Panel Data and Multilevel Models for Categorical Outcomes: Introduction

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

Overview. This course is going to focus on categorical outcomes (e.g. panel data and multilevel logistic regression models) but many of the same ideas will also apply to linear models. Because the results from categorical outcome models can often be difficult to interpret, I will also talk about how *adjusted predictions* and *marginal effects* can often make the substantive meaning of results clearer.

Many statistical analyses are done on samples of individuals (or institutions, or countries, etc.) that are sampled independently of each other with each case measured at only one point in time. For such data, methods like OLS regression, logistic regression, t-tests, and Poisson regression may be appropriate.

Other samples, hower, can be much more complex.

- With *panel data*, the same individuals or units are measured at multiple points in time. For example, the Wisconsin Longitudinal Study has periodically collected data from a sample of 1957 Wisconsin high school graduates in an effort to see what happens to them over their lifetimes. Economists sometimes collect information on the annual earnings of corporations, while political scientists might examine countries measured at multiple points over time.
- *Multilevel data* are collected from units organized or observed within units at a higher level (from which data are also obtained). For example, we might have data on students (level 1) who are clustered within classrooms (level 2). Or, we could have siblings clustered in families. In the United States, several studies have followed samples of students who started off in the same classrooms and then followed them as they moved through early grades into early adulthood.
- Panel data are actually a special type of multilevel data – records from multiple time points are clustered by individual.

Panel/ multilevel data offer special challenges. At a minimum the analysis must take into account that the records are not all independent of each other. An individual's response at time 1 will generally not be unrelated to his or her response at time 3. Forty students from the same school will share more in common than forty students from forty different schools. If, say, we had 200 individuals, each of whom was measured at 5 points in time, and we acted as though we had a sample of 1,000 independent cases, our standard errors would be too low and we would overstate the statistical significance of our results.

Getting the data set up correctly can also be a challenge. Data might be in wide format – one record for each case, with a caseid variable and different variables for each time point (e.g. inc1990, inc1992, inc1994). Such data will often have to be restructured to long format, with one record for each individual at each time point. The variables might then be caseid, year, and inc. Wide format will have fewer records but more variables. With long format, the software will have to know how the cases are connected, e.g. what the id variable is.

Beyond that, though, panel/multilevel data offer several unique analytical opportunities. A few that I will focus on include

- The effects of omitted variables can sometimes be controlled for. With *fixed effects models* and *conditional logit models*, individuals basically serve as their own controls. So, for example, if an important variable like gender or race is not included in the data set, you may be able to control for its effects anyway.
- With *multilevel models*, you can examine effects on a case's outcomes from each level, e.g. the parental Socio-Economic status of a child's parents and characteristics of the school that s/he attends. You can also examine interaction effects between levels, e.g. maybe the effect of parent's SES varies by the school attended. You will also often hear these referred to as *random-effects models*, *random-coefficients models*. *Mixed-effects models*, or *hierarchical linear models*.
- Sometimes we are interested more in the timing of events than whether or not the event occurs. For example, everybody dies sooner or later, but what causes some people to die more quickly than others? With the right kind of longitudinal data, you can use regular logistic regression for this. These are called *Discrete Time Methods for the Analysis of Event Histories*.

Course Outline

The first few handouts will focus on the analysis of panel data. With panel data, the same individuals (or countries, or businesses, etc.) are measured at multiple points in time. We will then transition into multilevel and (time permitting) event history models. As Hedeker notes, multilevel data are collected from units organized or observed within units at a higher level (from which data are also obtained). For example, we might have data on students (level 1) who are clustered within classrooms (level 2). Or, we could have siblings clustered in families. As we will see, the same techniques that are used for panel data can often be used with multilevel data, and vice-versa.

Course Web Page. The page may have additional materials not included here, including suggested readings, additional or revised handouts, and Stata do files used in these handouts.

https://www3.nd.edu/~rwilliam/Taiwan2018/index.html

Panel Data and Multilevel Models for Categorical Outcomes: Setting up the Data

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

We often have data where variables have been measured for the same subjects (or countries, or companies, or whatever) at multiple points in time. These are typically referred to as Panel Data or as Cross-Sectional Time Series Data. We need special techniques for analyzing such data, e.g. it would be a mistake to treat 200 individuals measured at 5 points in time as though they were 1,000 independent observations. Therefore, Stata has an entire manual and suite of XT commands devoted to panel data, e.g. `xtreg`, `xtlogit`, `xtpoisson`, etc. Some other commands, like `clogit`, can also sometimes be used. (Conversely, the `xt` commands can sometimes be used when you don't have panel data, e.g. you have data from students within a school. In such situations you might also use the `me`, mixed-effects, commands.)

In order to use these commands, though, the data set needs to be properly structured. This will sometimes require that the data be restructured from wide to long. In wide format, a data set has one record for each subject. This record has several variables, e.g. income1, income2, income3, where each of the income variables gives the value of income at a different time point. In long format, the data are restructured to have one record for each subject for each time point. I am going to give some examples of how to do this, but if in doubt be sure to read the Stata documentation for help on setting up your data.

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. Here is a listing of the values for the first three cases:

```
. version 13.1
. use https://www3.nd.edu/~rwilliam/statafiles/teenpov, clear
. rename inschool* school*
. list in 1/3
```

1.

id	pov1	mother1	spouse1	school1	hours1	pov2	mother2	spouse2	school2	hours2	pov3	mother3	spouse3	school3	hours3	pov4
22	1	0	0	1	21	0	0	0	1	15	0	0	0	1	3	0

mother4	spouse4	school4	hours4	age	black	pov5	mother5	spouse5	school5	hours5
0	0	1	0	16	0	0	0	0	1	0

2.

id	pov1	mother1	spouse1	school1	hours1	pov2	mother2	spouse2	school2	hours2	pov3	mother3	spouse3	school3	hours3	pov4
75	0	0	0	1	8	0	0	0	1	0	0	0	0	1	0	0

mother4	spouse4	school4	hours4	age	black	pov5	mother5	spouse5	school5	hours5
0	0	1	4	17	0	1	0	0	1	0

3.

id	pov1	mother1	spouse1	school1	hours1	pov2	mother2	spouse2	school2	hours2	pov3	mother3	spouse3	school3	hours3	pov4
92	0	0	0	1	30	0	0	0	1	27	0	0	0	1	24	1

mother4	spouse4	school4	hours4	age	black	pov5	mother5	spouse5	school5	hours5
1	0	0	31	16	0	1	1	0	0	0

The numbers at the ends of some variable names reflect the time period the variable refers to (1 = 1979, 2 = 1980, etc.) Variables without numbers in the names do not vary across time.

- id is the subject id number and is the same across each wave of the survey

- pov_t 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.
- $mother_t$ is coded 1 if the respondent currently has at least 1 child, 0 otherwise.
- $spouse_t$ is coded 1 if the respondent is currently living with a spouse, 0 otherwise.
- $school_t$ is coded 1 if the respondent is currently in school, 0 otherwise.
- $hours_t$ is the hours worked during the week of the survey.

The data are currently in wide format. There is one record per case with multiple variables representing values at different points in time. We need to get the data into long format instead. In Stata, we can do this with the `reshape` command.

```
. reshape long pov mother spouse school hours, i(id) j(year)
(note: j = 1 2 3 4 5)

Data                               wide   ->   long
-----------------------------------------------------------
Number of obs.                     1151   ->    5755
Number of variables                  28   ->       9
j variable (5 values)                     ->   year
xij variables:
                     pov1 pov2 ... pov5   ->   pov
            mother1 mother2 ... mother5   ->   mother
            spouse1 spouse2 ... spouse5   ->   spouse
            school1 school2 ... school5   ->   school
               hours1 hours2 ... hours5   ->   hours
-----------------------------------------------------------
```

The `reshape long` part of the command told Stata we wanted to reshape the data from wide to long. (There is also a `reshape wide` command for going from long to wide.) The variable list that followed was the list of variables (actually the stubnames of the variables) that varied across time (you should use a consistent naming convention, e.g. pov1, mother1, etc. pov79, mother79, pov80, mother80, would have also been ok. Be careful about doing something like inc2, inc79, inc80, inc81, where inc2 = income squared; Stata will think inc2 is another of the time-varying variables.) The variables not listed are those that do not vary across time; their values will be copied on to each of the new records for the case. `i(varlist)` specifies the variables whose unique values denote a logical observation. i() is required. In this case only `i(id)` was needed but in other cases multiple variables might define a case. `j(varname)` specifies the variable whose unique values denote a subobservation. Here is what the reshaped data for the first 3 (now 15) cases looks like.

```
. list in 1/15
```

	id	year	age	black	pov	mother	spouse	school	hours
1.	22	1	16	0	1	0	0	1	21
2.	22	2	16	0	0	0	0	1	15
3.	22	3	16	0	0	0	0	1	3
4.	22	4	16	0	0	0	0	1	0
5.	22	5	16	0	0	0	0	1	0
6.	75	1	17	0	0	0	0	1	8
7.	75	2	17	0	0	0	0	1	0
8.	75	3	17	0	0	0	0	1	0
9.	75	4	17	0	0	0	0	1	4
10.	75	5	17	0	1	0	0	1	0
11.	92	1	16	0	0	0	0	1	30
12.	92	2	16	0	0	0	0	1	27
13.	92	3	16	0	0	0	0	1	24
14.	92	4	16	0	1	1	0	0	31
15.	92	5	16	0	1	1	0	0	0

Each of the original cases now has 5 records, one for each year of the study. The value of year varies from 1 to 5. The values of age (age at first interview) and black have been duplicated on each of the 5 records. Instead of 5 poverty variables, we have 1, whose value can differ across the five records (e.g. the original value of pov2 for id 22 is now the value of pov for id 22 year 2). The same is true for the other time-varying variables.

The next thing we want to do is xtset the data. The `xtset` command tells Stata that these are Panel data. The usual format is

```
xtset panelvar
xtset panelvar timevar
```

That is, we must tell Stata what the panelvar is; in this case it is id. The timevar is optional and may or may not be necessary depending on our analysis. In the current case the timevar is year. `xtset` typed with no parameters tells us how the data are xtset.

```
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1 to 5
                delta:  1 unit

. xtset
       panel variable:  id (strongly balanced)
        time variable:  year, 1 to 5
                delta:  1 unit
```

NOTE (copied verbatim from the Stata 12 Manual): “The terms balanced and unbalanced are often used to describe whether a panel dataset is missing some observations. If a dataset does not contain a time variable, then panels are considered balanced if each panel contains the same number of observations; otherwise, the panels are unbalanced. When the dataset contains a time variable, panels are said to be strongly balanced if each panel contains the same time points, weakly balanced if each panel contains the same number of observations but not the same time points, and unbalanced otherwise.”

A data set might be unbalanced because data are missing for some years. If you were, say, analyzing countries, it might even be that the country did not exist during some time periods. Strongly balanced data are best but my understanding is that Stata can generally do a good job with unbalanced data.

Once the data are xtset, several commands are available to us; see `help xt`. For example, you can use the `xtsum` command, which is similar to the `summarize` command but contains some additional information.

```
. xtsum

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
id       overall |  6016.672   3298.064         22      12539 |     N =    5755
         between |             3299.211         22      12539 |     n =    1151
         within  |                   0   6016.672   6016.672 |     T =       5
                 |                                          |
year     overall |         3   1.414336          1          5 |     N =    5755
         between |                   0          3          3 |     n =    1151
         within  |             1.414336          1          5 |     T =       5
                 |                                          |
age      overall |  15.64639    1.04682         14         17 |     N =    5755
         between |             1.047184         14         17 |     n =    1151
         within  |                   0   15.64639   15.64639 |     T =       5
                 |                                          |
black    overall |  .5742832   .4944942          0          1 |     N =    5755
         between |             .4946661          0          1 |     n =    1151
         within  |                   0   .5742832   .5742832 |     T =       5
                 |                                          |
pov      overall |  .3768897    .484649          0          1 |     N =    5755
         between |             .3100424          0          1 |     n =    1151
         within  |             .3725925  -.4231103    1.17689 |     T =       5
                 |                                          |
mother   overall |  .1986099   .3989883          0          1 |     N =    5755
         between |             .3253864          0          1 |     n =    1151
         within  |             .2310605  -.6013901   .9986099 |     T =       5
                 |                                          |
spouse   overall |  .0992181   .2989806          0          1 |     N =    5755
         between |             .2206498          0          1 |     n =    1151
         within  |             .2018338  -.7007819   .8992181 |     T =       5
                 |                                          |
school   overall |  .6304083   .4827361          0          1 |     N =    5755
         between |               .32013          0          1 |     n =    1151
         within  |             .3614169  -.1695917   1.430408 |     T =       5
                 |                                          |
hours    overall |  8.671764   14.54341          0         90 |     N =    5755
         between |             9.363817          0       52.4 |     n =    1151
         within  |             11.13062 -43.72824   72.07176 |     T =       5
```

The different values for the standard deviations can sometimes be useful. For id, age and black, the within standard deviation is 0. This is because, within each subject, the value of these variables does not vary, i.e. for each of the five records the case has, the values of these variables are the same. For year, the between subjects standard deviation is 0. This is because all subjects have the same set of values on year. For poverty, the between and within standard deviations are nearly the same. This tells us that the variation in poverty across women is nearly equal to that observed within a woman over time. That is, if you were to draw two women randomly from the

data, the difference in poverty is expected to be nearly equal to the difference for the same woman in two randomly selected years.

