SOCY7708: Hierarchical Linear Modeling
Instructor: Natasha Sarkisian
Class notes: Non-continuous Dependent Variables

So far we've only dealt with continuous dependent variables, but HLM allows us to estimate models when the dependent variables are categorical. Your dependent variable can be dichotomous (0/1), categorical with multiple unordered categories, ordinal, or count variable. In such cases, linear models are inappropriate as there are no restrictions on the predicted values of level-1 outcome, the level-1 random effect (i.e. level 1 residual) cannot be normally distributed, and cannot have homogenous variance (the variance depends on the predicted value). Therefore, we need to use HGLM models for such variables. Like in non-hierarchical analysis, this is accomplished by specifying a link function that transforms the dependent variable so that the level-1 predicted values are constrained to be within a specific interval. Specifically, we use logit models for dichotomous variables, multinomial logit for categorical with unordered categories, ordered logit for ordinal variables, and Poisson models for count variables.

The following is an example of analysis with a dichotomous dependent variable. We'll use data from thai.dta file on 7,516 sixth graders nested within 356 primary schools from a national survey of primary education in Thailand, conducted in 1988. The dependent variable of interest is the probability that a child will repeat a grade during the primary years (REP1). The level-1 independent variables are whether a child attended pre-primary education (PPED) and child's gender (MALE). The level-2 variable is mean SES of school (MSESC). Since the dependent variable is binary, we will use melogit command. Let's do an unconditional model first:

```
. melogit rep1 || schoolid:

Mixed-effects logistic regression          Number of obs     =      7,516
Group variable:         schoolid           Number of groups  =        356
                                           Obs per group:
                                                         min =          2
                                                         avg =       21.1
                                                         max =         41

Integration method: mvaghermite            Integration pts.  =          7

                                           Wald chi2(0)      =          .
Log likelihood = -2768.6731                Prob > chi2       =          .
-------------------------------------------------------------------------
      rep1 |     Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----------+-------------------------------------------------------------
     _cons | -2.233851   .0876057   -25.50   0.000   -2.405555   -2.062147
-----------+-------------------------------------------------------------
schoolid   |
 var(_cons)|  1.732841   .2144817                     1.35957    2.208594
-------------------------------------------------------------------------
LR test vs. logistic model: chibar2(01) = 603.41     Prob >= chibar2 = 0.0000

. estat icc

Intraclass correlation
-------------------------------------------------------------------------
              Level |       ICC   Std. Err.     [95% Conf. Interval]
--------------------+----------------------------------------------------
           schoolid |  .3450013     .02797       .292416     .4016749
-------------------------------------------------------------------------
```

Note that the variance component does not contain an estimate of level 1 variance. That is because in logistic regression models, it is not possible to estimate both the coefficients and the error variance; therefore, in all logistic regression models, the error variance is always fixed to the same number which is $\pi^2/3 = 3.29$. That rule also applies to multilevel models, but only to their level 1 residuals. Knowing this means that we can calculate the intraclass correlation coefficient or the proportion of variance explained. For both, we can follow the procedures described on pp.224-227 of the Snijders and Bosker chapter on dichotomous outcomes. For instance, the ICC was calculated as:

$$\rho_I = \frac{\tau_0^2}{\tau_0^2 + \pi^2/3}$$

The model with level 1 and level2 that we'd like to try to estimate is:

LEVEL 1 MODEL

$\text{Prob}(REP1_{ij}=1|\beta_j) = \varphi_{ij}$

$\text{Log}[\varphi_{ij}/(1 - \varphi_{ij})] = \eta_{ij}$

$\eta_{ij} = \beta_{0j} + \beta_{1j}(MALE_{ij}) + \beta_{2j}(PPED_{ij})$

LEVEL 2 MODEL

$\beta_{0j} = \gamma_{00} + \gamma_{01}(MSESC_j - \overline{MSESC}_.) + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(MSESC_j - \overline{MSESC}_.) + u_{1j}$

$\beta_{2j} = \gamma_{20} + \gamma_{21}(MSESC_j - \overline{MSESC}_.) + u_{2j}$

Let's start with level 1 variables:

```
. melogit rep1 pped male || schoolid: pped male, cov(unstr)
```

The model fails to converge; I suspect it's the variance components so I look at those one by one:
```
. melogit rep1 pped male || schoolid: pped, cov(unstr)
```

```
Mixed-effects logistic regression              Number of obs      =       7,516
Group variable:          schoolid              Number of groups   =         356

                                               Obs per group:
                                                             min =           2
                                                             avg =        21.1
                                                             max =          41
Integration method: mvaghermite                Integration pts.   =           7

                                               Wald chi2(2)       =       63.85
Log likelihood =  -2720.987                    Prob > chi2        =      0.0000
-----------------------------------------------------------------------------
       rep1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+----------------------------------------------------------------
       pped | -.5577893    .1446756    -3.86   0.000    -.8413483   -.2742303
       male |  .5391314    .0763372     7.06   0.000     .3895132    .6887495
      _cons |  -2.27575    .1163893   -19.55   0.000    -2.503868   -2.047631
------------+----------------------------------------------------------------
schoolid    |
   var(pped)|  .0812087    .2653494                       .0001344    49.07981
  var(_cons)|  1.886038    .3054017                       1.373148    2.590499
------------+----------------------------------------------------------------
schoolid    |
   cov(pped,|
      _cons)| -.2396558    .2215629    -1.08   0.279    -.6739111    .1945996
-----------------------------------------------------------------------------
```

```
LR test vs. logistic model: chi2(3) = 574.23          Prob > chi2 = 0.0000

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

. est store pped_varies

. melogit rep1 pped male || schoolid: male, cov(unstr)

Mixed-effects logistic regression          Number of obs     =      7,516
Group variable:         schoolid           Number of groups  =        356

                                           Obs per group:
                                                         min =          2
                                                         avg =       21.1
                                                         max =         41

Integration method: mvaghermite            Integration pts.  =          7

                                           Wald chi2(2)      =      56.83
Log likelihood = -2720.1116                Prob > chi2       =     0.0000
-------------------------------------------------------------------------
        rep1 |     Coef.    Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------
        pped | -.6408643   .1002516    -6.39   0.000    -.8373538   -.4443747
        male |  .4495504   .1086671     4.14   0.000     .2365668    .662534
       _cons | -2.200042   .1134854   -19.39   0.000    -2.422469   -1.977615
-------------+-----------------------------------------------------------
schoolid     |
   var(male) |   .192939   .1597268                      .0380841    .9774554
  var(_cons) |  1.554265   .2629583                      1.115615    2.165387
-------------+-----------------------------------------------------------
schoolid     |
   cov(male, |
      _cons) |  .0913347   .1653605     0.55   0.581    -.2327659    .4154353
-------------------------------------------------------------------------
LR test vs. logistic model: chi2(3) = 575.98          Prob > chi2 = 0.0000

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

. est store male_varies

. melogit rep1 pped male || schoolid:

Mixed-effects logistic regression          Number of obs     =      7,516
Group variable:         schoolid           Number of groups  =        356

                                           Obs per group:
                                                         min =          2
                                                         avg =       21.1
                                                         max =         41

Integration method: mvaghermite            Integration pts.  =          7

                                           Wald chi2(2)      =      91.13
Log likelihood = -2721.7147                Prob > chi2       =     0.0000
-------------------------------------------------------------------------
        rep1 |     Coef.    Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------
        pped | -.6419783   .0995511    -6.45   0.000    -.8370949   -.4468616
        male |  .5362797   .0759943     7.06   0.000     .3873337    .6852258
       _cons |  -2.23709   .1066484   -20.98   0.000    -2.446117   -2.028063
-------------+-----------------------------------------------------------
schoolid     |
```

```
     var(_cons)|   1.703816    .2126685                       1.334064    2.176049
-----------------------------------------------------------------------------
LR test vs. logistic model: chibar2(01) = 572.78      Prob >= chibar2 = 0.0000

. est store novary

. est stats pped_varies male_varies novary

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
 pped_varies |      7,516          .  -2720.987       6   5453.974   5495.523
 male_varies |      7,516          .  -2720.112       6   5452.223   5493.772
      novary |      7,516          .  -2721.715       4   5451.429   5479.129
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

We can see that the log likelihoods are very close to each other (we could check with formal LR test but the outcome is predictable here), and BIC values clearly indicate that neither random slope is helpful. Therefor, our final model won't include random slopes. We will now add level 2 variable without cross-level interactions (I mean center MSESC first):

