Panel Data and Multilevel Models for Categorical Outcomes: Basic Multilevel Models

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Institute for Political Methodology, Taiwan, July 17 & 18, 2018

These notes borrow very heavily, often/usually verbatim, from the Stata 14.2 MULTILEVEL MIXED EFFECTS REFERENCE MANUAL, and from Paul Allison's book, *Fixed Effects Regression Models for Categorical Data*. I strongly encourage people to get their own copy. The Stata XT manual is also a good reference, as is *Microeconometrics Using Stata, Revised Edition*, by Cameron and Trivedi. Separate handouts examine fixed effects models and random effects models using commands like `clogit`, `xtreg`, and `xtlogit`. Some of the material here is repeated from those handouts.

Overview. Models estimated by `xt, re` commands (e.g. `xtreg, re` and `xtlogit, re`) can also often be estimated by `me` (mixed effect) commands (e.g. `mixed`, `melogit`). There are many types of data where either type of command will work – but these aren't necessarily panel data. For example, you might have a sample of schools, and within each school you have a sample of students. The latter might be more appropriately referred to as a multilevel data set. Quoting verbatim from the Stata 14.2 manual,

> Mixed-effects models are characterized as containing both fixed effects and random effects. The fixed effects are analogous to standard regression coefficients and are estimated directly. The random effects are not directly estimated (although they may be obtained postestimation) but are summarized according to their estimated variances and covariances. Random effects may take the form of either random intercepts or random coefficients, and the grouping structure of the data may consist of multiple levels of nested groups. As such, mixed-effects models are also known in the literature as multilevel models and hierarchical models. Mixed-effects commands fit mixed-effects models for a variety of distributions of the response conditional on normally distributed random effects.

A key thing to realize is that, in a panel or multilevel dataset, observations in the same cluster are correlated because they share common cluster-level random effects. Put another way, cases within a cluster are generally not independent of each other. The responses an individual gives at one point in time will not be unrelated to the responses given at another time. Students within a school will tend to be more similar than students from different schools. Failure to take into account the fact that cases within a cluster are not independent of each other and share common cluster-level random effects can distort parameter estimates and standard errors.

There are various reasons you might prefer `me` commands over `xt, re` commands.

- Commands like `mixed` and `melogit` can estimate much more complicated random effects models than can be done with `xtreg, re` and `xtlogit, re`. In this handout I am going to keep things fairly simple.
- You can have more levels in the `me` commands, e.g. you could have schools, students within schools, and multiple records for each student (e.g. exam performances across time). I will give an example like that for `melogit`.
- Unlike `xtreg` and `xtlogit` you can use the `svy:` prefix with `me` commands.

I will discuss linear models and logistic models in the rest of this handout.

Linear Mixed Effects Models – 2 Levels. `xtreg` random effects models can also be estimated using the `mixed` command in Stata.

The following is copied verbatim from pp. 357 & 367 of the Stata 14.2 manual entry for the `mixed` command.

