Module 7: Multilevel Models for Binary Responses

Stata Practical

George Leckie[1]
Centre for Multilevel Modelling

Pre-requisites
• Modules 1-6

Contents

---

[1] This Stata 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),[1] 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.

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

The file contains the following variables:

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

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

Load “7.1.dta” into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***Stata* Datasets and Do-files**
- Click “7.1.dta” to open the dataset

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

```
. describe

Contains data from 7.1.dta
  obs:          5,366
 vars:             17                          5 Sep 200909:38
 size:        177,078 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
comm            int    %9.0g                  Community ID
womid           int    %9.0g                  Woman ID
antemed         byte   %9.0g                  Antenatal from qualified medic
bord            byte   %9.0g                  Birth order
mage            byte   %9.0g                  Mother's age at birth
urban           byte   %9.0g                  Type of region of residence
meduc           byte   %9.0g                  Maternal education
islam           byte   %9.0g                  Religion
wealth          byte   %9.0g                  Wealth index (1=poorest)
magec           float  %9.0g
magecsq         float  %9.0g
meduc2          byte   %8.0g
meduc3          byte   %8.0g
wealth2         byte   %8.0g
wealth3         byte   %8.0g
wealth4         byte   %8.0g
wealth5         byte   %8.0g
-------------------------------------------------------------------------------
Sorted by:
```

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.

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

The intercept β_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 σ^2_{u0}.

Stata's main command for fitting multilevel models for binary response variables is the `xtmelogit` command.[2]The syntax for `xtmelogit` is similar to that for `xtmixed`. To fit the above model using the `xtmelogit` command, we type: `xtmelogit antemed || comm:, variance`.

The binary response variable (**antemed**) follows the command which is then followed by the list of fixed part explanatory variables (excluding the constant as this is included by default[3]).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 after two vertical bars `||`. The level 2 identifier (**comm**) is specified first followed by a colon and then the list of random part explanatory variables (again excluding the constant as this is included by default). Finally, the `variance` option reports the variances of the random intercept and any random coefficients included in the model (as opposed to the default of standard deviations).

Issuing the `xtmelogit` command gives the following output:

```
. xtmelogit antemed || comm:, variance

Refining starting values:

Iteration 0:   log likelihood = -3321.6208
Iteration 1:   log likelihood = -3313.2849
Iteration 2:   log likelihood = -3313.2818

Performing gradient-based optimization:

Iteration 0:   log likelihood = -3313.2818
Iteration 1:   log likelihood = -3313.2817

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(0)       =         .
Log likelihood = -3313.2817                     Prob > chi2        =         .

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .1486212   .0727516     2.04   0.041     .0060307    .2912118
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
```

[2] Note, two-level random intercept logit models can equally be fitted with the `xtlogit` command; see `help xtlogit`. To fit the equivalent model with the probit link function, see `help xtprobit`. We do not discuss the `xtlogit` or `xtprobit` commands as they cannot be used to fit more complicated multilevel models while `xtmelogit` can. However, we do note that `xtlogit` fits models considerably faster than `xtmelogit` and is therefore recommended for fitting two-level random intercept logit models. See Rabe-Hesketh and Skrondal (2008) for examples of two-level random intercept models fitted with both commands.

[3] Note, the `noconstant` option can be used to omit the constant from the fixed or the random part of the models; see `help xtmelogit`.

```
comm: Identity               |
                  var(_cons) |   1.502371   .1591921      1.220628    1.849145
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   808.64 Prob>=chibar2 = 0.0000
```

Before interpreting the model, we will discuss the estimation procedure that `xtmelogit` uses. As will be described in C7.7 (and in more detail in the Technical Appendix), there are several estimation procedures available for binary and other categorical response models. However, in Stata, only one procedure is implemented: maximum likelihood estimation using adaptive quadrature. As with the other procedures, this is an approximate method and so it is always important to assess whether the approximation is adequate. By default, `xtmelogit` uses adaptive quadrature with 7 integration points. To check that 7 integration points is adequate, the model can be refitted with a larger number of quadrature points (the `intpoints()` option is used to do this). If the two sets of model parameters are substantially the same, then7 integration points is adequate. It might also be the case that 7 integration points is more than adequate, in which case the model can be fitted with fewer points.

Table 7.1 gives the parameter estimates which are obtained for the above model when different numbers of integration points are specified: 1, 2, 3, 4, 5, 6, 7 and 15. The percentage difference between each parameter estimate and its most accurate estimate (i.e. when 15 integration points are used) is also reported.[4]The last row of the table reports the time (in seconds) that it takes for the model to converge.[5]

Table 7.1. Estimates for different numbers of integration points reported with the percentage difference between each estimate and that based on 15 integration points

Parameter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**15**
$\hat{\beta}_0$	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149
	-0.7%	-0.7%	-0.7%	0.0%	0.0%	0.0%	0.0%	
$\hat{\sigma}^2_{u0}$	1.464	1.464	1.483	1.501	1.500	1.502	1.502	1.503
	-2.6%	-2.6%	-1.3%	-0.1%	-0.2%	-0.1%	-0.1%	
Log likelihood	-3318	-3317	-3314	-3313	-3313	-3313	-3313	-3313
Seconds	3.8	3.3	3.4	2.8	2.9	3.1	2.4	3.3

The table shows that when 1 integration point is used, the constant is 0.7% smaller than when 15 points are used while the between-community variance is 2.6% smaller than its corresponding value. However, increasing the number of integration points to 4 gives an estimate for the variance which is only 0.1% smaller than when 15

[4] Note that using a higher number of integration points than 15 will lead to more accurate estimates. However, the results in Table 7.1 suggest that little will be gained by doing this.
[5] We used a 64bit 2-core multiprocessor version of Stata 11 on a 2.66Ghz Intel Xeon X7460 running on Windows Server 2008.

integration points are used. All 8 models took a similar length of time to converge. We will therefore continue to use the default setting of 7 integration points in the following random intercept models. However, when we come to specify random slope models (P7.5), we will revisit this issue to check that 7 integration points is still appropriate.

P7.1.2 Interpretation of the null two-level model

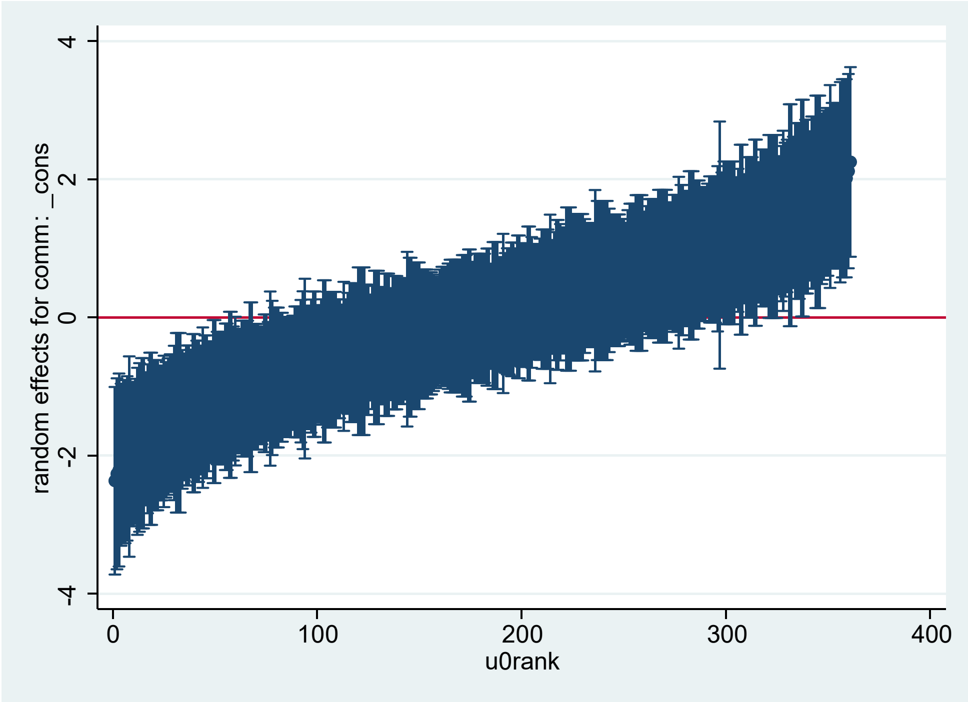
From the model estimates (using 7 integration points), we can say that the log-odds of receiving antenatal care from a medically-trained provider in an 'average' community (one with $u_{0j} = 0$) is estimated as $\hat{\beta}_0 = 0.149$. The intercept for community *j* is $0.149 + u_{0j}$, where the variance of $u_{0j}$is estimated as $\hat{\sigma}^2_{u0} = 1.502$.

The likelihood ratio statistic for testing the null hypothesis that $\sigma^2_{u0} = 0$ is reported in the final line of the output. The test statistic is 808.64 with a corresponding p-value of less than 0.00005 and so there is strong evidence that the between-community variance is non-zero.[6]

We will now examine estimates of the community effects or residuals, $\hat{u}_{0j}$, obtained from the null model. To calculate the residuals and produce a 'caterpillar plot' with the community effects shown in rank order together with 95% confidence intervals we can use the same commands as we used P5.1.2 for the continuous response two-level random intercepts model. (Note, the `predict` command's `reses` option is available as of Stata 11.):

```
. predict u0, reffects

. predict u0se, reses

. egen pickone = tag(comm)

. sort u0

. generate u0rank = sum(pickone)

. serrbar u0u0seu0rank if pickone==1, scale(1.96) yline(0)
```

[6]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. For this model, Stata automatically reports the correct p-value.



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.