```
. egen tag=tag(schoolid)

. sum msesc if tag==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
       msesc |        356    .0078371    .3806623      -.77       1.49

. gen msescm= msesc-r(mean)

. melogit rep1 pped male msescm || schoolid:

Mixed-effects logistic regression              Number of obs     =      7,516
Group variable:         schoolid               Number of groups  =        356

                                               Obs per group:
                                                             min =          2
                                                             avg =       21.1
                                                             max =         41

Integration method: mvaghermite                Integration pts.  =          7

                                               Wald chi2(3)      =      92.83
Log likelihood = -2720.7865                    Prob > chi2       =     0.0000
-----------------------------------------------------------------------------
        rep1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
        pped | -.6270442    .100192    -6.26   0.000    -.8234169   -.4306716
        male |  .5353579   .0759897     7.05   0.000     .3864209    .6842949
      msescm | -.2954074   .2163732    -1.37   0.172     -.719491    .1286762
       _cons | -2.244334   .1066351   -21.05   0.000    -2.453335   -2.035333
-------------+---------------------------------------------------------------
schoolid     |
   var(_cons)|  1.692804    .2115158                      1.325102    2.162541
-----------------------------------------------------------------------------
LR test vs. logistic model: chibar2(01) = 567.89      Prob >= chibar2 = 0.0000
```

4

I also tested cross-level interactions just in case, but, not surprisingly, they were not significant.

To calculate the proportion of variance explained, we use the following formula:

$$R^2_{\text{dicho}} = \frac{\sigma^2_F}{\sigma^2_F + \tau^2_0 + \sigma^2_R}$$

Note that in addition to the level 2 intercept variance $\tau 0$ and level 1 variance $\sigma^2_R = 3.29$, we need to know the variance of fitted values σ^2_F. That refers to the variance of linear predictions, which are the values that results if we multiply our coefficients by our variable values and add up these products. That is, we are talking about the predicted values of logits. To obtain the variance of fitted values, we can use predict command, predict fitted values (using xb option), and get their variance. Note that such R squared values are typically lower than values we are used to with OLS because σ^2_R is a fixed number.

```
. predict xb, xb

. sum xb, det

              Linear prediction, fixed portion only
-------------------------------------------------------------
        Percentiles      Smallest
 1%     -3.093573       -3.253093
 5%     -2.993134       -3.253093
10%     -2.919282       -3.253093      Obs              7,516
25%      -2.6416        -3.253093      Sum of Wgt.      7,516

50%     -2.283376                      Mean         -2.285421
                         Largest       Std. Dev.     .4486351
75%     -1.935602       -1.508738
90%     -1.632809       -1.508738      Variance      .2012735
95%     -1.594406       -1.508738      Skewness     -.0061893
99%     -1.538279       -1.479198      Kurtosis      2.036111

. di .2012735/(1.692804 + .2012735 + 3.29)
.03882533
```

That's our pseudo-R square for the model. .039.

Interpreting fixed effects

The interpretation of the fixed effects is very similar to the interpretation of the results of logistic regression—but be careful as we now have variables on multiple levels and can potentially have interactions across levels.

Interpreting coefficients themselves allows us to discuss the direction and significance of effects, but not their size. To talk about the size, we use odds ratios. Odds are ratios of two probabilities – probability of a positive outcome and a probability of a negative outcome (e.g. probability of voting divided by a probability of not voting). But since probabilities vary depending on values of X, such a ratio varies as well. What remains constant is the ratio of such odds – e.g. odds of repeating a grade for male child divided by odds of repeating a grade for a female child will be the same number regardless of the values of other variables on the model. Similarly, the odds ratio for age can be a ratio of the odds of repeating a grade for someone who is 12 y.o. to the

odds of an 11 y.o. child, or of a 17 y.o. to a 16 y.o. childs's odds – these will be the same regardless of what age values you pick, as long as they are one year apart.

Odds ratios are exponentiated logistic regression coefficients. They are sometimes called factor coefficients, because they are multiplicative coefficients. Odds ratios are equal to 1 if there is no effect, smaller than 1 if the effect is negative and larger than 1 if it is positive. Let's obtain them:

```
. melogit rep1 pped male msescm || schoolid:, or

Mixed-effects logistic regression              Number of obs     =       7,516
Group variable:          schoolid              Number of groups  =         356

                                               Obs per group:
                                                             min =           2
                                                             avg =        21.1
                                                             max =          41

Integration method: mvaghermite                Integration pts.  =           7

                                               Wald chi2(3)      =       92.83
Log likelihood = -2720.7865                    Prob > chi2       =      0.0000
-----------------------------------------------------------------------------
       rep1 | Odds Ratio  Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+----------------------------------------------------------------
       pped |  .5341683   .0535194    -6.26   0.000     .4389293    .6500723
       male |  1.708059   .1297948     7.05   0.000     1.471704    1.982373
     msescm |  .7442283    .161031    -1.37   0.172     .4870001    1.137322
      _cons |  .1059981   .0113031   -21.05   0.000     .0860063    .1306369
------------+----------------------------------------------------------------
schoolid    |
   var(_cons)|  1.692804   .2115158                    1.325102    2.162541
-----------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline odds (conditional on zero random effects).
LR test vs. logistic model: chibar2(01) = 567.89      Prob >= chibar2 = 0.0000
```

So for example, the odds ratio for 1.7 for males indicates that the odds of repeating a grade for males are 1.7 times higher than for females –or we can say 70% higher. And the odds ratio of 0.74 for MSESC indicates that each additional unit increase in mean SES of school is associated in a 26% decrease in a student's odds of repeating a grade (albeit non-significant). To get percent change, we subtract 1 from the odds ratio, and then multiply the result by 100.

Beware: if you would like to know what the increase would be per, say, 10 units increase in the independent variable – e.g. 10 years of age or education, you cannot simply multiple the odds ratio by 10! The coefficient, in fact, would be odds ratio to the power of 10. Or alternatively, you could take the regular logit coefficient, multiply it by 10 and then exponentiate it.

In addition, since odds ratios are multiplicative coefficients, when you want to interpret, for example, an interaction term, you would have to multiply rather than add the odds ratio numbers. Alternatively, you can add the numbers presented in the coefficient column and then exponentiate the result.

In addition to using odds ratios, we can use predicted probabilities (P) to interpret our results.

It is more interesting for interpretation purposes to calculate predicted probabilities for some hypothetical, strategically selected cases. For instance, if we want to calculate the predicted probabilities for males and females and for those who did and did not attend pre-primary education (and who are in a school with average SES – school SES at mean), we calculate:

```
. margins, at(male=(0 1) pped=(0 1)) atmeans

Adjusted predictions                           Number of obs    =      7,516
Model VCE    : OIM

Expression   : Marginal predicted mean, predict()

1._at        : pped            =            0
               male            =            0
               msescm          =     .001837 (mean)

2._at        : pped            =            0
               male            =            1
               msescm          =     .001837 (mean)

3._at        : pped            =            1
               male            =            0
               msescm          =     .001837 (mean)

4._at        : pped            =            1
               male            =            1
               msescm          =     .001837 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |    Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .1488415   .0109228    13.63   0.000     .1274331    .1702499
          2  |   .2122994   .0131279    16.17   0.000     .1865693    .2380296
          3  |   .0936595   .0083003    11.28   0.000     .0773912    .1099278
          4  |   .1395175   .0107515    12.98   0.000     .1184448    .1605901
------------------------------------------------------------------------------
```

Such strategically calculated predicted probabilities are very useful for a more intuitive presentation of results. Note, however, that the differences between probabilities are not constant for a given factor – e.g., the gender difference here is not the same for those who attended and did not attend preprimary education (PPED=1 vs 0). To better understand interpretation of coefficients in all of these model, you could review my SOCY7704 notes for each of these topics.

Interactions

Note that interactions as a method to compare two or more groups can be problematic in logit models because the coefficients are scaled according to the differences in residual dispersion. So it is not as appropriate to rely on the significance test of the interaction term to establish whether some process differs by group. This is especially problematic for HLM as it heavily relies on the use of cross-level interactions. The best approach to establish whether the two groups do differ is to examine differences in predicted probabilities. When doing that, you would have to decide which values to assign to the rest of the variables in your model. For more detail on these comparisons, see:

Long, J. Scott and Sarah A. Mustillo. 2021. "Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes." Sociological Methods & Research 50 (3): 1284-1320.