> `mixed` fits linear mixed-effects models. These models are also known as multilevel models or hierarchical linear models. The overall error distribution of the linear mixed-effects model is assumed to be Gaussian, and heteroskedasticity and correlations within lowest-level groups also may be modeled.
>
> Linear mixed models are models containing both fixed effects and random effects. They are a generalization of linear regression allowing for the inclusion of random deviations (effects) other than those associated with the overall error term. In matrix notation,
>
> $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \qquad (1)$$
>
> where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ design/covariate matrix for the fixed effects $\boldsymbol{\beta}$, and $\mathbf{Z}$ is the $n \times q$ design/covariate matrix for the random effects $\mathbf{u}$. The $n \times 1$ vector of errors $\boldsymbol{\epsilon}$ is assumed to be multivariate normal with mean 0 and variance matrix $\sigma^2_\epsilon\mathbf{R}$.
>
> The fixed portion of (1), $\mathbf{X}\boldsymbol{\beta}$, is analogous to the linear predictor from a standard OLS regression model with $\boldsymbol{\beta}$ being the regression coefficients to be estimated. For the random portion of (1), $\mathbf{Z}\mathbf{u}+\boldsymbol{\epsilon}$, we assume that $\mathbf{u}$ has variance–covariance matrix $\mathbf{G}$ and that $\mathbf{u}$ is orthogonal to $\boldsymbol{\epsilon}$ so that
>
> $$\text{Var}\begin{bmatrix}\mathbf{u}\\ \boldsymbol{\epsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & 0\\ 0 & \sigma^2_\epsilon\mathbf{R}\end{bmatrix}$$
>
> The random effects $\mathbf{u}$ are not directly estimated (although they may be predicted), but instead are characterized by the elements of $\mathbf{G}$, known as variance components, that are estimated along with the overall residual variance σ^2_ϵ and the residual-variance parameters that are contained within $\mathbf{R}$.
>
> The general forms of the design matrices $\mathbf{X}$ and $\mathbf{Z}$ allow estimation for a broad class of linear models: blocked designs, split-plot designs, growth curves, multilevel or hierarchical designs, etc. They also allow a flexible method of modeling within-cluster correlation. Subjects within the same cluster can be correlated as a result of a shared random intercept, or through a shared random slope on (say) age, or both. The general specification of $\mathbf{G}$ also provides additional flexibility—the random intercept and random slope could themselves be modeled as independent, or correlated, or independent with equal variances, and so forth. The general structure of $\mathbf{R}$ also allows for residual errors to be heteroskedastic and correlated, and allows flexibility in exactly how these characteristics can be modeled.

Here is how you can use `mixed` to replicate results from `xtreg, re`. Estimates differ slightly because different algorithms are being used. We also compare the results with what you get if you just use OLS regression instead.

Allison (starting on p. 7 of his book) gives an example using the National Longitudinal Survey of Youth. This subset of the data set has 581 children who were interviewed in 1990, 1992, and

1994. Variables with a t subscript were measured at each of the three points in time. Variables without a t subscript do not vary across time. Variables used in this example include

- id is the subject id number and is the same across each wave of the survey
- $anti_t$ is Antisocial behavior (scale ranges from 0 to 6)
- $self_t$ – Self esteem (scale ranges from 6 to 24)
- pov_t – coded 1 if family is in poverty, 0 otherwise
- black is coded 1 if the child is black, 0 otherwise
- hispanic is coded 1 if the child is Hispanic, 0 otherwise
- childage is child's age in 1990
- married is coded 1 if the child's mother was currently married in 1990, 0 otherwise
- gender is coded 1 if the child is female, 0 if male
- momage is the mother's age at birth of child
- momwork is coded 1 if the mother was employed in 1990, 0 otherwise