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

```
. xtmelogit antemed magec || comm:, variance

Refining starting values:

Iteration 0:   log likelihood = -3302.4674
Iteration 1:   log likelihood = -3294.2858
Iteration 2:   log likelihood =   -3294.28

Performing gradient-based optimization:

Iteration 0:   log likelihood =   -3294.28
Iteration 1:   log likelihood = -3294.2799

Mixed-effects logistic regression               Number of obs      =      5366
```

```
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(1)       =     37.44
Log likelihood = -3294.2799                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
magec |   -.0323352   .0052842    -6.12   0.000    -.0426921   -.0219783
       _cons |   .1451359    .072752     1.99   0.046     .0025447    .2877272
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |     1.5001   .1592894      1.218245    1.847165
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   796.58 Prob>=chibar2 = 0.0000
```

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:

$$\log\left(\frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}}\right) = 0.145 - 0.032\ \mathrm{magec}_{ij}$$

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 `predict` command with no options specified (no options are required as the default prediction type `mu` gives the predicted probabilities for each woman based on the fixed combined with the random parts of the model).

```
. predict predprob
(option mu assumed; predicted means)
```

We then transform these predicted probabilities to predicted log odds using the `logit()` function (unfortunately the `predict` command does not have an option to compute these community specific log-odds directly).

```
. generate predlogit = logit(predprob)
```

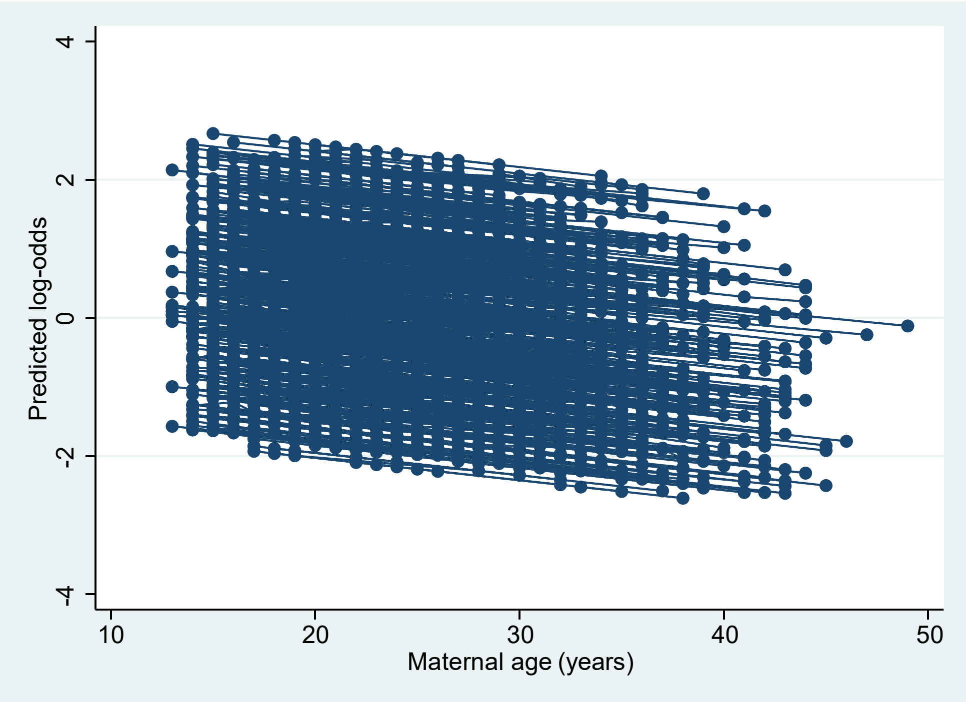
Next we drop the pre-existing **pickone** variable and then create a new **pickone** variable which picks one observation for each distinct value of **mage** found within each community.

```
. drop pickone
```

```
. egen pickone = tag(comm mage)
```

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

```
. twoway connected predlogit mage if pickone==1, connect(ascending) ///
>         ytitle(Predicted log-odds) xtitle(Maternal age (years))
```



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**:

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

```
. xtmelogit antemed magecmagecsq || comm:, variance

Refining starting values:

Iteration 0:   log likelihood = -3302.0625
Iteration 1:   log likelihood =  -3293.176
Iteration 2:   log likelihood = -3293.0189

Performing gradient-based optimization:

Iteration 0:   log likelihood = -3293.0189
Iteration 1:   log likelihood = -3293.0188

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361
```

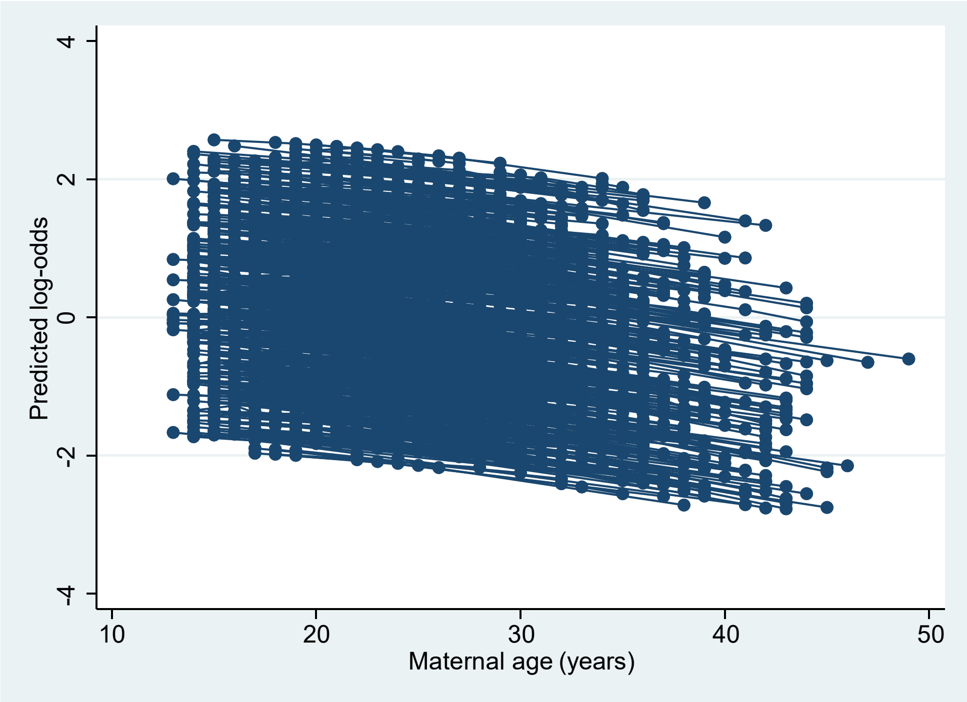
```
                                                Obs per group: min =          3
                                                               avg =       14.9
                                                               max =         25

Integration points =   7                        Wald chi2(2)       =      39.35
Log likelihood = -3293.0188                     Prob > chi2        =     0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
magec        |  -.0275969   .0060727    -4.54   0.000    -.0394991   -.0156947
magecsq      |  -.0010567   .0006684    -1.58   0.114    -.0023668    .0002533
       _cons |    .185073   .0768065     2.41   0.016      .034535    .3356111
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |   1.489691   .1585163      1.209263     1.83515
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   784.87 Prob>=chibar2 = 0.0000
```

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. You will first need to drop the old **predprob** and **predlogit** variables. However, there is no need to drop and recreate the old **pickone** variable.

```
. drop predprob predlogit

. predict predprob
(option mu assumed; predicted means)

. generate predlogit = logit(predprob)

. twoway connected predlogit mageif pickone==1, connect(ascending) ///
>         ytitle(Predicted log-odds) xtitle(Maternal age (years))
```

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.

Load “7.2.dta” into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***Stata* Datasets and Do-files**
- Click “7.2.dta” to open the dataset

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 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**:

```
. summarize wealth

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      wealth |      5366      3.0082    1.463163          1          5

. generate wealthc = wealth - 3.008
```

Next, fit the model:

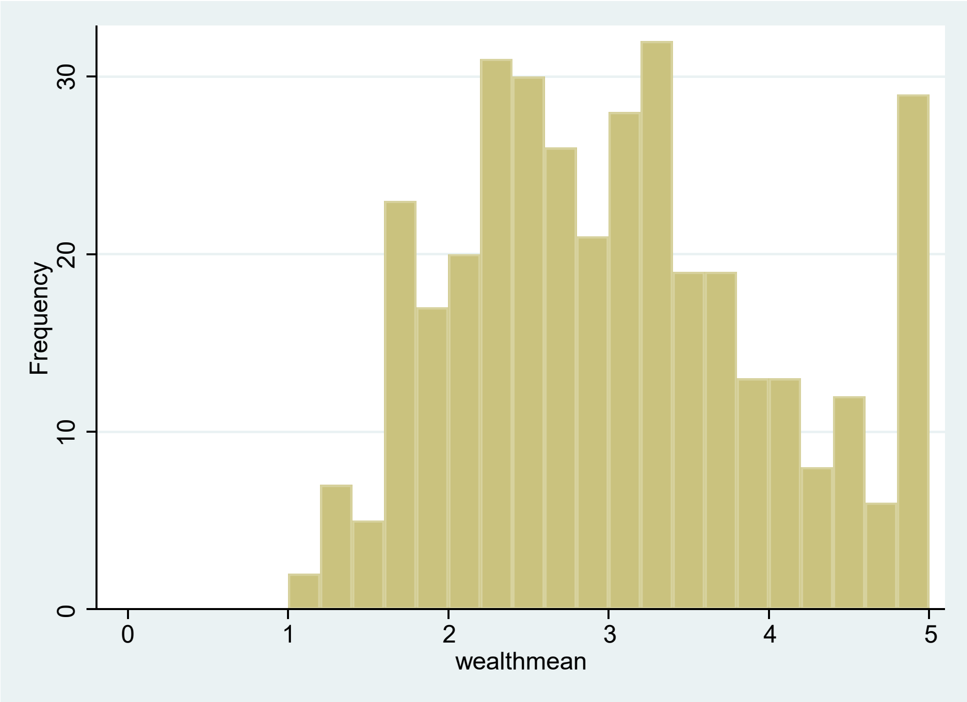
$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 \mathrm{magec}_{ij} + \beta_2 \mathrm{magecsq}_{ij}$$
$$+\beta_3 \mathrm{meduc2}_{ij} + \beta_4 \mathrm{meduc3}_{ij} + \beta_5 \mathrm{wealthc}_{ij} + u_{0j}$$

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 wealthc ///
>          || comm:, variance

Refining starting values:

Iteration 0:   log likelihood =  -2994.175
Iteration 1:   log likelihood = -2987.3033
```

```
Iteration 2:   log likelihood = -2987.2587

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2987.2587
Iteration 1:   log likelihood = -2987.2586

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(5)       =    591.90
Log likelihood = -2987.2586                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0002284   .0066466    -0.03   0.973    -.0132555    .0127987
     magecsq |  -.0010115   .0006909    -1.46   0.143    -.0023656    .0003426
      meduc2 |   .5509808   .0863579     6.38   0.000     .3817224    .7202392
      meduc3 |   1.315807   .1002754    13.12   0.000     1.119271    1.512343
     wealthc |   .3977915   .0302409    13.15   0.000     .3385204    .4570627
       _cons |  -.4559899   .0842324    -5.41   0.000    -.6210824   -.2908973
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |   .8886033   .1105534      .6963182    1.133987
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   323.64 Prob>=chibar2 = 0.0000
```

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 `histogram` command, we specify the `width()` option to set the width of the bins to 0.2 while we set the `start()` option to set the theoretical minimum mean community income (**wealthmean**) to 0.