As shown elsewhere, the amount of within-subject variability will impact which types of models work best for a particular problem.

Panel Data and Multilevel Models for Categorical Outcomes: Fixed effects and conditional logit 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 from Paul Allison's book, *Fixed Effects Regression Models for Categorical Data*. The Stata XT and ME manuals are also good references. See Allison's book for a more detailed explanations of why assertions made here are true and what the technical details behind the models are.

Overview. In experimental research, unmeasured differences between subjects are often controlled for via random assignment to treatment and control groups. Hence, even if a variable like Socio-Economic Status is not explicitly measured, because of random assignment, we can be reasonably confident that the effects of SES are approximately equal for all groups. Of course, random assignment is usually not possible with most survey research. If we want to control for the effect of a variable, we must explicitly measure it. If we don't measure it, we can't control for it. In practice, there will almost certainly be some variables we have failed to measure (or have measured poorly), so our models will likely suffer from some degree of omitted variable bias.

Allison notes, however, that when we have panel data (the same subjects measured at two or more points in time) another alternative presents itself: we can use the subjects as their own controls. With binary dependent variables, this can be done via the use of *conditional logit/fixed effects logit models*. With panel data we can control for stable characteristics (i.e. characteristics that do not change across time) whether they are measured or not. These include such things as sex, race, and ethnicity, as well as more difficult to measure variables such as intelligence, parents' child-rearing practices, and genetic makeup. This does not control for time-varying variables, but such variables can be explicitly included in the model, e.g. employment status, income.

Examples (from Allison): Suppose you want to know whether marriage reduced recidivism among chronic offenders. We could compare an individual's arrest rate when he is married with his arrest rate when he is not. The difference in arrest rates between the two periods is an estimate of the marriage effect for that individual. Or, you might see how a child's performance in school differs depending on how much time s/he spends playing video games. So, you could compare how the child does when not spending much time on video games versus when s/he does.

Allison notes there are two conditions for using fixed effects methods.

- The dependent variable must be measured on at least two occasions for each individual.
- The independent variables must change across time for some substantial portion of the individuals. Fixed effects models are not much good for looking at the effects of variables that do not change across time, like race and sex.

There are several other points to be aware of with fixed effects logit models.

- The good thing is that the effects of stable characteristics, such as race and gender, are controlled for, whether they are measured or not. The bad thing is that the effects of these variables are not estimated. Again, it is similar to an experiment with random assignment. The effects of variables not explicitly measured are controlled for (because random assignment makes the groups more or less similar on these characteristics) but their effects are not estimated.
- Other methods (e.g. random effects) can be used when we want to estimate the effects of variables like sex and race, but then the method is no longer controlling for omitted variables.
- Fixed effects estimates *use only within-individual differences*, essentially discarding any information about differences between individuals. If predictor variables vary greatly across individuals but have little variation over time for each individual, then fixed effects estimates will be imprecise and have large standard errors.
    - Why tolerate the higher errors? Allison says there is a trade-off between bias and efficiency. Other methods, e.g. random effects, will suffer from omitted variable bias; fixed effects methods help to control for omitted variable bias by having individuals serve as their own controls.
    - Keep in mind, however, that fixed effects doesn't control for unobserved variables that change over time. So, for example, a failure to include income in the model could still cause fixed effects coefficients to be biased.
    - Allison likes fixed effects models because they are less vulnerable to omitted variable bias. But he cautions that "in applications where the within-person variation is small relative to the between-person variation, the standard errors of the fixed effects coefficients may be too large to tolerate."
- Conditional logit/fixed effects models can be used for things besides Panel Studies. For example, Long & Freese show how conditional logit models can be used for alternative-specific data. If you read both Allison's and Long & Freese's discussion of the `clogit` command, you may find it hard to believe they are talking about the same command!

Example. 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.

We can use either Stata's `clogit` command or the `xtlogit, fe` command to do a fixed effects logit analysis. Both give the same results. (In fact, I believe `xtlogit, fe` actually calls `clogit`.) First we will use `xtlogit` with the `fe` option.

```
. use https://www3.nd.edu/~rwilliam/statafiles/teenpovxt, clear
. xtlogit pov i.mother i.spouse i.school hours i.year, fe nolog
note: multiple positive outcomes within groups encountered.
note: 324 groups (1,620 obs) dropped because of all positive or
      all negative outcomes.

Conditional fixed-effects logistic regression   Number of obs     =      4,135
Group variable: id                              Number of groups  =        827

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

                                                LR chi2(8)        =      97.28
Log likelihood  = -1520.1139                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   .5824322   .1595831     3.65   0.000      .269655    .8952094
    1.spouse |  -.7477585   .1753466    -4.26   0.000    -1.091431   -.4040854
    1.school |   .2718653   .1127331     2.41   0.016     .0509125    .4928181
       hours |  -.0196461   .0031504    -6.24   0.000    -.0258208   -.0134714
             |
        year |
          2  |   .3317803   .1015628     3.27   0.001      .132721    .5308397
          3  |   .3349777   .1082496     3.09   0.002     .1228124     .547143
          4  |   .4327654   .1165144     3.71   0.000     .2044013    .6611295
          5  |   .4025012   .1275277     3.16   0.002     .1525514     .652451
------------------------------------------------------------------------------
```

Here is how we interpret the results. The note "multiple positive outcomes within groups encountered" is a warning that you may need to check your data, because with some analyses there should be no more than one positive outcome. In the present case, that is not a problem, i.e. there is no reason that respondents cannot be in poverty at multiple points in time.

The note "324 groups (1620 obs) dropped because of all positive or all negative outcomes" means that 324 subjects were either in poverty during all 5 time periods or were not in poverty during all 5 time periods. Fixed-effects models are looking at the determinants of within-subject variability. If there is no variability within a subject, there is nothing to examine. Put another way, in the 827 groups that remained, sometime during the 5 year period the subject went from being in poverty to being out of poverty; or else switched from being out of poverty to being in poverty. If poverty status were something that hardly ever changed across time, or if very few people were ever in poverty, there would not be many cases left for a fixed effects analysis. Even as it is, more than a fourth of the sample has been dropped from the analysis. (Other techniques, like `xtreg, fe`, won't cost you so many cases.)

In terms of interpreting the coefficients, it may also be helpful to have the odds ratios.

```
. xtlogit, or

Conditional fixed-effects logistic regression   Number of obs     =      4,135
Group variable: id                              Number of groups  =        827

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

                                                LR chi2(8)        =      97.28
Log likelihood  = -1520.1139                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         pov |         OR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   1.790388   .2857157     3.65   0.000     1.309513    2.447848
    1.spouse |   .4734266   .0830137    -4.26   0.000     .3357355    .6675871
    1.school |    1.31241   .1479521     2.41   0.016     1.052231    1.636923
       hours |   .9805456   .0030891    -6.24   0.000     .9745098    .9866189
             |
        year |
          2  |   1.393447   .1415223     3.27   0.001     1.141931    1.700359
          3  |   1.397909   .1513231     3.09   0.002     1.130672    1.728308
          4  |   1.541515   .1796087     3.71   0.000      1.22679    1.936979
          5  |   1.495561   .1907255     3.16   0.002     1.164802    1.920242
------------------------------------------------------------------------------
```

The OR for mother is 1.79. This means that, if a girl switches from not having children to having children, her odds of being in poverty are multiplied by 1.79. Remember, these are teenagers at the start of the study, so having a baby while you are still very young is not good in terms of avoiding poverty. Conversely, if a girl switches from being unmarried to married, her odds of being in poverty get multiplied by .47, i.e. getting married helps you to stay out of poverty. Being in school multiplies the odds of poverty by 31 percent, while each additional hour you work reduces the odds of poverty by 2 percent. The year coefficients are all comparisons with year 1 and are all positive and significant; on an all other things equal basis, teens are more likely to be in poverty in the later years.

Notice that we did NOT include the time-invariant variables for age and black. Let's see what happens when we do.

```
. xtlogit pov i.mother i.spouse i.school hours i.year age i.black, fe nolog
note: multiple positive outcomes within groups encountered.
note: 324 groups (1,620 obs) dropped because of all positive or
      all negative outcomes.
note: age omitted because of no within-group variance.
note: 1.black omitted because of no within-group variance. [Rest of output deleted]
```

The two variables get dropped because their values do not vary within each group. Something that is a constant cannot explain variability in a dependent variable. (Allison, however, demonstrates that interactions between time-varying and time-constant variables can be included in the model.)

To do the same thing with `clogit`,

```
. use https://www3.nd.edu/~rwilliam/statafiles/teenpovxt, clear
. xtset, clear
. clogit pov i.mother i.spouse i.school hours i.year, group(id) nolog
note: multiple positive outcomes within groups encountered.
note: 324 groups (1,620 obs) dropped because of all positive or
      all negative outcomes.

Conditional (fixed-effects) logistic regression

                                                Number of obs     =      4,135
                                                LR chi2(8)        =      97.28
                                                Prob > chi2       =     0.0000
Log likelihood = -1520.1139                     Pseudo R2         =     0.0310

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   .5824322   .1595831     3.65   0.000      .269655    .8952094
    1.spouse |  -.7477585   .1753466    -4.26   0.000    -1.091431   -.4040854
    1.school |   .2718653   .1127331     2.41   0.016     .0509125    .4928181
       hours |  -.0196461   .0031504    -6.24   0.000    -.0258208   -.0134714
             |
        year |
          2  |   .3317803   .1015628     3.27   0.001      .132721    .5308397
          3  |   .3349777   .1082496     3.09   0.002     .1228124     .547143
          4  |   .4327654   .1165144     3.71   0.000     .2044013    .6611295
          5  |   .4025012   .1275277     3.16   0.002     .1525514     .652451
------------------------------------------------------------------------------
```

I did not need to clear the xtsettings; but I did so to illustrate that with `clogit`, it isn't necessary to `xtset` the data. Instead, the panelvar is specified by using the `group` option. Further, with neither method was the timevar actually needed. Instead of years, these could have been children within schools. The `xt` labeling of commands can be deceptive in that you do not necessarily need to have longitudinal data to use some of the commands.

WARNING!!! As I will explain later, marginal effects and adjusted predictions can often provide a great way to make the results from Categorical outcomes models more interpretable. But, Marginal effects and predicted values after `xtlogit, fe` and `clogit` can be problematic. By default, margins is giving you "the probability of a positive outcome assuming that the fixed effect is zero." This may be an unreasonable assumption. For a discussion of the problem and possible solutions, see Steve Samuels' comments at

http://www.statalist.org/forums/forum/general-stata-discussion/general/1304704-cannot-estimate-marginal-effect-after-xtlogit

Panel Data and Multilevel Models for Categorical Outcomes: Fixed effects versus random effects models

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

These notes borrow heavily from Paul Allison's book, *Fixed Effects Regression Models for Categorical Data*. The Stata XT manual is also a good reference.

Overview. With panel/cross sectional time series data, the most commonly estimated models are probably fixed effects and random effects models. Population-Averaged Models and Mixed Effects models are also sometime used. In this handout we will focus on the major differences between fixed effects and random effects models.

Several considerations will affect the choice between a fixed effects and a random effects model.