```
. melogit rep1 i.pped##c.msescm i.male##c.msescm || schoolid:

Mixed-effects logistic regression              Number of obs     =       7,516
Group variable:         schoolid               Number of groups  =         356

                                               Obs per group:
                                                             min =           2
                                                             avg =        21.1
                                                             max =          41

Integration method: mvaghermite                Integration pts.  =           7

                                               Wald chi2(5)      =       94.34
Log likelihood = -2719.6352                    Prob > chi2       =      0.0000
------------------------------------------------------------------------------
        rep1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.pped |  -.6265808   .1006481    -6.23   0.000    -.8238475   -.4293141
      msescm |   -.468646    .269038    -1.74   0.082    -.9959508    .0586588
             |
       pped# |
    c.msescm |
           1 |  -.0012548   .2610455    -0.00   0.996    -.5128946     .510385
             |
      1.male |   .5459253   .0764003     7.15   0.000     .3961834    .6956671
             |
       male# |
    c.msescm |
           1 |   .2988648   .1975823     1.51   0.130    -.0883894    .6861191
             |
       _cons |  -2.250515   .1071137   -21.01   0.000    -2.460454   -2.040576
-------------+----------------------------------------------------------------
schoolid     |
   var(_cons)|   1.691005   .2121482                      1.322379     2.16239
------------------------------------------------------------------------------
LR test vs. logistic model: chibar2(01) = 560.27      Prob >= chibar2 = 0.0000

. sum msesc

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       msesc |      7,516     .009674    .3759242       -.77       1.49

. global max=r(max)

. global min=r(min)

. global mean=r(mean)

. margins, at(msesc=($max $mean $min) male=(0 1) pped=(0 1))

Adjusted predictions                           Number of obs     =       7,516
Model VCE    : OIM

Expression   : Marginal predicted mean, predict()

1._at        : pped            =            0
               msesc           =         1.49
```

```
                male            =              0

2._at          : pped           =              0
                 msesc          =           1.49
                 male           =              1

3._at          : pped           =              0
                 msesc          =        .009674
                 male           =              0

4._at          : pped           =              0
                 msesc          =        .009674
                 male           =              1

5._at          : pped           =              0
                 msesc          =           -.77
                 male           =              0

6._at          : pped           =              0
                 msesc          =           -.77
                 male           =              1

7._at          : pped           =              1
                 msesc          =           1.49
                 male           =              0

8._at          : pped           =              1
                 msesc          =           1.49
                 male           =              1

9._at          : pped           =              1
                 msesc          =        .009674
                 male           =              0

10._at         : pped           =              1
                 msesc          =        .009674
                 male           =              1

11._at         : pped           =              1
                 msesc          =           -.77
                 male           =              0

12._at         : pped           =              1
                 msesc          =           -.77
                 male           =              1
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |     .088422   .0292065     3.03   0.002     .0311783    .1456657
          2  |    .1810887   .0478308     3.79   0.000     .0873421    .2748353
          3  |    .1481178   .0109159    13.57   0.000      .126723    .1695127
          4  |    .2128655   .0131584    16.18   0.000     .1870754    .2386556
          5  |    .1896291   .0277656     6.83   0.000     .1352094    .2440488
          6  |    .2309484   .0300885     7.68   0.000      .171976    .2899208
          7  |    .0529285   .0195964     2.70   0.007     .0145203    .0913366
          8  |    .1164426   .0358417     3.25   0.001     .0461941    .1866911
          9  |    .0931805    .008324    11.19   0.000     .0768657    .1094954
         10  |    .1399748   .0108341    12.92   0.000     .1187403    .1612093
         11  |     .122897   .0232624     5.28   0.000     .0773034    .1684905
         12  |    .1537191   .0265344     5.79   0.000     .1017127    .2057255
------------------------------------------------------------------------------
```

9

Adjusted Predictions with 95% CIs
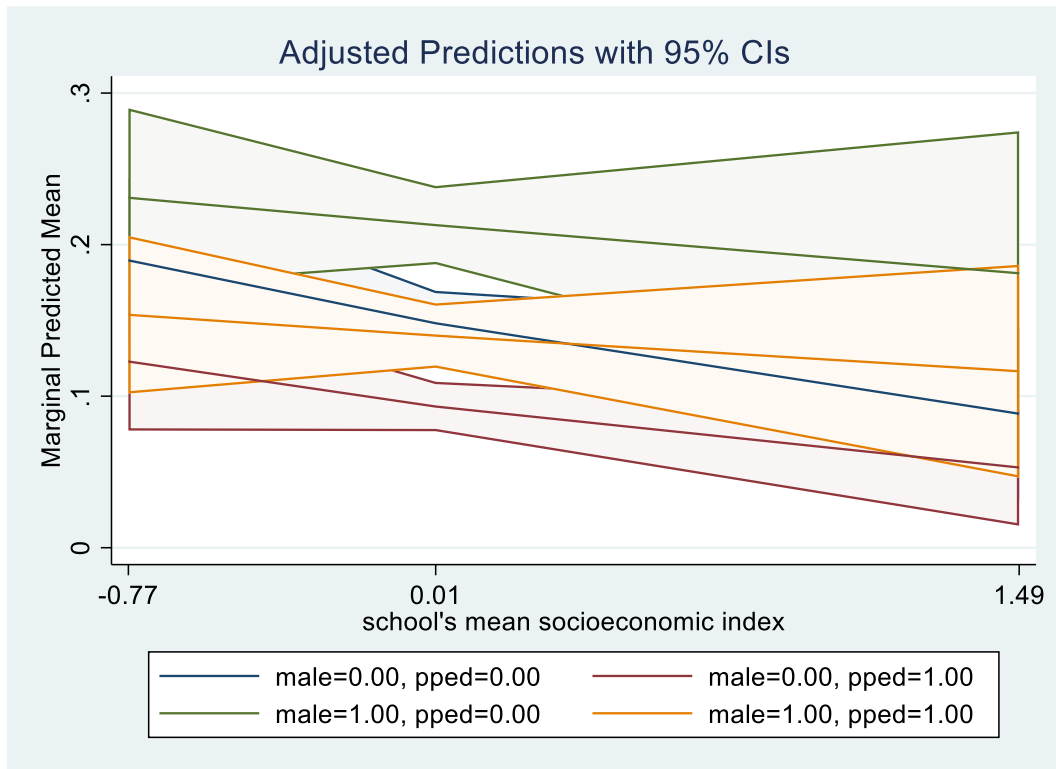
```
. margins, dydx(male) at(msesc=($min(.3)$max) pped=0)

Conditional marginal effects                          Number of obs     =       7,516
Model VCE    : OIM

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

1._at        : pped            =           0
               msesc           =        -.77

2._at        : pped            =           0
               msesc           =        -.47

3._at        : pped            =           0
               msesc           =        -.17

4._at        : pped            =           0
               msesc           =         .13

5._at        : pped            =           0
               msesc           =         .43

6._at        : pped            =           0
               msesc           =         .73

7._at        : pped            =           0
               msesc           =        1.03

8._at        : pped            =           0
               msesc           =        1.33

------------------------------------------------------------------------------
```