```
. bysort comm: egen wealthmean = mean(wealth)

. egen pickone = tag(comm)

. histogram wealthmean if pickone==1, width(0.2) start(0) frequency
(bin=25, start=0, width=.2)
```



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) but first we will store the results from the multilevel model as we will wish to return to these later.

```
. estimates store multilevel
```

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

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

```
. logit antemed magecmagecsq meduc2 meduc3 wealthc

Iteration 0:   log likelihood = -3717.6012
Iteration 1:   log likelihood = -3149.4921
Iteration 2:   log likelihood = -3149.0774
Iteration 3:   log likelihood = -3149.0774

Logistic regression                               Number of obs   =       5366
                                                  LR chi2(5)      =    1137.05
                                                  Prob > chi2     =     0.0000
Log likelihood = -3149.0774                       Pseudo R2       =     0.1529

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0046213   .0059102    -0.78   0.434    -.0162051    .0069625
     magecsq |  -.0013503   .0006307    -2.14   0.032    -.0025864   -.0001143
      meduc2 |   .3968979   .0757902     5.24   0.000     .2483518    .5454439
```

```
     meduc3 |   1.072526   .0849994    12.62   0.000     .9059303    1.239122
    wealthc |   .4819516   .0239126    20.15   0.000     .4350838    .5288194
      _cons |  -.3685847   .0593246    -6.21   0.000    -.4848588   -.2523106
------------------------------------------------------------------------------
```

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 1.388 for **meduc2** and 1.226 for **meduc3**. From equation (7.8) in C7.2.2 we would expect the ratio to be $\sqrt{(0.889 \ + 3.29)/3.29} =$ 1.127, 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 `estimates replay` command to redisplay that model's results without having to reestimate it.

```
. estimates replay multilevel, variance

--------------------------------------------------------------------------------
Model multilevel
--------------------------------------------------------------------------------
Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(5)       =    591.90
Log likelihood = -2987.2586                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
    antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      magec |  -.0002284   .0066466    -0.03   0.973    -.0132555    .0127987
    magecsq |  -.0010115   .0006909    -1.46   0.143    -.0023656    .0003426
     meduc2 |   .5509808   .0863579     6.38   0.000     .3817224    .7202392
     meduc3 |   1.315807   .1002754    13.12   0.000     1.119271    1.512343
    wealthc |   .3977915   .0302409    13.15   0.000     .3385204    .4570627
      _cons |  -.4559899   .0842324    -5.41   0.000    -.6210824   -.2908973
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |   .8886033   .1105534      .6963182    1.133987
------------------------------------------------------------------------------
```

```
LR test vs. logistic regression: chibar2(01) =   323.64 Prob>=chibar2 = 0.0000
```

The variance partition coefficient (VPC) is calculated as 0.889/(0.889+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 1.502, giving a VPC of 1.502/(1.502+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

Load “7.4.dta” into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***Stata*** **Datasets and Do-files**
- Click “7.4.dta” to open the dataset

First we will refit the multilevel model estimated at the beginning of P7.2 with mother’s age, education and wealth as predictors:[7]

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

If you have fitted the above model in the current session of Stata, there is no need to refit the model. Instead of running the `xtmelogit` command below, simply type `estimates replay multilevel` followed by `estimates restore multilevel` to redisplay the model’s results and to then restore it to being the active model.

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 wealthc ///
>         || comm:, mle variance

Refining starting values:

Iteration 0:   log likelihood =  -2994.175
Iteration 1:   log likelihood = -2987.3033
Iteration 2:   log likelihood = -2987.2587

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2987.2587
Iteration 1:   log likelihood = -2987.2586

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(5)       =    591.90
Log likelihood = -2987.2586                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0002284   .0066466    -0.03   0.973    -.0132555    .0127987
     magecsq |  -.0010115   .0006909    -1.46   0.143    -.0023656    .0003426
```

[7] Note for ease of exposition in this section, we have removed the 0 subscript from the level 2 random intercept effect.

```
     meduc2 |   .5509808   .0863579     6.38   0.000     .3817224    .7202392
     meduc3 |   1.315807   .1002754    13.12   0.000     1.119271    1.512343
    wealthc |   .3977915   .0302409    13.15   0.000     .3385204    .4570627
      _cons |    -.45599   .0842324    -5.41   0.000    -.6210825   -.2908974
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |   .8886033   .1105534      .6963182    1.133987
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   323.64 Prob>=chibar2 = 0.0000
```

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

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.[8] 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):

$$\begin{aligned} z_{ij} &= -0.456 - 0.0002\ \mathrm{magec}_{ij} - 0.001\ \mathrm{magecsq}_{ij} \\ &+0.551\ \mathrm{meduc2}_{ij} + 1.316\ \mathrm{meduc3}_{ij} + 0.398\ \mathrm{wealthc}_{ij} \end{aligned}$$

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 **magec** is zero by definition because **magec** is maternal age centred on its sample mean. The variable **magecsq** is the square of **magec** so **magecsq** must be fixed at the square of the value we fix **magec** at. We therefore set **magec** =0 and **magecsq** =0 for all women in the sample:

```
. replace magec = 0
(5366 real changes made)

. replace magecsq = 0
(5366 real changes made)
```

Next we use the `predict` command with the `fixedonly` option to calculate the predicted probability for each woman in the sample based only on the fixed part of the model. The random part of the model is ignored making this equivalent to setting $u_j = 0$.

[8] For ease we have dropped the 0 subscript from the level 2 random intercept effect.

```
. predict medianpredprob, fixedonly
(option mu assumed; predicted means)
```

Finally we list the predicted probabilities for each combination of wealth and education. In the `list` command, we specify the `abbreviate()` option with the value 14 to prevent variable names of up to 14 characters from being abbreviated (the default option is to abbreviate variable names longer than 8 characters). The `sepby()` option is specified to draw a separator line whenever **wealth** changes value.

```
. egen pickone = tag(wealth meduc)

. sort wealth meduc

. list wealth meduc medianpredprob if pickone==1, abbreviate(14) sepby(wealth)

       +---------------------------------+
       | wealth   meduc   medianpredprob |
       |---------------------------------|
  405. |      1       1         .2218643 |
  960. |      1       2         .3309583 |
 1096. |      1       3         .5152382 |
       |---------------------------------|
 1447. |      2       1         .2979575 |
 1690. |      2       2         .4240761 |
 2183. |      2       3         .6127208 |
       |---------------------------------|
 2450. |      3       1         .3871633 |
 2651. |      3       2         .5229162 |
 2867. |      3       3         .7019404 |
       |---------------------------------|
 3257. |      4       1         .4846398 |
 3651. |      4       2         .6199935 |
 3947. |      4       3         .7780514 |
       |---------------------------------|
 4211. |      5       1          .583299 |
 4363. |      5       2         .7083352 |
 4646. |      5       3         .8391798 |
       +---------------------------------+
```

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 $u^{(1)}$, $u^{(2)}$ . . ., $u^{(M)}$

ii) For each simulated value (m = 1,…,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 Stata with M = 1000 for each combination of wealth and education. We start by calculating the logit of the median predicted probabilities. This gives the linear predictor for each woman in the sample (fixing **magec** =0 and **magecsq** =0 rather than at their observed values):

```
. generatemedianpredlogit = logit(medianpredprob)
```

Next we reduce the dataset to 15 rows, one for each combination of wealth and education. We do this with the `keep` command which works the same way as `drop`, except that you specify the variables or observations to be kept rather than the variables or observations to be deleted:

```
. keep if pickone==1
(5351 observations deleted)
```

Then we keep only those variables required to calculate the population-averaged predictions and the variable **medianpredprob** which we need to compare the predictions from the two methods:

```
. keep wealth meduc medianpredprob medianpredlogit
```

At this point it is helpful to `list` the data to confirm that it has the 15 rows and 4 variables that we specified above:

```
. list, abbreviate(15) sepby(wealth)

     +---------------------------------------------------+
     | meduc   wealth   medianpredprob   medianpredlogit |
     |---------------------------------------------------|
  1. |     1        1         .2218643         -1.254835 |
  2. |     2        1         .3309583          -.703854 |
  3. |     3        1         .5152382          .0609718 |
     |---------------------------------------------------|
  4. |     1        2         .2979575         -.8570433 |
  5. |     2        2         .4240761         -.3060625 |
  6. |     3        2         .6127208          .4587634 |
     |---------------------------------------------------|
  7. |     1        3         .3871633         -.4592518 |
  8. |     2        3         .5229162          .0917291 |
  9. |     3        3         .7019404          .8565548 |
     |---------------------------------------------------|
 10. |     1        4         .4846398         -.0614603 |
 11. |     2        4         .6199935          .4895207 |
 12. |     3        4         .7780514          1.254347 |
     |---------------------------------------------------|
 13. |     1        5          .583299          .3363312 |
 14. |     2        5         .7083352          .8873121 |
 15. |     3        5         .8391798          1.652138 |
     +---------------------------------------------------+
```

Next we will `expand` the data to have 1000 rows for each combination of wealth and education. The command replaces each observation in the dataset with 1000 copies of the observation. This results in a dataset with 15,000 rows. We need to do this as, for each combination of wealth and education, we will calculate the predicted probability 1000 times, where for each of these times we insert a different simulated value of the random effect in the calculation.

```
. expand 1000
(14985 observations created)
```