The data used here have already been converted into long format.

```
. use https://www3.nd.edu/~rwilliam/statafiles/nlsyxt.dta, clear
. * Two level linear model, preceded by single-level OLS regression model
. reg anti self pov i.year i.black i.hispanic childage i.married i.gender momage i.momwork

      Source |       SS           df       MS      Number of obs   =     1,743
-------------+----------------------------------   F(11, 1731)     =     15.16
       Model |   380.85789        11  34.6234446   Prob > F        =    0.0000
    Residual |  3952.25743     1,731  2.28322208   R-squared       =    0.0879
-------------+----------------------------------   Adj R-squared   =    0.0821
       Total |  4333.11532     1,742  2.48743704   Root MSE        =     1.511

------------------------------------------------------------------------------
        anti |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        self |  -.0741425   .0109632    -6.76   0.000     -.095645   -.0526401
         pov |   .4354025   .0855275     5.09   0.000     .2676544    .6031505
             |
        year |
          92 |   .0521538   .0887138     0.59   0.557    -.1218437    .2261512
          94 |   .2255775   .0888639     2.54   0.011     .0512856    .3998694
             |
     1.black |   .1678622   .0881839     1.90   0.057    -.0050959    .3408204
  1.hispanic |  -.2483772   .0948717    -2.62   0.009    -.4344523   -.0623021
    childage |    .087056   .0622121     1.40   0.162    -.0349628    .2090747
   1.married |  -.0888875    .087227    -1.02   0.308    -.2599689     .082194
    1.gender |  -.4950259   .0728886    -6.79   0.000     -.637985   -.3520668
      momage |  -.0166933   .0173463    -0.96   0.336    -.0507153    .0173287
   1.momwork |   .2120961   .0800071     2.65   0.008     .0551754    .3690168
       _cons |   2.675312   .7689554     3.48   0.001     1.167132    4.183491
------------------------------------------------------------------------------

. est store reg
```

```
. * 2 level linear model
. xtreg anti self pov i.year i.black i.hispanic childage i.married i.gender momage i.momwork, re

Random-effects GLS regression                   Number of obs     =      1,743
Group variable: id                              Number of groups  =        581

R-sq:                                           Obs per group:
     within  = 0.0320                                         min =          3
     between = 0.1067                                         avg =        3.0
     overall = 0.0853                                         max =          3

                                                Wald chi2(11)     =     104.53
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
        anti |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        self |  -.0620586    .009518    -6.52   0.000    -.0807135   -.0434036
         pov |    .246818   .0804041     3.07   0.002     .0892288    .4044072
             |
        year |
          92 |   .0473322   .0587008     0.81   0.420    -.0677193    .1623836
          94 |   .2163669   .0588738     3.68   0.000     .1009763    .3317575
             |
     1.black |   .2268535   .1255617     1.81   0.071     -.019243    .4729499
  1.hispanic |  -.2181591   .1380795    -1.58   0.114      -.48879    .0524718
    childage |   .0884583   .0909947     0.97   0.331     -.089888    .2668047
   1.married |   -.049499   .1262863    -0.39   0.695    -.2970156    .1980176
    1.gender |  -.4834304   .1064056    -4.54   0.000    -.6919815   -.2748793
      momage |  -.0219284   .0252608    -0.87   0.385    -.0714386    .0275818
   1.momwork |   .2612145   .1145722     2.28   0.023     .0366571     .485772
       _cons |   2.531237   1.094669     2.31   0.021     .3857254    4.676749
-------------+----------------------------------------------------------------
     sigma_u |  1.1355938
     sigma_e |  .99707353
         rho |  .56467881   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. est store xtreg
```

```
. mixed anti self pov i.year i.black i.hispanic childage i.married i.gender momage i.momwork || id:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -2927.1991
Iteration 1:   log likelihood = -2927.1991

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =      1,743
Group variable: id                              Number of groups  =        581

                                                Obs per group:
                                                              min =          3
                                                              avg =        3.0
                                                              max =          3

                                                Wald chi2(11)     =     105.36
Log likelihood = -2927.1991                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
        anti |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        self |  -.0620764   .0094874    -6.54   0.000    -.0806715   -.0434814
         pov |   .2471376    .080136     3.08   0.002     .0900739    .4042013
             |
        year |
          92 |   .0473396   .0585299     0.81   0.419    -.0673769     .162056
          94 |   .2163811   .0587023     3.69   0.000     .1013267    .3314355
             |
     1.black |   .2267537   .1249996     1.81   0.070     -.018241    .4717483
  1.hispanic |  -.2182088   .1374561    -1.59   0.112    -.4876177    .0512001
    childage |   .0884559   .0905831     0.98   0.329    -.0890837    .2659956
   1.married |  -.0495647   .1257172    -0.39   0.693     -.295966    .1968365
    1.gender |  -.4834488   .1059246    -4.56   0.000    -.6910572   -.2758405
      momage |  -.0219197   .0251467    -0.87   0.383    -.0712064    .0273669
   1.momwork |   .2611318   .1140581     2.29   0.022      .037582    .4846816
       _cons |   2.531431    1.08976     2.32   0.020     .3955417    4.667321
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   1.282674   .0960323      1.107612    1.485404
-----------------------------+------------------------------------------------
               var(Residual) |   .9928691   .0412577      .9152108    1.077117
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 518.98        Prob >= chibar2 = 0.0000

. est store mixed

. lrtest mixed reg, force

Likelihood-ratio test                                 LR chi2(2)  =    518.98
(Assumption: reg nested in mixed)                     Prob > chi2 =    0.0000
```