```
                   |              Delta-method
                   |      dy/dx    Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0.male       |    (base outcome)
-------------+----------------------------------------------------------------
1.male       |
         _at |
          1  |    .0413193    .0217315    1.90   0.057    -.0012736     .0839122
          2  |    .0510788    .0147275    3.47   0.001     .0222134     .0799442
          3  |    .0599054    .0099883    6.00   0.000     .0403288      .079482
          4  |    .0678063    .0098567    6.88   0.000     .0484876     .0871251
          5  |    .0747985    .0138159    5.41   0.000     .0477198     .1018772
          6  |    .0809072    .0191838    4.22   0.000     .0433077     .1185067
          7  |    .0861653    .0249034    3.46   0.001     .0373556      .134975
          8  |     .090612    .0306991    2.95   0.003     .0304428     .1507812
-------------------------------------------------------------------------------
```

Note: dy/dx for factor levels is the discrete change from the base level.

. marginsplot, x(msesc) plotop(msymbol(i)) recastci(rarea) ciopts(fintensity(5))

  Variables that uniquely identify margins: msesc

Conditional Marginal Effects of 1.male with 95% CIs
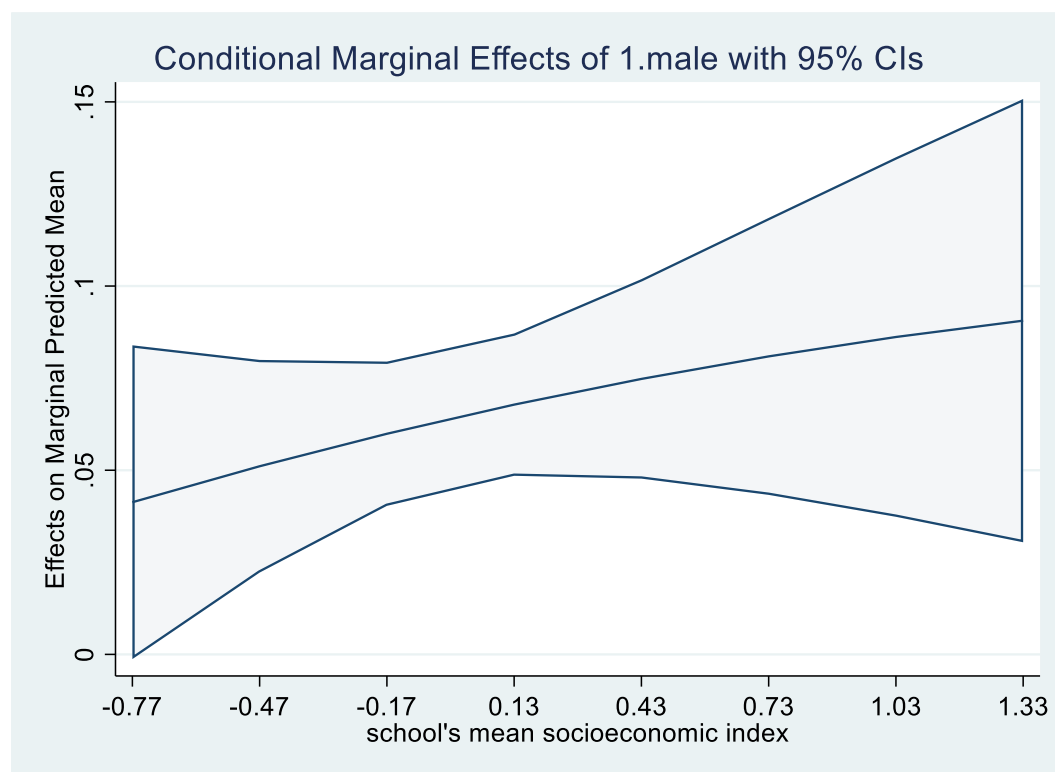
Ordered Logit

When the outcome variable is categorical but not binary – that is, either an ordinal variable or a nominal one with more than 2 categories—we can also use logit models, but need to modify them.

If your dependent variable has ordered categories (i.e. the order of categories is meaningful but the distances between them are arbitrary), you can use ordered logit. Here, we assume a latent dependent variable, and it is divided into intervals – those are categories we actually observe:
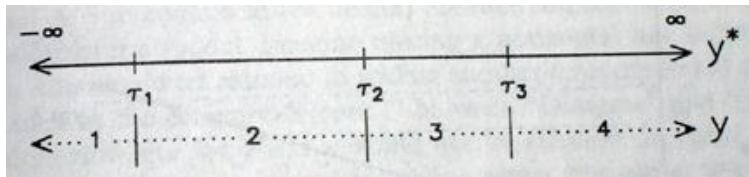
Then, our regression model of latent Y on X is assumed to look like this (you can see how the categories are mapped onto the latent variable – they are not equal).

Panel B: Regression of Observed y
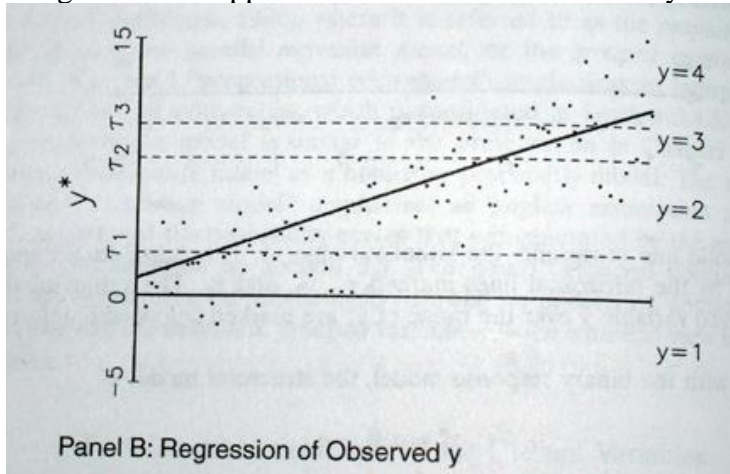
This is one interpretation of ordered logit model. Another one is that it combines a set of binary logits by constraining them to be the same equation. We could estimate binary logit models for each category to predict probability of belonging to that group or any group below it. We could then require all of these logits to have the same slopes and we could estimate them simultaneously – the result is the ordered logit model. To understand why they have to be the same (this is called parallel slopes assumption), we can return to our latent Y model – the slope of the line is the same across all categories – for the entire span of the latent variable. That is how this assumption looks when we examine probabilities:
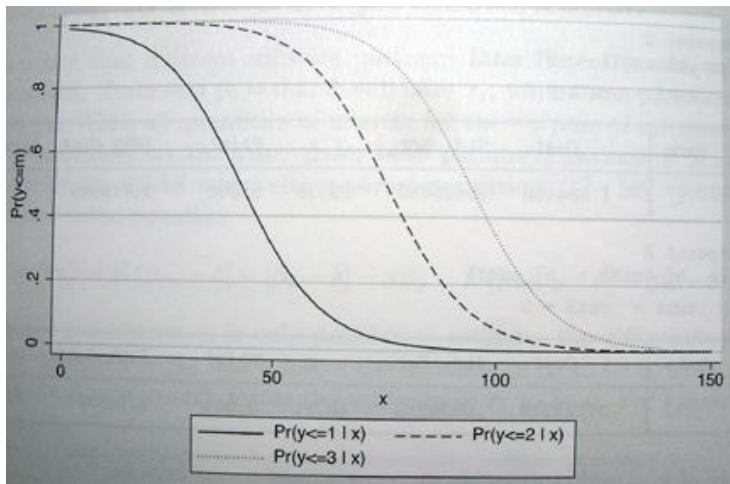
We'll use a subset of longitudinal data from the Health and Retirement Study (hrs_hours_reshaped.dta) and predict hours of work based on hours spent helping one's parents as well as various controls. We will recode hours of work into 4 categories for this example. I will right away estimate a model with odds ratios.