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 variance equal to 0.889 (our estimate of the random intercept variance).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 `rnormal()` random number generating function where we specify `0` for the mean and `sqrt(0.889)` (the square root of 0.889) for the standard deviation.

```
. generate u = rnormal(0,sqrt(0.889))
```

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:

```
. generatemeanpredprob = invlogit(medianpredlogit + u)
```

Finally, we calculate the mean of these probabilities for each combination of wealth and education in the data.

```
. collapse (mean) meanpredprob, by(wealth meduc medianpredprob)
```

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

```
. list wealth meduc medianpredprob meanpredprob, abbreviate(14) sepby(wealth)

     +-----------------------------------------------+
     | wealth   meduc   medianpredprob   meanpredprob |
     |-----------------------------------------------|
  1. |      1       1         .2218643       .2591737 |
  2. |      1       2         .3309583       .3659121 |
  3. |      1       3         .5152382        .508471 |
     |-----------------------------------------------|
  4. |      2       1         .2979575       .3308483 |
  5. |      2       2         .4240761       .4362706 |
  6. |      2       3         .6127208       .6000571 |
     |-----------------------------------------------|
  7. |      3       1         .3871633       .4118967 |
  8. |      3       2         .5229162       .5317675 |
  9. |      3       3         .7019404       .6676902 |
     |-----------------------------------------------|
 10. |      4       1         .4846398       .4784383 |
 11. |      4       2         .6199935       .6023493 |
 12. |      4       3         .7780514       .7431008 |
     |-----------------------------------------------|
 13. |      5       1          .583299       .5658181 |
 14. |      5       2         .7083352       .6774453 |
 15. |      5       3         .8391798       .8030654 |
```

```
+-----------------------------------------------+
```

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.

Load "7.5.dta" into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment
- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***Stata* Datasets and Do-files**
- Click "7.5.dta" to open the dataset

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.

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

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, σ^2_{u5} and σ_{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 Ω_u.

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

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

To fit this model:

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 wealthc ///
>         || comm: wealthc, covariance(unstructured) ///
>         mle variance

Refining starting values:

Iteration 0:   log likelihood = -3088.6682  (not concave)
Iteration 1:   log likelihood =  -2993.895
Iteration 2:   log likelihood = -2984.5014

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2984.5014
Iteration 1:   log likelihood = -2981.8785
Iteration 2:   log likelihood = -2981.8289
Iteration 3:   log likelihood = -2981.8286

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   7                        Wald chi2(5)       =    528.82
Log likelihood = -2981.8286                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0001365    .006661    -0.02   0.984    -.0131919    .0129189
     magecsq |  -.0010593   .0006936    -1.53   0.127    -.0024188    .0003002
      meduc2 |   .5491784   .0870629     6.31   0.000     .3785383    .7198185
      meduc3 |   1.311499   .0999473    13.12   0.000     1.115606    1.507392
     wealthc |   .4069847   .0326996    12.45   0.000     .3428947    .4710746
       _cons |  -.4567603   .0844117    -5.41   0.000    -.6222042   -.2913164
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Unstructured           |
              var(wealth~m)  |   .0249108   .0238409      .0038172    .1625657
                 var(_cons)  |   .8652498   .1103462      .6738868    1.110954
        cov(wealth~m,_cons)  |  -.1159435   .0380789     -.1905767   -.0413102
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(3) =   334.50   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

Note, we have specified the `covariance(unstructured)` option to allow the random intercepts and slopes to covary (as opposed to the default that they are independent).

At this point we revisit the issue of how the accuracy of the parameter estimates is affected by the chosen number of integration points. Table 7.2 gives the parameter estimates which are obtained for the above model when different numbers of integration points are specified: 1, 2, 3, 4, 5, 6, 7 and 15. The percentage difference between each parameter estimate and its most accurate estimate (i.e. when 15 integration points are used) is also reported.

Table 7.2. Estimates for different numbers of integration points reported with the percentage difference between each estimate and that based on 15 integration points

Parameter	1	2	3	4	5	6	7	15
$\hat{\beta}_0$	-0.456	-0.457	-0.457	-0.457	-0.457	-0.457	-0.457	-0.457
	-0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
$\hat{\beta}_1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
$\hat{\beta}_2$	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
$\hat{\beta}_3$	0.548	0.549	0.549	0.549	0.549	0.549	0.549	0.549
	-0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
$\hat{\beta}_4$	1.310	1.311	1.311	1.311	1.311	1.311	1.311	1.311
	-0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
$\hat{\beta}_5$	0.406	0.408	0.408	0.407	0.407	0.407	0.407	0.407
	-0.2%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%	
$\hat{\sigma}^2_{u0}$	0.843	0.844	0.857	0.865	0.865	0.865	0.865	0.865
	-2.5%	-2.4%	-0.9%	0.0%	0.0%	0.0%	0.0%	
$\hat{\sigma}_{u05}$	-0.110	-0.110	-0.115	-0.117	-0.116	-0.116	-0.116	-0.116
	-5.4%	-5.3%	-1.1%	0.4%	-0.3%	-0.1%	0.0%	
$\hat{\sigma}^2_{u5}$	0.016	0.022	0.027	0.025	0.025	0.025	0.025	0.025
	-36.0%	-12.0%	8.0%	0.0%	0.0%	0.0%	0.0%	
Log likelihood	-2985	-2984	-2982	-2982	-2982	-2982	-2982	-2982
Seconds	50	41	48	55	57	70	84	277

The table shows that, even when 1 integration point is used, the fixed part parameter estimates are very similar to those for when 15 integration points are used. However, for low numbers of integration points, the random part parameter estimates are substantially smaller than their corresponding estimates based on 15 integration points. In particular, when 1 integration point is used, the slope variance $\hat{\sigma}^2_{u5}$ is 36.0% smaller than when 15 points are used. However, as with the random intercept model in P7.1.1, 4 integration points is sufficient to reduce the difference between each estimate and its value based on 15 integration points to less than 1%.

The table also shows that models with more integration points are slower to converge. For example, the model with 7 integration points took 50% longer to converge than the model with 4 integration points for no obvious increase in precision.[9]This is in contrast to the results for the random intercept model in P7.1.1 where all models converged in approximately the same length of time. In subsequent models, we will therefore reduce the number of integration points to 4.However, we advise users to always fit their own final models with more integration points. As before, when we come to specify models with more complex random parts, we will revisit this issue to check that 4 integration points is still appropriate.

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 σ^2_{u5} and σ_{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(-2981.8286 - -2987.2586) = 10.86 on 2 d.f.

The 5% point of a chi-squared distribution on 1 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 $0.407 + \hat{u}_{5j}$, and the between-community variance in the effect of wealth is estimated as 0.025. Because wealth has been centred about its sample mean, the intercept variance $\hat{\sigma}^2_{u0}$= 0.865 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

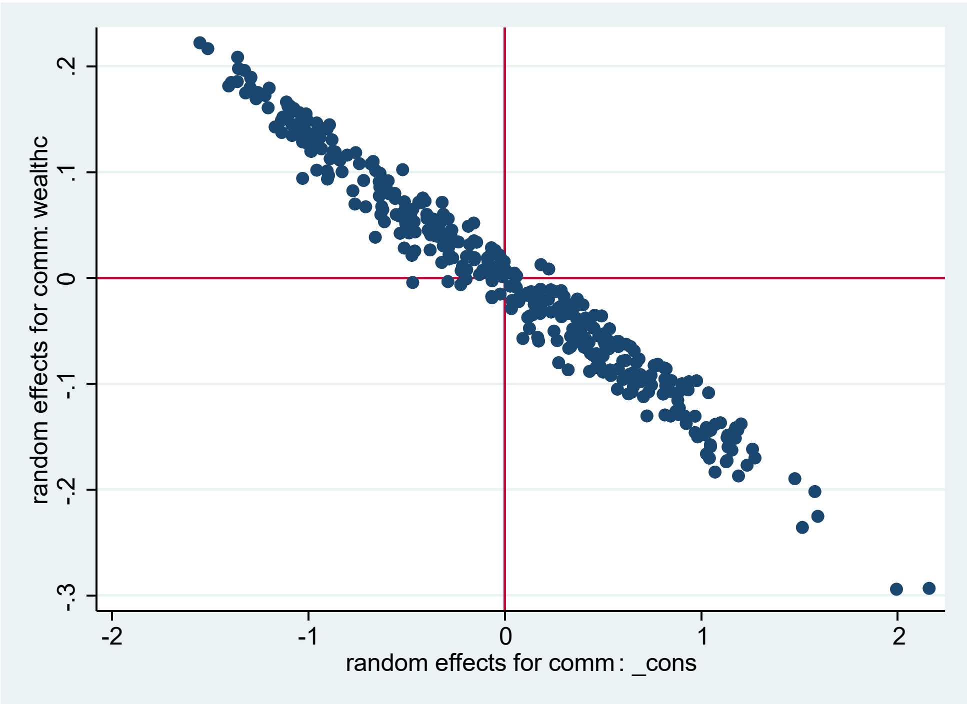
The negative intercept-slope covariance estimate ($\hat{\sigma}_{u5}$= -0.116) implies that communities with above-average antenatal care uptake (intercept residual $\hat{u}_{0j}$> 0) tend also to have below-average effects of wealth (slope residual $\hat{u}_{5j}$< 0). Put another way, there is less of an income gradient in use of antenatal care in communities with high uptake. We can obtain the estimated correlation as follows:

[9]See Rabe-Hesketh and Skrondal (2008) for further details on speed considerations when fitting models with the `xtmelogit` command.

```
. estat recov, corr

Random-effects correlation matrix for level comm

             | wealthc      _cons
-------------+----------------------
   wealthc |          1
      _cons | -.7897351          1
```

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

```
. predict u1 u0, reffects

. egen pickone = tag(comm)

. scatter u1 u0 if pickone==1, yline(0) xline(0)
```



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 (**magec** = 0, **magecsq** = 0) and no education (the reference category of **meduc**: **meduc2** = 0, **meduc3** = 0) is:

$$\log\left(\frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}}\right) = (-0.457 + \hat{u}_{0j}) + (0.407 + \hat{u}_{5j})\ \text{wealthc}_{ij}$$

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 -0.457 to -0.457 + 0.549 = 0.092.