At the bottom of the mixed output, you see `LR test vs. linear model: chibar2(01) = 518.98.` This is the same as the `lrtest` of the mixed model versus the OLS regression model. If the test statistic were not significant, it would mean that it was ok to use OLS regression.

```
. esttab reg xtreg mixed, nobaselevels mtitles
```

```
------------------------------------------------------------
                      (1)             (2)             (3)
                      reg           xtreg           mixed
------------------------------------------------------------
main
self              -0.0741***      -0.0621***      -0.0621***
                  (-6.76)         (-6.52)         (-6.54)

pov                 0.435***        0.247**         0.247**
                   (5.09)          (3.07)          (3.08)

92.year            0.0522          0.0473          0.0473
                   (0.59)          (0.81)          (0.81)

94.year             0.226*          0.216***        0.216***
                   (2.54)          (3.68)          (3.69)

1.black             0.168           0.227           0.227
                   (1.90)          (1.81)          (1.81)

1.hispanic         -0.248**        -0.218          -0.218
                  (-2.62)         (-1.58)         (-1.59)

childage           0.0871          0.0885          0.0885
                   (1.40)          (0.97)          (0.98)

1.married         -0.0889         -0.0495         -0.0496
                  (-1.02)         (-0.39)         (-0.39)

1.gender           -0.495***       -0.483***       -0.483***
                  (-6.79)         (-4.54)         (-4.56)

momage            -0.0167         -0.0219         -0.0219
                  (-0.96)         (-0.87)         (-0.87)

1.momwork           0.212**         0.261*          0.261*
                   (2.65)          (2.28)          (2.29)

_cons               2.675***        2.531*          2.531*
                   (3.48)          (2.31)          (2.32)
------------------------------------------------------------
lns1_1_1
_cons                                                0.124***
                                                    (3.33)
------------------------------------------------------------
lnsig_e
_cons                                              -0.00358
                                                    (-0.17)
------------------------------------------------------------
N                    1743            1743            1743
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

As you can see, the `mixed` and `xtreg` regression coefficients are virtually identical. Using OLS regression would cause some effects to be mis-estimated, especially poverty. Among other things, the multilevel model shows us that higher self-esteem tends to reduce anti-social behavior while being in poverty tends to increase it. Also girls have lower levels of anti-social behavior while anti-social behavior tends to be a little higher for those children with working mothers.

Logistic Mixed Effects Models – 2 Levels. `xtlogit` random effects models can also be estimated using the `melogit` command in Stata. At least for simpler models, the procedures are very similar to what you do with `mixed`.

Here is an example from Allison's 2009 book *Fixed Effects Regression Models*. Data are from the National Longitudinal Study of Youth (NLSY). The data set has 1151 teenage girls who were interviewed annually for 5 years beginning in 1979. The data have already been reshaped and `xtset` so they can be used for panel data analysis. That is, each of the 1151 cases has 5 different records, one for each year of the study. The variables are

- id is the subject id number and is the same across each wave of the survey
- year is the year the data were collected in. 1 = 1979, 2 = 1980, etc.
- pov is coded 1 if the subject was in poverty during that time period, 0 otherwise.
- age is the age at the first interview.
- black is coded 1 if the respondent is black, 0 otherwise.
- mother is coded 1 if the respondent currently has at least 1 child, 0 otherwise.
- spouse is coded 1 if the respondent is currently living with a spouse, 0 otherwise.
- school is coded 1 if the respondent is currently in school, 0 otherwise.
- hours is the hours worked during the week of the survey.

Similar to before, we estimate models using `logit`, `xtlogit`, and `melogit`, and note the similarities and differences between them.