1. *What is the nature of the variables that have been omitted from the model?*
    a. If you think there are no omitted variables – or if you believe that the omitted variables are uncorrelated with the explanatory variables that are in the model – then a random effects model is probably best. It will produce unbiased estimates of the coefficients, use all the data available, and produce the smallest standard errors. More likely, however, is that omitted variables will produce at least some bias in the estimates.
    b. If there are omitted variables, and these variables are correlated with the variables in the model, then fixed effects models may provide a means for controlling for omitted variable bias. In a fixed-effects model, subjects serve as their own controls. The idea/hope is that whatever effects the omitted variables have on the subject at one time, they will also have the same effect at a later time; hence their effects will be constant, or "fixed." HOWEVER, in order for this to be true, the omitted variables must have time-invariant values with time-invariant effects.
        i. By time-invariant values, we mean that the value of the variable does not change across time. Gender and race are obvious examples, but this can also include things like the Educational Level of the Respondent's Father.
        ii. By time-invariant effects, we mean the variable has the same effect across time, e.g. the effect of gender on the outcome at time 1 is the same as the effect of gender at time 5.
        iii. If either of these assumptions is violated, we need to have explicit measurements of the variables in question and include them in our models. In the case of time-varying effects, we can include things like the interaction of gender with time. We also need explicit measurements of time-invariant variables if they are thought to interact with other variables in the model, e.g. we think the effect of SES differs by race.
2. *How much variability is there within subjects?*
    a. If subjects change little, or not at all, across time, a fixed effects model may not work very well or even at all. There needs to be within-subject variability in the variables if we are to use subjects as their own controls. If there is little variability

within subjects then the standard errors from fixed effects models may be too large to tolerate.
    b. Conversely, random effects models will often have smaller standard errors. But, the trade-off is that their coefficients are more likely to be biased.
3. *Do we wish to estimate the effects of variables whose values do not change across time, or do we merely wish to control for them?*
    a. With fixed effects models, we do not estimate the effects of variables whose values do not change across time. Rather, we control for them or "partial them out." This is similar to an experiment with random assignment. We may not measure variables like SES, but whatever effects those variable have are (subject to sampling variability) assumed to be more or less the same across groups because of random assignment.
    b. Random effects models will estimate the effects of time-invariant variables, but the estimates may be biased because we are not controlling for omitted variables.
4. *Does the study design already control for omitted variables and differences across groups?*
    a. Many clinical studies randomly assign people to treatment and control groups, or rely on some sort of matching procedure when selecting subjects.
    b. As a result you will often see more emphasis on random effects models and less on fixed effects.

Fixed effects models. Allison says "In a fixed effects model, the unobserved variables are allowed to have any associations whatsoever with the observed variables." Fixed effects models control for, or partial out, the effects of time-invariant variables with time-invariant effects. This is true whether the variable is explicitly measured or not. Exactly how they do so varies by the statistical technique being used. The optional appendix discusses these methods further. Unfortunately, the effects of time-invariant variables that are measured cannot be estimated.

```
. use https://www3.nd.edu/~rwilliam/statafiles/teenpovxt, clear
. *fixed effects
. xtlogit pov i.mother i.spouse i.school hours i.year i.black age, fe nolog
note: multiple positive outcomes within groups encountered.
note: 324 groups (1,620 obs) dropped because of all positive or
      all negative outcomes.
note: 1.black omitted because of no within-group variance.
note: age omitted because of no within-group variance.

Conditional fixed-effects logistic regression   Number of obs     =      4,135
Group variable: id                              Number of groups  =        827

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

                                                LR chi2(8)        =      97.28
Log likelihood  = -1520.1139                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   .5824322   .1595831     3.65   0.000      .269655    .8952094
    1.spouse |  -.7477585   .1753466    -4.26   0.000    -1.091431   -.4040854
    1.school |   .2718653   .1127331     2.41   0.016     .0509125    .4928181
       hours |  -.0196461   .0031504    -6.24   0.000    -.0258208   -.0134714
             |
        year |
          2  |   .3317803   .1015628     3.27   0.001      .132721    .5308397
          3  |   .3349777   .1082496     3.09   0.002     .1228124     .547143
          4  |   .4327654   .1165144     3.71   0.000     .2044013    .6611295
          5  |   .4025012   .1275277     3.16   0.002     .1525514     .652451
             |
     1.black |           0  (omitted)
         age |           0  (omitted)
------------------------------------------------------------------------------
```

Random Effects Models. Quoting Allison, "In a random effects model, the unobserved variables are assumed to be uncorrelated with (or, more strongly, statistically independent of) all the observed variables." That assumption will often be wrong but, for the reasons given above (e.g. standard errors may be very high with fixed effects, RE lets you estimate effects for time-invariant variables), an RE model may still be desirable under some circumstances. RE models can be estimated via Generalized Least Squares (GLS). Here is an example of a random effects logistic regression model.

```
. *random effects
. 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
```

Among other things, according to this model, blacks are significantly more likely to be in poverty than are whites. The highly significant likelihood ratio test at the end tells us it would not be appropriate to use regular logistic regression instead. Note too that there are some major differences in the coefficients for the fixed and random effects models, which might reflect the importance of omitted variable bias in the latter.

Mixed Effects Model. Give or take a few decimal places, a mixed-effects model (aka multilevel model or hierarchical model) replicates the above results. Again, it is ok if the data are `xtset` but it is not required. We will explain mixed effects models more later.

```
. * Equivalent mixed-effects model
. xtset, clear
. 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
```

Suggested Exercise. Try running the following commands. What do you think they tell you? How are they related to the above melogit command?

```
use https://www3.nd.edu/~rwilliam/statafiles/teenpovxt, clear
logit pov i.mother i.spouse i.school hours i.year i.black age
est store logit
melogit pov i.mother i.spouse i.school hours i.year i.black age || id:, nolog
est store melogit
lrtest logit melogit, stats force
```

Appendix (Optional): Estimation methods for fixed-effects models

Fixed effects models control for, or partial out, the effects of time-invariant variables with time-invariant effects. This is true whether the variable is explicitly measured or not. Exactly how they do so varies by the statistical technique being used. Some of the methods used include

- *Demeaning variables*. The within-subject means for each variable (both the Xs and the Y) are subtracted from the observed values of the variables. Hence, within each subject, the demeaned variables all have a mean of zero. For time-invariant variables, e.g. gender, the demeaned variables will have a value of 0 for every case, and since they are constants they will drop out of any further analysis. This basically gets rid of all between-subject variability (which may be contaminated by omitted variable bias) and leaves only the within-subject variability to analyze. This method works for linear regression models but does not work for things like logistic regression.
- *Unconditional maximum likelihood*. With UML, dummy variables are created for each subject (except one) and included in the model. So, for example, if you had 2000 subjects each of whom was measured at 5 points in time, you would include 1,999 dummy variables in the model. Needless to say, this can be pretty time consuming, and can produce a lot of coefficients that you aren't really interested in! However, Allison argues that it is better to use `nbreg` with UML than it is to use Stata's `xtnbreg, fe`. The latter, he claims, uses a flawed approach and does not, in fact control for all stable predictors. UML can also be used for linear regression but produces biased estimates with logistic regression.
- *Conditional maximum likelihood*. This is used for logistic regression and some other statistical techniques. Quoting Allison (p. 32; α_i refers to the fixed effects parameters),

> The solution is to do conditional maximum likelihood, which *conditions* the α_i parameters out of the likelihood function (Chamberlain, 1980). This is accomplished by conditioning the likelihood function on the total number of events observed for each person. In effect, each person's contribution to the likelihood function is the answer to a question such as the following: Given that a girl was in poverty for 2 out of the 5 years, what is the probability that this happened in, say, Years 2 and 4 (when it actually occurred) rather than in one of the nine other possible pairs of years? These conditional probabilities do not contain the α_i parameters. This conditioning approach only works for the logistic regression model for dichotomous response variables, not for other "link" functions such as probit or complementary log-log.

Note that, with the conditional logit model, for all subjects where the dependent variable is a constant (e.g. at all five time periods the subject has a value of 1 on the dependent variable, or a value of zero) the case is dropped from the statistical analysis. Basically, there is no alternative possibility to compare to, e.g. the only way you can have 5 ones is by being a one at every time period.

Before proceeding, we will show examples of UML (the dummy variable for each case approach). This will show that `regress` using UML gives the same results as `xtreg,fe` but different results when using `logit` and `xtlogit, fe`. The data sets used here are also used in Allison's book.

```
. set more off
. use https://www3.nd.edu/~rwilliam/statafiles/nlsy.dta, clear
. des anti* self* pov*

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
------------------------------------------------------------
anti90          byte   %8.0g                  child antisocial behavior in 1990
anti92          byte   %8.0g                  child antisocial behavior in 1992
anti94          byte   %8.0g                  child antisocial behavior in 1994
self90          byte   %8.0g                  child self-esteem in 1990
pov90           byte   %8.0g                  family poverty status in 1990
[some output deleted]

. gen id=_n
. reshape long anti pov self, i(id) j(year)
(note: j = 90 92 94)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                      581   ->    1743
Number of variables                  17   ->      12
j variable (3 values)                     ->   year
xij variables:
                  anti90 anti92 anti94   ->   anti
                     pov90 pov92 pov94   ->   pov
                  self90 self92 self94   ->   self
-----------------------------------------------------------------------------

. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 90 to 94, but with gaps
                delta:  1 unit

. * UML works fine with linear regression model
. xtreg anti self pov i.year, fe

Fixed-effects (within) regression               Number of obs      =      1743
Group variable: id                              Number of groups   =       581

R-sq:  within  = 0.0331                         Obs per group: min =         3
       between = 0.0418                                        avg =       3.0
       overall = 0.0359                                        max =         3

                                                F(4,1158)          =      9.92
corr(u_i, Xb)  = 0.0683                         Prob > F           =    0.0000
```

```
------------------------------------------------------------------------------
        anti |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        self |  -.0551514   .0105258    -5.24   0.000    -.0758031   -.0344997
         pov |   .1124749   .0934099     1.20   0.229    -.0707967    .2957464
             |
        year |
          92 |   .0443934    .058584     0.76   0.449    -.0705493     .159336
          94 |   .2107366   .0587978     3.58   0.000     .0953744    .3260987
             |
       _cons |   2.637156   .2173038    12.14   0.000     2.210803     3.06351
-------------+----------------------------------------------------------------
     sigma_u |  1.3218868
     sigma_e |  .99707353
         rho |  .63737335   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(580, 1158) =     5.16           Prob > F = 0.0000

. set matsize 2000
. reg anti self pov i.year i.id

      Source |       SS       df       MS              Number of obs =    1743
-------------+------------------------------           F(584,  1158) =    5.48
       Model |  3181.88311   584  5.44842999           Prob > F      =  0.0000
    Residual |  1151.23221  1158  .994155619           R-squared     =  0.7343
-------------+------------------------------           Adj R-squared =  0.6003
       Total |  4333.11532  1742  2.48743704           Root MSE      =  .99707

------------------------------------------------------------------------------
        anti |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        self |  -.0551514   .0105258    -5.24   0.000    -.0758031   -.0344997
         pov |   .1124749   .0934099     1.20   0.229    -.0707967    .2957464
             |
        year |
          92 |   .0443934    .058584     0.76   0.449    -.0705493     .159336
          94 |   .2107366   .0587978     3.58   0.000     .0953744    .3260987
             |
          id |
           2 |  -.8875251   .8194485    -1.08   0.279    -2.495295    .7202448
           3 |   4.130859   .8194591     5.04   0.000     2.523068    5.738649
 [Rest of coefficients for dummy variables for ids are deleted]

. * UML does not work fine with logit -- Need conditional model instead
. xtlogit pov mother spouse school hours i.year, fe nolog
note: multiple positive outcomes within groups encountered.
note: 324 groups (1620 obs) dropped because of all positive or
      all negative outcomes.

Conditional fixed-effects logistic regression   Number of obs      =      4135
Group variable: id                              Number of groups   =       827

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

                                                LR chi2(8)         =     97.28
Log likelihood  = -1520.1139                    Prob > chi2        =    0.0000
```

```
------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      mother |   .5824322   .1595831     3.65   0.000      .269655    .8952094
      spouse |  -.7477585   .1753466    -4.26   0.000    -1.091431   -.4040854
      school |   .2718653   .1127331     2.41   0.016     .0509125    .4928181
       hours |  -.0196461   .0031504    -6.24   0.000    -.0258208   -.0134714
             |
        year |
          2  |   .3317803   .1015628     3.27   0.001      .132721    .5308397
          3  |   .3349777   .1082496     3.09   0.002     .1228124     .547143
          4  |   .4327654   .1165144     3.71   0.000     .2044013    .6611295
          5  |   .4025012   .1275277     3.16   0.002     .1525514     .652451
------------------------------------------------------------------------------
```

```
. logit pov mother spouse school hours i.year i.id, nolog
note: 141.id != 0 predicts failure perfectly
      141.id dropped and 5 obs not used
note: 298.id != 0 predicts success perfectly
      298.id dropped and 5 obs not used

[Other similar warnings deleted – these are the 324 cases where the outcome is the
same at all 5 time periods for the case]

Logistic regression                               Number of obs   =       4135
                                                  LR chi2(834)    =     998.93
                                                  Prob > chi2     =     0.0001
Log likelihood = -2304.2196                       Pseudo R2       =     0.1781

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      mother |   .7341873    .179498     4.09   0.000     .3823778    1.085997
      spouse |  -.9407072   .1971326    -4.77   0.000     -1.32708   -.5543344
      school |   .3410341   .1264389     2.70   0.007     .0932184    .5888497
       hours |  -.0246849   .0035439    -6.97   0.000    -.0316308   -.0177391
             |
        year |
          2  |   .4196558   .1142231     3.67   0.000     .1957827     .643529
          3  |   .4218788    .121389     3.48   0.001     .1839608    .6597968
          4  |   .5452897   .1306011     4.18   0.000     .2893163    .8012631
          5  |   .5071969   .1427835     3.55   0.000     .2273463    .7870475
             |
          id |
          75 |   -.107972   1.592235    -0.07   0.946    -3.228695    3.012751
          92 |   1.206116   1.476275     0.82   0.414     -1.68733    4.099562
[Coefficients for other id dummies not shown]
```

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,
>
> $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \tag{1}$$
>
> 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.