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.

Preserve the data:

```
. preserve
```

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

```
. replace magec = 0
(5366 real changes made)

. replace magecsq = 0
(5366 real changes made)

. replace meduc2 = 0
(1649 real changes made)

. replace meduc3 = 0
(1851 real changes made)
```

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

```
. predict predprob, fixedonly
(option mu assumed; predicted means)
```

Transform these predicted probabilities back to predicted log-odds:

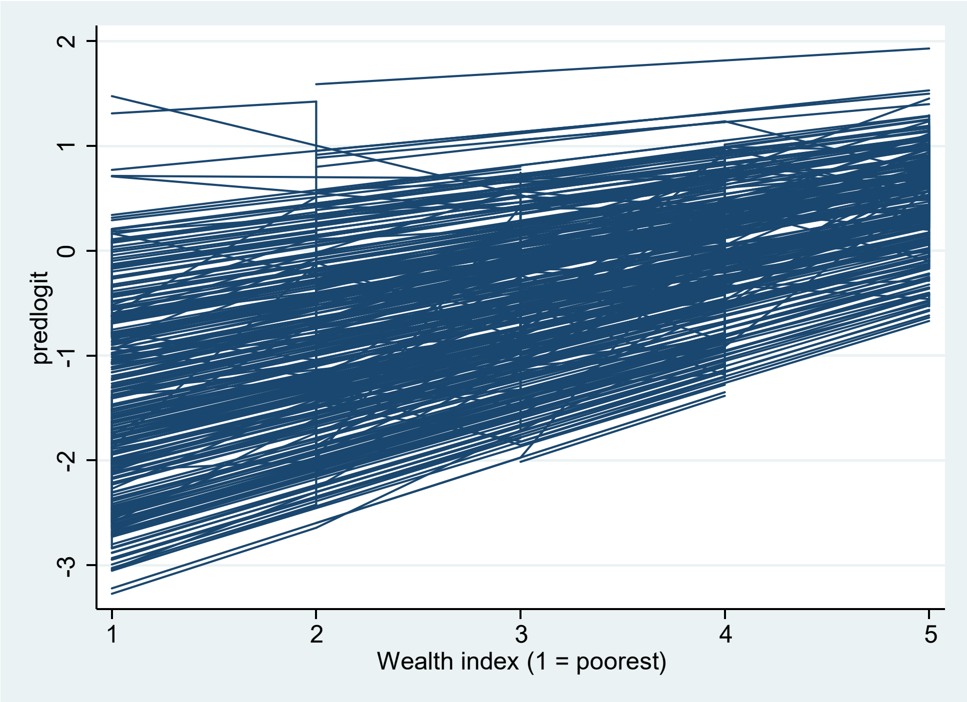
```
. generate predlogit = logit(predprob)
```

Add on the random part prediction:

```
. replace predlogit = predlogit + u0 + u1*wealthc
(5366 real changes made)
```

Plot the predicted community lines:

```
. drop pickone

. egen pickone = tag(commwealthc)

. line predlogit wealth if pickone==1, connect(ascending)
```



Notice that some of the lines are not drawn properly. This can be corrected by removing from the plot those communities where there is no variation in wealth (in quintiles) across households. To do this, we first `generate` a new variable **multiplewealth** and initially set its values equal to those of the **pickone** variable.

```
. generate multiplewealth = pickone
```

We then replace **multiplewealth** with the value 0 for those communities where all women have the same level of wealth. This is achieved by sorting the observations within each community by **wealth** and then looking to see whether the value of the last observation of **wealth** in each community is the same as the first observation. If it is, the women in the community all have the same level of wealth and so we set **multiplewealth** equal to 0. The relevant command is (see P5.2 for further details about this command for a similar graphing problem):

```
. bysort comm (wealth): replace multiplewealth = 0 if wealth[_N]==wealth[1]
(16 real changes made)
```

The output indicates that there are 16 communities where there is no variation in wealth across households.

```
. line predlogit wealth if multiplewealth==1, connect(ascending)
```

The plot is now drawn correctly. 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.

Now restore the data to its original state:

```
. restore
```

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:

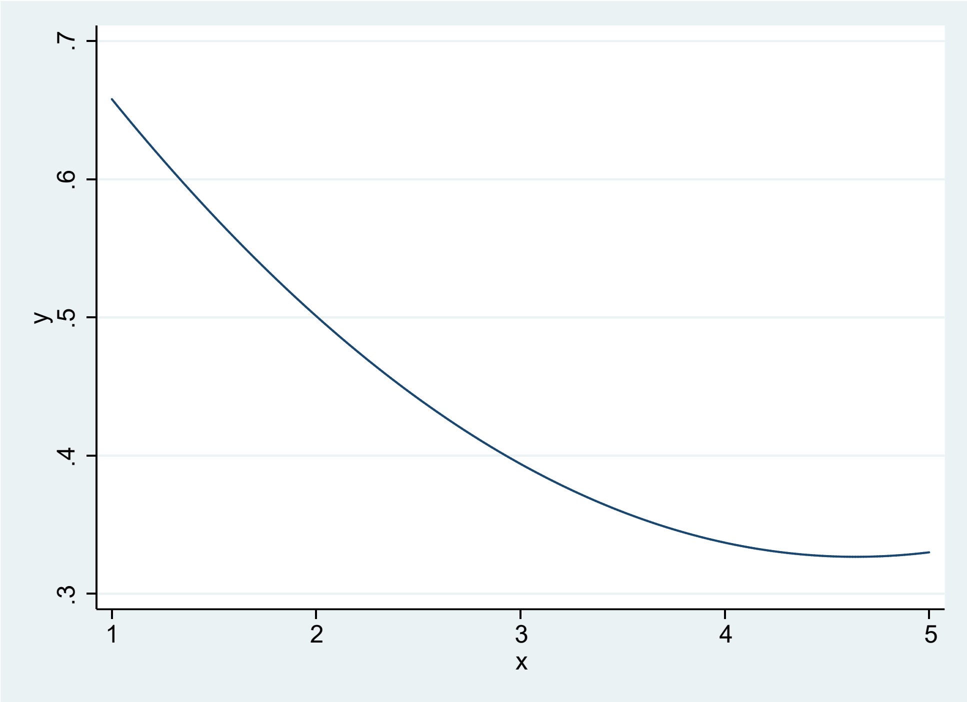
$$\mathrm{var}(u_{0j} + u_{5j}\mathrm{wealth}_{ij}) = \mathrm{var}(u_{0j}) + 2\mathrm{cov}(u_{0j}, u_{5j})\ \mathrm{wealth}_{ij} + \mathrm{var}(u_{5j})\ \mathrm{wealth}_{ij}^2$$
$$= \sigma_{u0}^2 + 2\sigma_{u05}\ \mathrm{wealth}_{ij} + \sigma_{u5}^2\ \mathrm{wealth}_{ij}^2$$

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

$$0.865 - 0.232\ \mathrm{wealth}_{ij} + 0.025\ \mathrm{wealth}_{ij}^2.$$

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

```
. twoway function 0.865 + -0.232*x + 0.025*x^2, range(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 ($2\hat{\sigma}_{u05}$ = 0.232) dominates over the coefficient of the quadratic term ($\hat{\sigma}^2_{u5}$=0.025), to the extent that $\hat{\sigma}^2_{u5}$ 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 (note that we use the `intpoints(4)` option to specify that the model is to be fitted with 4 integration points):

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

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 ///
>         wealth2 wealth3 wealth4 wealth5 ///
>         || comm:, variance intpoints(4)

Refining starting values:

Iteration 0:   log likelihood = -2990.2696
Iteration 1:   log likelihood = -2982.4163
Iteration 2:   log likelihood = -2982.2975

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2982.2975
Iteration 1:   log likelihood = -2982.2968
Iteration 2:   log likelihood = -2982.2968

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   4                        Wald chi2(8)       =    597.78
Log likelihood = -2982.2968                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0011905   .0066498    -0.18   0.858    -.0142239    .0118429
     magecsq |     -.001    .0006903    -1.45   0.147    -.0023531     .000353
      meduc2 |   .5533821   .0863128     6.41   0.000     .3842121    .7225521
      meduc3 |    1.30319   .1002437    13.00   0.000     1.106716    1.499664
     wealth2 |   .4673151   .1076433     4.34   0.000      .256338    .6782922
     wealth3 |   .6838284   .1100168     6.22   0.000     .4681994    .8994574
     wealth4 |   1.053761   .1162236     9.07   0.000     .8259666    1.281555
     wealth5 |    1.76677   .1361834    12.97   0.000     1.499855    2.033684
       _cons |  -1.248439   .1031217   -12.11   0.000    -1.450554   -1.046325
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Identity               |
                  var(_cons) |   .8471886   .1073127      .6609355    1.085928
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =   301.99 Prob>=chibar2 = 0.0000
```

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 0.467 while the difference between quintiles 3 and 2 is 0.684-0.467 = 0.217. 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 and preserve the data:

```
. estimates store model1

. preserve
```