```
. recode  rworkhours80 (0=0) (1/30=1) (31/50=2) (51/80=3), gen(rworkhours4)
(23290 differences between rworkhours80 and rworkhours4)

. meologit rworkhours4 rpoorhealth  rmarried rtotalpar rsiblog hchildlg rallparhelptw
female raedyrs age minority || hhidpn: rpoorhealth , or

Mixed-effects ologit regression                 Number of obs      =      30,541
Group variable:           hhidpn                 Number of groups   =       6,243

                                                 Obs per group:
                                                               min =           1
                                                               avg =         4.9
                                                               max =           9

Integration method: mvaghermite                  Integration pts.   =           7

                                                 Wald chi2(10)      =     2886.84
Log likelihood = -29889.765                      Prob > chi2        =      0.0000
------------------------------------------------------------------------------
    rworkhours4 | Odds Ratio  Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
    rpoorhealth |   .2024437   .0142883   -22.63   0.000     .1762897    .2324778
       rmarried |   .7595365   .0495799    -4.21   0.000     .6683211    .8632014
      rtotalpar |   2.472556   .0587662    38.09   0.000     2.360018     2.59046
        rsiblog |   1.092916   .0567357     1.71   0.087     .9871859     1.20997
       hchildlg |   .8388503    .046101    -3.20   0.001     .7531903    .9342524
   rallparhelptw |   .9464933   .0038434   -13.54   0.000     .9389902    .9540563
         female |   .2541045   .0169547   -20.53   0.000     .2229551    .2896058
        raedyrs |   1.108991   .0129775     8.84   0.000     1.083845     1.13472
            age |   .8665796   .0093365   -13.29   0.000     .8484722    .8850735
        minority |   1.020861   .0821003     0.26   0.797     .8719888    1.195151
----------------+-------------------------------------------------------------
          /cut1 |  -7.457429    .642393                     -8.716496    -6.198362
          /cut2 |  -6.453125   .6418843                     -7.711196    -5.195055
          /cut3 |  -2.937608   .6407136                     -4.193384    -1.681833
----------------+-------------------------------------------------------------
hhidpn          |
 var(rpoorhealth)|   2.840702   .3109701                      2.292161    3.520516
      var(_cons)|   4.743435   .1555185                      4.448212    5.058252
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
LR test vs. ologit model: chi2(2) = 8925.32              Prob > chi2 = 0.0000

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

Briefly, the odds ratios for ordered logit are cumulative odds of belonging to a certain category or lower versus belonging to one of the higher categories. For example, if our dependent variable is the level of agreement with some statement and the categories are agree=3, not sure=2, and disagree=1, and if the odds ratio for gender as a predictor of that agreement is 2.00, we can say that the odds of disagreeing rather than agreeing or being not sure are 2 times higher for women than for men. Similarly, the odds of disagreeing or being not sure are also twice as high for women than for men.

What this means is that ologit assumes that these two odds ratios are essentially the same and thus uses the average. That is called the parallel slopes assumption. So we are assuming these two odds ratios are the same – if they differ significantly, the assumption is violated.
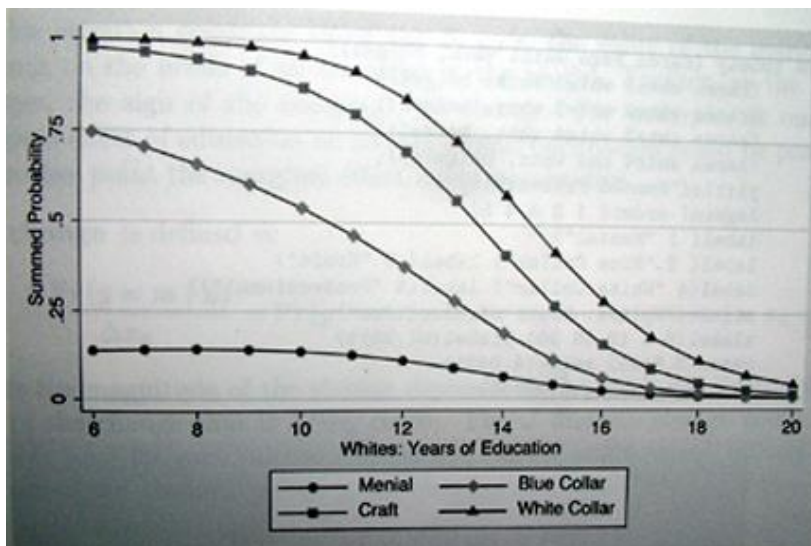
Stata does not provide diagnostic tools for panel data for testing the parallel slopes assumption, so in order to obtain a rough test, you might want to run your model without taking panel nature of the data into account (using regular ologit command) and test that assumption that way, even though such a test will be approximate. Another way to do so would be to create the corresponding dichotomies (here, rworkhours4=1 vs 2 3 4, rworkhours4= 1 2 vs 3 4, and rworkhours4=1 2 3 vs 4) and estimate models separately for each dichotomy using melogit – we can then examine whether odds ratios indeed look similar in such models (that's what parallel slopes assumption presumes).

To interpret coefficients here, for example, if a respondent reports poor health, their odds of working full time or overtime are almost 80% lower.

Multinomial Logit

We use multinomial logit models when we have multiple categories but cannot order them (or we can, but the parallel regression assumption does not hold). Here the order of categories is unimportant. Multinomial logit model is equivalent to simultaneous estimation of multiple logits where each of the categories is compared to one selected so-called base category. But if we would estimate them separately, we would lose information, as each logit would be estimated on a different sample (selected category plus base category, with all other categories omitted from analyses). To avoid that, we use multinomial logit.

Multinomial logit does not assume parallel slopes – so if we estimate it for ordinal level variable and then plot cumulative probabilities, we would see something like this (note the variation in slope!):

