the right-hand side of the “`~`” sign.

```

comm      (Intercept) 1.464      1.21
Number of obs: 5366, groups:  comm, 361

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.14809     0.07178   2.063   0.0391 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Before interpreting the model, we will discuss the estimation procedure that `glmer` uses<sup>7</sup>. The estimation procedure optimizes a function of the log likelihood using penalized iteratively re-weighted least squares. The log-likelihood is evaluated using the Laplacian approximation<sup>8</sup>. This approximation method may lead to unsatisfactory estimates, so we suggest that the reader uses the parameter estimates provided by R as a starting point and check their results using MCMC (for example with MLwiN or WinBUGS).

### P7.1.2 Interpretation of the null two-level model

From the model estimates (using Laplacian approximation), 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  $\hat{\beta}_0 = 0.148$ . The intercept for community  $j$  is  $0.148 + u_{0j}$ , where the variance of  $u_{0j}$  is estimated as  $\hat{\sigma}_{u_0}^2 = 1.464$ .

The likelihood ratio statistic for testing the null hypothesis, that  $\sigma_{u_0}^2 = 0$ , can be calculated by comparing the two-level model, with the corresponding single-level model without the level 2 random effects.

```

> fita <- glm(antemed ~ 1, data = mydata, family = binomial("logit"))

> logLik(fita) - logLik(fit)
'log Lik.' -399.8391 (df=1)

```

The test statistic is 799.8 ( $-2 \times (-399.83)$ ) with 1 degree of freedom, so there is strong evidence that the between-community variance is non-zero.<sup>9</sup>

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

<sup>7</sup>For further details see the PDF vignettes available on the `lme4` website <http://cran.r-project.org/web/packages/lme4>, in particular the vignette entitled “Computational Methods” which deals with the statistical theory.

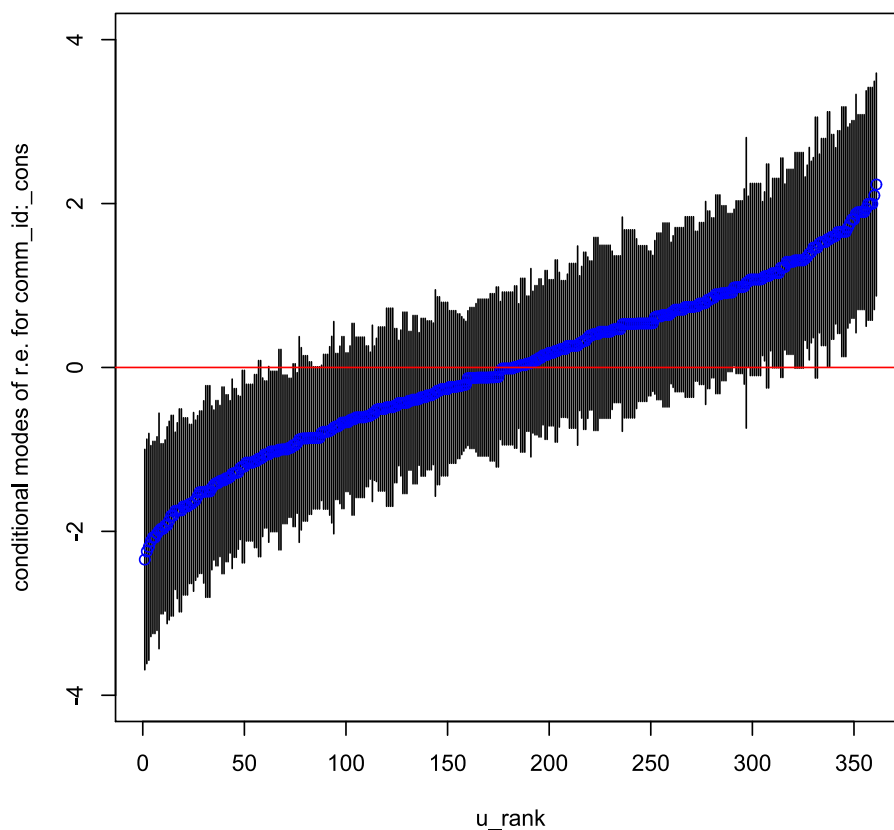
<sup>8</sup>In the `glmerhelp` file, it is noted that by specifying the option `nAGQ=n`, with `n` greater than 1, changes the approximation method to adaptive Gauss-Hermite approximation, with a greater value of `n` leading to more accurate evaluation of the log-likelihood. However, when we tried this in this and other examples we found that specifying different values of `n` did not lead to any change in our estimates. This apparent bug may be fixed in later versions of `lme4`. We refer the reader to the following blog for a discussion of the different approximation methods available in R (`glmer`) and Stata (`xtmelogit` command):

[http://www.stat.columbia.edu/~cook/movabletype/archives/2010/09/r\\_vs\\_stata\\_or\\_d.html](http://www.stat.columbia.edu/~cook/movabletype/archives/2010/09/r_vs_stata_or_d.html).

<sup>9</sup>Note that the test statistic has a non-standard sampling distribution as the null hypothesis of a zero variance is on the boundary of the parameter space; we do not envisage a negative variance. In this case the correct p-value is half the one obtained from the tables of chi-squared distribution with 1 degree of freedom.

the community effects shown in rank order together with 95% confidence intervals we can use the same commands as we used P5.1.2 for the continuous response two-level random intercepts model:

```
> u0 <- ranef(fit, condVar = TRUE)
> u0se <- sqrt(attr(u0[[1]], "postVar")[1, , ])
> commid <- as.numeric(rownames(u0[[1]]))
> u0tab <- cbind("commid" = commid, "u0" = u0[[1]], "u0se" = u0se)
> colnames(u0tab)[2] <- "u0"
> u0tab <- u0tab[order(u0tab$u0), ]
> u0tab <- cbind(u0tab, c(1:dim(u0tab)[1]))
> u0tab <- u0tab[order(u0tab$commid), ]
> colnames(u0tab)[4] <- "u0rank"
> plot(u0tab$u0rank, u0tab$u0, type = "n", xlab = "u_rank", ylab = "conditional
modes of r.e. for comm_id:_cons", ylim = c(-4, 4))
> segments(u0tab$u0rank, u0tab$u0 - 1.96*u0tab$u0se, u0tab$u0rank, u0tab$u0 +
1.96*u0tab$u0se)
> points(u0tab$u0rank, u0tab$u0, col = "blue")
> abline(h = 0, col = "red")
```



The plot shows the estimated residuals for all 361 communities in the sample. For a substantial number of communities, the 95% confidence interval does not overlap the horizontal line at zero, indicating that uptake of antenatal care in these communities is significantly above average (above the zero line) or below average (below the zero line). Compared to the plot for the US election data (C7.2), the confidence intervals are quite wide. This is because the sample size within a community is much smaller than the sample size within a state, leading to larger standard errors for the estimated community residuals  $\hat{u}_{0j}$ .

### P7.1.3 Adding an explanatory variable

Next we will include maternal age as an explanatory variable in the model. Although we know from our single-level analysis (P6.1 and P6.6) that there is a curvilinear relationship between the log-odds of antenatal care and age, we will start by fitting a linear age effect.

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + u_{0j}$$

```
> (fit2 <- glmer(antemed ~ magec + (1|comm), family = binomial("logit"), data =
mydata))
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magec + (1 | comm)
Data: mydata
      AIC      BIC    logLik deviance df.resid
6603.407 6623.171 -3298.704  6597.407     5363
Random effects:
Groups Name      Std.Dev.
comm  (Intercept) 1.209
Number of obs: 5366, groups:  comm, 361
Fixed Effects:
(Intercept)      magec
  0.14460      -0.03236
```

Note that there is little change in the estimate of the between-community variance, suggesting that the distribution of maternal age is similar across communities.

The equation of the average fitted regression line, expressing the relationship between the log-odds of receiving antenatal care and maternal age, is:

$$\log\left(\frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}}\right) = 0.144 - 0.032 \text{magec}_{ij}$$

The fitted line for a given community will differ from the average line in its intercept, by an amount  $\hat{u}_j$  for community  $j$ . A plot of the predicted community lines will therefore show a set of parallel lines. To produce this plot, we first need to calculate the predicted log-odds of antenatal care for each woman, based on her age at survey and community of residence. To do this we compute the predicted probability of antenatal care for each woman using the `fitted` command. This command extracts the predicted (or fitted) values and is an alternative to the



`predict` command which is not available for the models fitted using `lmer` and `glmer`.

```
> predprob <- fitted(fit2)
```

We then transform these predicted probabilities to predicted log odds using the `logit()` function from the R additional `VGAM` library.

```
> library(VGAM)
Loading required package: stats4
Loading required package: splines

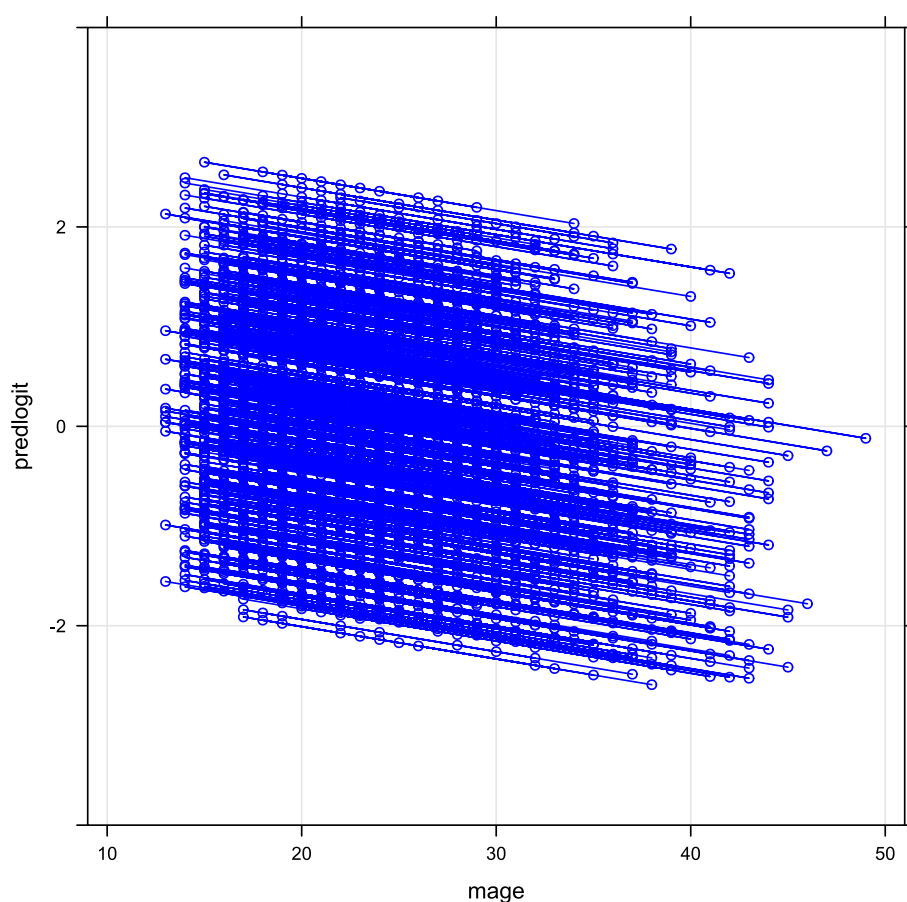
> predlogit <- logit(predprob)
```

Next we create a new dataframe containing only one observation for each distinct value of `mage` found within each community:

```
> datapred <- unique(data.frame(cbind(predlogit = predlogit, comm = mydata$comm,
mage = mydata$mage)))
```

Finally, the `xyplot` command plots the predicted probabilities against maternal age.