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

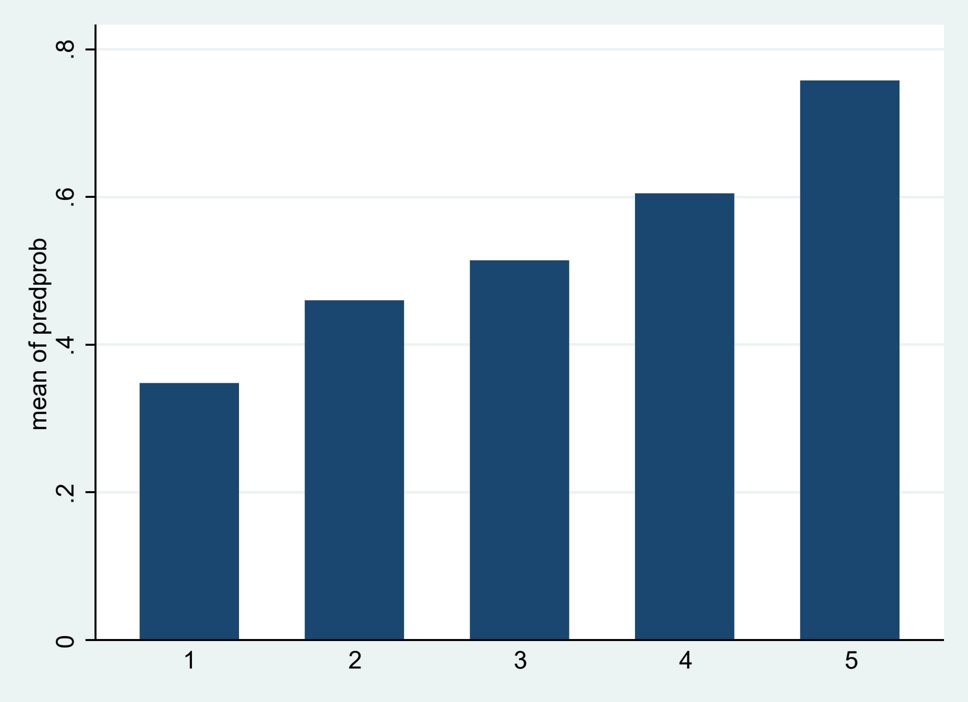
```
. replace magec = 0
(5366 real changes made)

. replace magecsq = 0
(5366 real changes made)

. summarize meduc2

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      meduc2 |      5366    .3073053      .46142          0          1

. replace meduc2 = 0.307
meduc2 was byte now float
(5366 real changes made)

. summarize meduc3

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      meduc3 |      5366    .3449497    .4753962          0          1

. replace meduc3 = 0.345
meduc3 was byte now float
(5366 real changes made)
```

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

```
. predict predprob, fixedonly
(option mu assumed; predicted means)
```

Next we use the `graph bar` command to draw a vertical bar chart. We request the mean of **predprob** to be plotted for each value of the categorical variable **wealth**:

```
. graph bar (mean) predprob, over(wealth)
```



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

Restore the data:

```
. restore
```

In the most general random coefficients[10] 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 × 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 5th quintile and the other quintiles. For example, the difference between the 1st and 2nd quintiles and between the 2nd and 3rd 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:

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

[10]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.

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 ///
>         wealth2 wealth3 wealth4 wealth5 ///
>         || comm: wealth5, covariance(unstructured) ///
>         mle variance intpoints(4)

Refining starting values:

Iteration 0:   log likelihood = -2991.5477
Iteration 1:   log likelihood = -2979.9921
Iteration 2:   log likelihood = -2977.8202

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2977.8202
Iteration 1:   log likelihood = -2977.7947
Iteration 2:   log likelihood = -2977.7947

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   4                        Wald chi2(8)       =    576.97
Log likelihood = -2977.7947                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0006968   .0066866    -0.10   0.917    -.0138023    .0124087
     magecsq |  -.0010292    .000694    -1.48   0.138    -.0023894     .000331
      meduc2 |    .552911   .0870039     6.36   0.000     .3823865    .7234356
      meduc3 |   1.312186   .1006184    13.04   0.000     1.114978    1.509395
     wealth2 |   .4669261   .1086655     4.30   0.000     .2539455    .6799066
     wealth3 |   .6817284   .1111109     6.14   0.000     .4639552    .8995017
     wealth4 |   1.055917   .1177571     8.97   0.000     .8251173    1.286717
     wealth5 |   1.848519    .143718    12.86   0.000     1.566837    2.130201
       _cons |  -1.260604   .1058378   -11.91   0.000    -1.468043   -1.053166
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Unstructured           |
                var(wealth5) |   .4446906   .2535677       .145443    1.359637
                  var(_cons) |   .9592973   .1261599      .7413261    1.241358
          cov(wealth5,_cons) |  -.4538797   .1657395      -.778723   -.1290363
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(3) =   311.00   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

And store the results:

```
. estimates store model2
```

The model assumes that the coefficients of **wealth2**, **wealth3** and **wealth4** are fixed across communities, estimated as 0.467, 0.682 and 1.056 respectively. The coefficient of **wealth5** (i.e. the difference between the top and bottom quintiles) varies across communities, and is estimated as 1.849 + $\hat{u}_{8j}$ 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: $\sigma^2_{u8} = \sigma_{u08} = 0$).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(-2977.7947 - -2982.2968) = 9.0042 on 2 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, Stata provides the `lrtest` command:

```
. lrtest model1 model2

Likelihood-ratio test                                 LR chi2(2)  =      9.00
(Assumption: model1 nested in model2)           Prob > chi2 =    0.0111

Note: The reported degrees of freedom assumes the null hypothesis is not on the
boundary of the parameter space. If this is not
      true, then the reported test is conservative.
```

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 9.00 while the p-value of 0.0111 confirms that the random effect for **wealth5** is indeed significant. The last three lines of the output inform us that the likelihood ratio test is conservative.[11]This means that the reported p-value is not the correct p-value; rather it is an upper bound for the correct p-value. The test is described as conservative as when the correct p-value is 0.05 (i.e. the random slope model is just preferred to the random intercepts model), the reported p-value will be slightly higher leading us to incorrectly favour the simpler model. However, as long as the reported p-value is less than 0.05 then the same will be the case for the correct p-value and so it is safe to infer that the random slopes model is preferred to the random intercepts model.

The between-community variance is now estimated as:

$$\begin{aligned}\mathrm{var}\left(u_{0j} + u_{8j}\mathrm{wealth5}_{ij}\right) &= \sigma^2_{u0} + 2\sigma_{u08}\mathrm{wealth5}_{ij} + \sigma^2_{u8}\mathrm{wealth5}^2_{ij} \\ &= 0.959 - 0.908\ \mathrm{wealth5}_{ij} + 0.445\ \mathrm{wealth5}^2_{ij}\end{aligned}$$

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

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

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

[11]See `help j_xtmixedlr` for further details.

We will now extend the model to include a random coefficient on **wealth4**. This model additionally allows the difference between the 4th quintile and the first three quintiles to vary across communities.

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

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 ///
>         wealth2 wealth3 wealth4 wealth5 ///
>         || comm: wealth4 wealth5, covariance(unstructured) ///
>         mle variance intpoints(4)

Refining starting values:

Iteration 0:   log likelihood = -2999.4893
Iteration 1:   log likelihood =  -2977.039
Iteration 2:   log likelihood = -2974.4054

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2974.4054
Iteration 1:   log likelihood = -2973.7633
Iteration 2:   log likelihood = -2973.6831
Iteration 3:   log likelihood = -2973.6768
Iteration 4:   log likelihood = -2973.6767

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   4                        Wald chi2(8)       =    558.41
Log likelihood = -2973.6767                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0005901   .0067076    -0.09   0.930    -.0137368    .0125565
     magecsq |  -.0010538   .0006968    -1.51   0.130    -.0024195     .000312
      meduc2 |   .5589059   .0874136     6.39   0.000     .3875783    .7302334
      meduc3 |   1.316547   .1005916    13.09   0.000     1.119392    1.513703
     wealth2 |   .4771014    .110339     4.32   0.000     .2608409    .6933619
     wealth3 |   .6879957   .1129888     6.09   0.000     .4665417    .9094498
     wealth4 |   1.016506   .1199063     8.48   0.000     .7814938    1.251518
     wealth5 |   1.844475   .1448421    12.73   0.000      1.56059     2.12836
       _cons |  -1.261404   .1095108   -11.52   0.000    -1.476041   -1.046766
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Unstructured           |
                var(wealth4) |   .1923405   .1555373      .0394223    .9384257
                var(wealth5) |   .4219141   .2461365      .1344774    1.323728
                  var(_cons) |    1.13218   .1600053      .8582607    1.493521
        cov(wealth4,wealth5) |   .0585771   .1326277     -.2013684    .3185226
          cov(wealth4,_cons) |  -.3640967   .1492776     -.6566753   -.0715181
          cov(wealth5,_cons) |  -.5339521   .1880823     -.9025865   -.1653176
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(6) =   319.24   Prob > chi2 = 0.0000
```

```
Note: LR test is conservative and provided only for reference.
```

And store the results:

```
. estimates store model3
```

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

As we have now extended the model to include an additional random effect, we take this opportunity to once again compare how the parameter estimates vary when we specify different numbers of integration points. Table 7.3 gives the parameter estimates which are obtained for the above model when different numbers of integration points are specified: 1, 2, 3, 4, 5, 6, 7 and 15. As before, the percentage difference between each parameter estimate and its most accurate estimate (i.e. when 15 integration points are used) is also reported. Note, we have not shown the fixed part parameter estimates as even when just 1 integration point is specified, each estimate is within 1% of the estimate obtained when 15 integration points are used.

Table 7.3. Estimates for different numbers of integration points reported with the percentage difference between each estimate and that based on 15 integration points

Parameter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**15**
$\hat{\sigma}^2_{u0}$	1.088	1.095	1.12	1.132	1.132	1.133	1.133	1.133
	-3.97%	-3.35%	-1.15%	-0.09%	-0.09%	0.00%	0.00%	
$\hat{\sigma}_{u07}$	-0.833	-0.779	-0.775	-0.78	-0.78	-0.78	-0.78	-0.78
	6.79%	-0.13%	-0.64%	0.00%	0.00%	0.00%	0.00%	
$\hat{\sigma}^2_{u7}$	0.155	0.177	0.189	0.192	0.192	0.192	0.192	0.192
	-19.27%	-7.81%	-1.56%	0.00%	0.00%	0.00%	0.00%	
$\hat{\sigma}_{u08}$	-0.764	-0.774	-0.772	-0.773	-0.772	-0.773	-0.773	-0.773
	-1.16%	0.13%	-0.13%	0.00%	-0.13%	0.00%	0.00%	
$\hat{\sigma}_{u78}$	0.278	0.205	0.196	0.206	0.205	0.206	0.206	0.206
	34.95%	-0.49%	-4.85%	0.00%	-0.49%	0.00%	0.00%	

$\hat{\sigma}^2_{u8}$	0.399	0.391	0.414	0.422	0.422	0.422	0.422	0.422
	-5.45%	-7.35%	-1.90%	0.00%	0.00%	0.00%	0.00%	
Log likelihood	-2977	-2976	-2974	-2974	-2974	-2974	-2974	-2974
seconds	122	144	213	310	515	813	1250	11468
minutes	2	2	4	5	9	14	21	191

Once again we see that when 1 integration points is used the random part parameter estimates are substantially smaller than when 15 points are used. However, specifying 4 integration points is again sufficient to lead all estimates to be within 1% of those when 15 points are used. We will therefore continue to specify 4 integration points each time we fit a new multilevel model.

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