14

The best way to estimate a random effects multinomial logit in Stata is using SEM module, and specifically gsem command with mlogit option. You can either specify a version estimating a single random effect or a separate random effect for each equation (we have m-1 equations in multinomial logit with m alternatives). The second option is less restrictive and usually preferred. Here is an example using rworkhours4 we generated above (its values are 0 -3). We can either use model builder to build the following GSEM model, or enter a command:


```
. gsem (1.rworkhours4 <- rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw
female raedyrs age minority M1[hhidpn])  (2.rworkhours4 <- rpoorhealth rmarried
rtotalpar rsiblog hchildlg rallparhelptw female raedyrs age minority M2[hhidpn])
(3.rworkhours4 <- rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw female
raedyrs age minority M3[hhidpn]), mlogit

Generalized structural equation model          Number of obs    =      30541
Log likelihood = -27850.963

 ( 1)   [1.rworkhours4]M1[hhidpn] = 1
 ( 2)   [2.rworkhours4]M2[hhidpn] = 1
 ( 3)   [3.rworkhours4]M3[hhidpn] = 1
-------------------------------------------------------------------------------
               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+---------------------------------------------------------------
0.rworkhours4  |  (base outcome)
---------------+---------------------------------------------------------------
1.rworkhours4 <- |
   rpoorhealth |  -1.277058   .0828982   -15.41   0.000    -1.439536   -1.114581
      rmarried |  -.0611107   .1034527    -0.59   0.555    -.2638743    .141653
      rtotalpar |   .7099914   .0435582    16.30   0.000     .6246189    .7953638
        rsiblog |   .1661102    .077469     2.14   0.032     .0142738    .3179466
       hchildlg |   -.096628   .0827913    -1.17   0.243    -.2588961     .06564
  rallparhelptw |  -.0141175   .0064669    -2.18   0.029    -.0267924   -.0014426
        female  |  -.1735534   .0970918    -1.79   0.074    -.3638498    .016743
        raedyrs |   .1371333   .0171756     7.98   0.000     .1034697    .1707969
```

```
           age |  -.0867121    .0155403    -5.58   0.000    -.1171705   -.0562537
      minority |  -.0228635     .116838    -0.20   0.845    -.2518618    .2061348
               |
      M1[hhidpn] |         1   (constrained)
               |
          _cons |   .9628375    .9322679     1.03   0.302     -.864374    2.790049
---------------+----------------------------------------------------------------
2.rworkhours4 <- |
    rpoorhealth |  -1.750017    .0704374   -24.85   0.000    -1.888072   -1.611963
      rmarried |  -.4516769    .0962687    -4.69   0.000    -.6403599   -.2629938
      rtotalpar |   1.442584    .0412238    34.99   0.000     1.361786    1.523381
        rsiblog |   .1445202    .0763168     1.89   0.058    -.0050579    .2940982
        hchildlg |  -.3232379     .080663    -4.01   0.000    -.4813346   -.1651413
    rallparhelptw |  -.0809335    .0061454   -13.17   0.000    -.0929784   -.0688887
         female |  -1.543764    .1004039   -15.38   0.000    -1.740552   -1.346976
         raedyrs |    .130158     .017173     7.58   0.000     .0964996    .1638165
           age |  -.2034751    .0160609   -12.67   0.000    -.2349538   -.1719963
      minority |   .1211653    .1181552     1.03   0.305    -.1104146    .3527453
               |
      M2[hhidpn] |         1   (constrained)
               |
          _cons |   9.600093    .9520152    10.08   0.000     7.734177    11.46601
---------------+----------------------------------------------------------------
3.rworkhours4 <- |
    rpoorhealth |  -1.916659    .1124895   -17.04   0.000    -2.137134   -1.696184
      rmarried |  -.5334537    .1407739    -3.79   0.000    -.8093655   -.2575419
      rtotalpar |   1.748915    .0562372    31.10   0.000     1.638692    1.859137
        rsiblog |   .2218076    .1072693     2.07   0.039     .0115636    .4320515
        hchildlg |  -.2066806    .1144466    -1.81   0.071    -.4309919    .0176307
    rallparhelptw |  -.1073193    .0103503   -10.37   0.000    -.1276055   -.0870331
         female |  -2.875229    .1459855   -19.70   0.000    -3.161355   -2.589102
         raedyrs |   .2372999    .0241533     9.82   0.000     .1899602    .2846395
           age |  -.2478874    .0225544   -10.99   0.000    -.2920931   -.2036816
      minority |  -.0932506     .166257    -0.56   0.575    -.4191083    .2326072
               |
      M3[hhidpn] |         1   (constrained)
               |
          _cons |   8.004163    1.330828     6.01   0.000     5.395788    10.61254
---------------+----------------------------------------------------------------
  var(M1[hhidpn])|   7.277483    .3634233                     6.598935    8.025803
  var(M2[hhidpn])|   10.11641    .4346091                     9.299468    11.00512
  var(M3[hhidpn])|   16.62133    .7747086                     15.17022    18.21125
---------------+----------------------------------------------------------------
  cov(M2[hhidpn],|
   M1[hhidpn])|   5.774199     .324637    17.79   0.000     5.137922    6.410476
  cov(M3[hhidpn],|
   M1[hhidpn])|   6.872237      .41161    16.70   0.000     6.065496    7.678978
  cov(M3[hhidpn],|
   M2[hhidpn])|   10.77613      .50847    21.19   0.000     9.779549    11.77272
-------------------------------------------------------------------------------
```

16
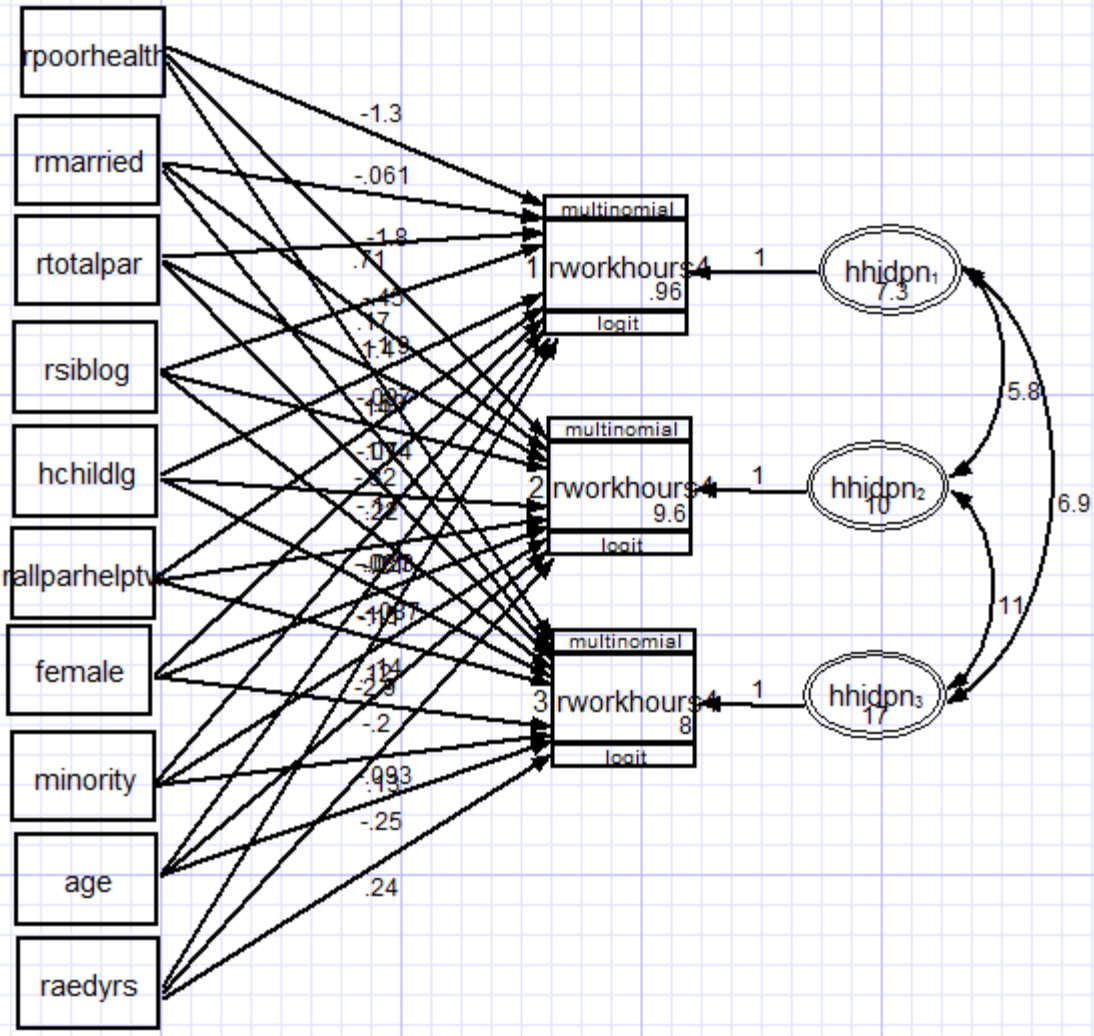
You would need to exponentiate coefficients to get odds ratios. Note that variables predict the membership in each group as compared to membership in the omitted group (rworkgroup4=0), so you can interpret odds ratios as the effect of a given predictor on the odds of being in a given group vs in the omitted group.

```
. est store gsem

. est table gsem, eform star b(%4.3f)

-----------------------------
    Variable |      gsem
-------------+---------------
0b.rworkho~4 |
       _cons | (omitted)
-------------+---------------
1.rworkhou~4 |
 rpoorhealth |      0.279***
    rmarried |      0.941
    rtotalpar |     2.034***
      rsiblog |     1.181*
     hchildlg |     0.908
 rallparhel~w |     0.986*
       female |     0.841
```

```
      raedyrs |     1.147***
          age |     0.917***
     minority |     0.977
              |
   M1[hhidpn] |     2.718
              |
        _cons |     2.619
--------------+--------------
2.rworkhou~4  |
 rpoorhealth  |     0.174***
    rmarried  |     0.637***
    rtotalpar |     4.232***
      rsiblog |     1.155
     hchildlg  |     0.724***
rallparhel~w  |     0.922***
       female |     0.214***
      raedyrs |     1.139***
          age |     0.816***
     minority |     1.129
              |
   M2[hhidpn] |     2.718
              |
        _cons |   1.5e+04***
--------------+--------------
3.rworkhou~4  |
 rpoorhealth  |     0.147***
    rmarried  |     0.587***
    rtotalpar |     5.748***
      rsiblog |     1.248*
     hchildlg  |     0.813
rallparhel~w  |     0.898***
       female |     0.056***
      raedyrs |     1.268***
          age |     0.780***
     minority |     0.911
              |
   M3[hhidpn] |     2.718
              |
        _cons |  2993.395***
--------------+--------------
         var( |
  M1[hhidpn])| 1447.340***
         var( |
  M2[hhidpn])|   2.5e+04***
         var( |
  M3[hhidpn])|   1.7e+07***
         cov( |
  M1[hhidpn],|
  M2[hhidpn])|  321.886***
         cov( |
  M1[hhidpn],|
  M3[hhidpn])|  965.105***
         cov( |
  M2[hhidpn],|
  M3[hhidpn])|   4.8e+04***
------------------------------
legend: * p<0.05; ** p<0.01; *** p<0.001
```

So for example, the odds of working part time rather than not working are 72.1% lower for those in poor health than for those in good health, while the odds of working full time rather than not

working are 82.6% lower and the odds of working overtime rather than not working are 85.3% lower for the same health comparison.