```
> xyplot(predlogit ~ mage, data = datapred, groups = comm, type = c("p", "l", "g"),
col = "blue", xlim = c(9, 51), ylim = c(-4, 4))
```



For a woman aged 22, the log-odds of receiving antenatal care ranges from about -2.2 to 2.5 depending on which community she lives in. This translates to a range in probabilities of  $\exp(-2.2)/[1+\exp(-2.2)] = 0.10$  to  $\exp(2.5)/[1+\exp(2.5)] = 0.92$ , so there are strong community effects.

In Module 6, we found that the age effect showed some curvature, where the probability of uptake increases until the mid 20s, then starts to decrease (see P6.6.2). To fit a quadratic function in **mage**:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + u_{0j}$$

```
> (fit3 <- glmer(antedmed ~ magec + magecsq + (1|comm), family = Generalized linear
mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antedmed ~ magec + magecsq + (1 | comm)
Data: mydata
      AIC      BIC    logLik deviance df.resid
6602.850 6629.201 -3297.425  6594.850     5362
Random effects:
Groups Name      Std.Dev.
comm (Intercept) 1.205
Number of obs: 5366, groups: comm, 361
Fixed Effects:
(Intercept)      magec      magecsq
  0.184787    -0.027587    -0.001063
```

The squared age term is not significant at the 5% level, but we will retain it for now.

Rerun the commands given earlier to calculate the predicted log-odds of receiving antenatal care as a function of age and community.

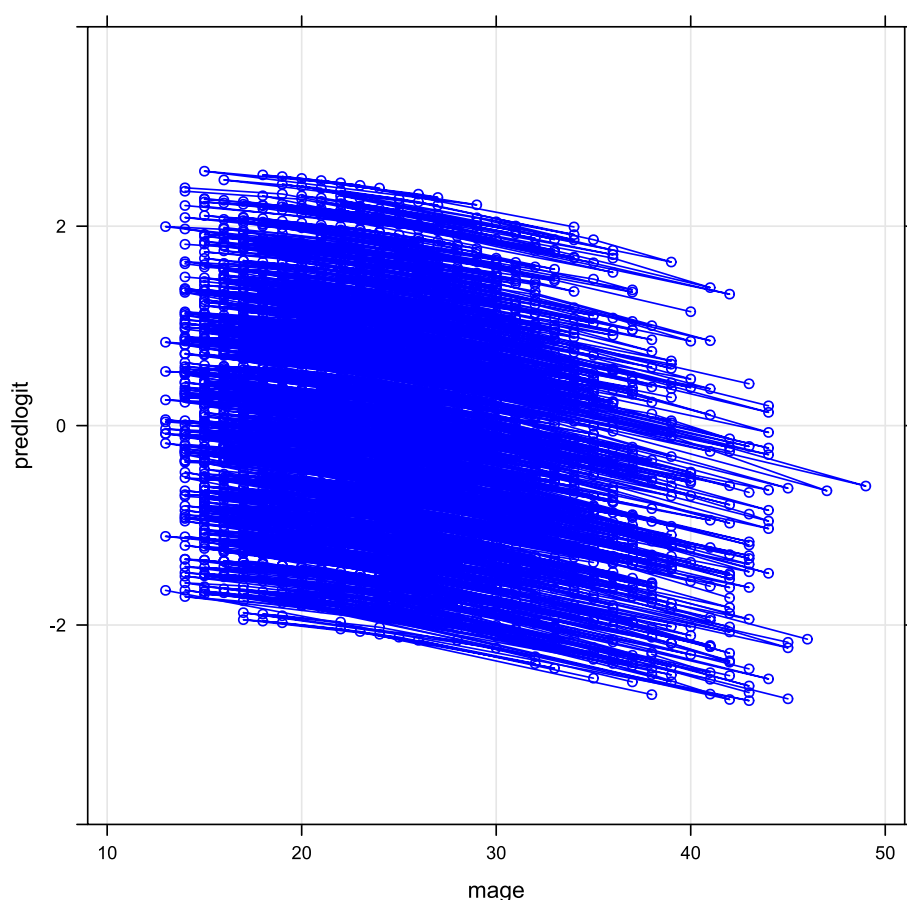
```
> predprob <- fitted(fit3)

> predlogit <- logit(predprob)

> datapred <- unique(data.frame(cbind(predlogit = predlogit, comm = mydata$comm,
mage = mydata$mage)))

> xyplot(predlogit ~ mage, data = datapred, groups = comm, type = c("p", "l",
"g"), col = "blue", xlim = c(9, 51), ylim = c(-4, 4))
```

The plot of the new **predlogit** variable versus **mage** should look like this:



Notice that the prediction 'lines' are now slightly curved because of the age-squared term, but the curves are still parallel because the relationship with age is assumed to be the same in each community.

**Don't forget to take the online quiz!**

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 [Q1](#) to open the first question

## P7.2 Latent Variable Representation of a Random Intercept Model

In this exercise, we will compare a single-level and multilevel model to see the impact of adding the community-level random effect. From the comparison of the latent variable representations of a single-level and multilevel model (C7.2.2), we would expect the coefficients to increase in magnitude when a random effect is added to the model. We then examine the impact of adding woman-level characteristics to the model. Finally we calculate the variance partition coefficient, using the formula derived from the latent variable representation of a random intercept model.

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to **R Datasets and R files**
- Right click “7.2.txt” and select **Save Link As...** to save the dataset to your computer

Read the dataset into R and create a dataframe object named **mydata**:

```
> mydata <- read.table("7.2.txt", header = TRUE, sep = ",")
```

### P7.2.1 Comparison of a single-level and multilevel threshold model

The last model we fitted in P7.1 was a random intercept logit model with a quadratic function for maternal age. We will extend this model to include two further woman-level predictors: maternal education (with dummies **meduc1** for the ‘primary’ and **meduc2** for the ‘secondary or higher’ categories) and household wealth index (**wealth**, in quintiles and treated as continuous).

First we generate a grand-mean centred version of **wealth**:

```
> summary(mydata$wealth)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   2.000   3.000   3.008   4.000   5.000

> mydata$wealthc <- mydata$wealth - mean(mydata$wealth)
```

Next, we load the **lme4** library and fit the model:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealthc}_{ij} + u_{0j}$$

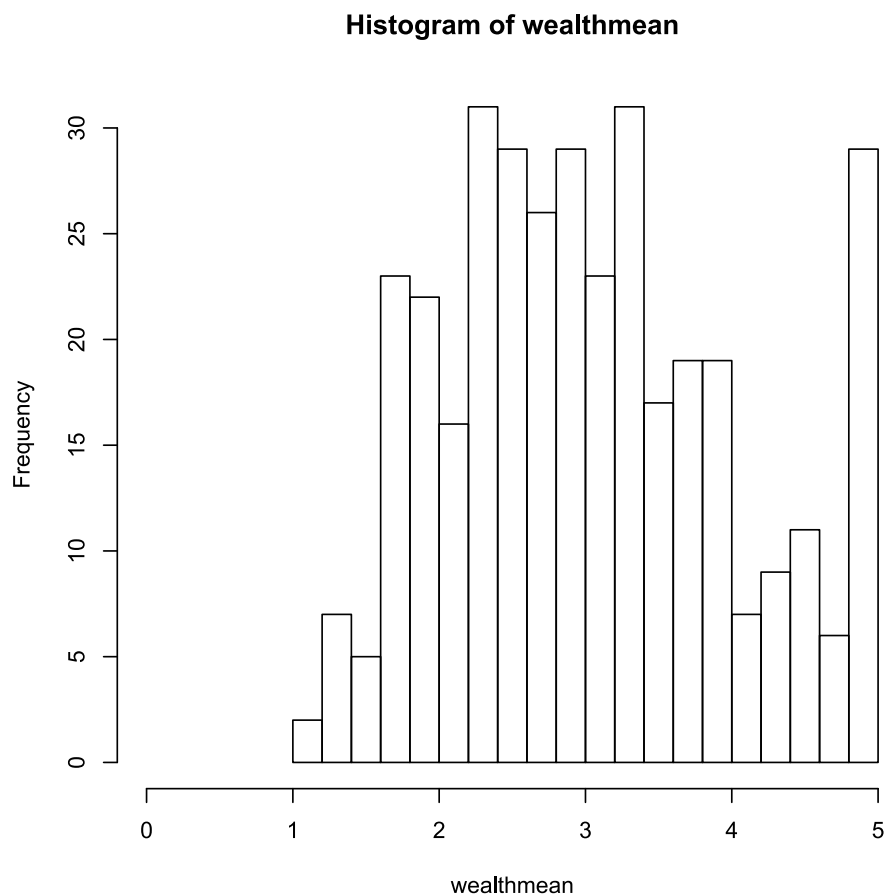
```
> library(lme4)
Loading required package: Matrix
```

## Module 7 (R Practical): Multilevel Models for Binary Responses

```
> (fit <- glmer(antedmed ~ magec + magecsq + meduc2 + meduc3 + wealthc + (1 |  
comm), data = mydata, family = binomial("logit"))  
Generalized linear mixed model fit by maximum likelihood (Laplace  
Approximation) [glmerMod]  
Family: binomial ( logit )  
Formula: antedmed ~ magec + magecsq + meduc2 + meduc3 + wealthc + (1 |  
comm)  
Data: mydata  
      AIC      BIC    logLik deviance df.resid  
5993.303 6039.418 -2989.652  5979.303     5359  
Random effects:  
Groups Name      Std.Dev.  
comm (Intercept) 0.929  
Number of obs: 5366, groups: comm, 361  
Fixed Effects:  
(Intercept)      magec      magecsq      meduc2      meduc3      wealthc  
-0.4559070  -0.0002799  -0.0010140   0.5501056   1.3148245   0.3994231
```

Notice that the addition of **meduc2**, **meduc3** and **wealthc** has substantially reduced the between-community variance, suggesting that the distribution of one or more variables varies across communities. This is expected because some communities will have higher proportions of educated women and relatively wealthier households than others. To illustrate this, we will calculate the mean of **wealth** for each community and look at its distribution across the 361 communities. In the `hist` command, we specify the `breaks` option to set the number of bins to 25 and the `xlim` option to set the range of the x axis.

```
> wealthmean <- tapply(mydata$wealth, mydata$comm, mean)  
  
> hist(wealthmean, xlim = c(0, 5), breaks = 25)
```



As expected, there is a large amount of between-community variation in the mean wealth index. We could repeat this exercise plotting, for example, the proportion with secondary or higher education (i.e. with **meduc** = 3) in a community.

We will now fit the single-level version of this model (i.e. with the same predictors but no community random effects).

We can fit the single-level model using the `glm` command (see Module 6):

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealthc}_{ij}$$

```
> fitb <- glm(antemed ~ magec + magecsq + meduc2 + meduc3 + wealthc, data = mydata, family = binomial("logit"))
```

```
> summary(fitb)
```

Call:

```
glm(formula = antemed ~ magec + magecsq + meduc2 + meduc3 + wealthc,
     family = binomial("logit"), data = mydata)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9187	-0.9782	0.5882	0.9377	2.1982

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.3685848  0.0593246  -6.213 5.20e-10 ***
magec       -0.0046213  0.0059102  -0.782  0.4343
magecsq     -0.0013503  0.0006307  -2.141  0.0323 *
meduc2       0.3968979  0.0757902   5.237 1.63e-07 ***
meduc3      1.0725261  0.0849994  12.618 < 2e-16 ***
wealthc     0.4819516  0.0239126  20.155 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7435.2  on 5365  degrees of freedom
Residual deviance: 6298.2  on 5360  degrees of freedom
AIC: 6310.2

Number of Fisher Scoring iterations: 4

```

Comparing the two sets of results, the coefficients of the education dummies, **meduc2** and **meduc3**, increase when the random effect is added. The ratio of the multilevel to single-level estimate is 1.387 for **meduc2** and 1.226 for **meduc3**. From equation (7.8) in C7.2.2 we would expect the ratio to be  $\sqrt{(0.864 + 3.29)/3.29} = 1.123$ , but remember that this relationship is approximate and only applies when the variable in question has exactly the same distribution in each community. In contrast, the coefficient of **wealthc** decreases when the community random effect is added. We would not expect the 1.127 ratio to apply here because we have already seen that the mean of the household wealth index varies substantially from community to community. Furthermore, we might expect that household wealth is associated with unobserved community-level determinants of antenatal care uptake, for example the availability and quality of maternal health services. If better services are offered in less-deprived areas, and these areas have higher use of antenatal care from medically-trained providers, we would expect that controlling for unobserved community characteristics in the multilevel model will reduce the effect of household wealth.

## P7.2.2 Variance partition coefficient

We will first revert to the multilevel model. We use the `summary` command to redisplay that model's results without having to reestimate it.