```
. lrtest model2 model3

Likelihood-ratio test                                 LR chi2(3)  =      8.24
(Assumption: model2 nested in model3)           Prob > chi2 =    0.0414

Note: The reported degrees of freedom assumes the null hypothesis is not on the
boundary of the parameter space. If this is not true, then the reported test is
conservative.
```

The test statistic is 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:

$$\begin{aligned}
&\mathrm{var}\left(u_{0j} + u_{7j}\mathrm{wealth4}_{ij} + u_{8j}\mathrm{wealth5}_{ij}\right) = \sigma^2_{u0} + 2\sigma_{u07}\mathrm{wealth4}_{ij} + \sigma^2_{u7}\mathrm{wealth4}^2_{ij}\\
&+2\sigma_{u08}\mathrm{wealth5}_{ij} + 2\sigma_{u78}\mathrm{wealth4}_{ij}\mathrm{wealth5}_{ij}\\
&+\sigma^2_{u8}\mathrm{wealth5}^2_{ij}\\
&= 1.132 - 0.728\ \mathrm{wealth4}_{ij} + 0.192\ \mathrm{wealth4}^2_{ij}\\
&-1.068\ \mathrm{wealth5}_{ij} + 0.117\ \mathrm{wealth4}_{ij}\mathrm{wealth5}_{ij}\\
&+0.422\ \mathrm{wealth5}^2_{ij}
\end{aligned}$$

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

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

1.132 -0.728 + 0.192 = 0.596 in quintile 4 (**wealth4** = 1, **wealth5** = 0) and

1.132 -1.068 + 0.422 = 0.486 in quintile 5 (**wealth4** = 0, **wealth5** = 1)

Thus, the between-community variance is estimated as 1.132 for quintiles 1-3, 0.596 for the 4^{th} quintile, and 0.486 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.

Load "7.6.dta" into memory and open the do-file for this lesson:

From within the LEMMA Learning Environment

- Go to **Module 7: Multilevel Models for Binary Responses**, and scroll down to ***Stata* Datasets and Do-files**
- Click " 7.6.dta" to open the dataset

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.

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

```
. xtmelogit antemed magecmagecsq meduc2 meduc3 ///
>         wealth2 wealth3 wealth4 wealth5 ///
>         urban ///
>         || comm: wealth4 wealth5, covariance(unstructured) ///
>         mle variance intpoints(4)

Refining starting values:

Iteration 0:   log likelihood = -2973.7722
Iteration 1:   log likelihood =   -2943.18
Iteration 2:   log likelihood = -2942.6495

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2942.6495
Iteration 1:   log likelihood = -2940.9051
Iteration 2:   log likelihood =  -2940.736
Iteration 3:   log likelihood = -2940.7294
Iteration 4:   log likelihood = -2940.7294
Iteration 5:   log likelihood = -2940.7294
```

```
Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   4                        Wald chi2(9)       =    643.10
Log likelihood = -2940.7294                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |    .000228   .0067088     0.03   0.973    -.0129211    .0133771
     magecsq |  -.0010775   .0006975    -1.54   0.122    -.0024445    .0002895
      meduc2 |   .5789762   .0874215     6.62   0.000     .4076332    .7503193
      meduc3 |   1.372047   .1004864    13.65   0.000     1.175098    1.568997
     wealth2 |   .4588008   .1102916     4.16   0.000     .2426332    .6749685
     wealth3 |   .6503777   .1128832     5.76   0.000     .4291307    .8716248
     wealth4 |   .9717542   .1199754     8.10   0.000     .7366068    1.206902
     wealth5 |   1.527697   .1412455    10.82   0.000     1.250861    1.804533
       urban |   .9796354    .119197     8.22   0.000     .7460135    1.213257
       _cons |  -1.528537   .1122899   -13.61   0.000    -1.748621   -1.308453
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Unstructured           |
               var(wealth4) |   .2254125   .1664349      .0530258    .9582273
               var(wealth5) |   .4387414   .2124431      .1698432    1.133363
                 var(_cons) |   .9828335   .1425214      .7396845     1.30591
        cov(wealth4,wealth5) |   .1179589   .1296125     -.1360769    .3719946
          cov(wealth4,_cons) |  -.3505832   .1382717     -.6215908   -.0795756
          cov(wealth5,_cons) |  -.5896267   .1730921     -.9288809   -.2503724
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(6) =   269.77   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

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 to 1.132 to 0.983. To see how the between-community variance has changed in the top two quintiles:

$$
\begin{aligned}
&\mathrm{var}\left(u_{0j} + u_{7j}\mathrm{wealth4}_{ij} + u_{8j}\mathrm{wealth5}_{ij}\right) = \sigma^2_{u0} + 2\sigma_{u07}\mathrm{wealth4}_{ij} + \sigma^2_{u7}\mathrm{wealth4}^2_{ij} \\
&+2\sigma_{u08}\mathrm{wealth5}_{ij} + 2\sigma_{u78}\mathrm{wealth4}_{ij}\mathrm{wealth5}_{ij} \\
&+\sigma^2_{u8}\mathrm{wealth5}^2_{ij} \\
&= 0.983 - 0.702\ \mathrm{wealth4}_{ij} + 0.225\ \mathrm{wealth4}^2_{ij} \\
&-1.180\ \mathrm{wealth5}_{ij} + 0.236\ \mathrm{wealth4}_{ij}\mathrm{wealth5}_{ij} \\
&+0.439\ \mathrm{wealth5}^2_{ij}
\end{aligned}
$$

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

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

0.983 – 0.702 + 0.225 = 0.506 in quintile 4 (**wealth4** = 1, **wealth5** = 0) and

0.983 – 1.180 + 0.439 = 0.242 in quintile 5 (**wealth4** = 0, **wealth5** = 1)

For ease of comparison, the between-community variances for each model are given in Table 7.4.

Table 7.4. 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%)	1.132	0.983
4	0.596	0.506
5 (top 20%)	0.486	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:

```
. tabulate wealth urban, row

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

    Wealth |
     index |   Type of region of
(1=poorest |       residence
         ) |         0          1 |     Total
-----------+----------------------+----------
         1 |       997        170 |     1,167
           |     85.43      14.57 |    100.00
-----------+----------------------+----------
         2 |       833        184 |     1,017
           |     81.91      18.09 |    100.00
-----------+----------------------+----------
         3 |       771        219 |       990
           |     77.88      22.12 |    100.00
-----------+----------------------+----------
         4 |       722        267 |       989
           |     73.00      27.00 |    100.00
-----------+----------------------+----------
         5 |       359        844 |     1,203
           |     29.84      70.16 |    100.00
-----------+----------------------+----------
     Total |     3,682      1,684 |     5,366
           |     68.62      31.38 |    100.00
```

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.

```
. bysort comm: egen wealth5mean = mean(wealth5)
```

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.

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

```
. xtmelogit antemed magec magecsq meduc2 meduc3 ///
>         wealth2 wealth3 wealth4 wealth5 ///
>         urban wealth5mean ///
>         || comm: wealth4 wealth5, covariance(unstructured) ///
>         mle variance intpoints(4)

Refining starting values:

Iteration 0:   log likelihood = -2963.3214
Iteration 1:   log likelihood = -2936.9666
Iteration 2:   log likelihood = -2926.1163

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2926.1163
Iteration 1:   log likelihood = -2925.4816
Iteration 2:   log likelihood = -2925.4184
Iteration 3:   log likelihood = -2925.4176
Iteration 4:   log likelihood = -2925.4176

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   4                        Wald chi2(10)      =    651.07
Log likelihood = -2925.4176                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |    .000212   .0067046     0.03   0.975    -.0129287    .0133528
     magecsq |  -.0011096    .000698    -1.59   0.112    -.0024777    .0002586
      meduc2 |   .5872605   .0873712     6.72   0.000     .4160162    .7585049
      meduc3 |   1.389801   .1005332    13.82   0.000      1.19276    1.586843
     wealth2 |   .4449457   .1099988     4.05   0.000      .229352    .6605394
     wealth3 |   .6216429   .1126599     5.52   0.000     .4008335    .8424522
     wealth4 |   .9140961   .1201685     7.61   0.000     .6785701    1.149622
     wealth5 |   1.191916   .1498124     7.96   0.000     .8982891    1.485543
```

```
       urban |   .4963382   .1424018     3.49   0.000     .2172358    .7754406
 wealth5mean |    1.48004   .2697773     5.49   0.000     .9512861    2.008794
       _cons |  -1.630005   .1126784   -14.47   0.000     -1.85085   -1.409159
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Unstructured           |
                var(wealth4) |   .2346589   .1727506      .0554373    .9932813
                var(wealth5) |   .3690885   .1847585       .138368     .984522
                  var(_cons) |   .8965674   .1342691      .6685099    1.202425
        cov(wealth4,wealth5) |   .1270336   .1252611     -.1184736    .3725407
          cov(wealth4,_cons) |  -.3413827   .1344797      -.604958   -.0778073
          cov(wealth5,_cons) |  -.5313693   .1596948     -.8443654   -.2183733
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(6) =   236.60   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

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 1.480) 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:

```
. generate wealth2Xwealth5mean = wealth2*wealth5mean

. generate wealth3Xwealth5mean = wealth3*wealth5mean

. generate wealth4Xwealth5mean = wealth4*wealth5mean

. generate wealth5Xwealth5mean = wealth5*wealth5mean
```

Now fit the model:

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

```
. xtmelogit antemed magec magecsq meduc2 meduc3 ///
>         wealth2 wealth3 wealth4 wealth5 ///
```