Panel Data and Multilevel Models for Categorical Outcomes: Discrete Time Methods for the Analysis of Event histories

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

Often, we are interested not only in whether an event occurs, but how quickly it happens (if at all). What factors speed up or delay death? Why do some friendships last longer than others? What causes some conflicts to be resolved quickly, while others drag on for years or even decades? Why do some individuals get tenure sooner than do others? Why do some people marry at young ages while others wait until they are much older?

Stata has a whole manual and suite of commands devoted to Survival Time Analysis. As Allison (1982, 1984; 2014; see exact citations later) points out, however, in some situations basic logistic regression techniques can be used. He refers to such approaches as *Discrete Time Methods for the Analysis of Event Histories*. To use such methods, you have to have Panel Data, e.g. repeated measures on the same individuals collected at multiple points in time on a regular basis, such as annually. At each time point, the dependent variable of interest is either coded 0 (the event has not happened yet) or 1 (the event occurred during the current interval, although you may not know exactly when). *After the event occurs no additional records are included for that case*. The coefficients for the logistic regression then tell you what factors speed up or slow down the pace at which the event in question occurs.

Allison explains how his procedure addresses problems that would be difficult to deal with via conventional regression techniques. First, the event may not occur (if it occurs at all) until after the data collection has ended; that is, the data may be *right censored*. (Somewhat more problematic is *left-censoring*, e.g. you don't know when exposure to risk began. For example, you might not know when a friendship or marriage started or when a person began an academic career. Still, Allison offers some ideas on what to do.) Second, his method allows the use of *time-varying covariates*, i.e. independent variables whose values change across time. For example, if somebody suddenly starts publishing more papers, that could speed up the rate at which they get tenure; or if they start smoking they might die more quickly. I will give two examples that illustrate the strategy.

Example 1. Allison (1999) analyzes a data set of 301 male and 177 female biochemists. The units of analysis are person-years rather than persons. Each person has one record for each year they were an assistant professor, for up to ten years; once a person achieves tenure no further records are added. This results in 1,741 person-years for men and 1,056 person-years for women. The dependent variable in his analysis, tenure, is promotion to associate professor, coded 1 if the person was promoted in that year, 0 otherwise. For the independent variables, year is the number of years since the beginning of the assistant professorship, yearsq is years squared, select is a measure of the selectivity of the colleges where scientists received their bachelor's degrees, articles is the cumulative number of articles published by the end of each person-year, and prestige is a measure of prestige of the department in which scientists were employed. The primary substantive interest of the analysis is whether the determinants of tenure differ for men (group 0) and women (group 1). Here is how we can conduct an EHA with these data.

```
. use https://www3.nd.edu/~rwilliam/statafiles/xtenure, clear
(Gender differences in receipt of tenure (Scott Long 06Jul2006))
. quietly logit tenure i.female year c.year#c.year select articles prestige
. est store baseline
. quietly logit tenure i.female year c.year#c.year select articles prestige
i.female#c.articles
. est store interaction
. est tab baseline interaction, b(%7.4f) star

----------------------------------------
    Variable |  baseline    interact~n
-------------+--------------------------
      female |
     Female  | -0.3538**     0.0100
             |
        year |  1.7232***    1.7201***
             |
      c.year#|
      c.year | -0.1253***   -0.1253***
             |
      select |  0.1544***    0.1521***
    articles |  0.0548***    0.0722***
    prestige | -0.4136***   -0.3935***
             |
      female#|
  c.articles |
     Female  |              -0.0375*
             |
       _cons | -6.8127***   -7.0004***
----------------------------------------
legend: * p<0.05; ** p<0.01; *** p<0.001
```

Several things stand out. The baseline model shows that women receive tenure more slowly than do men. The longer you have been an assistant professor, the more likely you are to receive tenure soon. Those at more prestigious universities receive tenure more slowly, while those who went to a more selective undergraduate institution get tenure faster. Not surprisingly, the more articles someone publishes, the more quickly they get tenure.

Perhaps the most concerning result from the baseline model is that women get tenure more slowly than men. This concern gets amplified in the 2nd model, when we add an interaction for female * articles. The coefficients suggest that each article published helps women only half as much as it did men. Does this reflect discrimination against women? Do we need additional measures, such as indicators of paper quality? Do women face obstacles not measured here, such as family-unfriendly workplaces? Those questions are not answered here, but the results do suggest the need for more study.

Having said that, there has been a lot of controversy over whether the above models are valid. If interested, see https://www3.nd.edu/~rwilliam/oglm/index.html for an extended discussion.

Example 2. The rest of this handout actually consists of of references to the classic 1987 ASR paper, *The Stability of Students' Interracial Friendships*, by Maureen Hallinan and Richard Williams. Alas, I was apparently less well organized 30+ years ago, and I can't find any of the original materials or data sets. If I can ever find the data I will try to rework some of these analyses, but if not we'll just have to trust my much younger self.

Excerpts from “The Stability of Students’ Interracial Friendships”, by Maureen Hallinan and Richard Williams

In 1976-77, a large, longitudinal data set was obtained from 1,477 students in 48 classes in six public and four private schools in northern California. The sample contains 229 black students and 226 non-black students. The students were given a sociometric questionnaire six times during the school year at approximately six-week intervals. The students were given a list of their classmates and, next to each name, were the categories: "Best Friend”, "Friend", "Know", "Don't Know", and "My Name". They were asked to circle the appropriate category for each student and encouraged to name as many best friends and friends as they wished.

To examine the determinants of interracial friendship stability, a dyadic-level analysis is required. In each dyad, P is designated the chooser and O the student who can be chosen. We examine those dyads in which P chooses 0 as Best Friend at some time during the course of the school year. Our interest is the stability of that choice. The dependent variable for the descriptive analysis in Table 2 is the termination of P's choice of O (Dissol), coded as unity if the friendship dissolved and zero if the friendship continued. The dependent variable is the same for the inferential analyses reported in Tables 3, 4, and 5, except that coding is reversed (1 = continuation, 0 = dissolution) to facilitate interpretation of parameter estimates. The best friend choices are used instead of the weaker friend choices because the latter are likely to contain more response error.

To obtain the dyadic-level data file for the analysis, records were created for all possible dyadic combinations of students within each of the 16 classrooms. Each dyad is included in the sample twice; in the first case, one member of the dyad is designated as ***P,*** the chooser, and the other member as *0,* the person chosen. In the second case, the chooser and chosen designation is reversed. This redundancy is necessary because friendship choices need not be mutual. To prevent standard errors from being inflated, each dyad is weighted by one-half in the inferential analysis.

Analyzing the stability of dyadic friendship choices is not straightforward. It is tempting to do a conventional regression analysis in which the observed duration of the friendship is the dependent variable. However, Allison (1984) has outlined a number of reasons why such a strategy is inappropriate for individual-level data. The basic problems are the same for dyadic-level data.

First, the ultimate duration of a friendship choice is not known for choices that were still in existence at the end of the school year. These observations are said to be "right-censored." Simply using the observed duration clearly underestimates the true duration and can produce substantial biases. Further, it has been shown that excluding the censored observations is also highly problematic (Sorensen 1977; Tuma and Hannan 1978).

Second, even during the school year, it is not known exactly when the friendship choices began or ended. Only the status of the friendship at each of the six observational periods is known. Assumptions of methods that require precise interval-level measurement may be violated.

Third, the values of some explanatory variables of interest can change across time (e.g., whether or not both members of the dyad are in the same reading group, or whether or not friendship choices are reciprocated). Changes in the values of variables might affect the stability of the friendship choice. Conventional regression techniques do not provide any convenient means of incorporating time-varying explanatory variables in the analysis.

Finally, many of the dyads are not only right-censored, but left-censored as well. Over half of the friendship choices already existed by the first observational period. These choices were made either extremely early in the school year or before school begun, but it is impossible to tell exactly when. Thus, again, the true value of duration is not known. Further, it seems reasonable to suspect that friendship choices made prior to the school year may differ substantially from those formed during it.

Allison (1982, 1984) has proposed a technique for dealing with the first three of these problems. The strategy treats each discrete time unit for each dyad as a separate observation or unit of analysis. If the friendship choice ended after four time periods, four different observations would be created. On the first three observations, dissolution would be coded 0 while on the last observation it would be coded unity. Time periods in which the friendship choice did not yet exist, was just being reported for the first time, or after the friendship choice had already terminated, are excluded from the analysis because the friendship choice was not at risk of dissolving at those times. Explanatory variables for each of these new observations are assigned whatever values they had at that particular unit of time. The final step is to pool the observations and compute maximum likelihood estimates for the logistic regression model.

Allison's technique addresses each of the first three concerns we presented. Dyads in which duration of a friendship choice is censored contribute exactly what is known about them – that the friendship choice did not end in any of the time periods in which they were observed. The method does not require that the duration be precisely measured; simply knowing the status of the friendship choice at each of the different observational periods is sufficient. Time-varying explanatory variables are easily incorporated into the analysis because each six-week interval the friendship choice is at risk is treated as a distinct observation.

The final problem of left-censoring is not so easily dealt with. One approach is to simply discard the initially censored intervals (Allison 1984). However, an examination of differences between friendship choices formed before the school year and those formed during it may be of interest. Therefore, we perform analyses on the total sample and separate analyses for the left-censored and non-left-censored observations.

Since there are only two possible outcomes for each friendship choice (continuation or dissolution), we analyze the data using a logistic regression model.

A positive beta coefficient implies that the friendship choice dyads that have a higher value on the independent variable X will tend to survive longer, while a negative coefficient implies that a higher value on the independent variable will lead to shorter friendship choices.

Table 3. Multivariate Logistic Regression of Friendship Stability on Organizational and Dyadic-Level Variables for Full Sample

Variable	Total ($N = 3,103$)	Bl–Wh ($N = 586$)	Wh–Bl ($N = 366$)	Bl–Bl ($N = 1,358$)	Wh–Wh ($N = 793$)
Intercept	1.34**	.04	4.15*	.71	.75
	(.50)	(1.23)	(1.75)	(.83)	(1.15)
Recip	1.01***	.72**	1.14***	.88***	1.44***
	(.10)	(.24)	(.31)	(.14)	(.21)
Sex-P	−.25**	.00	−.52	−.49***	−.13
	(.09)	(.20)	(.30)	(.13)	(.19)
Samesex	.80***	1.05***	.98**	.73**	.99***
	(.10)	(.24)	(.40)	(.14)	(.29)
Rankdiff	.01	.03**	.01	−.00	.01
	(.01)	(.01)	(.02)	(.01)	(.01)
Grade	.02	.21	−.36	.14	.06
	(.06)	(.15)	(.20)	(.13)	(.14)
Classize	−.02	−.01	−.07*	.00	−.02
	(.01)	(.02)	(.03)	(.02)	(.02)
Read Same	.07	−.01	.49	−.09	.21
	(.10)	(.24)	(.34)	(.15)	(.21)
Prop Black	.64***	−.27	1.89**	.05	1.80**
	(.19)	(.48)	(.65)	(.44)	(.70)
Climate	−.31***	−.53**	−.17	−.51**	−.19
	(.08)	(.18)	(.25)	(.18)	(.14)
Period 1	−1.02***	−.92***	−1.45***	−.84***	−1.14***
	(.09)	(.19)	(.28)	(.13)	(.18)

Note: Standard errors are in parentheses.
* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.