```
. * 2 level logit models, preceded by single-level logit model
. use https://www3.nd.edu/~rwilliam/statafiles/teenpovxt, clear

. logit pov i.mother i.spouse i.school hours i.year i.black age, nolog

Logistic regression                             Number of obs     =      5,755
                                                LR chi2(10)       =     490.47
                                                Prob > chi2       =     0.0000
Log likelihood = -3567.5752                     Pseudo R2         =     0.0643

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   .9122333   .0852721    10.70   0.000     .7451031    1.079364
    1.spouse |  -1.169479   .1174809    -9.95   0.000    -1.399737   -.9392206
    1.school |  -.3099841   .0778067    -3.98   0.000    -.4624824   -.1574859
       hours |  -.0254242   .0023527   -10.81   0.000    -.0300355    -.020813
             |
        year |
          2  |   .2132299   .0888648     2.40   0.016     .0390581    .3874017
          3  |   .1310815   .0916184     1.43   0.153    -.0484873    .3106504
          4  |   .1277693   .0947098     1.35   0.177    -.0578586    .3133972
          5  |   .0207599   .0994805     0.21   0.835    -.1742183     .215738
             |
     1.black |   .4848109   .0586833     8.26   0.000     .3697937     .599828
         age |  -.0717551    .028906    -2.48   0.013    -.1284097   -.0151004
       _cons |   .5472231   .4735445     1.16   0.248    -.3809071    1.475353
------------------------------------------------------------------------------

. est store logit
```

```
. xtlogit pov i.mother i.spouse i.school hours i.year i.black age, re nolog

Random-effects logistic regression              Number of obs     =      5,755
Group variable: id                              Number of groups  =      1,151

Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =          5
                                                              avg =        5.0
                                                              max =          5

Integration method: mvaghermite                 Integration pts.  =         12

                                                Wald chi2(10)     =     266.60
Log likelihood  = -3403.7655                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   1.009877    .118372     8.53   0.000     .7778724    1.241882
    1.spouse |  -1.171833   .1512544    -7.75   0.000    -1.468286   -.8753802
    1.school |  -.1145721   .0990775    -1.16   0.248    -.3087604    .0796163
       hours |  -.0259014   .0028771    -9.00   0.000    -.0315403   -.0202624
             |
        year |
          2  |   .2830958   .1000437     2.83   0.005     .0870138    .4791778
          3  |    .213423   .1040523     2.05   0.040     .0094842    .4173618
          4  |   .2415184   .1090094     2.22   0.027     .0278639     .455173
          5  |   .1447937   .1161395     1.25   0.212    -.0828355     .372423
             |
     1.black |   .6093942   .0975653     6.25   0.000     .4181698    .8006186
         age |  -.0627952   .0472163    -1.33   0.184    -.1553373     .029747
       _cons |  -.0045847   .7620829    -0.01   0.995     -1.49824     1.48907
-------------+----------------------------------------------------------------
    /lnsig2u |   .3086358   .1008833                      .1109083    .5063634
-------------+----------------------------------------------------------------
     sigma_u |   1.166862   .0588584                     1.057021    1.288117
         rho |   .2927197   .0208864                      .2535175    .3352612
------------------------------------------------------------------------------
LR test of rho=0: chibar2(01) = 327.62                  Prob >= chibar2 = 0.000

. est store xtlogit
```

```
. melogit pov i.mother i.spouse i.school hours i.year i.black age || id:, nolog

Mixed-effects logistic regression               Number of obs     =      5,755
Group variable:              id                 Number of groups  =      1,151

                                                Obs per group:
                                                              min =          5
                                                              avg =        5.0
                                                              max =          5