```
>         urban wealth5mean ///
>         wealth2Xwealth5mean wealth3Xwealth5mean ///
>         wealth4Xwealth5mean wealth5Xwealth5mean ///
>         || comm: wealth4 wealth5, covariance(unstructured) ///
>         mle variance intpoints(4)

Refining starting values:

Iteration 0:   log likelihood = -2961.3746
Iteration 1:   log likelihood = -2933.3246
Iteration 2:   log likelihood = -2922.5619

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2922.5619
Iteration 1:   log likelihood = -2921.6589
Iteration 2:   log likelihood =  -2921.574
Iteration 3:   log likelihood = -2921.5728
Iteration 4:   log likelihood = -2921.5728

Mixed-effects logistic regression               Number of obs      =      5366
Group variable: comm                            Number of groups   =       361

                                                Obs per group: min =         3
                                                               avg =      14.9
                                                               max =        25

Integration points =   4                        Wald chi2(14)      =    666.13
Log likelihood = -2921.5728                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     antemed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       magec |  -.0003509   .0067109    -0.05   0.958     -.013504    .0128022
     magecsq |  -.0010712   .0006978    -1.54   0.125    -.0024388    .0002964
      meduc2 |   .5849272   .0874776     6.69   0.000     .4134743    .7563801
      meduc3 |   1.382202    .100506    13.75   0.000     1.185214     1.57919
     wealth2 |   .5527484   .1330517     4.15   0.000     .2919719     .813525
     wealth3 |    .666891   .1393838     4.78   0.000     .3937037    .9400783
     wealth4 |   1.084688   .1505235     7.21   0.000     .7896676    1.379709
     wealth5 |   1.509855   .2067066     7.30   0.000     1.104718    1.914993
       urban |   .4767094   .1412987     3.37   0.001     .1997691    .7536498
 wealth5mean |   2.990588   .6967129     4.29   0.000     1.625056     4.35612
wealth2Xwe~n |  -1.236317   .7906449    -1.56   0.118    -2.785952     .313319
wealth3Xwe~n |  -.6996157   .7764824    -0.90   0.368    -2.221493    .8222618
wealth4Xwe~n |   -1.59801   .7457976    -2.14   0.032    -3.059746   -.1362732
wealth5Xwe~n |  -1.808272   .7377477    -2.45   0.014    -3.254231   -.3623132
       _cons |  -1.776004   .1284617   -13.83   0.000    -2.027785   -1.524224
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
comm: Unstructured           |
                var(wealth4) |   .2353727   .1738067      .0553607    1.000715
                var(wealth5) |   .3917773   .1867019      .1539565    .9969661
                  var(_cons) |   .8944615   .1340329      .6668239    1.199809
        cov(wealth4,wealth5) |   .1502172   .1268216     -.0983485    .3987829
          cov(wealth4,_cons) |  -.3503154   .1348174     -.6145527   -.0860781
          cov(wealth5,_cons) |  -.5558316   .1600109     -.8694473    -.242216
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(6) =   232.49   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

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 set **magec**, **meduc** and **urban**equal to their mean values and set **magecsq** to the square of the sample mean for **magec**:

```
. replace magec = 0
(5366 real changes made)

. replace magecsq = 0
(5366 real changes made)

. summarize meduc2

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      meduc2 |      5366    .3073053      .46142          0          1

. replace meduc2 = 0.307
meduc2 was byte now float
(5366 real changes made)

. summarize meduc3

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      meduc3 |      5366    .3449497    .4753962          0          1

. replace meduc3 = 0.345
meduc3 was byte now float
(5366 real changes made)

. summarize urban

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       urban |      5366    .3138278    .4640906          0          1

. replace urban = 0.314
urban was byte now float
(5366 real changes made)
```

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. The recoding rules are specified within parentheses.

```
. recode wealth5mean (0/0.1 = 0) (0.1/0.3 = 0.2) (0.3/1.0 = 0.4)
(wealth5mean: 3418 changes made)
```

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

```
. recode wealth2Xwealth5mean wealth3Xwealth5mean ///
>        wealth4Xwealth5mean wealth5Xwealth5mean ///
```

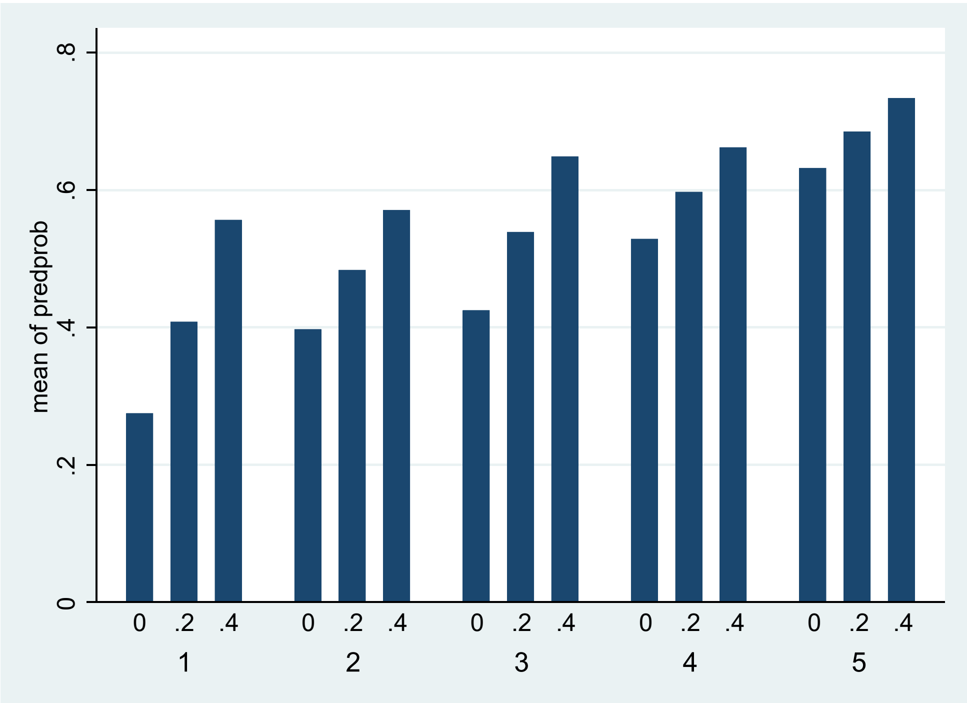
```
>          (0/0.1 = 0) (0.1/0.3 = 0.2) (0.3/1.0 = 0.4)
(wealth2Xwealth5mean: 493 changes made)
(wealth3Xwealth5mean: 591 changes made)
(wealth4Xwealth5mean: 686 changes made)
(wealth5Xwealth5mean: 1182 changes made)
```

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).

```
. predict predprob, fixedonly
(option mu assumed; predicted means)
```

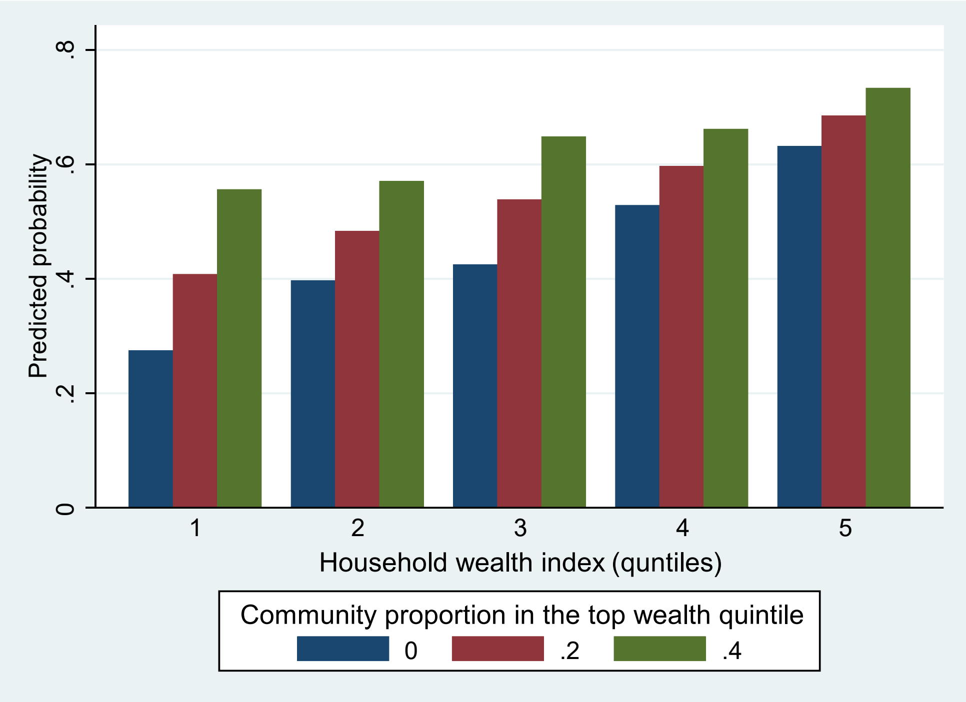
We will now plot the predictions. We use the `graph bar` command, this time specifying the `over()` option twice. The first `over()` option specifies that we want the mean of **predprob** to be plotted for each value of **wealth5mean**. The second `over()` option specifies that we want to repeat this for each of the 5 values of **wealth**. This gives a total of 15 vertical bars:

```
. graph bar (mean) predprob, over(wealth5mean) over(wealth)
```



The above graph is a little difficult to interpret as the bars for the three values of **wealth5mean**are all plotted using the same colour. 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 `graph bar` command. We use the `asyvars` option to display the bars for each value of **wealth5mean** in different colours while the `ytitle()`, `b1title()` and `subtitle()` options are used to add various titles to the graph.

```
. graph bar (mean) predprob, over(wealth5mean) over(wealth) asyvars ///
>         ytitle(Predicted probability) ///
>         b1title(Household wealth index (quntiles)) ///
>         legend(rows(1) ///
>         subtitle(Community proportion in the top wealth quintile))
```



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

It is not possible to fit models in Stata with MCMC methods.

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

Rabe-Hesketh, S. and Skrondal, A. (2008) *Multilevel and longitudinal modeling using Stata (Second Edition)*.College Station, TX: Stata Press.