DISCUSSION

One might think that because students' interracial friendships are fairly uncommon, they are also unstable. Our research shows that this is not the case. Interracial friendship choices in the desegregated classrooms in our sample were fairly stable. While they generally did not last the entire school year, they did continue for several weeks and often months. Indeed, students' interracial friendship choices were almost as stable as their same-race choices. This surprising result may be because interracial friendships are unlikely in the first place and are made only if there is a strong attraction between a black and white student that then sustains the relationship over time.

This research has several policy implications. Clearly, dyadic-level characteristics have the strongest impact on the stability of interracial friendship choices. However, it is also clear that schools are not powerless in this area. If school personnel wish to support interracial sociability in desegregated schools, they should try to provide a classroom environment that promotes stable interracial friendship choices. Our study shows that this can be done by paying attention to the racial composition of the class and to the class climate. The ratio of black to white students can afford opportunities for black and white students to interact with each other to foster positive sentiment between them. The classroom climate can decrease major status differences between black and white students by providing opportunities for all students to win the esteem of their peers. Thus, by manipulating the environmental and organizational factors that affect interpersonal attraction and the cohesiveness of relationships, school administrators and teachers can help sustain interracial friendship ties once they are made.

Using Stata's Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects

Richard Williams
rwilliam@ND.Edu
https://www.nd.edu/~rwilliam/
University of Notre Dame
Original version presented at the Stata User Group Meetings, Chicago, July 14, 2011
Published version available at http://www.stata-journal.com/article.html?article=st0260
Current presentation updates the article and was last revised January 20, 2018

Motivation for Paper

- Many journals place a strong emphasis on the sign and statistical significance of effects – but often there is very little emphasis on the substantive and practical significance
- Unlike scholars in some other fields, most Sociologists seem to know little about things like marginal effects or adjusted predictions, let alone use them in their work
- Many users of Stata seem to have been reluctant to adopt the margins command.
  - The manual entry is long, the options are daunting, the output is sometimes unintelligible, and the advantages over older and simpler commands like adjust and mfx are not always understood

- This presentation therefore tries to do the following
  - Briefly explain what adjusted predictions and marginal effects are, and how they can contribute to the interpretation of results
  - Explain what factor variables (introduced in Stata 11) are, and why their use is often critical for obtaining correct results
  - Explain some of the different approaches to adjusted predictions and marginal effects, and the pros and cons of each:
    - APMs (Adjusted Predictions at the Means)
    - AAPs (Average Adjusted Predictions)
    - APRs (Adjusted Predictions at Representative values)
    - MEMs (Marginal Effects at the Means)
    - AMEs (Average Marginal Effects)
    - MERs (Marginal Effects at Representative values)

Adjusted Predictions - New margins versus the old adjust

```
. version 11.1
. webuse nhanes2f, clear
. keep if !missing(diabetes, black, female, age, age2, agegrp)
(2 observations deleted)
. label variable age2 "age squared"
. * Compute the variables we will need
. tab1 agegrp, gen(agegrp)
. gen femage = female*age
. label variable femage "female * age interaction"
. sum diabetes black female age age2 femage, separator(6)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    diabetes |     10335    .0482825     .214373          0          1
       black |     10335    .1050798    .3066711          0          1
      female |     10335    .5250121    .4993982          0          1
         age |     10335    47.56584    17.21752         20         74
        age2 |     10335    2558.924    1616.804        400       5476
      femage |     10335    25.05031    26.91168          0         74
```

Model 1: Basic Model

```
. logit diabetes black female age , nolog

Logistic regression                                     Number of obs   =      10335
                                                        LR chi2(3)      =     374.17
                                                        Prob > chi2     =     0.0000
Log likelihood = -1811.9828                             Pseudo R2       =     0.0936

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .7179046   .1268061     5.66   0.000     .4693691      .96644
      female |   .1545569   .0942982     1.64   0.101    -.0302642    .3393779
         age |   .0594654   .0037333    15.93   0.000     .0521484    .0667825
       _cons |  -6.405437   .2372224   -27.00   0.000    -6.870384    -5.94049
------------------------------------------------------------------------------
```

- Among other things, the results show that getting older is bad for your health – but just how bad is it???
- Adjusted predictions (aka predictive margins) can make these results more tangible.
- With adjusted predictions, you specify values for each of the independent variables in the model, and then compute the probability of the event occurring for an individual who has those values.
- So, for example, we will use the adjust command to compute the probability that an “average” 20 year old will have diabetes and compare it to the probability that an “average” 70 year old will.

```
. adjust age = 20 black female, pr

------------------------------------------------------------------------------
     Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, female = .52501209
 Covariate set to value: age = 20
------------------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |    .006308
----------------------
     Key:  pr  =  Probability

. adjust age = 70 black female, pr

------------------------------------------------------------------------------
     Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, female = .52501209
 Covariate set to value: age = 70
------------------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |    .110438
----------------------
     Key:  pr  =  Probability
```

- The results show that a 20 year old has less than a 1 percent chance of having diabetes, while an otherwise-comparable 70 year old has an 11 percent chance.

- But what does “average” mean? In this case, we used the common, but not universal, practice of using the mean values for the other independent variables (female, black) that are in the model.

- The margins command easily (in fact more easily) produces the same results

```
. margins, at(age=(20 70)) atmeans vsquish

Adjusted predictions                                  Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
1._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          20
2._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          70

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0063084   .0009888     6.38   0.000     .0043703    .0082465
          2  |   .1104379    .005868    18.82   0.000     .0989369     .121939
------------------------------------------------------------------------------
```

Factor variables

- So far, we have not used factor variables (or even explained what they are)
- The previous problems were addressed equally well with both older Stata commands and the newer margins command
- We will now show how margin's ability to use factor variables makes it much more powerful and accurate than its predecessors

Model 2: Squared term added

```
. quietly logit diabetes black female age age2, nolog
. adjust age = 70 black female age2, pr

------------------------------------------------------------------------------
     Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, female = .52501209, age2 = 2558.9238
 Covariate set to value: age = 70
------------------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |    .373211
----------------------
     Key:  pr  =  Probability
```

- In this model, adjust reports a much higher predicted probability of diabetes than before – 37 percent as opposed to 11 percent!
- But, luckily, adjust is wrong. Because it does not know that age and age2 are related, it uses the mean value of age2 in its calculations, rather than the correct value of 70 squared.
- While there are ways to fix this, using the margins command and factor variables is a safer solution.
  - The use of factor variables tells margins that age and age^2 are not independent of each other and it does the calculations accordingly.
  - In this case it leads to a much smaller (and also correct) estimate of 10.3 percent.

```
. quietly logit diabetes i.black i.female age c.age#c.age, nolog
. margins, at(age = 70) atmeans

Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =          70

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .1029814   .0063178    16.30   0.000     .0905988     .115364
------------------------------------------------------------------------------
```

- The i.black and i.female notation tells Stata that black and female are categorical variables rather than continuous. As the Stata 11-15 User Manuald explain (section 11.4.3.1), “i.group is called a factor variable…When you type i.group, it forms the indicators for the unique values of group.”
- The # (pronounced cross) operator is used for interactions.
  - The use of # implies the i. prefix, i.e. unless you indicate otherwise Stata will assume that the variables on both sides of the # operator are categorical and will compute interaction terms accordingly.
  - Hence, we use the c. notation to override the default and tell Stata that age is a continuous variable.
  - So, c.age#c.age tells Stata to include age^2 in the model; we do not want or need to compute the variable separately.
  - By doing it this way, Stata knows that if age = 70, then age^2 = 4900, and it hence computes the predicted values correctly.

Model 3: Interaction Term

```
. quietly logit diabetes black female age femage, nolog
. * Although not obvious, adjust gets it wrong
. adjust female = 0 black age femage, pr

------------------------------------------------------------------------------
     Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, age = 47.565844, femage = 25.050314
 Covariate set to value: female = 0
------------------------------------------------------------------------------
----------------------
      All |         pr
----------+-----------
          |    .015345
----------------------
     Key:  pr  =  Probability
```

- Once again, adjust gets it wrong
- If female = 0, femage must also equal zero
- But adjust does not know that, so it uses the average value of femage instead.
- Margins (when used with factor variables) does know that the different components of the interaction term are related, and does the calculation right.

```
. quietly logit diabetes i.black i.female age i.female#c.age, nolog
. margins female, atmeans grand

Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =   47.56584 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |
          0  |   .0250225   .0027872     8.98   0.000     .0195597    .0304854
          1  |   .0372713   .0029632    12.58   0.000     .0314635    .0430791
             |
       _cons |   .0308641   .0020865    14.79   0.000     .0267746    .0349537
------------------------------------------------------------------------------
```

Model 4: Multiple dummies

```
. quietly logit diabetes black female agegrp2 agegrp3 agegrp4 agegrp5 agegrp6
. adjust agegrp6 = 1 black female agegrp2 agegrp3 agegrp4 agegrp5, pr

------------------------------------------------------------------------------
     Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, female = .52501209, agegrp2 = .15674891,
agegrp3 = .12278665, agegrp4 = .12472182, agegrp5 = .27595549
 Covariate set to value: agegrp6 = 1
------------------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |    .320956
----------------------
     Key:  pr  =  Probability
```

- More depressing news for old people: now adjust says they have a 32 percent chance of having diabetes

- But once again adjust is wrong: If you are in the oldest age group, you can't also have partial membership in some other age category. 0, not the means, is the correct value to use for the other age variables when computing probabilities.

- Margins (with factor variables) realizes this and does it right again.

```
. quietly logit diabetes i.black i.female i.agegrp, nolog
. margins agegrp, atmeans grand

Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               1.agegrp        =    .2244799 (mean)
               2.agegrp        =    .1567489 (mean)
               3.agegrp        =    .1227866 (mean)
               4.agegrp        =    .1247218 (mean)
               5.agegrp        =    .2759555 (mean)
               6.agegrp        =    .0953072 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      agegrp |
           1 |   .0061598   .0015891     3.88   0.000     .0030453    .0092744
           2 |   .0124985    .002717     4.60   0.000     .0071733    .0178238
           3 |   .0323541   .0049292     6.56   0.000     .0226932    .0420151
           4 |   .0541518   .0062521     8.66   0.000      .041898    .0664056
           5 |    .082505   .0051629    15.98   0.000     .0723859     .092624
           6 |   .1106978    .009985    11.09   0.000     .0911276     .130268
             |
       _cons |   .0303728   .0022281    13.63   0.000     .0260059    .0347398
------------------------------------------------------------------------------
```

Different Types of Adjusted Predictions

- There are at least three common approaches for computing adjusted predictions
  - APMs (Adjusted Predictions at the Means).
    - All of the examples so far have used this
  - AAPs (Average Adjusted Predictions)
  - APRs (Adjusted Predictions at Representative values)

- For convenience, we will explain and illustrate each of these approaches as we discuss the corresponding ways of computing marginal effects

Marginal Effects

- As Cameron & Trivedi note (p. 333), "An ME [marginal effect], or partial effect, most often measures the effect on the conditional mean of y of a change in one of the regressors, say X_k. In the linear regression model, the ME equals the relevant slope coefficient, greatly simplifying analysis. For nonlinear models, this is no longer the case, leading to remarkably many different methods for calculating MEs."

- Marginal effects are popular in some disciplines (e.g. Economics) because they often provide a good approximation to the amount of change in Y that will be produced by a 1-unit change in X_k. With binary dependent variables, they offer some of the same advantages that the Linear Probability Model (LPM) does – they give you a single number that expresses the effect of a variable on P(Y=1).