To further complicate this model and allow for random slopes, you need to introduce even more latent variables with hhidpn in them and have these variables interact with those predictors for which you want to have random slopes. In terms of syntax, it looks like this:

```
gsem (1.rworkhours4 <- rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw
female raedyrs age minority M1[hhidpn] 1.rpoorhealth#M2[hhidpn])  (2.rworkhours4 <-
rpoorhealth rmarried rtotalpar rsiblog hchildlg rallparhelptw female raedyrs age
minority M3[hhidpn] 1.rpoorhealth#M4[hhidpn]) (3.rworkhours4 <- rpoorhealth rmarried
rtotalpar rsiblog hchildlg rallparhelptw female raedyrs age minority M5[hhidpn]
1.rpoorhealth#M6[hhidpn]), mlogit
```

For a simpler case, a diagram would look like this:
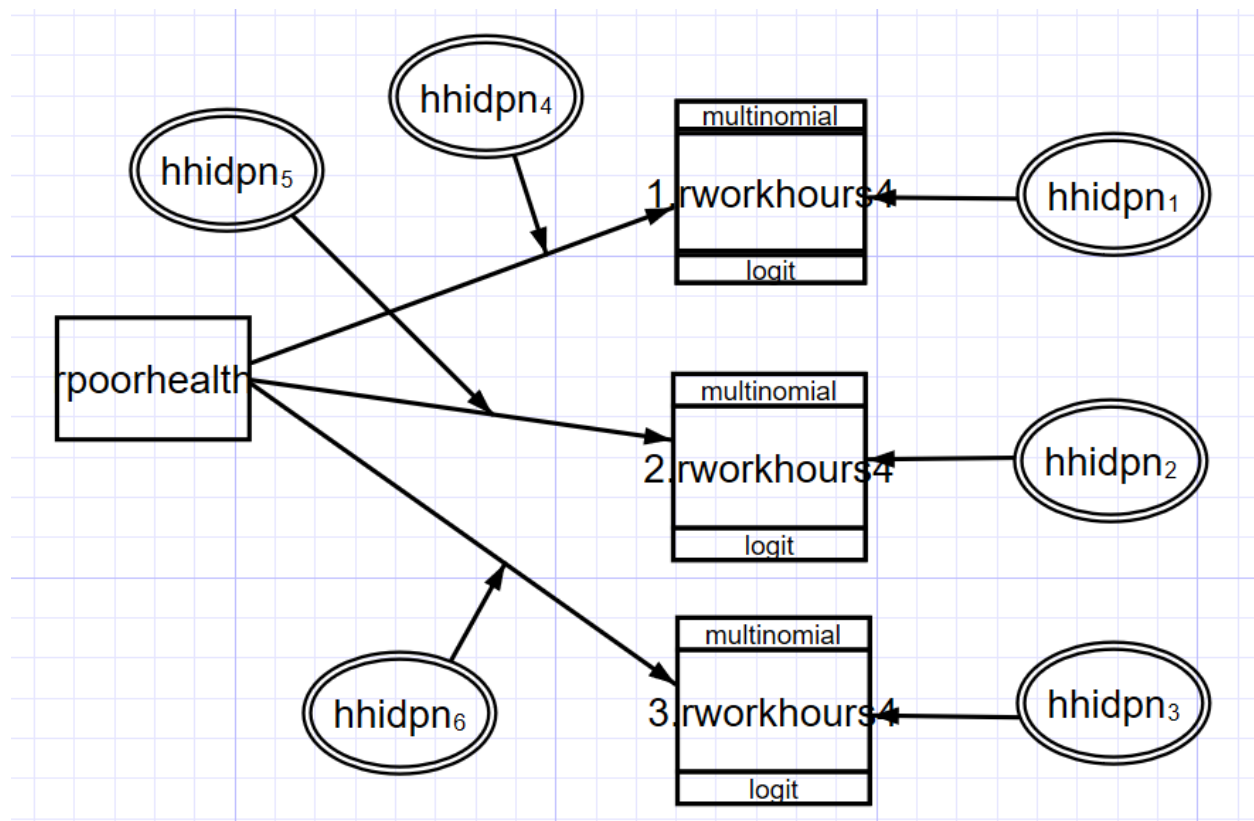
These kinds of models tend to estimate very slowly, and you may need to deal with convergence issues – see intro 12 at https://www.stata.com/manuals/sem.pdf

Count data models

Count variables are often treated as though they are continuous, and regular regression is used, but it can result in inefficient, inconsistent, and biased estimates. Need to use models that are developed specifically for count data. Poisson model is the most basic of them.

Characteristics of Poisson distribution:

1. $E(y) = \mu$

2. The variance equals the mean: $Var(y)=E(y)=\mu$ -- equidispersion. In practice, the variance is often larger than μ: this is called overdispersion. The main reason for overdispersion is heterogeneity – if there are different groups within data that have different means and all of them are actually equal to their variances, when you put all of these groups together, the resulting combination will have variance larger than the mean. Therefore, we need to control for all those sources of heterogeneity. Thus, when using Poisson regression, we need to ensure that the conditional variance equals to the mean – that is $Var(y|X)=E(y|X)$.

3. As μ increases, the probability of zeros decreases. But for many count variables, there are more observed zeros than would be predicted from Poisson distribution