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(10)     =     266.64
Log likelihood = -3403.7637                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
    1.mother |   1.009935   .1183721     8.53   0.000     .7779301     1.24194
    1.spouse |  -1.171859   .1512457    -7.75   0.000    -1.468295   -.8754231
    1.school |   -.114617   .0990711    -1.16   0.247    -.3087927    .0795587
       hours |  -.0259016   .0028769    -9.00   0.000    -.0315403   -.0202629
             |
        year |
          2  |   .2830838   .1000419     2.83   0.005     .0870052    .4791624
          3  |   .2134042     .10405     2.05   0.040       .00947    .4173385
          4  |   .2414921   .1090061     2.22   0.027      .027844    .4551401
          5  |    .144759   .1161351     1.25   0.213    -.0828617    .3723796
             |
     1.black |   .6094854   .0975621     6.25   0.000     .4182672    .8007036
         age |  -.0628037   .0472134    -1.33   0.183    -.1553403     .029733
       _cons |  -.0045483   .7620352    -0.01   0.995     -1.49811    1.489013
-------------+----------------------------------------------------------------
id           |
   var(_cons)|   1.361483   .1371712                      1.117513    1.658715
------------------------------------------------------------------------------
LR test vs. logistic model: chibar2(01) = 327.62      Prob >= chibar2 = 0.0000

. est store melogit

. lrtest melogit logit, force

Likelihood-ratio test                                 LR chi2(1)  =    327.62
(Assumption: logit nested in melogit)                 Prob > chi2 =    0.0000
```

Similar to before, `melogit` reports `LR test vs. logistic model: chibar2(01) = 327.62`. This is the same as the `lrtest` of the `melogit` vs `logit` models. This indicates that it would be a mistake to ignore the multilevel nature of the nature (i.e. assume cases were uncorrelated within clusters).

```
. * ln2sigu and var(_cons) are the same thing parameterized differently
. di exp(.309)
1.3620624
```

`xtlogit` reported ln2sigu equaled .309 while `melogit` reported var(cons) equaled 1.361483. These are actually the same number just parameterized differently, i.e. one is logged and the other is not.

```
. esttab logit xtlogit melogit, nobaselevels mtitles

------------------------------------------------------------
                      (1)             (2)             (3)
                    logit         xtlogit         melogit
------------------------------------------------------------
pov
1.mother            0.912***        1.010***        1.010***
                   (10.70)          (8.53)          (8.53)

1.spouse           -1.169***       -1.172***       -1.172***
                   (-9.95)         (-7.75)         (-7.75)

1.school           -0.310***       -0.115          -0.115
                   (-3.98)         (-1.16)         (-1.16)

hours              -0.0254***      -0.0259***      -0.0259***
                  (-10.81)         (-9.00)         (-9.00)

2.year              0.213*          0.283**         0.283**
                   (2.40)          (2.83)          (2.83)

3.year              0.131           0.213*          0.213*
                   (1.43)          (2.05)          (2.05)

4.year              0.128           0.242*          0.241*
                   (1.35)          (2.22)          (2.22)

5.year             0.0208           0.145           0.145
                   (0.21)          (1.25)          (1.25)

1.black             0.485***        0.609***        0.609***
                   (8.26)          (6.25)          (6.25)

age                -0.0718*        -0.0628         -0.0628
                   (-2.48)         (-1.33)         (-1.33)

_cons               0.547          -0.00458        -0.00455
                   (1.16)          (-0.01)         (-0.01)
------------------------------------------------------------
lnsig2u
_cons                               0.309**
                                   (3.06)
------------------------------------------------------------
var(_cons[~)
_cons                                               1.361***
                                                   (9.93)
------------------------------------------------------------
N                    5755            5755            5755
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

The `xtlogit` and `melogit` results are identical other than some very slight differences caused by using different algorithms. Both differ somewhat from the `logit` results, which ignore the multilevel nature of the data. Among other things the multilevel model results show that having a spouse and working more hours tend to reduce the likelihood of being in poverty, while having a child or being black tend to increase the likelihood.