- Personally, I find marginal effects for categorical independent variables easier to understand and also more useful than marginal effects for continuous variables
- The ME for categorical variables shows how P(Y=1) changes as the categorical variable changes from 0 to 1, after controlling in some way for the other variables in the model.
  - With a dichotomous independent variable, the marginal effect is the difference in the adjusted predictions for the two groups, e.g. for blacks and whites.
- There are different ways of controlling for the other variables in the model. We will illustrate how they work for both Adjusted Predictions & Marginal Effects.

```
. * Back to basic model
. logit diabetes i.black i.female age , nolog

Logistic regression                                     Number of obs   =      10335
                                                        LR chi2(3)      =     374.17
                                                        Prob > chi2     =     0.0000
Log likelihood = -1811.9828                             Pseudo R2       =     0.0936

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .7179046   .1268061     5.66   0.000     .4693691      .96644
    1.female |   .1545569   .0942982     1.64   0.101    -.0302642    .3393779
         age |   .0594654   .0037333    15.93   0.000     .0521484    .0667825
       _cons |  -6.405437   .2372224   -27.00   0.000    -6.870384    -5.94049
------------------------------------------------------------------------------
```

APMs - Adjusted Predictions at the Means

```
. margins black female, atmeans

Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =    47.56584 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |
           0 |   .0294328   .0020089    14.65   0.000     .0254955    .0333702
           1 |   .0585321   .0067984     8.61   0.000     .0452076    .0718566
             |
      female |
           0 |   .0292703   .0024257    12.07   0.000      .024516    .0340245
           1 |   .0339962   .0025912    13.12   0.000     .0289175    .0390748
------------------------------------------------------------------------------
```

MEMs – Marginal Effects at the Means

```
. * MEMs - Marginal effects at the means
. margins, dydx(black female) atmeans

Conditional marginal effects                       Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =    47.56584 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .0290993   .0066198     4.40   0.000     .0161246    .0420739
    1.female |   .0047259   .0028785     1.64   0.101    -.0009158    .0103677
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

- The results tell us that, if you had two otherwise-average individuals, one white, one black, the black's probability of having diabetes would be 2.9 percentage points higher (Black APM = .0585, white APM = .0294, MEM = .0585 - .0294 = .029).

- And what do we mean by average? With APMs & MEMs, average is defined as having the mean value for the other independent variables in the model, i.e. 47.57 years old, 10.5 percent black, and 52.5 percent female.

- So, if we didn't have the margins command, we could compute the APMs and the MEM for race as follows. Just plug in the values for the coefficients from the logistic regression and the mean values for the variables other than race.

```
. * Replicate results for black without using margins
. scalar female_mean = .5250121
. scalar age_mean = 47.56584
. scalar wlogodds = _b[1.black]*0 + _b[1.female]*female_mean + _b[age]*age_mean + _b[_cons]
. scalar wodds = exp(wlogodds)
. scalar wapm = wodds/(1 + wodds)
. di "White APM = " wapm
White APM = .02943284

. scalar blogodds = _b[1.black]*1 + _b[1.female]*female_mean + _b[age]*age_mean + _b[_cons]
. scalar bodds = exp(blogodds)
. scalar bapm = bodds/(1 + bodds)
. di "Black APM = " bapm
Black APM = .05853209

. di "MEM for black = " bapm - wapm
MEM for black = .02909925
```

- MEMs are easy to explain. They have been widely used. Indeed, for a long time, MEMs were the only option with Stata, because that is all the old mfx command supported.

- But, many do not like MEMs. While there are people who are 47.57 years old, there is nobody who is 10.5 percent black or 52.5 percent female.

- Further, the means are only one of many possible sets of values that could be used – and a set of values that no real person could actually have seems troublesome.

- For these and other reasons, many researchers prefer AAPs & AMEs.

AAPs - Average Adjusted Predictions

```
. * Average Adjusted Predictions (AAPs)
. margins black female

Predictive margins                                Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |
           0 |   .0443248   .0020991    21.12   0.000     .0402107    .0484389
           1 |    .084417   .0084484     9.99   0.000     .0678585    .1009756
             |
      female |
           0 |   .0446799   .0029119    15.34   0.000     .0389726    .0503871
           1 |   .0514786    .002926    17.59   0.000     .0457436    .0572135
------------------------------------------------------------------------------
```

AMEs – Average Marginal Effects

```
. margins, dydx(black female)

Average marginal effects                        Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .0400922   .0087055     4.61   0.000     .0230297    .0571547
    1.female |   .0067987   .0041282     1.65   0.100    -.0012924    .0148898
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

- Intuitively, the AME for black is computed as follows:
  - Go to the first case. Treat that person as though s/he were white, regardless of what the person's race actually is. Leave all other independent variable values as is. Compute the probability this person (if he or she were white) would have diabetes
  - Now do the same thing, this time treating the person as though they were black.
  - The difference in the two probabilities just computed is the marginal effect for that case
  - Repeat the process for every case in the sample
  - Compute the average of all the marginal effects you have computed. This gives you the AME for black.

```
. * Replicate AME for black without using margins
. clonevar xblack = black
. quietly logit diabetes i.xblack i.female age, nolog
. margins, dydx(xblack)

Average marginal effects                        Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.xblack

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.xblack |   .0400922   .0087055     4.61   0.000     .0230297    .0571547
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. replace xblack = 0
. predict adjpredwhite
. replace xblack = 1
. predict adjpredblack
. gen meblack = adjpredblack - adjpredwhite
. sum adjpredwhite adjpredblack meblack

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
adjpredwhite |     10335    .0443248    .0362422    .005399   .1358214
adjpredblack |     10335     .084417    .0663927   .0110063   .2436938
     meblack |     10335    .0400922    .0301892   .0056073   .1078724
```

- In effect, you are comparing two hypothetical populations – one all white, one all black – that have the exact same values on the other independent variables in the model.

- Since the only difference between these two populations is their race, race must be the cause of the differences in their likelihood of diabetes.

- Many people like the fact that all of the data is being used, not just the means, and feel that this leads to superior estimates.

- Others, however, are not convinced that treating men as though they are women, and women as though they are men, really is a better way of computing marginal effects.

- The biggest problem with both of the last two approaches, however, may be that they only produce a single estimate of the marginal effect. However "average" is defined, averages can obscure difference in effects across cases.

- In reality, the effect that variables like race have on the probability of success varies with the characteristics of the person, e.g. racial differences could be much greater for older people than for younger.

- If we really only want a single number for the effect of race, we might as well just estimate an OLS regression, as OLS coefficients and AMEs are often very similar to each other.

- APRs (Adjusted Predictions at Representative values) & MERs (Marginal Effects at Representative Values) may therefore often be a superior alternative.

- APRs/MERs can be both intuitively meaningful, while showing how the effects of variables vary by other characteristics of the individual.

- With APRs/MERs, you choose ranges of values for one or more variables, and then see how the marginal effects differ across that range.

APRs – Adjusted Predictions at Representative values

```
. * APRs - Adjusted Predictions at Representative Values (Race Only)
. margins black, at(age=(20 30 40 50 60 70)) vsquish

Predictive margins                                 Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
1._at        : age             =          20
2._at        : age             =          30
3._at        : age             =          40
4._at        : age             =          50
5._at        : age             =          60
6._at        : age             =          70

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _at#black |
        1 0  |   .0058698   .0009307     6.31   0.000     .0040457    .0076938
        1 1  |   .0119597   .0021942     5.45   0.000     .0076592    .0162602
        2 0  |   .0105876   .0013063     8.11   0.000     .0080273    .0131479
        2 1  |    .021466   .0033237     6.46   0.000     .0149517    .0279804
        3 0  |   .0190245   .0017157    11.09   0.000     .0156619    .0223871
        3 1  |   .0382346   .0049857     7.67   0.000     .0284628    .0480065
        4 0  |   .0339524   .0021105    16.09   0.000     .0298159    .0380889
        4 1  |   .0671983   .0075517     8.90   0.000     .0523972    .0819994
        5 0  |   .0598751   .0028793    20.79   0.000     .0542318    .0655184
        5 1  |   .1154567   .0118357     9.75   0.000     .0922591    .1386544
        6 0  |   .1034603   .0057763    17.91   0.000     .0921388    .1147817
        6 1  |   .1912405    .019025    10.05   0.000     .1539522    .2285289
------------------------------------------------------------------------------
```

MERs – Marginal Effects at Representative values

```
. margins, dydx(black female) at(age=(20 30 40 50 60 70)) vsquish

Average marginal effects                        Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female
1._at        : age             =          20
2._at        : age             =          30
3._at        : age             =          40
4._at        : age             =          50
5._at        : age             =          60
6._at        : age             =          70

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.black      |
         _at |
          1  |   .0060899   .0016303     3.74   0.000     .0028946    .0092852
          2  |   .0108784   .0027129     4.01   0.000     .0055612    .0161956
          3  |   .0192101   .0045185     4.25   0.000     .0103541    .0280662
          4  |   .0332459   .0074944     4.44   0.000      .018557    .0479347
          5  |   .0555816   .0121843     4.56   0.000     .0317008    .0794625
          6  |   .0877803   .0187859     4.67   0.000     .0509606    .1245999
-------------+----------------------------------------------------------------
1.female     |
         _at |
          1  |   .0009933   .0006215     1.60   0.110    -.0002248    .0022114
          2  |     .00178   .0010993     1.62   0.105    -.0003746    .0039345
          3  |    .003161   .0019339     1.63   0.102    -.0006294    .0069514
          4  |   .0055253   .0033615     1.64   0.100     -.001063    .0121137
          5  |   .0093981   .0057063     1.65   0.100     -.001786    .0205821
          6  |   .0152754   .0092827     1.65   0.100    -.0029184    .0334692
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

- Earlier, the AME for black was 4 percent, i.e. on average blacks' probability of having diabetes is 4 percentage points higher than it is for whites.

- But, when we estimate marginal effects for different ages, we see that the effect of black differs greatly by age. It is less than 1 percentage point for 20 year olds and almost 9 percentage points for those aged 70.

- Similarly, while the AME for gender was only 0.6 percent, at different ages the effect is much smaller or much higher than that.

- In a large model, it may be cumbersome to specify representative values for every variable, but you can do so for those of greatest interest.
  - For other variables you have to set them to their means, or use average adjusted predictions, or use some other approach.

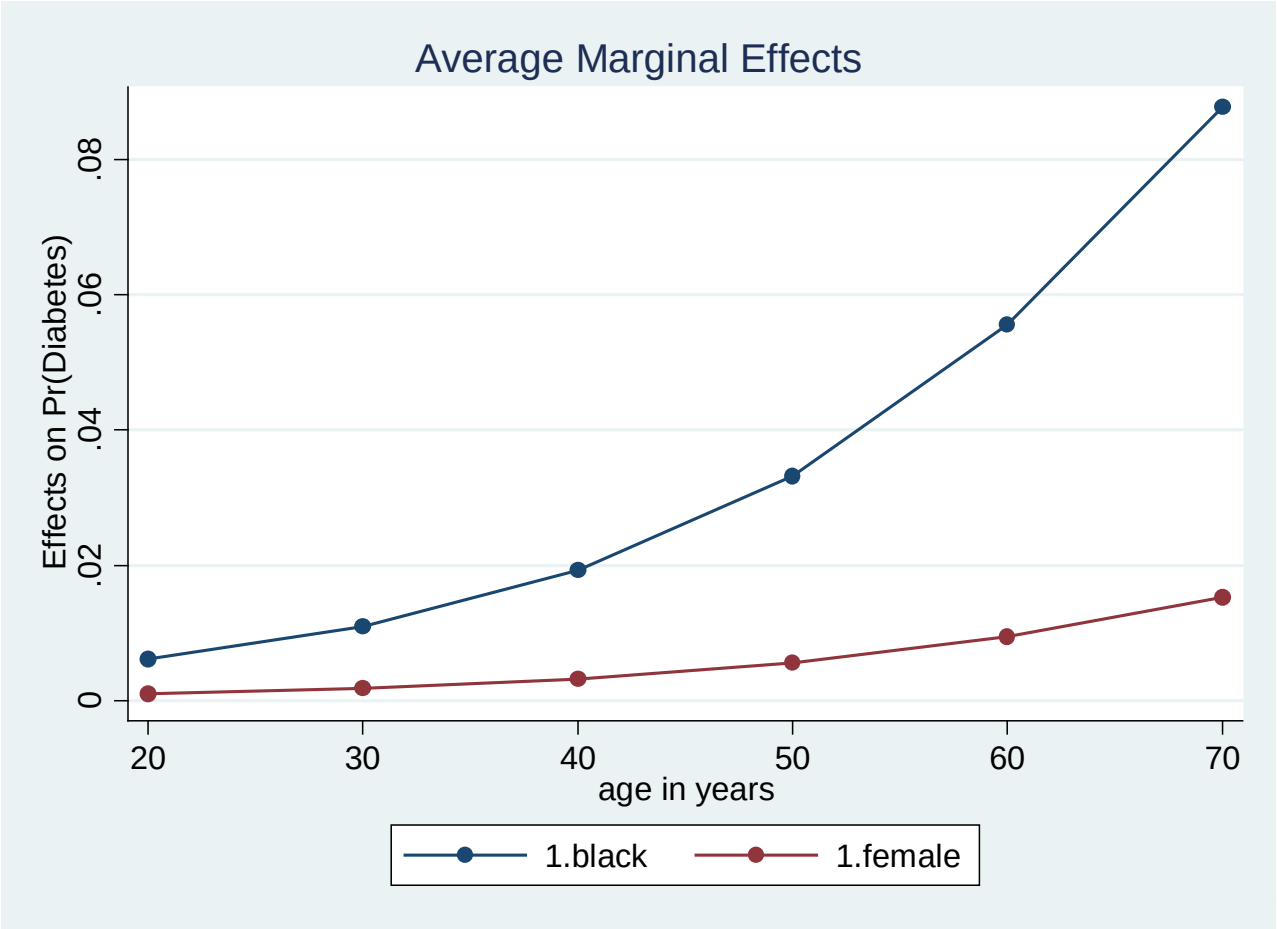
Graphing results

- The output from the margins command can be very difficult to read. It can be like looking at a 5 dimensional crosstab where none of the variables have value labels
- The marginsplot command introduced in Stata 12 makes it easy to create a visual display of results.