```

> summary(fit)
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealthc + (1 |
  comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)

      AIC      BIC    logLik deviance df.resid
5993.3   6039.4  -2989.7   5979.3     5359

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.5037 -0.6443  0.2568  0.6514  3.7578

Random effects:

```

## Module 7 (R Practical): Multilevel Models for Binary Responses

```
Groups Name      Variance Std.Dev.
comm  (Intercept) 0.863    0.929
Number of obs: 5366, groups: comm, 361

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.4559070  0.0826099  -5.519 3.41e-08 ***
magec       -0.0002799  0.0065337  -0.043  0.966
magecsq     -0.0010140  0.0006808  -1.489  0.136
meduc2       0.5501056  0.0843274   6.523 6.87e-11 ***
meduc3       1.3148245  0.0974009  13.499 < 2e-16 ***
wealthc      0.3994231  0.0294778  13.550 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) magec  magcsq meduc2 meduc3
magec  -0.024
magecsq -0.277 -0.489
meduc2  -0.542  0.208 -0.048
meduc3  -0.543  0.311 -0.072  0.549
wealthc  0.176 -0.123  0.071 -0.167 -0.360
```

The variance partition coefficient (VPC) is calculated as  $0.864 / (0.864 + 3.29) = 0.21$ . Thus 21% of the residual variation in the propensity to use antenatal care services ( $y^*$ ) is attributable to unobserved community characteristics. In P7.1.1 the between-community variance for the null model (i.e. with no predictors) was estimated as 1.465, giving a VPC of  $1.465 / (1.465 + 3.29) = 0.31$  or 31%. However, changes in the VPC should be interpreted with caution because the addition of a level 1 variable  $x$  can increase the between-community variance even when the distribution of  $x$  is the same in each community (see C7.2.3 for an illustration using simulated data).

### Don't forget to take the online quiz!

From within the LEMMA learning environment

- Go down to the section for **Module 7: Multilevel Models for Binary Responses**
- Click "[7.2 Latent Variable Representations of a Random Intercept Model](#)" to open Lesson 7.2
- Click [Q1](#) to open the first question



## **P7.3 Population-Averaged and Cluster-Specific Effects**

There is no practical for this lesson, but don't forget to take the online quiz!  
Please continue to C7.4.

## P7.4 Predicted Probabilities from a Multilevel Model

Download the Rdataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to **RDatasets and Rfiles**
- Right click “7.4.txt” and select **Save Link As...** to save the dataset onto your computer

Read the dataset into R and create a dataframe object named **mydata**:

```
> mydata <- read.table("7.3.txt", header = TRUE, sep = ",")
```

First we will refit the multilevel model estimated at the beginning of P7.2 with mother’s age, education and wealth as predictors:<sup>10</sup>

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} \\ + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealthc}_{ij} + u_j$$

If you have fitted the above model in the current session of R, there is no need to refit the model; you can simply redisplay the model’s results by typing `summary(fit)`. If you are running a new R session then you will have to load the **lme4** library and fit the model in the usual way:

```
> library(lme4)
Loading required package: Matrix

> (fit <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + wealthc + (1 |
  comm), data = mydata, family = binomial("logit"))
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial (logit)
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealthc + (1 |
  comm)
Data: mydata
      AIC      BIC    logLik deviance df.resid
5993.303 6039.418 -2989.652  5979.303     5359
Random effects:
Groups Name      Std.Dev.
comm (Intercept) 0.929
Number of obs: 5366, groups: comm, 361
Fixed Effects:
(Intercept)      magec      magecsq      meduc2      meduc3      wealthc
-0.4559036  -0.0002802  -0.0010140   0.5501052   1.3148223   0.3994216
```

In C7.4 we compared two methods for computing predicted probabilities from a multilevel model.

<sup>10</sup> Note for ease of exposition in this section, we have removed the 0 subscript from the level 2 random intercept effect.

**Method A: substituting the mean of the level 2 residuals ( $u_j = 0$ )**

Method A involves substituting  $u_j = 0$  in the formula for the response probability.<sup>11</sup> For the above model, for example, the predicted probability of receiving antenatal care would be calculated using:

$$\hat{\pi}_{ij} = \frac{\exp(z_{ij})}{1 + \exp(z_{ij})}$$

where  $z$  is the linear predictor (i.e. the predictions on the logit scale):

$$z_{ij} = -0.456 - 0.0002 \text{ magedc}_{ij} - 0.001 \text{ magedcsq}_{ij} \\ + 0.550 \text{ meduc2}_{ij} + 1.315 \text{ meduc3}_{ij} + 0.399 \text{ wealthc}_{ij}$$

The resulting predictions represent probabilities for the median community, and are sometimes referred to as *cluster-specific* predictions. We will use this approach to compute median probabilities for each combination of wealth and education found in the data (there are 5 values of wealth and three categories of education giving 15 combinations), holding mother's age at its sample mean.

First note that the sample mean of **magedc** is zero by definition because **magedc** is maternal age centred on its sample mean. The variable **magedcsq** is the square of **magedc** so **magedcsq** must be fixed at the square of the value we fix **magedc** at. We therefore set **magedc** = 0 and **magedcsq** = 0 for all women in the sample. We start by creating a new data frame **mydatapred** and setting the value of **magedc** and **magedcsq** to 0:

```
> mydatapred <- mydata
> mydatapred$magedc <- 0
> mydatapred$magedcsq <- 0
```

Next we extract the model matrix from the fitted model using the command **model.matrix**. This returns a data frame containing only the variables involved in the fixed part of the model. It takes two arguments, the first representing the fixed part of the model, and the second being a data frame containing at least the variables appearing in the fixed part of the model:

```
> X <- model.matrix(terms(fit), mydatapred)
```

We then extract the values of the parameter estimates for the fixed effects with the **fixef** command:

```
> b <- fixef(fit)
```

---

<sup>11</sup> For ease we have dropped the 0 subscript from the level 2 random intercept effect.

Finally to obtain the predicted probabilities, we simply apply the parameter estimates to the data frame we created, using a matrix multiplication "`%*%`", and then take the inverse logit using the VGAM library's `logit` command:

```
> medianpredproblogit <- X %*% b

> library(VGAM)
Loading required package: stats4
Loading required package: splines

> medianpredprob <- logit(medianpredproblogit, inverse = TRUE)
```

Finally we list the predicted probabilities for each combination of wealth and education. This is done by combining, in a new data frame, the predicted probabilities, the logit values of these predicted probabilities and the values for **wealth** and **meduc** extracted from **mydatapred**. The final steps are to sort the data frame by **wealth** and **meduc** and to rename the first two columns:

```
> mydatapred2 <- unique(data.frame(cbind("medianpredprob" = medianpredprob,
"medianpredproblogit" = medianpredproblogit, wealth = mydatapred$wealth, meduc =
mydata$meduc)))

> mydatapred2 <- mydatapred2[order(mydatapred2$wealth, mydatapred2$meduc), ]

> colnames(mydatapred2)[1:2] <- c("medianpredprob", "medianpredproblogit")

> mydatapred2
  medianpredprob medianpredproblogit wealth meduc
13      0.2213146      -1.25802203      1      1
24      0.3300593      -0.70791687      1      2
97      0.5141962       0.05680023      1      3
4       0.2976318      -0.85860043      2      1
3       0.4234821      -0.30849528      2      2
7       0.6121175       0.45622183      2      3
18      0.3871807      -0.45917875      3      1
1       0.5227160       0.09092640      3      2
8       0.7017497       0.85564351      3      3
55      0.4850652      -0.05975716      4      1
5       0.6201884       0.49034800      4      2
2       0.7781754       1.25506511      4      3
478     0.5841090       0.33966436      5      1
66      0.7088426       0.88976952      5      2
36      0.8394965       1.65448662      5      3
```

### **Method B: averaging over simulated values of $u_j$**

The second method we consider is to compute predicted probabilities, averaging over values of  $u_j$  drawn from a normal distribution with variance equal to the estimated level 2 variance, i.e.  $N(0, \hat{\sigma}_u^2)$ . Predictions that average over the random effect distribution are sometimes referred to as *population-averaged* probabilities. To recap from C7.4, the procedure for a model with one predictor  $x$  is as follows:

- i) Generate  $M$  values for random effect  $u$  from  $N(0, \hat{\sigma}_u^2)$  and denote the generated values by  $u^{(1)}, u^{(2)}, \dots, u^{(M)}$
- ii) For each simulated value ( $m = 1, \dots, M$ ) compute, for a given value of  $x$ ,

$$\pi^{(m)} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x + u^{(m)})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x + u^{(m)})}$$

iii) Calculate the mean of the probabilities computed in ii):

$$\pi = \frac{1}{M} \sum_{m=1}^M \pi^{(m)}$$

Steps i)-iii) can then be repeated for different values of  $x$ .

We will now implement this method in R with  $M = 1000$  for each combination of wealth and education. We use the data frame **mydatapred2** created in Method A:

```
> mydatapred2
  medianpredprob medianpredproblogit wealth meduc
13      0.2213146      -1.25802203      1      1
24      0.3300593      -0.70791687      1      2
97      0.5141962       0.05680023      1      3
4       0.2976318      -0.85860043      2      1
3       0.4234821      -0.30849528      2      2
7       0.6121175       0.45622183      2      3
18      0.3871807      -0.45917875      3      1
1       0.5227160       0.09092640      3      2
8       0.7017497       0.85564351      3      3
55      0.4850652      -0.05975716      4      1
5       0.6201884       0.49034800      4      2
2       0.7781754       1.25506511      4      3
478     0.5841090       0.33966436      5      1
66      0.7088426       0.88976952      5      2
36      0.8394965       1.65448662      5      3
```

Next we will expand the data to have 1000 rows for each combination of wealth and education. To do this, we use the `rep` command to create four new versions of the four variables in the data frame (**medianpredprob**, **medianpredproblogit**, **wealth** and **meduc**), where each new version is 15,000 rows long rather than the original 15 rows. Finally, we combine these four new variables into a new data frame **simdata**:

```
> swealth <- rep(mydatapred2$wealth, 1000)
> smeduc <- rep(mydatapred2$meduc, 1000)
> smedpredprob <- rep(mydatapred2$medianpredprob, 1000)
> smedpredproblogit <- rep(mydatapred2$medianpredproblogit, 1000)
> simdata <- data.frame(cbind(wealth = swealth, meduc = smeduc, medianpredprob =
smedpredprob, medpredproblogit = smedpredproblogit))
```

We need to do this as, for each combination of wealth and education, we will calculate the predicted probability 1000 times, where foreach of these times we insert a different simulated value of the random effect in the calculation.

For each of the 15,000 rows we draw a value of the random effect from a normal distribution with mean equal to zero and standard deviation equal to 0.92964 (our estimate of the random intercept standard deviation). This leads to 1000 simulations, repeated 15 times for each combination of values for **meduc** and **wealth**, so the total number of simulations is 15,000. To generate these simulated values, we use

the `rnorm()` random number generating function where we specify 0 for the mean and 0.92964 for the standard deviation.

```
> u <- rnorm(dim(simdata)[1], mean = 0, sd = 0.92964)
```

To calculate the 15,000 predicted probabilities we take the inverse logit (i.e. the antilogit) of the sum of the linear predictor and simulated random effect:

```
> simdata$meanpredprob <- logit(simdata$medpredproblogit + u, inverse = TRUE)
```

Finally, we calculate the mean of these probabilities for each combination of wealth and education in the data by aggregating the data using the `aggregate` command with the `by` option.

```
> simdata_agg <- aggregate(simdata$meanpredprob, by = list(simdata$wealth,
simdata$meduc, simdata$medianpredprob), FUN = mean)
```

To compare the median (cluster-specific) and mean (population-averaged) predictions:

```
> colnames(simdata_agg) <- c("wealth", "meduc", "medianpredprob", "meanpredprob")
```

```
> simdata_agg <- simdata_agg[order(simdata_agg$wealth, simdata_agg$meduc), ]
```

```
> simdata_agg
  wealth meduc medianpredprob meanpredprob
1      1     1      0.2213146      0.2568933
3      1     2      0.3300593      0.3510958
7      1     3      0.5141962      0.5189003
2      2     1      0.2976318      0.3260429
5      2     2      0.4234821      0.4307060
10     2     3      0.6121175      0.5905228
4      3     1      0.3871807      0.4110342
8      3     2      0.5227160      0.5227829
12     3     3      0.7017497      0.6703297
6      4     1      0.4850652      0.4895699
11     4     2      0.6201884      0.6140723
14     4     3      0.7781754      0.7384886
9      5     1      0.5841090      0.5665994
13     5     2      0.7088426      0.6778359
15     5     3      0.8394965      0.8095898
```

The first point to make is that the substantive conclusions regarding the effects of education and wealth are similar whichever set of probabilities is used. This is because the predictions are in the 0.2 to 0.8 range where the logistic transformation is fairly linear. The median and mean are closest around values of 0.5 (e.g. row 8) and become further apart as we move towards zero or 1 (e.g. rows 1 and 15). Had the level 2 variance been larger (as is often the case with longitudinal data), or the predicted probabilities were more extreme, we would have seen more marked differences between the median and mean estimates.

**Don't forget to take the online quiz!**

From within the LEMMA learning environment

- Go down to the section for **Module 7: Multilevel Models for Binary Responses**
- Click "[7.4 Predicted Probabilities from a Multilevel Model](#)" to open Lesson 7.4
- Click [Q1](#) to open the first question

## P7.5 Two-Level Random Slope Model

The models fitted in previous exercises have allowed the probability of receiving antenatal care from a medically-trained provider to depend on the community of residence (as well as individual characteristics). This was achieved by allowing the model intercept to vary randomly across communities in a *random intercept* model. We have assumed, however, that the effects of individual characteristics such as age and education are the same in each community, i.e. the coefficients of all explanatory variables are fixed across communities. We will now extend the random intercept model from which we calculated predicted probabilities in P7.4 to allow both the intercept and the coefficient of one of the explanatory variables to vary randomly across communities.

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to **RDatasets and Rfiles**
- Right click “7.5.txt” and select **Save Link As...** to save the dataset on to your computer

Read the dataset into R and create a data frame object named **mydata**:

```
> mydata <- read.table("7.5.txt", header = TRUE, sep = ",")
```

Load the two additional libraries needed for this section:

```
> library(VGAM)
Loading required package: stats4
Loading required package: splines

> library(lme4)
Loading required package: Matrix
```

### P7.5.1 Allowing the effect of wealth to vary across communities

In this model, the effects of maternal age, education and household wealth are assumed to be the same in each community. We will now fit a random slope (coefficient) for wealth to allow its effect to vary across communities.

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} - \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealthc}_{ij} + u_{0j} + u_{5j} \text{wealthc}_{ij}$$

Note that a new term  $u_{5j}$  has been added to the model, so that the coefficient of **wealthc** has become  $\beta_{5j} = \beta_5 + u_{5j}$ , and so the community-level variance has been replaced by a matrix with two new parameters,  $\sigma_{u5}^2$  and  $\sigma_{u05}$ . The two community



level residuals  $u_{0j}$  and  $u_{5j}$  are assumed to follow a bivariate normal distribution with mean vector 0 and variance-covariance matrix  $\Omega_u$ .

$$\begin{pmatrix} u_{0j} \\ u_{5j} \end{pmatrix} \sim \text{MVN}(0, \Omega_u), \quad 0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Omega_u = \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u05} & \sigma_{u5}^2 \end{pmatrix}$$

Note that the slope residual, and associated variance and covariance, have a subscript of '5' because **wealth** is the 5<sup>th</sup> explanatory variable in the model (not including the constant).

To fit this model:<sup>12</sup>

```
> fit <- glmer(antemed ~ magedc + magedcsq + meduc2 + meduc3 + wealthc + (1 +
wealthc | comm), data = mydata, family = binomial("logit"),
glmerControl(calc.derivs = FALSE))

> summary(fit)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magedc + magedcsq + meduc2 + meduc3 + wealthc + (1 +
wealthc | comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)

            AIC          BIC    logLik deviance df.resid
5987.0      6046.3   -2984.5   5969.0      5357

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.3332 -0.6531  0.2904  0.6410  4.3984

Random effects:
 Groups Name      Variance Std.Dev. Corr
 comm   (Intercept) 0.84206  0.9176
        wealthc     0.01427  0.1194  -1.00
Number of obs: 5366, groups:  comm, 361

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.4547848  0.0823493  -5.523 3.34e-08 ***
magedc       -0.0001221  0.0065341  -0.019  0.985
magedcsq     -0.0010562  0.0006816  -1.549  0.121
meduc2        0.5471131  0.0846713   6.462 1.04e-10 ***
meduc3        1.3092827  0.0968300  13.521 < 2e-16 ***
wealthc       0.4050941  0.0301318  13.444 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) magedc  magedcsq meduc2 meduc3
magedc    -0.022
magedcsq  -0.279 -0.489
meduc2    -0.545  0.206 -0.047
meduc3    -0.543  0.307 -0.069  0.553
wealthc    0.008 -0.120  0.070 -0.159 -0.344
```

<sup>12</sup>Note: here we have added the calc.derivs=FALSE option. This eliminates warning messages related to this calculation and saves some estimation time. For further details see <https://cran.r-project.org/web/packages/lme4/vignettes/lmerperf.html>

Note, that the way we specified the random term `((1 + wealthc | comm))` implies that the random intercepts and slopes covary<sup>13</sup>.

### *Testing for a random slope*

We can use a likelihood ratio test to test whether the effect of wealth varies across communities. The null hypothesis is that the two new parameters  $\sigma_{u5}^2$  and  $\sigma_{u05}$  are simultaneously equal to zero. The likelihood ratio test statistic is calculated as two times the difference in the log likelihood values between the model with and without the random slope for wealth (the model without the random slope was estimated in P7.4):

$$LR = 2(-2984.494 - -2989.644) = 10.31 \text{ on } 2 \text{ d.f.}$$

The 5% point of a chi-squared distribution on 2 d.f. is 5.99. We therefore conclude that the effect of wealth does indeed vary across communities.

### **P7.5.2 Interpretation of a random slope model**

The effect of wealth on the log-odds of receiving antenatal care in community  $j$  is estimated as  $0.405 + \hat{u}_{5j}$ , and the between-community variance in the effect of wealth is estimated as 0.016. Because wealth has been centred about its sample mean, the intercept variance  $\hat{\sigma}_{u0}^2 = 0.843$  is interpreted as the between-community variance in the log-odds of antenatal care at the mean of the wealth index.

### *Examining the intercept and slope residuals for communities*

The negative intercept-slope covariance estimate ( $\hat{\sigma}_{u5} = -0.110$ ) implies that communities with above-average antenatal care uptake (intercept residual  $\hat{u}_{0j} > 0$ ) tend also to have below-average effects of wealth (slope residual  $\hat{u}_{5j} < 0$ ). Put another way, there is less of an income gradient in use of antenatal care in communities with high uptake. We can obtain the estimated correlation as follows:

```
> VarCorr(fit)
Groups Name      Std.Dev. Corr
comm      (Intercept) 0.91764
          wealthc     0.11945 -0.999
```

To obtain a plot of the community slopes versus the community intercepts for wealth,  $\hat{u}_{5j}$  vs.  $\hat{u}_{0j}$ :

```
> reffects <- ranef(fit, condVar = TRUE)

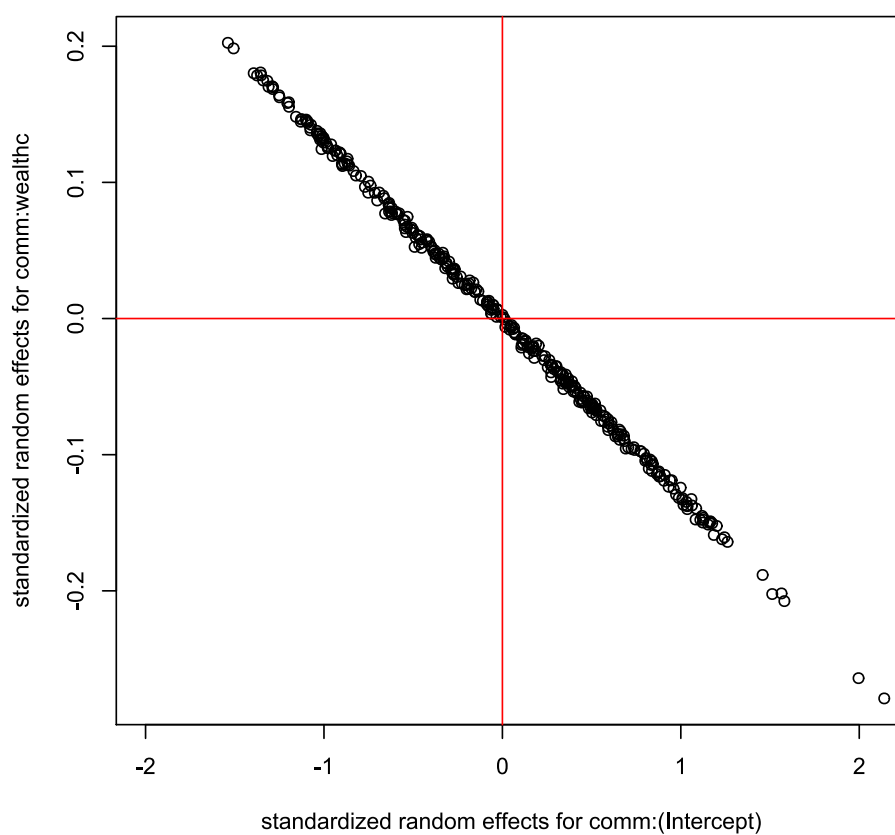
> plot(reffects[[1]], xlab = "standardized random effects for comm:(Intercept)",
       ylab = "standardized random effects for comm:wealthc", xlim = c(-2, 2.1))

> abline(v = 0, col = "red")

> abline(h = 0, col = "red")
```

---

<sup>13</sup>We refer the reader to Module 5 (P5.3) to see how to fit a model where random intercepts and slopes are independent.



If we knew the geographical location of communities, it might be of interest to use this plot to identify communities that had low uptake and steep income gradients (i.e. communities in the top left hand quadrant). Efforts to improve maternal health services might then be targeted towards such areas.

### *Community prediction lines*

The equation of the fitted regression line for community  $j$ , for a woman of mean age ( $\text{magec} = 0$ ,  $\text{magecsq} = 0$ ) and no education (the reference category of  $\text{meduc}$ :  $\text{meduc2} = 0$ ,  $\text{meduc3} = 0$ ) is:

$$\log\left(\frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}}\right) = (-0.456 + \hat{u}_{0j}) + (0.399 + \hat{u}_{5j}) \text{wealthc}_{ij}$$

To obtain the fitted line for women with different ages or levels of education, only the intercept would change. For a woman with primary education, for example, the intercept would increase from -0.456 to  $-0.456 + 0.39 = 0.083$ .

To produce a plot of the predicted community lines, we first need to compute the predicted log-odds,  $\text{logit}(\hat{\pi}_{ij})$ , for each woman, based on their value of **wealthc** and their community of residence.

We first copy the data into a new data frame **mydatapred**:

```
> mydatapred <- mydata
```

Change all values of **magec**, **magecsq**, **meduc2** and **meduc3** to zero:

```
> mydatapred$magec <- 0
> mydatapred$magecsq <- 0
> mydatapred$meduc2 <- 0
> mydatapred$meduc3 <- 0
```

Next, calculate the predicted log-odds for each woman in the data based only on the fixed part of the model.

```
> X <- model.matrix(terms(fit), mydatapred)
> b <- fixef(fit)
> predlogit <- X %*% b
```

Combine these predicted log-odds with the values of **wealthc**, **wealth** and **comm** in a new dataframe:

```
> mydatapred2 <- unique(data.frame(cbind(predlogit = predlogit, wealthc =
mydatapred$wealthc, comm = mydatapred$comm, wealth = mydatapred$wealth)))
> colnames(mydatapred2)[1] <- c("predlogit")
```

Define a community (or random effects) indicator:

```
> re_id <- as.integer(rownames(reffects$comm))
```

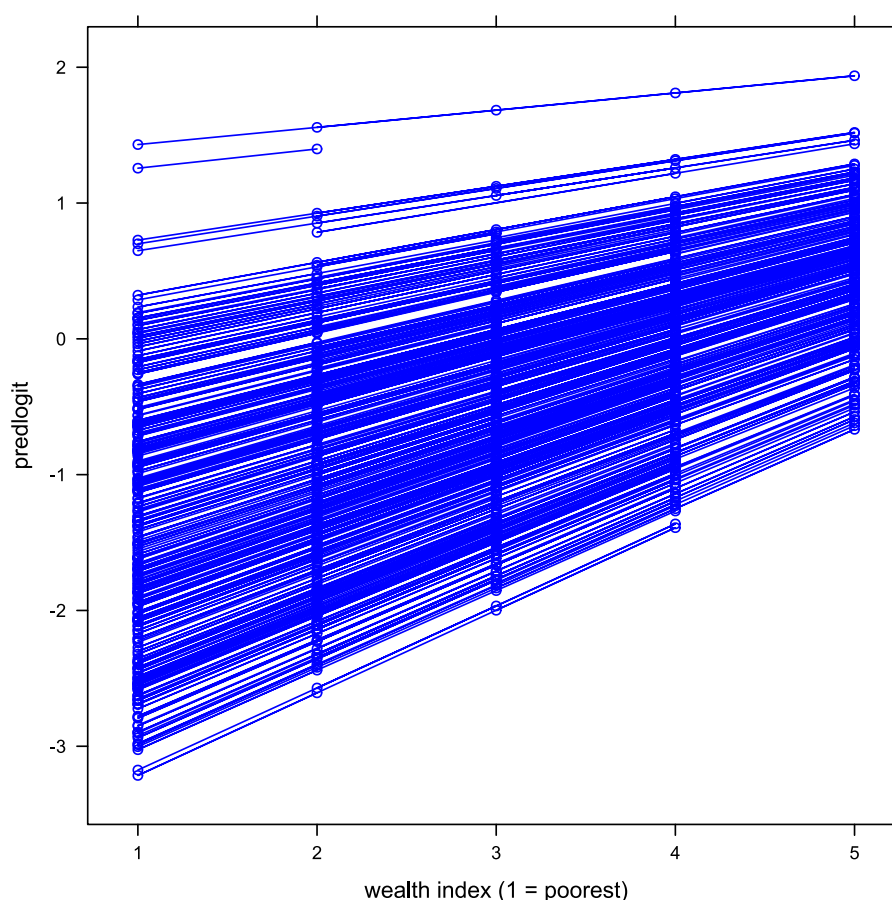
Add on the random part prediction:

```
> u0 <- data.frame(cbind(commid = re_id, u0 = reffects[[1]][, 1]))
> u1 <- data.frame(cbind(commid = re_id, u1 = reffects[[1]][, 2]))
> for (i in 1:dim(mydatapred2)[1]){
  mydatapred2$predlogit[i] <- mydatapred2$predlogit[i] + u0$u0[u0$commid ==
mydatapred2$comm[i]] + u1$u1[u1$commid ==
mydatapred2$comm[i]]*mydatapred2$wealthc[i]
}
```

Note, that to add the random part, we needed to loop over all the log-odds and add the **u0** and **u1** corresponding to the associated community.

Plot the predicted community lines:

```
> xyplot(predlogit ~ wealth, data = mydatapred2, groups = comm, type =
c("p","l"), col = "blue", xlab = "wealth index (1 = poorest)")
```



Notice that some lines are shorter than others because not all communities contain women in the higher wealth quintiles. We can also see that the community lines are ‘fanning in’ as wealth increases. This is expected because of the negative correlation between the intercept and slope residuals.

### ***Between-community variance as a function of wealth***

From the plot of the predicted lines for each community, we can see that the lines are more spread out for the lower quintiles of the wealth index than at the higher quintiles. In other words, the variability in the log-odds of receiving antenatal care decreases as **wealth** increases. Fitting a random slope for **wealth** implies that the between-community variance is a function of **wealth**, rather than constant as in the random intercept model.

The community-level variance function takes the following form:

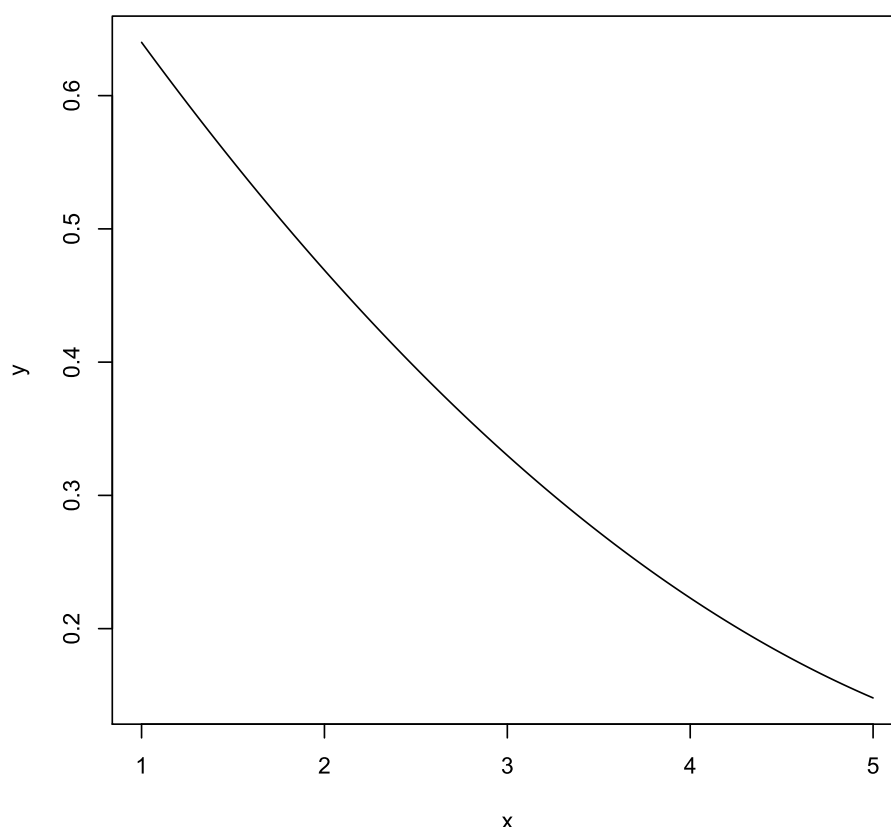
$$\begin{aligned} \text{var}(u_{0j} + u_{5j}\text{wealth}_{ij}) &= \text{var}(u_{0j}) + 2\text{cov}(u_{0j}, u_{5j}) \text{wealth}_{ij} + \text{var}(u_{5j}) \text{wealth}_{ij}^2 \\ &= \sigma_{u0}^2 + 2\sigma_{u05} \text{wealth}_{ij} + \sigma_{u5}^2 \text{wealth}_{ij}^2 \end{aligned}$$

which is estimated as (substituting estimates of  $\sigma_{u0}^2$ ,  $\sigma_{u05}$  and  $\sigma_{u5}^2$ ):

$$0.843 - 0.219 \text{wealth}_{ij} + 0.016 \text{wealth}_{ij}^2.$$

We can now plot the between-community variance as a function of wealth:

```
> x <- seq(1, 5, 0.01)
> y <- 0.843 + (- 0.219) *x +0.016*x^2
> plot(x, y, type = "l", xlim = c(1, 5))
```



As expected from the ‘fanning in’ pattern of the community prediction lines, the between-community variance decreases as a function of wealth. Although the variance is a quadratic function of wealth, the plot mostly shows a linear decrease. This is because the coefficient of the linear term in the variance function ( $2\hat{\sigma}_{u05} = 0.219$ ) dominates over the coefficient of the quadratic term ( $\hat{\sigma}_{u5}^2 = 0.016$ ), to the extent that  $\hat{\sigma}_{u5}^2$  contributes relatively little to the total between-community variance.

### P7.5.3 Fitting random coefficients to categorical wealth

So far in this module, we have fitted a linear effect for **wealth**. Because **wealth** is a categorical variable, we might instead create dummy variables for four of the five categories and include these as explanatory variables. This is exactly what we did in our single-level analysis of P6.6.1, and our decision to treat **wealth** as a continuous variable was based on the finding that the coefficients of the **wealth** dummies showed an approximately linear increase. In the random slopes analysis of this

lesson, we have assumed that wealth has a linear relationship with the log-odds of antenatal care in all communities, but we have allowed the slope of this relationship to vary across communities. This implies that the between-community variance is a quadratic function of wealth, although our plot of the variance function shows that the variance depends linearly on wealth.

In this exercise, we will reassess the assumption that wealth has a linear relationship by once again fitting dummy variables. This allows for a more flexible community variance function, but comes at the cost of adding many more parameters to the model. We will therefore investigate whether we can simplify the model by having random coefficients on a subset of the **wealth** dummies.

Fit the following random intercept model:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealth2}_{ij} + \beta_6 \text{wealth3}_{ij} + \beta_7 \text{wealth4}_{ij} + \beta_8 \text{wealth5}_{ij} + u_j$$

```
> fit2 <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
+ wealth4 + wealth5 + (1 | comm), family = binomial("logit"), data = mydata,
glmerControl(calc.derivs = FALSE))
```

```
> summary(fit2)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
+ wealth4 + wealth5 + (1 | comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)
```

AIC	BIC	logLik	deviance	df.resid
5989.2	6055.1	-2984.6	5969.2	5356

```
Scaled residuals:
      Min       1Q   Median       3Q      Max
-3.7414 -0.6450  0.2408  0.6600  3.7267
```

```
Random effects:
Groups Name      Variance Std.Dev.
comm  (Intercept) 0.8228   0.9071
Number of obs: 5366, groups: comm, 361
```

```
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.2506911  0.1016554 -12.303  < 2e-16 ***
magec        -0.0012355  0.0065363  -0.189    0.850
magecsq      -0.0010014  0.0006804  -1.472    0.141
meduc2        0.5528446  0.0843463   6.554 5.58e-11 ***
meduc3        1.3024320  0.0973692  13.376 < 2e-16 ***
wealth2       0.4677772  0.1058749   4.418 9.95e-06 ***
wealth3       0.6853019  0.1080508   6.342 2.26e-10 ***
wealth4       1.0563029  0.1140943   9.258 < 2e-16 ***
wealth5       1.7730695  0.1330305  13.328 < 2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
      (Intr) magec  magecsq meduc2 meduc3 welth2 welth3 welth4
magec      0.035
magecsq   -0.251 -0.488
```

```

meduc2 -0.317 0.205 -0.045
meduc3 -0.228 0.312 -0.072 0.547
wealth2 -0.507 -0.010 -0.012 -0.098 -0.113
wealth3 -0.510 -0.030 0.021 -0.123 -0.194 0.544
wealth4 -0.483 -0.067 0.038 -0.165 -0.281 0.525 0.572
wealth5 -0.453 -0.119 0.060 -0.147 -0.325 0.462 0.507 0.550

```

As noted in the single-level analysis of P6.6.1, the log-odds of antenatal care increases with wealth quintile. The change is not completely linear because, for example, the difference between quintiles 2 and 1 is 0.468 while the difference between quintiles 3 and 2 is  $0.686 - 0.468 = 0.218$ . Nevertheless, there is a monotonic increasing relationship. We will now explore the relationship between the *probability* of antenatal care and wealth, adjusting for age and education. First, however, we will store the current results by copying the data into a new dataframe **mydatapred3**:

```
> mydatapred3 <- mydata
```

Change all values of **magec**, **meduc2** and **meduc3** to their mean values and set **magecsq** to the square of the mean value of **magec**:

```

> mydatapred3$magec <- 0
> mydatapred3$magecsq <- 0
> mean(mydatapred3$meduc2)
[1] 0.3073053
> mydatapred3$meduc2 <- mean(mydatapred3$meduc2)
> mean(mydatapred3$meduc3)
[1] 0.3449497
> mydatapred3$meduc3 <- mean(mydatapred3$meduc3)

```

Calculate the predicted probability for each woman based only on the fixed part of the model.

```

> X <- model.matrix(terms(fit2), mydatapred3)
> b <- fixef(fit2)
> predlogit <- X %*% b
> predprob <- logit(predlogit, inverse = TRUE)

```

Next we combine the predicted probability with the categorical variable **wealth** into a new dataframe:

```

> mydatapred3.2 <- unique(data.frame(cbind(predprob = predprob, wealth =
mydatapred3$wealth)))
> colnames(mydatapred3.2)[1] <- "predprob"

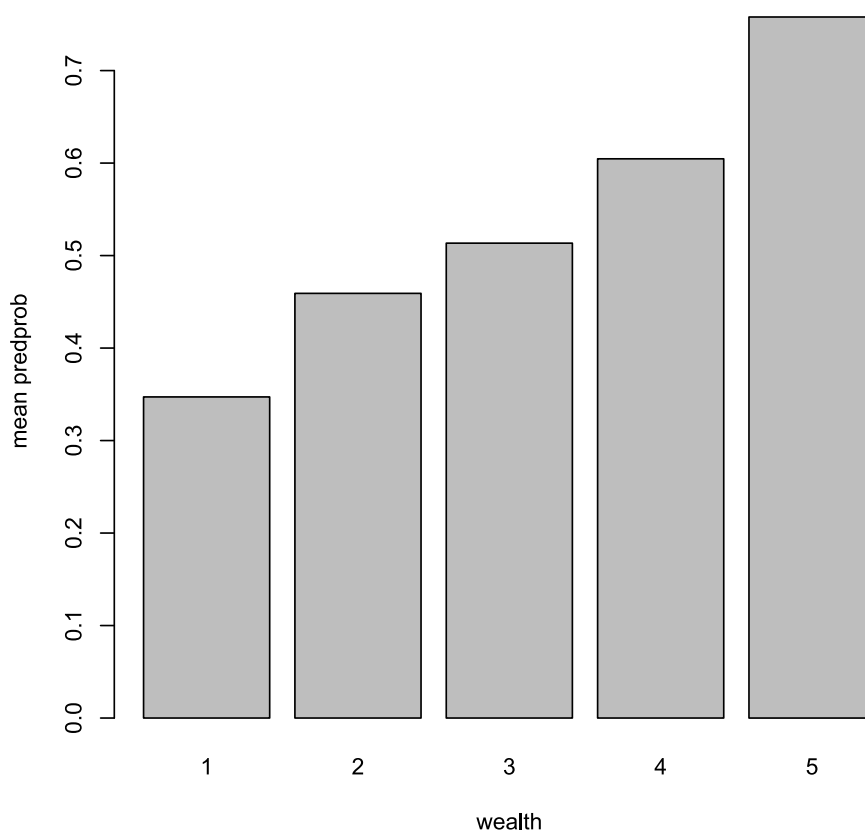
```

We then use the **tapply** command to calculate the mean predicted probability for each wealth category and use the **barplot** command to draw a vertical bar chart:

```
> mean.table <- tapply(mydatapred3.2$predprob, mydatapred3.2$wealth, mean)
```



```
> barplot(mean.table, ylab = "mean predprob", xlab = "wealth")
```



We can see that the relationship between the probability of antenatal care and wealth is fairly linear.

In the most general random coefficients<sup>14</sup> model, we could allow the coefficients of all four **wealth** dummies to vary across communities. However, this is a complex model which will lead to a  $5 \times 5$  covariance matrix at the community level with 14 more parameters than the random intercept model. We will start with a simpler model with a random coefficient for only **wealth5**. This model allows the probability of antenatal care to be different for each wealth category (as in the random intercept model above), but assumes that the only community variance in the relationship with wealth is in the difference between the 5<sup>th</sup> quintile and the other quintiles. For example, the difference between the 1<sup>st</sup> and 2<sup>nd</sup> quintiles and between the 2<sup>nd</sup> and 3<sup>rd</sup> quintiles is assumed to be the same in each community. A model with a random coefficient for only **wealth5** also implies that the between-community variance is the same for quintiles 1-4, but different for the top quintile.

Fit the model:

---

<sup>14</sup>Notice the switch in terminology from 'slope' to 'coefficient'. The two terms are often used interchangeably, but 'slope' is really only appropriate for a linear relationship. 'Coefficient' is a more general term that is more appropriate for a dummy variable, where the coefficient represents the difference between two groups rather than a linear effect.

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} \\ + \beta_5 \text{wealth2}_{ij} + \beta_6 \text{wealth3}_{ij} + \beta_7 \text{wealth4}_{ij} + \beta_8 \text{wealth5}_{ij} \\ + u_{0j} + u_{8j} \text{wealth5}_{ij}$$

```
> fit3 <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3
+ wealth4 + wealth5 + (1 + wealth5 | comm), family = binomial("logit"), data = mydata,
glmerControl(calc.derivs = FALSE))
```

```
> summary(fit3)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
wealth4 + wealth5 + (1 + wealth5 | comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)
```

AIC	BIC	logLik	deviance	df.resid
5985.0	6064.1	-2980.5	5961.0	5354

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.2977	-0.6402	0.2723	0.6335	3.7978

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
comm	(Intercept)	0.9291	0.9639	
	wealth5	0.3927	0.6267	-0.70

Number of obs: 5366, groups: comm, 361

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.2590740	0.1039234	-12.115	< 2e-16 ***
magec	-0.0005881	0.0065610	-0.090	0.929
magecsq	-0.0010319	0.0006827	-1.512	0.131
meduc2	0.5535044	0.0848292	6.525	6.80e-11 ***
meduc3	1.3135877	0.0975821	13.461	< 2e-16 ***
wealth2	0.4689487	0.1066669	4.396	1.10e-05 ***
wealth3	0.6808669	0.1089064	6.252	4.06e-10 ***
wealth4	1.0547594	0.1153322	9.145	< 2e-16 ***
wealth5	1.8447066	0.1338826	13.779	< 2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	magec	magecsq	meduc2	meduc3	welth2	welth3	welth4
magec		0.033						
magecsq	-0.245		-0.487					
meduc2	-0.311	0.206	-0.046					
meduc3	-0.223	0.312	-0.071	0.549				
wealth2	-0.500	-0.009	-0.014	-0.098	-0.113			
wealth3	-0.503	-0.029	0.020	-0.123	-0.194	0.544		
wealth4	-0.478	-0.066	0.037	-0.165	-0.279	0.523	0.571	
wealth5	-0.520	-0.117	0.058	-0.144	-0.314	0.457	0.497	0.529

The model assumes that the coefficients of **wealth2**, **wealth3** and **wealth4** are fixed across communities, estimated as 0.468, 0.681 and 1.055 respectively. The coefficient of **wealth5** (i.e. the difference between the top and bottom quintiles) varies across communities, and is estimated as  $1.845 + \hat{u}_{8j}$  in community  $j$ .

We can carry out a likelihood ratio test to test the null hypothesis of no community-variation in the difference between the top and other four quintiles ( $H_0: \sigma_{u8}^2 = \sigma_{u08} = 0$ ). The likelihood ratio test statistic is calculated as twice the difference in the log likelihood values between the model with and without the random slope for **wealth5**:

$$LR = 2(-2980.522 - -2984.586) = 8.156 \text{ on } 2 \text{ d.f.}$$

The 5% point of a chi-squared distribution on 2 d.f. is 5.99. We therefore conclude that the difference between the top and the other four quintiles of **wealth** does indeed vary across communities.

Rather than always calculating the likelihood ratio test by hand, R provides the `anova` command:

```
> anova(fit2, fit3)
Data: mydata
Models:
fit2: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
fit2:      wealth4 + wealth5 + (1 | comm)
fit3: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
fit3:      wealth4 + wealth5 + (1 + wealth5 | comm)
      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
fit2 10 5989.2 6055.1 -2984.6   5969.2
fit3 12 5985.0 6064.1 -2980.5   5961.0 8.1454      2    0.01703 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Where, the names of the restricted and unrestricted models are listed immediately after the command. The command confirms that the likelihood ratio test statistic is 8.156 while the p-value of 0.01694 confirms that the random effect for **wealth5** is indeed significant.

The between-community variance is now estimated as:

$$\begin{aligned} \text{var}(u_{0j} + u_{8j}\text{wealth5}_{ij}) &= \sigma_{u0}^2 + 2\sigma_{u08}\text{wealth5}_{ij} + \sigma_{u8}^2\text{wealth5}_{ij}^2 \\ &= 0.927 - 0.847\text{wealth5}_{ij} + 0.400\text{wealth5}_{ij}^2 \end{aligned}$$

which, because **wealth5** can only take values of 0 and 1, simplifies to:

0.927 in quintiles 1-4 (**wealth5** = 0) and

0.480 in quintile 5 (**wealth5** = 1)

We will now extend the model to include a random coefficient on **wealth4**. This model additionally allows the difference between the 4<sup>th</sup> quintile and the first three quintiles to vary across communities. Note this model takes a little longer to reach convergence.

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealth2}_{ij} + \beta_6 \text{wealth3}_{ij} + \beta_7 \text{wealth4}_{ij} + \beta_8 \text{wealth5}_{ij} + u_{0j} + u_{7j} \text{wealth4}_{ij} + u_{8j} \text{wealth5}_{ij}$$

```
> fit4 <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
+ wealth4 + wealth5 + (1 + wealth4 + wealth5 | comm), family = binomial("logit"),
data = mydata, glmerControl(calc.derivs = FALSE))
```

```
> summary(fit4)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
+ wealth4 + wealth5 + (1 + wealth4 + wealth5 | comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)
```

AIC	BIC	logLik	deviance	df.resid
5983.3	6082.1	-2976.6	5953.3	5351

```
Scaled residuals:
      Min       1Q   Median       3Q      Max
-3.3690 -0.6384  0.2695  0.6326  4.0748
```

```
Random effects:
Groups Name          Variance Std.Dev. Corr
comm  (Intercept)  1.0854    1.0418
      wealth4      0.1522    0.3901   -0.83
      wealth5      0.3982    0.6310   -0.76  0.28
Number of obs: 5366, groups: comm, 361
```

```
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.2639349  0.1071939 -11.791 < 2e-16 ***
magec        -0.0005964  0.0065704  -0.091  0.928
magecsq      -0.0010631  0.0006846  -1.553  0.120
meduc2        0.5586699  0.0850909   6.566 5.18e-11 ***
meduc3        1.3150983  0.0974114  13.500 < 2e-16 ***
wealth2       0.4785905  0.1079892   4.432 9.34e-06 ***
wealth3       0.6909265  0.1104151   6.258 3.91e-10 ***
wealth4       1.0203670  0.1161600   8.784 < 2e-16 ***
wealth5       1.8468572  0.1354939  13.631 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
      (Intr) magec  magcsq meduc2 meduc3 welth2 welth3 welth4
magec      0.032
magecsq    -0.238 -0.486
meduc2     -0.303  0.206 -0.045
meduc3     -0.218  0.310 -0.069  0.551
wealth2    -0.493 -0.008 -0.014 -0.097 -0.111
wealth3    -0.498 -0.028  0.020 -0.120 -0.190  0.544
wealth4    -0.552 -0.066  0.036 -0.167 -0.277  0.525  0.569
wealth5    -0.537 -0.114  0.056 -0.141 -0.308  0.458  0.498  0.550
```

The three new parameters are:  $\sigma_{u7}^2$  (the variance of the **wealth4** random effect),  $\sigma_{u07}$  (the covariance between the intercept and **wealth4** random effects) and  $\sigma_{u78}$  (the covariance between the **wealth4** and **wealth5** random effects).

We will carry out a joint significance test of the null hypothesis that all three new parameters are zero:

```

> anova(fit3, fit4)
Data: mydata
Models:
fit3: antemed ~ magedc + magedcsq + meduc2 + meduc3 + wealth2 + wealth3 +
fit3:      wealth4 + wealth5 + (1 + wealth5 | comm)
fit4: antemed ~ magedc + magedcsq + meduc2 + meduc3 + wealth2 + wealth3 +
fit4:      wealth4 + wealth5 + (1 + wealth4 + wealth5 | comm)
      Df      AIC      BIC    logLik deviance  Chisq Chi Df Pr(>Chisq)
fit3 12 5985.0 6064.1 -2980.5    5961.0
fit4 15 5983.3 6082.1 -2976.6    5953.3 7.7883      3 0.05059 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The test statistic is borderline significant at the 5% level. We therefore conclude that the difference between quintile 4 and the bottom 3 quintiles also varies across communities.

To see how the between-community variance depends on wealth:

$$\begin{aligned}
 \text{var}(u_{0j} + u_{7j}\text{wealth4}_{ij} + u_{8j}\text{wealth5}_{ij}) &= \sigma_{u0}^2 + 2\sigma_{u07}\text{wealth4}_{ij} + \sigma_{u7}^2\text{wealth4}_{ij}^2 \\
 &+ 2\sigma_{u08}\text{wealth5}_{ij} + 2\sigma_{u78}\text{wealth4}_{ij}\text{wealth5}_{ij} \\
 &+ \sigma_{u8}^2\text{wealth5}_{ij}^2 \\
 &= 1.088 - 0.684\text{wealth4}_{ij} + 0.155\text{wealth4}_{ij}^2 \\
 &- 1.006\text{wealth5}_{ij} + 0.138\text{wealth4}_{ij}\text{wealth5}_{ij} \\
 &+ 0.399\text{wealth5}_{ij}^2
 \end{aligned}$$

which, because **wealth5** can only take values of 0 and 1, simplifies to:

1.088 in quintiles 1-3 (**wealth4** = 0, **wealth5** = 0),

1.088-0.684 + 0.155 = 0.559 in quintile 4 (**wealth4** = 1, **wealth5** = 0) and

1.088-1.006 + 0.399 = 0.481 in quintile 5 (**wealth4** = 0, **wealth5** = 1)

Thus, the between-community variance is estimated as 1.088 for quintiles 1-3, 0.559 for the 4<sup>th</sup> quintile, and 0.481 for the top quintile. The greater amount of variation for the bottom three quintiles indicates that the community of residence has the strongest effect on the probability of receiving antenatal care for women whose households are in the bottom 60% of the distribution on the wealth index.

**Don't forget to take the online quiz!**

From within the LEMMA learning environment

- Go down to the section for **Module 7: Multilevel Models for Binary Responses**
- Click "[7.5 Two-level Random Slope Model](#)" to open Lesson 7.5
- Click [Q1](#) to open the first question

## P7.6 Adding Level 2 Explanatory Variables: Contextual Effects

So far in these exercises we have considered the effects of level 1 explanatory variables and fitted random slope models that allow their effects to vary across level 2 units (communities). As in the analysis of continuous responses, however, a major benefit of a multilevel modelling approach is the ability to include level 1 *and* level 2 explanatory variables. In particular, we are often interested in whether level 2 variables can explain level 2 variation.

Download the R dataset for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to **RDatasets and Rfiles**
- Right click “7.6.txt” and select **Save Link As...** to save the dataset onto your computer

Read the dataset into R and create a data frame object named **mydata**:

```
> mydata <- read.table("7.6.txt", header = TRUE, sep = ",")
```

Load the additional libraries needed for this session:

```
> library(VGAM)
Loading required package: stats4
Loading required package: splines