Logistic Mixed Effects Models – 3 Levels. In the examples presented so far there has been no compelling reason to favor `me` commands over `xt` commands. All of these have involved two-level datasets. However the Stata 14 Mixed Effects manual gives several other interesting examples. Here we reproduce an example given for a three-level dataset (again, much of the following material is copied verbatim from the manual with a few little tweaks here and there). From p. 120 of the me manual

> Rabe-Hesketh, Toulopoulou, and Murray (2001) analyzed data from a study measuring the cognitive ability of patients with schizophrenia compared with their relatives and control subjects. Cognitive ability was measured as the successful completion of the "Tower of London", a computerized task, measured at three levels of difficulty. For all but one of the 226 subjects, there were three measurements (one for each difficulty level). Because patients' relatives were also tested, a family identifier, family, was also recorded.

```
. * 3 level logit model, preceded by single-level logit model
. webuse towerlondon, clear
(Tower of London data)

. des

Contains data from http://www.stata-press.com/data/r14/towerlondon.dta
  obs:           677                          Tower of London data
 vars:             5                          31 May 2014 10:41
 size:         4,739                          (_dta has notes)
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
family          int     %8.0g                 Family ID
subject         int     %9.0g                 Subject ID
dtlm            byte    %9.0g                 1 = task completed
difficulty      byte    %9.0g                 Level of difficulty: -1, 0, or 1
group           byte    %8.0g                 1: controls; 2: relatives; 3:
schizophrenics
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Sorted by: family  subject

. fre group

group -- 1: controls; 2: relatives; 3: schizophrenics
-----------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+--------------------------------------------
Valid   1     |        194      28.66      28.66      28.66
        2     |        294      43.43      43.43      72.08
        3     |        189      27.92      27.92     100.00
        Total |        677     100.00     100.00
-----------------------------------------------------------
```

Since each subject (except 1 of the controls) takes 3 tests, we see that the sample consists of 63 schizophrenics, 98 relatives, and 65 controls. (Later output will show that there are 118 families.)

We will list the records for three different families to provide a clearer feel for how the data set is structured.

```
. list if family == 1 | family == 3 | family == 60
```

	family	subject	dtlm	diffic~y	group
1.	1	19	1	-1	3
2.	1	19	0	0	3
3.	1	19	0	1	3
4.	1	20	0	-1	3
5.	1	20	1	0	3
6.	1	20	0	1	3
7.	1	21	1	-1	3
8.	1	21	0	0	3
9.	1	21	0	1	3
10.	1	70	0	-1	2
11.	1	70	0	0	2
12.	1	70	0	1	2
13.	1	71	0	-1	2
14.	1	71	0	0	2
15.	1	71	0	1	2
16.	1	72	1	-1	2
17.	1	72	1	0	2
18.	1	72	0	1	2
19.	1	73	1	-1	2
20.	1	73	0	0	2
21.	1	73	0	1	2
22.	1	74	1	-1	2
23.	1	74	0	0	2
24.	1	74	0	1	2
25.	1	75	0	-1	2
26.	1	75	1	0	2
27.	1	75	0	1	2
49.	3	17	1	-1	3
50.	3	17	0	0	3
51.	3	17	0	1	3
52.	3	18	0	-1	3
53.	3	18	0	0	3
54.	3	18	0	1	3
55.	3	66	0	-1	2
56.	3	66	0	0	2
57.	3	66	0	1	2
58.	3	68	1	-1	2
59.	3	68	0	0	2
60.	3	68	0	1	2
484.	60	186	1	-1	1
485.	60	186	0	0	1
486.	60	186	0	1	1

As we see, family 1 has 27 records. These records are produced by 9 different individuals (subject id #s 19, 20, 21, 70, 71, 72, 73, 74, and 75). All 9 individuals took all 3 versions of the Tower of London test. Three of the individuals were schizophrenics (group = 3) while the other 6 were other family members (group = 2). None of the individuals in this family were classified as controls.