```
. quietly logit diabetes i.black i.female age, nolog

. quietly margins, dydx(black female) at(age=(20 30 40 50 60 70)) vsquish

. marginsplot, noci
```


Average Marginal Effects
Effects on Pr(Diabetes)
0
.02
.04
.06
.08
20
30
40
50
60
70
age in years
1.black
1.female

A more complicated example

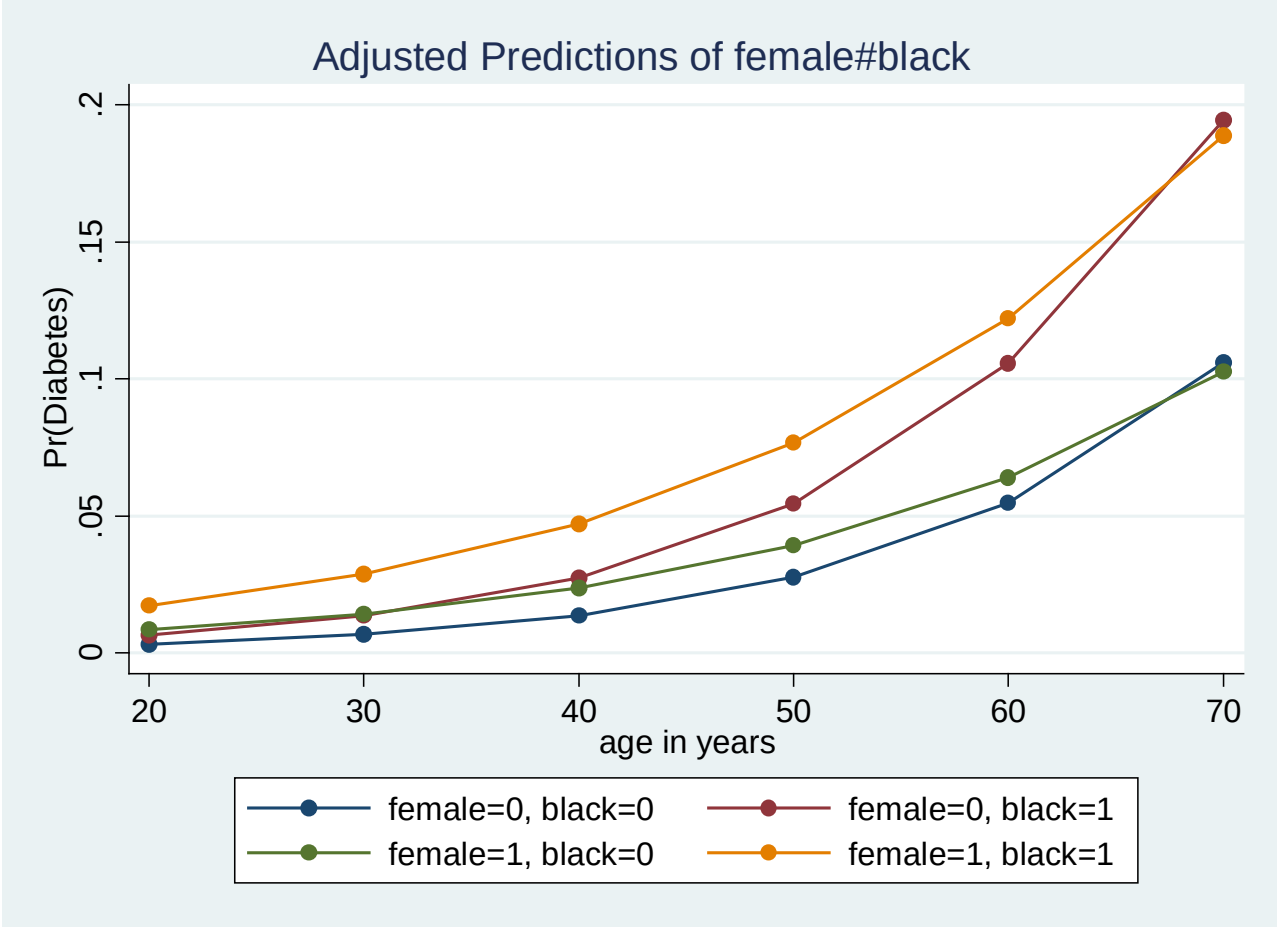
```
. quietly logit diabetes i.black i.female age i.female#c.age, nolog
. margins female#black, at(age=(20 30 40 50 60 70)) vsquish

Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
1._at        : age             =          20
2._at        : age             =          30
3._at        : age             =          40
4._at        : age             =          50
5._at        : age             =          60
6._at        : age             =          70

------------------------------------------------------------------------------
                 |            Delta-method
                 |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
_at#female#black |
          1 0 0  |    .003304      .0009     3.67   0.000       .00154    .0050681
          1 0 1  |    .006706   .0019396     3.46   0.001    .0029044    .0105076
          1 1 0  |   .0085838    .001651     5.20   0.000      .005348    .0118196
          1 1 1  |   .0173275   .0036582     4.74   0.000    .0101576    .0244974
          2 0 0  |   .0067332   .0014265     4.72   0.000    .0039372    .0095292
          2 0 1  |   .0136177   .0031728     4.29   0.000    .0073991    .0198362
          2 1 0  |   .0143006   .0021297     6.71   0.000    .0101264    .0184747
          2 1 1  |    .028699   .0049808     5.76   0.000    .0189368    .0384613
          3 0 0  |   .0136725   .0020998     6.51   0.000    .0095569    .0177881
          3 0 1  |   .0274562   .0049771     5.52   0.000    .0177013     .037211
          3 1 0  |   .0237336   .0025735     9.22   0.000    .0186896    .0287776
          3 1 1  |   .0471751   .0066696     7.07   0.000    .0341029    .0602473
          4 0 0  |   .0275651   .0028037     9.83   0.000       .02207    .0330603
          4 0 1  |   .0545794   .0075901     7.19   0.000    .0397031    .0694557
          4 1 0  |   .0391418   .0029532    13.25   0.000    .0333537    .0449299
          4 1 1  |   .0766076   .0090659     8.45   0.000    .0588388    .0943764
          5 0 0  |   .0547899   .0038691    14.16   0.000    .0472066    .0623733
          5 0 1  |   .1055879   .0121232     8.71   0.000    .0818269    .1293489
          5 1 0  |   .0638985   .0039287    16.26   0.000    .0561983    .0715986
          5 1 1  |   .1220509   .0131903     9.25   0.000    .0961985    .1479034
          6 0 0  |   .1059731   .0085641    12.37   0.000    .0891878    .1227584
          6 0 1  |   .1944623   .0217445     8.94   0.000    .1518439    .2370807
          6 1 0  |   .1026408   .0075849    13.53   0.000    .0877747    .1175069
          6 1 1  |   .1889354   .0206727     9.14   0.000    .1484176    .2294532
------------------------------------------------------------------------------

. marginsplot, noci
```


Adjusted Predictions of female#black
Pr(Diabetes)
0
.05
.1
.15
.2
20
30
40
50
60
70
age in years
female=0, black=0
female=0, black=1
female=1, black=0
female=1, black=1

Marginal effects of interaction terms

- People often ask what the marginal effect of an interaction term is. Stata's margins command replies: there isn't one. You just have the marginal effects of the component terms. The value of the interaction term can't change independently of the values of the component terms, so you can't estimate a separate effect for the interaction.

```
. quietly logit diabetes i.black i.female age i.female#c.age, nolog
. margins, dydx(*)

Average marginal effects                          Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female age

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .0396176   .0086693     4.57   0.000      .022626    .0566092
    1.female |   .0067791   .0041302     1.64   0.101     -.001316    .0148743
         age |   .0026632   .0001904    13.99   0.000     .0022901    .0030364
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

- For more on marginal effects and interactions, See Vince Wiggins' excellent discussion at

  http://www.stata.com/statalist/archive/2013-01/msg00293.html

A few other points

- Margins would also give the wrong answers if you did not use factor variables. You should use margins because older commands, like adjust and mfx, do not support the use of factor variables

- Margins supports the use of the svy: prefix with svyset data. Some older commands, like adjust, do not.

- With older versions of Stata, margins is, unfortunately, more difficult to use with multiple-outcome commands like ologit or mlogit. But this is also true of many older commands like adjust. Stata 14 made it much easier to use margins with multiple outcome commands.

- In the past the xi: prefix was used instead of factor variables. In most cases, *do not use xi: anymore*. The output from xi: looks horrible. More critically, the xi: prefix will cause the same problems that computing dummy variables yourself does, i.e. margins will not know how variables are inter-related.

- Long & Freese's spost13 commands were rewritten to take advantage of margins. Commands like mtable and mchange basically make it easy to execute several margins commands at once and to format the output. From within Stata type `findit spost13_ado`. Their highly recommended book can be found at

  http://www.stata.com/bookstore/regression-models-categorical-dependent-variables/

- Patrick Royston's mcp command (available from SSC) provides an excellent means for using margins with continuous variables and graphing the results. From within Stata type findit mcp. For more details see

  http://www.stata-journal.com/article.html?article=gr0056

References

- Williams, Richard. 2012. “Using the margins command to estimate and interpret adjusted predictions and marginal effects.” The Stata Journal 12(2):308-331.
- Available for free at

  http://www.stata-journal.com/article.html?article=st0260

- This handout is adapted from the article. The article includes more information than in this presentation. However, this presentation also includes some additional points that were not in the article.
- Please cite the above article if you use this material in your own research.