> library(lme4)
Loading required package: Matrix
```

The BDHS04 dataset contains one community-level variable: an indicator of whether the community is classified as urban or rural (**urban**). We can also derive community-level variables by aggregating woman-level variables, for example the proportion of women in a community whose households are in the top quintile of the wealth index.

### P7.6.1 Contextual effects

We will begin by adding **urban** to the model estimated at the end of P7.5.

$$\begin{aligned} \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = & \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} \\ & + \beta_5 \text{wealth2}_{ij} + \beta_6 \text{wealth3}_{ij} + \beta_7 \text{wealth4}_{ij} + \beta_8 \text{wealth5}_{ij} \\ & + \beta_9 \text{urban}_j \\ & + u_{0j} + u_{7j} \text{wealth4}_{ij} + u_{8j} \text{wealth5}_{ij} \end{aligned}$$

```
> fit <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
wealth4 + wealth5 + urban + (1 + wealth4 + wealth5 | comm), family = binomial("logit"),
data = mydata, glmerControl(calc.derivs = FALSE))
>
```

## Module 7 (R Practical): Multilevel Models for Binary Responses

```
> summary(fit)
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
  wealth4 + wealth5 + urban + (1 + wealth4 + wealth5 | comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)

      AIC      BIC   logLik deviance df.resid

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.7277 -0.6363  0.2638  0.6388  4.1211

Random effects:
Groups Name      Variance Std.Dev. Corr
comm  (Intercept) 0.9421   0.9706
      wealth4     0.1816   0.4261  -0.79
      wealth5     0.4116   0.6415  -0.89  0.43
Number of obs: 5366, groups:  comm, 361

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.5249571  0.1099292 -13.872 < 2e-16 ***
magec        0.0002553  0.0065784   0.039  0.969
magecsq      -0.0010937  0.0006858  -1.595  0.111
meduc2        0.5750461  0.0853002   6.741 1.57e-11 ***
meduc3        1.3674577  0.0978308  13.978 < 2e-16 ***
wealth2       0.4609123  0.1079850   4.268 1.97e-05 ***
wealth3       0.6530712  0.1103781   5.917 3.29e-09 ***
wealth4       0.9715344  0.1165770   8.334 < 2e-16 ***
wealth5       1.5312780  0.1366650  11.205 < 2e-16 ***
urban         0.9761124  0.1162618   8.396 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) magec  magcsq meduc2 meduc3 welth2 welth3 welth4 welth5
magec    0.024
magecsq -0.232 -0.486
meduc2   -0.312  0.207 -0.045
meduc3   -0.245  0.311 -0.068  0.554
wealth2  -0.472 -0.008 -0.014 -0.097 -0.112
wealth3  -0.470 -0.029  0.021 -0.121 -0.192  0.543
wealth4  -0.514 -0.066  0.036 -0.167 -0.280  0.522  0.565
wealth5  -0.435 -0.115  0.056 -0.152 -0.328  0.456  0.498  0.551
urban    -0.291  0.022 -0.002  0.053  0.111 -0.022 -0.044 -0.050 -0.278
```

Note this model may take a couple of minutes to fit. We find that urban women are more likely than rural women to use antenatal care services from a medically-trained provider. The intercept variance, representing the between-community variance for households in the bottom 60% (quintiles 1-3) of the wealth distribution, has decreased slightly from 1.088 to 0.942. To see how the between-community variance has changed in the top two quintiles:

$$\begin{aligned}
\text{var}(u_{0j} + u_{7j}\text{wealth4}_{ij} + u_{8j}\text{wealth5}_{ij}) &= \sigma_{u0}^2 + 2\sigma_{u07}\text{wealth4}_{ij} + \sigma_{u7}^2\text{wealth4}_{ij}^2 \\
&+ 2\sigma_{u08}\text{wealth5}_{ij} + 2\sigma_{u78}\text{wealth4}_{ij}\text{wealth5}_{ij} \\
&+ \sigma_{u8}^2\text{wealth5}_{ij}^2 \\
&= 0.942 - 0.652\text{wealth4}_{ij} + 0.179\text{wealth4}_{ij}^2 \\
&- 1.109\text{wealth5}_{ij} + 0.235\text{wealth4}_{ij}\text{wealth5}_{ij} \\
&+ 0.409\text{wealth5}_{ij}^2
\end{aligned}$$

which, because **wealth4** and **wealth5** can only take values of 0 and 1, simplifies to:

0.942 in quintiles 1-3 (**wealth4** = 0, **wealth5** = 0),

0.942 - 0.652 + 0.179 = 0.469 in quintile 4 (**wealth4** = 1, **wealth5** = 0) and

0.942 - 1.109 + 0.409 = 0.242 in quintile 5 (**wealth4** = 0, **wealth5** = 1)

For ease of comparison, the between-community variances for each model are given in Table 7.1.

*Table 7.1. Estimates of between-community variance from models with random coefficients for wealth*

Wealth quintile	Model without urban dummy	Model with urban dummy
1-3 (bottom 60%)	1.088	0.942
4	0.559	0.469
5 (top 20%)	0.481	0.242

The addition of **urban** has explained some community-level variation in antenatal care uptake for women in each wealth quintile, but the greatest reduction is in the top quintile: much of the community effect found among the richest women is explained by urban-rural differences in uptake. To shed further light on this finding, we will look at the relationship between **wealth** and **urban**.

Tabulate the two variables:

```

> table(mydata$wealth, mydata$urban)

      0      1
1  997  170
2  833  184
3  771  219
4  722  267
5  359  844

> table(mydata$wealth, mydata$urban)/cbind(rowSums(table(mydata$wealth,
mydata$urban)), rowSums(table(mydata$wealth, mydata$urban)))

      0      1
1 0.8543273 0.1456727
2 0.8190757 0.1809243
3 0.7787879 0.2212121
4 0.7300303 0.2699697
5 0.2984206 0.7015794

```



We find a strong relationship between wealth and type of region of residence: only 30% of the richest women live in rural areas, compared to between 73% and 85% of women in the bottom four wealth quintiles.

Next, we test whether there is a contextual effect of wealth. Does living in a better-off community (with a high proportion of households in the top wealth quintile) have an effect on a woman's chance of receiving antenatal care that is over and above the effect of her own economic status? We first need to aggregate the **wealth5** dummy to the community level after having sorted the data by **comm**.

```
> mydata <- mydata[order(mydata$comm), ]
> wealth5mean <- aggregate(mydata$wealth5, by = list(mydata$comm), mean)
> colnames(wealth5mean) <- c("commid", "wealth5mean")
```

We now add this new variable to our data frame **mydata**:

```
> mydata$wealth5mean <- rep(0, dim(mydata)[1])
> for (i in 1:dim(mydata)[1]){
  mydata$wealth5mean[i] <- wealth5mean$wealth5mean[wealth5mean$commid ==
mydata$comm[i]]
}
```

Our new variable, **wealth5mean**, contains the proportion of women in the community whose households are in the top wealth quintile. We will now add this variable to the model.

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} \\ + \beta_5 \text{wealth2}_{ij} + \beta_6 \text{wealth3}_{ij} + \beta_7 \text{wealth4}_{ij} + \beta_8 \text{wealth5}_{ij} \\ + \beta_9 \text{urban}_j + \beta_{10} \text{wealth5mean}_j \\ + u_{0j} + u_{7j} \text{wealth4}_{ij} + u_{8j} \text{wealth5}_{ij}$$

```
> fit2 <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3
+ wealth4 + wealth5 + urban + wealth5mean + (1 + wealth4 + wealth5 | comm),
family = binomial("logit"), data = mydata, glmerControl(calc.derivs = FALSE))
>
> summary(fit2)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + wealth2 + wealth3 +
wealth4 + wealth5 + urban + wealth5mean + (1 + wealth4 + wealth5 |
comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)

            AIC          BIC    logLik deviance df.resid
5889.8      6001.8    -2927.9   5855.8      5349

Scaled residuals:
      Min       1Q   Median       3Q      Max
-4.3556 -0.6359  0.2265  0.6391  4.0993

Random effects:
```

```

Groups Name          Variance Std.Dev. Corr
comm  (Intercept) 0.8592  0.9269
      wealth4    0.1902  0.4362  -0.80
      wealth5    0.3685  0.6071  -0.91  0.48
Number of obs: 5366, groups:  comm, 361

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.6102643  0.1098354 -14.661 < 2e-16 ***
magec       0.0002245  0.0065711  0.034 0.972750
magecsq     -0.0011805  0.0006871 -1.718 0.085763 .
meduc2      0.5844597  0.0852986  6.852 7.29e-12 ***
meduc3      1.3888135  0.0979696 14.176 < 2e-16 ***
wealth2     0.4287884  0.1076486  3.983 6.80e-05 ***
wealth3     0.6068659  0.1100949  5.512 3.54e-08 ***
wealth4     0.8888360  0.1165323  7.627 2.40e-14 ***
wealth5     1.1721891  0.1461471  8.021 1.05e-15 ***
urban       0.5043029  0.1396170  3.612 0.000304 ***
wealth5mean  1.4809148  0.2639049  5.612 2.01e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) magec  magecsq meduc2 meduc3 welth2 welth3 welth4 welth5 urban
magec        0.023
magecsq      -0.232 -0.483
meduc2       -0.318  0.206 -0.044
meduc3       -0.255  0.312 -0.069  0.556
wealth2      -0.464 -0.009 -0.014 -0.099 -0.115
wealth3      -0.456 -0.030  0.022 -0.124 -0.196  0.541
wealth4      -0.489 -0.067  0.037 -0.172 -0.289  0.520  0.563
wealth5      -0.323 -0.111  0.056 -0.156 -0.333  0.432  0.480  0.543
urban        -0.119  0.014  0.002  0.025  0.056 -0.005 -0.009  0.008  0.021
wealth5mean  -0.183  0.006 -0.009  0.033  0.064 -0.023 -0.048 -0.088 -0.389 -
0.582

```

We find that there is indeed a positive contextual effect of wealth that is over and above the positive effect of household wealth. The coefficient of **wealth5mean** (estimated as 1.483) is the difference in the expected log-odds of antenatal care for a woman in a community where *all* households are in the top wealth quintile and a woman in a community where *none* of the households are in the top quintile.

### P7.6.2 Cross-level interactions

The current model assumes that the contextual effect of wealth is the same for all women, regardless of their own wealth. We will modify this assumption to allow the effect of community wealth on a woman's chance of using antenatal services to depend on her own economic status. We do this by including in the model the interaction between individual wealth (represented by the dummies **wealth2** to **wealth5**) and community wealth (**wealth5mean**), a *cross-level interaction*. First we must generate the four interaction terms<sup>15</sup>:

```

> mydata$wealth2Xwealth5mean <- mydata$wealth2*mydata$wealth5mean
> mydata$wealth3Xwealth5mean <- mydata$wealth3*mydata$wealth5mean

```

<sup>15</sup> An alternative approach to specifying interaction terms in R is to use the factor command:

```

> mydata$wealthf<-factor(mydata$wealth)

```

And to use the interaction symbol ':' in the model specification, that is `wealthf:wealth5mean`.

## Module 7 (R Practical): Multilevel Models for Binary Responses

```
> mydata$wealth4Xwealth5mean <- mydata$wealth4*mydata$wealth5mean
> mydata$wealth5Xwealth5mean <- mydata$wealth5*mydata$wealth5mean
```

Now fit the model:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 \text{magec}_{ij} + \beta_2 \text{magecsq}_{ij} + \beta_3 \text{meduc2}_{ij} + \beta_4 \text{meduc3}_{ij} + \beta_5 \text{wealth2}_{ij} + \beta_6 \text{wealth3}_{ij} + \beta_7 \text{wealth4}_{ij} + \beta_8 \text{wealth5}_{ij} + \beta_9 \text{urban}_j + \beta_{10} \text{wealth5mean}_j + \beta_{11} \text{wealth2Xwealth5mean}_{ij} + \beta_{12} \text{wealth3Xwealth5mean}_{ij} + \beta_{13} \text{wealth4Xwealth5mean}_{ij} + \beta_{14} \text{wealth5Xwealth5mean}_{ij} + u_{0j} + u_{7j} \text{wealth4}_{ij} + u_{8j} \text{wealth5}_{ij}$$

```
> fit3 <- glmer(antemed ~ magec + magecsq + meduc2 + meduc3 + urban + wealth2 +
wealth3 + wealth4 + wealth5 + wealth5mean
+ + wealth2Xwealth5mean + wealth3Xwealth5mean + wealth4Xwealth5mean +
wealth5Xwealth5mean + (1 + wealth4 + wealth5 | comm), family =
binomial("logit"), data = mydata, glmerControl(calc.derivs = FALSE))
>
> summary(fit3)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
Family: binomial ( logit )
Formula: antemed ~ magec + magecsq + meduc2 + meduc3 + urban + wealth2 +
wealth3 + wealth4 + wealth5 + wealth5mean + wealth2Xwealth5mean +
wealth3Xwealth5mean + wealth4Xwealth5mean + wealth5Xwealth5mean + (1 +
wealth4 + wealth5 | comm)
Data: mydata
Control: glmerControl(calc.derivs = FALSE)