By way of contrast, family 3 had 12 records produced by 4 individuals (subjects 17, 18, 66 and 68) all of whom took all three versions of the Tower of London test. Two were schizophrenic while the other two were other family members.

Family 60 only had 1 individual who had 3 records. The individual was classified as a control. Looking at the data set, there seem to be several families like this, i.e. it appears all the controls came from single-person families with no schizophrenics in them.

We will now do a `logit` and `melogit` analysis of the data. The syntax/ procedure is almost identical to before, except (a) there is no corresponding `xtlogit` command, and (b) individuals are nested within families so the syntax reflects that.

```
. logit dtlm difficulty i.group, nolog

Logistic regression                                     Number of obs     =        677
                                                        LR chi2(3)        =     119.58
                                                        Prob > chi2       =     0.0000
Log likelihood = -313.89079                             Pseudo R2         =     0.1600

------------------------------------------------------------------------------
        dtlm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  difficulty |  -1.313382   .1409487    -9.32   0.000    -1.589636   -1.037127
             |
       group |
          2  |  -.1396641   .2282452    -0.61   0.541    -.5870164    .3076883
          3  |  -.8313329   .2742339    -3.03   0.002    -1.368822   -.2938443
             |
       _cons |  -1.160498   .1824503    -6.36   0.000    -1.518094   -.8029023
------------------------------------------------------------------------------

. est store logit

. melogit dtlm difficulty i.group || family: || subject:, nolog

Mixed-effects logistic regression               Number of obs     =        677

------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+-------------------------------------------
         family |        118          2        5.7         27
        subject |        226          2        3.0          3
------------------------------------------------------------

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(3)      =      74.90
Log likelihood = -305.12041                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
          dtlm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
    difficulty |  -1.648505   .1932075    -8.53   0.000    -2.027185   -1.269826
               |
         group |
            2  |  -.2486841   .3544076    -0.70   0.483    -.9433102     .445942
            3  |  -1.052306   .3999921    -2.63   0.009    -1.836276   -.2683357
               |
         _cons |  -1.485863   .2848455    -5.22   0.000     -2.04415   -.9275762
```

```
--------------+------------------------------------------------------------------
family        |
     var(_cons)|   .5692105   .5215654                      .0944757    3.429459
--------------+------------------------------------------------------------------
family>subject |
     var(_cons)|   1.137917   .6854853                      .3494165    3.705762
------------------------------------------------------------------------------
LR test vs. logistic model: chi2(2) = 17.54              Prob > chi2 = 0.0002

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

. est store melogit

. lrtest logit melogit, force

Likelihood-ratio test                                 LR chi2(2)  =     17.54
(Assumption: logit nested in melogit)                 Prob > chi2 =    0.0002

. esttab logit melogit, nobaselevels mtitles

--------------------------------------------
                      (1)             (2)
                    logit         melogit
--------------------------------------------
dtlm
difficulty         -1.313***       -1.649***
                  (-9.32)         (-8.53)

2.group            -0.140          -0.249
                  (-0.61)         (-0.70)

3.group            -0.831**        -1.052**
                  (-3.03)         (-2.63)

_cons              -1.160***       -1.486***
                  (-6.36)         (-5.22)
--------------------------------------------
var(_cons[~)
_cons                               0.569
                                   (1.09)
--------------------------------------------
var(_cons[~)
_cons                               1.138
                                   (1.66)
--------------------------------------------
N                     677             677
--------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Not surprisingly, the more difficult the test, the less likely individuals are to complete it. Schizophrenics have more difficulty passing the tests than do controls or relatives. The likelihood ratio tests tell us that it would be a mistake to treat these cases as independent observations, and hence logit should not be used.