Panel Data and Multilevel Models for Categorical Outcomes: AAPs, AMEs and APRs for Multilevel Models

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

Results from logistic regression and many other methods can often be hard to interpret. For example, what does a coefficient of .2 for female (coded 0 = male, 1 = female) mean? Does it mean females are a little more likely to experience the event, a lot more likely, or what? As with regular logistic regression, adjusted predictions and marginal effects can help with the interpretation of multilevel random effects models. Margins with Fixed effects models are not so straightforward though, and should be approached with caution. For a discussion, see

http://www.statalist.org/forums/forum/general-stata-discussion/general/1304704-cannot-estimate-marginal-effect-after-xtlogit

Example. Consider a modified version of our earlier poverty example. This time, we will include an interaction between black and hours. This allows for the possibility that blacks benefit more (or less) than do whites for each hour worked.

```
. melogit pov i.mother i.spouse i.school hours i.year i.black i.black#c.hours 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(11)     =     277.16
Log likelihood = -3399.6342                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   1.023185   .1183556     8.65   0.000     .7912122    1.255157
    1.spouse |  -1.172154   .1509384    -7.77   0.000    -1.467988   -.8763204
    1.school |  -.1123479   .0989736    -1.14   0.256    -.3063327    .0816368
       hours |  -.0170478    .004132    -4.13   0.000    -.0251464   -.0089492
             |
        year |
          2  |   .2861683   .1000751     2.86   0.004     .0900246     .482312
          3  |    .219169   .1040961     2.11   0.035     .0151444    .4231936
          4  |   .2497039   .1090519     2.29   0.022     .0359661    .4634416
          5  |   .1488229   .1161253     1.28   0.200    -.0787785    .3764243
             |
     1.black |   .7280679   .1057163     6.89   0.000     .5208678     .935268
             |
black#c.hours |
          1  |  -.0155339     .00538    -2.89   0.004    -.0260785   -.0049892
             |
         age |  -.0602152   .0470168    -1.28   0.200    -.1523664     .031936
       _cons |  -.1248085   .7601223    -0.16   0.870    -1.614621    1.365004
-------------+----------------------------------------------------------------
id           |
   var(_cons)|    1.33912   .1358071                      1.097728    1.633595
------------------------------------------------------------------------------
LR test vs. logistic model: chibar2(01) = 319.42      Prob >= chibar2 = 0.0000
```

It is obvious from the output, and not too surprising, that those who are mothers at the time of the survey, do not have a spouse, are black, and work more hours, are more likely to be in poverty. But how much more likely? One percent? 50 percent? Or what? Further complicating matters is that the interaction between black and hours is significantly negative, suggesting that working more hours reduces poverty more for blacks than it does whites. But how much? AAPs (Average Adjusted Predictions), AMEs (Average Marginal Effects), APRs (Adjusted Predictions at Representative values) and MERs (Marginal Effects at Representative values) can give us some guidance.

```
. margins mother spouse black, grand

Predictive margins                              Number of obs     =      5,755
Model VCE    : OIM

Expression   : Marginal predicted mean, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      mother |
           0 |   .3420837   .0090595    37.76   0.000     .3243275    .3598399
           1 |   .5293023   .0198266    26.70   0.000      .490443    .5681617
             |
      spouse |
           0 |   .3975583   .0087569    45.40   0.000      .380395    .4147216
           1 |   .2132171   .0188862    11.29   0.000     .1762008    .2502333
             |
       black |
           0 |    .314291   .0124135    25.32   0.000     .2899609    .3386211
           1 |   .4223253   .0112836    37.43   0.000     .4002098    .4444407
             |
       _cons |   .3778618   .0082933    45.56   0.000     .3616072    .3941164
------------------------------------------------------------------------------
```

These results are, I think, much easier to get a substantive feel for. The constant (which we got because we added the grand option) tells us that 37.8 percent of the subjects are in poverty at the time of the interview. But, for those who are mothers, the figure is almost 53 percent. Similarly, about 42 percent of blacks (compared to 31.4 percent of whites) are in poverty, as are about 40 percent of those without a spouse (compared with 21.3 percent of those who do). Keep in mind that these are the estimated differences AFTER all other variables in the model have been controlled for, e.g. even after controlling for hours worked and motherhood status, differences between whites and blacks remain.)

You may also find it helpful to compute the AMEs, which, in the case of a dichotomous independent variable, are simply the differences between the adjusted predictions.

```
. margins, dydx(mother spouse black)

Average marginal effects                        Number of obs     =      5,755
Model VCE    : OIM

Expression   : Marginal predicted mean, predict()
dy/dx w.r.t. : 1.mother 1.spouse 1.black

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   .1872186   .0217485     8.61   0.000     .1445923    .2298448
    1.spouse |  -.1843413   .0203254    -9.07   0.000    -.2241783   -.1445042
     1.black |   .1080343   .0168314     6.42   0.000     .0750453    .1410233
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

We again see that those who are black are about 11 percentage points more likely on average to be in poverty than whites, but we do not see what the predicted probabilities were for blacks and whites separately.

What about hours worked, which is a continuous variable? We can estimate AMEs for it:

```
. margins, dydx(hours)

Average marginal effects                        Number of obs     =      5,755
Model VCE    : OIM

Expression   : Marginal predicted mean, predict()
dy/dx w.r.t. : hours

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hours |  -.0046285   .0004939    -9.37   0.000    -.0055965   -.0036605
------------------------------------------------------------------------------
```

However, I personally do not finds AMEs for continuous variables at all helpful. Instead, I prefer APRs.

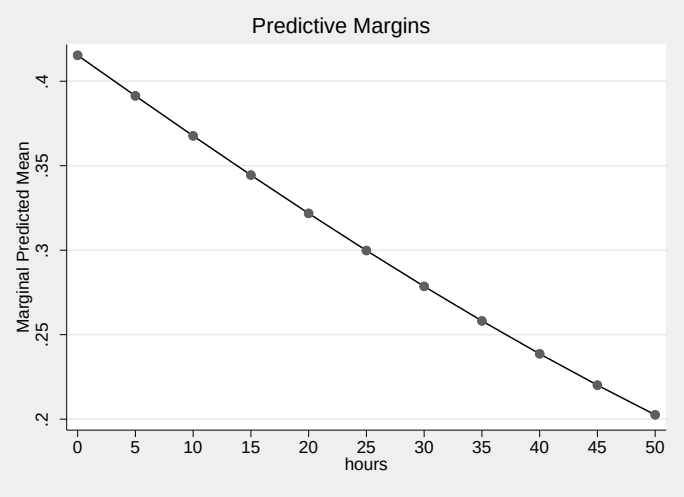
```
. margins, at(hours = (0(5)50)) vsquish

Predictive margins                              Number of obs     =      5,755
Model VCE    : OIM

Expression   : Marginal predicted mean, predict()
1._at        : hours           =           0
2._at        : hours           =           5
3._at        : hours           =          10
4._at        : hours           =          15
5._at        : hours           =          20
6._at        : hours           =          25
7._at        : hours           =          30
8._at        : hours           =          35
9._at        : hours           =          40
10._at       : hours           =          45
11._at       : hours           =          50
```

```
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .4153367   .0095408    43.53   0.000      .396637    .4340364
          2  |   .3913385    .008655    45.22   0.000      .374375     .408302
          3  |   .3676703   .0085009    43.25   0.000     .3510089    .3843318
          4  |   .3444526    .009041    38.10   0.000     .3267326    .3621727
          5  |   .3217977   .0100689    31.96   0.000     .3020631    .3415324
          6  |   .2998081     .01135    26.41   0.000     .2775625    .3220536
          7  |   .2785748   .0127067    21.92   0.000      .25367    .3034795
          8  |   .2581765   .0140241    18.41   0.000     .2306897    .2856633
          9  |   .2386786   .0152318    15.67   0.000     .2088248    .2685324
         10  |   .2201328   .0162884    13.51   0.000     .1882081    .2520575
         11  |   .2025766   .0171718    11.80   0.000     .1689206    .2362326
------------------------------------------------------------------------------
```

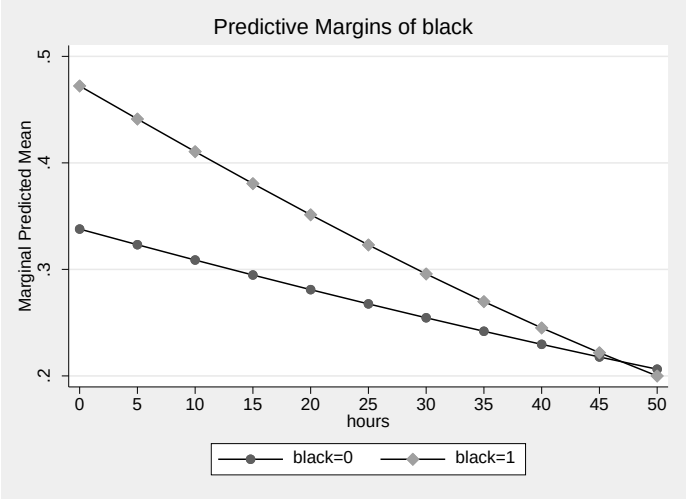
```
. marginsplot, noci scheme(sj) name(hours)
```



The output from margins and the graph produced by marginsplot provide a much clearer impact of the effect of hours worked. Those who do not work at all are predicted to have a 41.5% chance of being in poverty. Conversely, those who work 40 hours a week are predicted to have only a 23.8% chance.

It is also often helpful to get APRs for a combination of categorical and continuous variables:

```
. quietly margins black, at(hours = (0(5)50))
. marginsplot, noci scheme(sj) name(blackhours)
```



Remember that the validity of any results you get are contingent on the model being correct. But if this model is correct, it suggests that (at least in this sample) working provides a more powerful means for blacks to get out of poverty than it does for whites. When the average white or black do not work any hours, the predicted difference in poverty is about 13 percentage points. But, for those who work 40 or more hours a week, the predicted difference is almost zero.

Additional Material. Much more on margins can be found on my website at

https://www3.nd.edu/~rwilliam/stats3/index.html

As those notes show, I am a big fan of the spost13 commands by Long and Freese. Many are basically shells for margins, and are easier to use and produce more aesthetically output. mtable seems to work with melogit, but other commands might not work with panel/multilevel models. To get a copy, from within Stata type `findit spost13_ado`. For more, see

https://www3.nd.edu/~rwilliam/stats3/Margins04.pdf

I am also a huge fan of Patrick Royston's mcp command, available from SSC. It is great for making the effects of continuous variables more interpretable. Unfortunately, it doesn't seem to work with melogit, but it does work after many other commands. See

https://www3.nd.edu/~rwilliam/stats3/Margins03.pdf

I'm primarily focusing on binary dependent variables in this course. To see how marginal effects can be used with ordinal models, check out

https://www3.nd.edu/~rwilliam/stats3/Margins05.pdf

What do marginal effects for continuous variables mean, and why am I not a fan of them? For a discussion, see

https://www3.nd.edu/~rwilliam/stats3/Margins02.pdf

Panel Data and Multilevel Models for Categorical Outcomes: Sample Assignment

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

This assignment focuses on basic panel data and multilevel methods.

1. This example is adapted from the Stata 14 documentation on the `xtprobit` command. We have (synthetic) data on whether workers complain to managers at fast-food restaurants. The covariates are age (in years of the worker), grade (years of schooling completed by the worker), south (equal to 1 if the restaurant is located in the South, 0 otherwise), tenure (the number of years spent on the job by the worker), gender (of the worker; 1 = male, 0 = female), race (of the worker; 1 = Other, 2 = Black, 3 = White), income (in thousands of dollars by the restaurant), genderm (gender of the manager; 1 = male, 0 = female), chicken (1 = specializes in chicken, 0 = specializes in other types of food).

Note that we do not have multiple years of data for each restaurant. Instead, we have data for multiple employees for each restaurant. The term "cross-sectional time series," or xt, is a little misleading because the xt commands work fine in many cases when the data are not longitudinal. For example, you could have a sample of schools, with multiple students from each school.

Run the following code. You can add other commands if you wish.

```
webuse chicken, clear
label define sex 0 "Female" 1 "Male"
label values gender genderm sex
label define race 1 "Other" 2 "Black" 3 "White"
label values race race
keep complain age grade south tenure gender race income genderm chicken restaurant
xtset
xtsum
xtlogit complain age grade i.south tenure i.gender i.race income i.genderm i.chicken, nolog fe
est store fe
xtlogit complain age grade i.south tenure i.gender i.race income i.genderm i.chicken, nolog re
est store re
estimates table fe re, star
estimates restore re
margins south gender race genderm chicken
margins, dydx(south gender race genderm chicken)
```

Now answer the following questions.

a. Suppose that you were primarily concerned with omitted variable bias. What model might you favor, and why?
b. Suppose your primary concern was in assessing whether or not restaurants that specialize in chicken have more complaints than other types of restaurants. What model would you prefer then?

c. Even though they were specified on the command line, several variables are not included in the fixed effects model. Several cases are dropped too. Explain why. Use the results from the xtsum command to support your argument.
d. Interpret the results from the random effects model. What factors affect the likelihood of workers complaining? Use the results from both the `xtlogit` and `margins` commands. Run additional analyses if you think it would be helpful.

2. Use the `clogit` and `melogit` commands to replicate and compare the results you got above for the fixed effects and random effects models. The `melogit` estimates will differ slightly from the `xtlogit` results. Specifically, run something like

```
clogit … [finish the command]
est store clogit
melogit … [finish the command]
est store melogit
estimates table fe clogit re melogit, star
```

3. (Optional but recommended if you are using Panel Data or multilevel data in your own work.) Do similar analyses using a data set of your choice. You don't have to perfectly mirror the above analysis but see if a random effects and/or fixed effects model can offer you any helpful insights.