            AIC          BIC      logLik deviance df.resid
5890.0      6028.3    -2924.0    5848.0      5345

Scaled residuals:
    Min       1Q   Median       3Q      Max
-4.0727 -0.6323  0.2449  0.6375  4.3211

Random effects:
Groups Name      Variance Std.Dev. Corr
comm  (Intercept) 0.8590   0.9268
      wealth4     0.1885   0.4341  -0.81
      wealth5     0.3682   0.6068  -0.94  0.56
Number of obs: 5366, groups:  comm, 361

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.7734198   0.1253549  -14.147 < 2e-16 ***
magec          -0.0003918   0.0065826   -0.060 0.952540
magecsq        -0.0010739   0.0006863   -1.565 0.117607
meduc2          0.5843343   0.0854684    6.837 8.10e-12 ***
meduc3          1.3796250   0.0979783   14.081 < 2e-16 ***
urban           0.4763595   0.1391269    3.424 0.000617 ***
wealth2         0.5520197   0.1307482    4.222 2.42e-05 ***
wealth3         0.6673339   0.1368063    4.878 1.07e-06 ***
wealth4         1.0850417   0.1467749    7.393 1.44e-13 ***
wealth5         1.5080871   0.2039792    7.393 1.43e-13 ***
wealth5mean     2.9771258   0.6771692    4.396 1.10e-05 ***
wealth2Xwealth5mean -1.2262980   0.7682569   -1.596 0.110442
wealth3Xwealth5mean -0.6934160   0.7563297   -0.917 0.359238
wealth4Xwealth5mean -1.5971458   0.7214243   -2.214 0.026837 *
wealth5Xwealth5mean -1.7935757   0.7175061   -2.500 0.012429 *
```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation matrix not shown by default, as p = 15 > 12.
Use print(x, correlation=TRUE) or
vcov(x) if you need it

```

Comparing the estimated coefficient of each interaction term with its standard error we find that only **wealth4Xwealth5mean** and **wealth5Xwealth5mean** are statistically significant at the 5% level. To interpret the interaction effect, we will compute predicted probabilities of receiving antenatal care for different combinations of **wealth** and **wealth5mean**, holding **mage**, **meduc** and **urban** fixed at their sample means. We will consider community proportions in the top wealth quintile of 0, 0.2, and 0.4 (note that many communities have no sample women in this quintile).

First we create a new data frame, **mydatapred**, and set **magec**, **meduc** and **urbanequal** to their mean values and set **magecsq** to the square of the sample mean for **magec**:

```

> mydatapred <- mydata
> mydatapred$magec <- 0
> mydatapred$magecsq <- 0
> mydatapred$meduc2 <- mean(mydatapred$meduc2)
> mydatapred$meduc3 <- mean(mydatapred$meduc3)
> mydatapred$urban <- mean(mydatapred$urban)

```

Next we recode **wealth5mean** to take one of the three proportions of interest: 0, 0.2, or 0.4. One way of doing this is to recode all values of **wealth5mean** in the range 0 to 0.1 to the first value, all values of **wealth5mean** in the range 0.1 to 0.3 to the second value and all values of **wealth5mean** in the range 0.3 to 1.0 to the last value:

```

> mydatapred$wealth5mean[mydatapred$wealth5mean >= 0 & mydatapred$wealth5mean <
0.1] <- 0
> mydatapred$wealth5mean[mydatapred$wealth5mean >= 0.1 & mydatapred$wealth5mean
< 0.3] <- 0.2
> mydatapred$wealth5mean[mydatapred$wealth5mean >= 0.3 & mydatapred$wealth5mean
<= 1] <- 0.4

```

We must also change the values of the four interaction terms to reflect the changes we have just made to the **wealth5mean** variable:

```

> mydatapred$wealth2Xwealth5mean <- mydatapred$wealth2*mydatapred$wealth5mean
> mydatapred$wealth3Xwealth5mean <- mydatapred$wealth3*mydatapred$wealth5mean
> mydatapred$wealth4Xwealth5mean <- mydatapred$wealth4*mydatapred$wealth5mean
> mydatapred$wealth5Xwealth5mean <- mydatapred$wealth5*mydatapred$wealth5mean

```

Now that we have set all the explanatory variables to their specified values, we can calculate the median (i.e cluster specific) predicted probability for each woman (see P7.4 for a comparison of population average and cluster specific predicted probabilities).

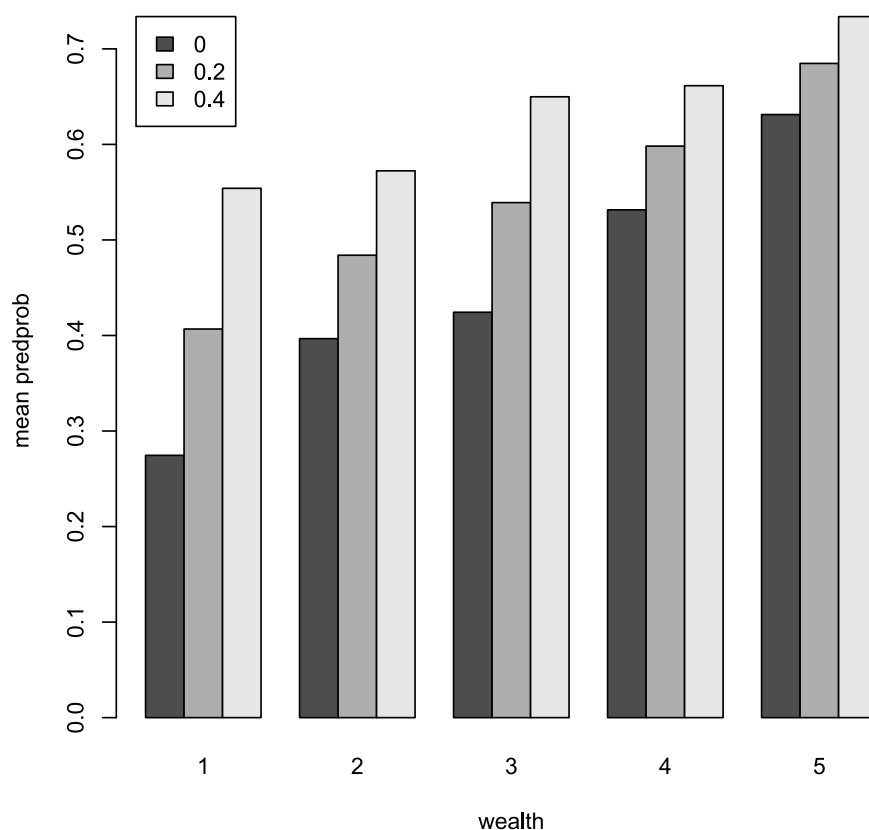
```
> X <- model.matrix(terms(fit3), mydatapred)
> b <- fixef(fit3)
> predlogit <- X %*% b
> predprob <- logit(predlogit, inverse = TRUE)
```

We then combine the predicted probabilities and the values of **wealth5mean** into the **mydatapred2** dataframe, and calculate the mean probability for each **wealth5mean** value using the **tapply** command:

```
> mydatapred2 <- unique(data.frame(cbind(predprob = predprob, wealth5mean =
mydatapred$wealth5mean, wealth = mydatapred$wealth)))
> colnames(mydatapred2)[1] <- c("predprob")
> mean.table <- tapply(mydatapred2$predprob, list(mydatapred2$wealth5mean,
mydatapred2$wealth), mean)
```

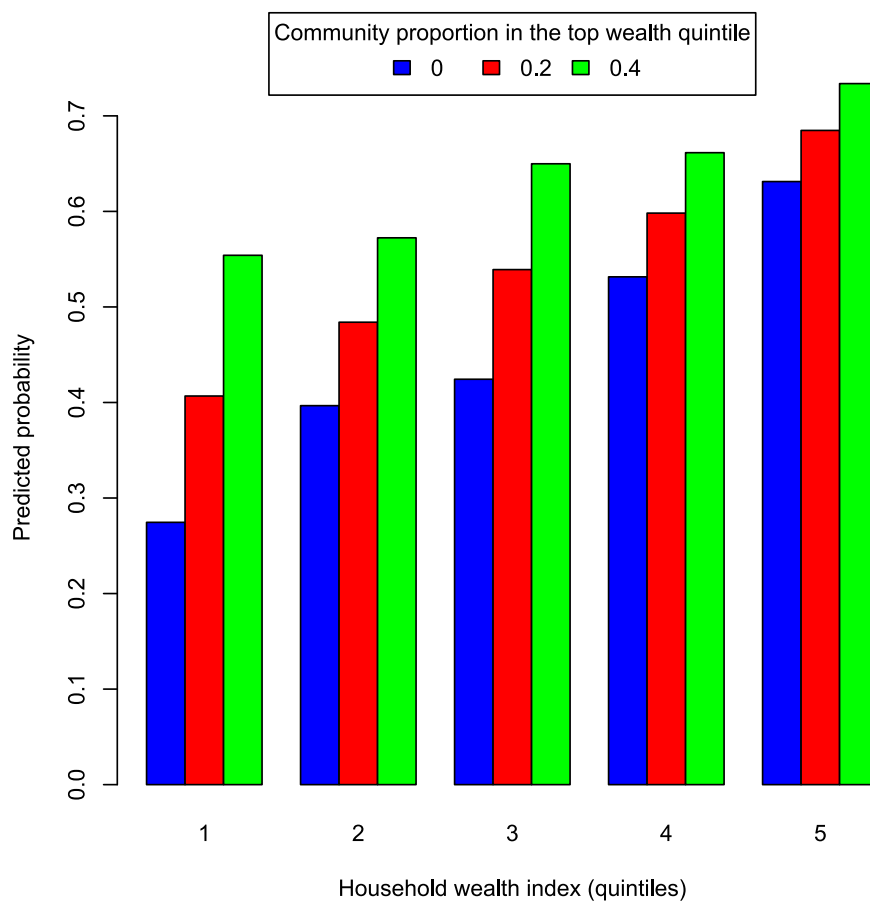
We will now plot the predictions, using the **barplot** command. This gives a total of 15 vertical bars:

```
> barplot(mean.table, ylab = "mean predprob", xlab = "wealth", beside = TRUE,
axes = TRUE, legend = TRUE, args.legend = list(x = "topleft", inset = c(0.025,
0)))
```



The above graph is not very attractive as the three values of `wealth5mean` are all plotted using different shades of grey. There are also no titles indicating what the values on the x-axis refer to. We can improve this graph by adding extra options to the `barplot` command. For example, we add the `title` option to the `args.legend` option to display a title for the legend.

```
> barplot(mean.table, ylab = "Predicted probability", xlab = "Household wealth
index (quintiles)", beside = TRUE, axes = TRUE, col = c("blue", "red", "green"),
legend = TRUE, args.legend = list(x = "top", inset = c(0, -0.1), title =
"Community proportion in the top wealth quintile", horiz = TRUE))
```



We can see from the predictions plot that the contextual effect of wealth (i.e. the difference in predicted probabilities for **wealth5mean** values of 0, 0.2 and 0.4) is weaker among women in the top 40% of the wealth distribution (**wealth4** and **wealth5**). Living in a deprived community (as indicated by a low value on **wealth5mean**) as opposed to a better-off community is more of a barrier to using antenatal services for poorer women. Alternatively, but equivalently, we can say that the effect of individual wealth is stronger in poorer communities. The two interpretations are both consistent with the predictions plot; they are just different ways of viewing the same information.

**Don't forget to take the online quiz!**

From within the LEMMA learning environment

- Go down to the section for **Module 7: Multilevel Models for Binary Responses**
- Click "[7.6 Adding Level 2 Explanatory Variables: Contextual Effects](#)" to open Lesson 7.6
- Click [Q1](#) to open the first question

## **P7.7 Estimation of Binary Response Models: MCMC Methods**

Binary response models can be fitted using MCMC methods in R using the additional library `MCMCglmm` which can be downloaded in the usual way. Description of how to fit such models is out of the scope of this document but we refer the interested reader to Hadfield (2010) and the `MCMCglmm` course vignettes on the following webpage: <http://cran.r-project.org/web/packages/MCMCglmm/>.

**Don't forget to take the online quizzes for this module if you haven't already done so! (see page 1 for details for how to find the quizzes)**

## **References**

Hadfield, J.D. (2010) Journal of Statistical Software, 33, 2, 1-22.