4. As μ increases, the Poisson distribution approximates normal.

5. The assumption of independence of events – past outcomes don't affect future outcomes.

Poisson distributions:
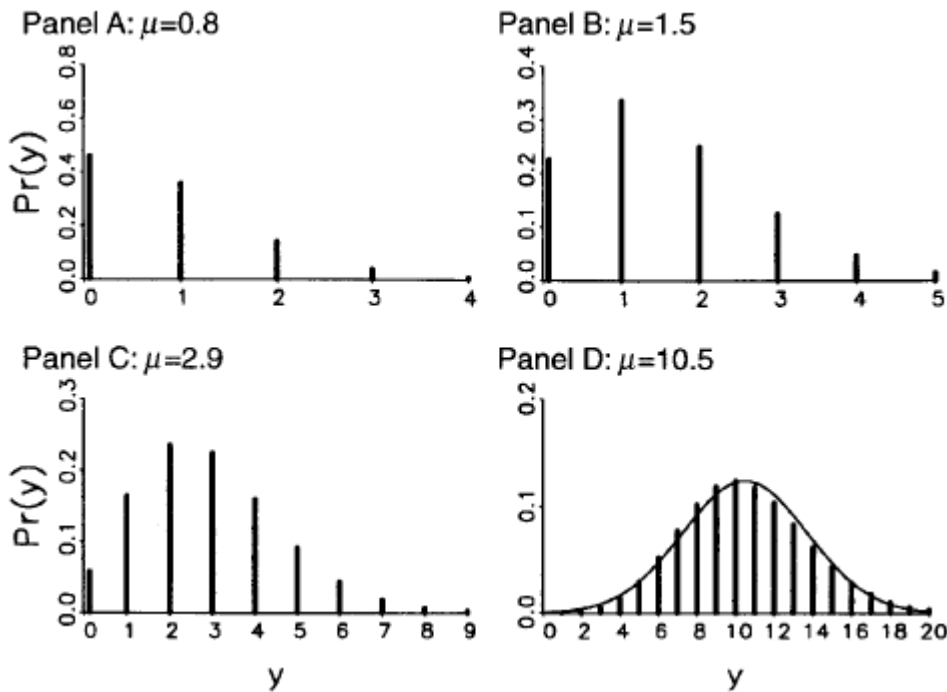
Luckily, we can estimate a mixed effects Poisson model or negative binomial model. Let's examine an example of mixed Poisson model (there is also a corresponding negative binomial regression command, menbreg):

```
. mepoisson  rworkhours80 rpoorhealth  rmarried rtotalpar rsiblog hchildlg
rallparhelptw female raedyrs age minority || hhidpn:

Mixed-effects Poisson regression                    Number of obs      =      30541
Group variable: hhidpn                              Number of groups   =       6243

                                                    Obs per group: min =          1
                                                                   avg =        4.9
                                                                   max =          9

Integration points =    7                           Wald chi2(10)      =   30723.86
```

```
Log likelihood = -237592.87                         Prob > chi2         =     0.0000

------------------------------------------------------------------------------
rworkhours80 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 rpoorhealth |   -.317193   .0052424   -60.50   0.000    -.327468    -.3069181
    rmarried |  -.0465037   .0082793    -5.62   0.000    -.0627308   -.0302767
   rtotalpar |   .3166688   .0022498   140.75   0.000     .3122593    .3210784
     rsiblog |   .1548763   .0113293    13.67   0.000     .1326713    .1770813
    hchildlg |  -.1326281    .010343   -12.82   0.000    -.1529001   -.1123561
 rallparhel~w | -.0194976   .0004307   -45.27   0.000    -.0203418   -.0186534
      female |  -1.088872   .0594299   -18.32   0.000    -1.205353    -.972392
     raedyrs |   .1293615    .009998    12.94   0.000     .1097657    .1489573
         age |  -.1223611   .0096613   -12.67   0.000     -.141297   -.1034253
    minority |  -.1271817   .0701653    -1.81   0.070    -.2647031    .0103396
       _cons |   7.385015   .5605793    13.17   0.000      6.2863    8.483731
------------------------------------------------------------------------------


------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
hhidpn: Identity             |
                  sd(_cons)  |   2.247979   .0274216     2.194871    2.302372
------------------------------------------------------------------------------
LR test vs. Poisson regression:  chibar2(01) =  3.9e+05 Prob>=chibar2 = 0.0000
```

With regular coefficients presented here, we can interpret sign and significance, but to interpret the size, we exponentiate the coefficients – these are called incidence rate ratios. They are also multiplicative coefficients, like odds ratios, and can be interpreted as percent change in the number of events.

```
. mepoisson  rworkhours80 rpoorhealth  rmarried rtotalpar rsiblog hchildlg
rallparhelptw female raedyrs age minority || hhidpn:, irr

Mixed-effects Poisson regression           Number of obs      =      30,541
Group variable:          hhidpn            Number of groups   =       6,243

                                           Obs per group:
                                                         min =          1
                                                         avg =        4.9
                                                         max =          9

Integration method: mvaghermite            Integration pts.   =          7

                                           Wald chi2(10)      =   30835.46
Log likelihood =  -237571.3                Prob > chi2        =     0.0000
------------------------------------------------------------------------------
rworkhours80 |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 rpoorhealth |   .7282366   .0038166   -60.51   0.000     .7207944    .7357557
    rmarried |   .9545705   .0078981    -5.62   0.000     .9392154    .9701767
   rtotalpar |   1.372556   .0030877   140.77   0.000     1.366517    1.378621
     rsiblog |    1.16752   .0131903    13.71   0.000     1.141952    1.193661
    hchildlg |   .8757708   .0090403   -12.85   0.000     .8582302    .8936699
 rallparhel~w |  .9806913   .0004224   -45.27   0.000     .9798638    .9815195
      female |   .3359979   .0189861   -19.30   0.000     .3007725    .3753487
     raedyrs |   1.138186   .0107564    13.70   0.000     1.117298    1.159464
         age |   .8847575   .0080896   -13.39   0.000     .8690434    .9007556
    minority |   .8809494   .0585982    -1.91   0.057     .7732706    1.003623
       _cons |   1610.439   855.6806    13.90   0.000      568.424    4562.642
-------------+----------------------------------------------------------------
hhidpn       |
```

```
     var(_cons)|   5.119811   .1233827                        4.883608    5.367439
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
Note: _cons estimates baseline incidence rate (conditional on zero random
      effects).
LR test vs. Poisson model: chibar2(01) = 3.9e+05      Prob >= chibar2 = 0.0000
```

So for example, if the respondent is currently married, they work 4.5% fewer hours than if they are not married. Each extra year of education increases the number of hours worked by 14%.

In practice, Poisson regression model rarely fits due to overdispersion. One key process that often creates overdispersion is known as contagion – violation of the assumption of the independence of events. This assumption is often unrealistic; e.g. if you have your first child, that increases your chances of having your second.

To better model overdispersion from this and other sources, we can use negative binomial model. It allows taking into account unobserved heterogeneity. To do so, it introduces an additional parameter – alpha, known as the dispersion parameter. Increasing alpha increases the conditional variance of our count variable. If alpha is zero, the model becomes regular Poisson model.

```
. menbreg  rworkhours80 rpoorhealth  rmarried rtotalpar rsiblog hchildlg rallparhelptw
female raedyrs age minority || hhidpn:, irr

Mixed-effects nbinomial regression            Number of obs     =      30,541
Overdispersion:            mean
Group variable:           hhidpn              Number of groups  =       6,243

                                              Obs per group:
                                                            min =           1
                                                            avg =         4.9
                                                            max =           9

Integration method: mvaghermite               Integration pts.  =           7

                                              Wald chi2(10)     =     2400.45
Log likelihood = -110234.72                   Prob > chi2       =      0.0000
------------------------------------------------------------------------------
rworkhours80 |      IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 rpoorhealth |  .4204695   .0156794   -23.23   0.000     .3908346    .4523515
    rmarried |  .8480047   .0445282    -3.14   0.002     .7650718    .9399274
   rtotalpar |  1.671643   .0306495    28.02   0.000     1.612638    1.732808
     rsiblog |  1.124322   .0469158     2.81   0.005     1.036028     1.22014
    hchildlg |  .8842482   .0392321    -2.77   0.006     .8106031    .9645841
rallparhel~w |  .9687546   .0030539   -10.07   0.000     .9627874    .9747587
      female |  .3541711    .019064   -19.28   0.000     .3187098    .3935779
     raedyrs |  1.106181   .0103088    10.83   0.000     1.086159    1.126571
         age |  .8974593   .0078051   -12.44   0.000     .8822911    .9128882
    minority |  1.002555   .0645823     0.04   0.968     .8836405    1.137472
       _cons |  1431.585   739.6125    14.07   0.000      520.06     3940.769
-------------+----------------------------------------------------------------
     /lnalpha |  .8295796   .0117374                       .8065748    .8525844
-------------+----------------------------------------------------------------
hhidpn       |
   var(_cons)|  3.850327   .1192165                       3.623615    4.091222
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
```

```
Note: _cons estimates baseline incidence rate (conditional on zero random
      effects).
LR test vs. nbinomial model: chibar2(01) = 2030.20    Prob >= chibar2 = 0.0000
```

The extra parameter here, expressed as logarithm alpha, is clearly different from 0 (based on the confidence interval), so this model would be preferable over poisson. Of course, we can still explore random slopes and cross-level interactions, etc.

Overall, some of the same concerns apply here as was the case for logistic regression – for instance, we have to be cautious when interpreting interactions and examine predicted counts. In terms of variance, the level 1 residuals variance, which we assumed to be 3.29 in logit-based models, is assumed to be equal to predicted mean, so you need to find the value of average predicted count by generating predicted values, exponentiating them, and calculating their mean. You can use that number as level 1 variance in the formulas for percent of variance explained discussed above as well as when calculating the intraclass correlation coefficient.