Multilevel Modeling Day 2
Intermediate and Advanced Issues: Multilevel Models as Mixed Models

Jian Wang

September 18, 2012

What are mixed models

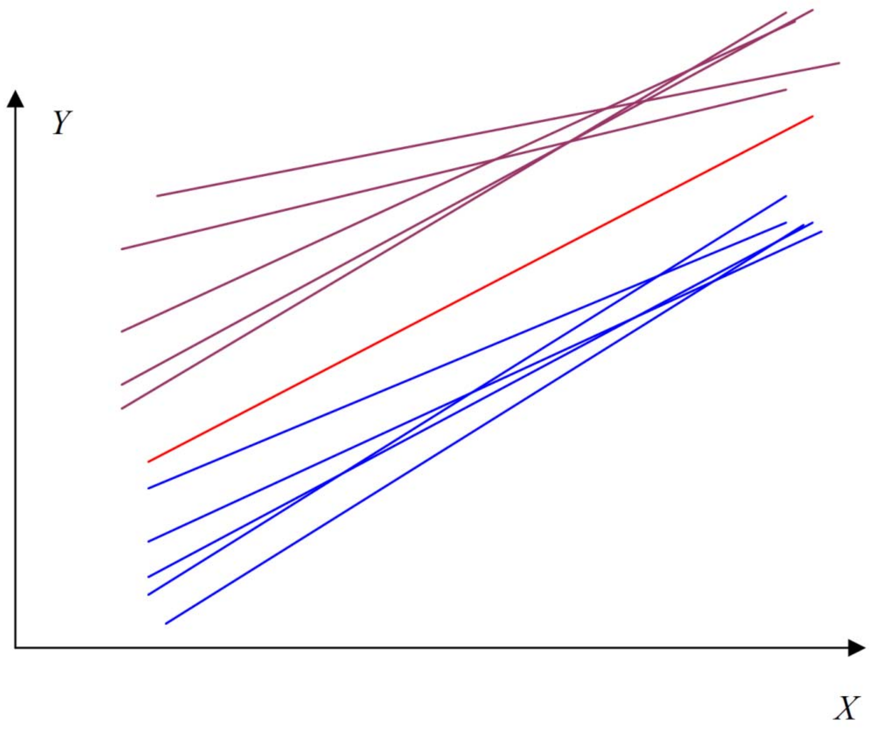
- The simplest multilevel models are in fact mixed models: have fixed and random parameters or effects

  (3.3) $$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \ .$$
- γ_{00} is a fixed parameter and τ_{00} and σ^2 are random parameters
- The theory of mixed models is directly applicable

- Mixed effect models: regression analysis models with two types of effects:
  - Fixed effects (intercepts, slopes): describe the population studied as a whole. These effects are just like intercepts and slopes in conventional regression models.
  - Random effects: “bumps” up and down on the population intercepts and slopes, which are used to describe subpopulations. These effects can vary across subpopulations

Illustration of the Fixed and Random Effects

- X-socioeconomic status (SES); Y-achievement
- 10 schools
- Red line: population relationship line
  - describes an overall pattern of relationship
  - A "stable" relationship, i.e., its intercept and slope are fixed.
- Within each school, the intercepts and slopes vary
  - Typically schools are randomly drawn
  - Random effects



Mixed Model

- Mixed model are used for multilevel modeling, growth curve analysis, panel analysis or cross-sectional time series analysis
- Multilevel modeling: GPA related to study hours, but this relationship may differ by major
  - Fixed effects: overall relationship that GPA increases with study hours
  - Random effects: for a given major, the way this relationship differs from the overall one
- Longitudinal modeling: fluid intelligence declines with age in different nursing homes
  - Fixed effects: general trend
  - Random effects: for a given nursing home, it may have its own intercept and slope parameters. The amounts by which they differ from those in the overall population are represented by the random effects.
- More complex models: further nesting
- STATA comments:
  - xtmixed: fit mixed regression models
  - xtmelogit: fit mixed models with binary outcomes
  - xtmepoisson: fit mixed models to count data

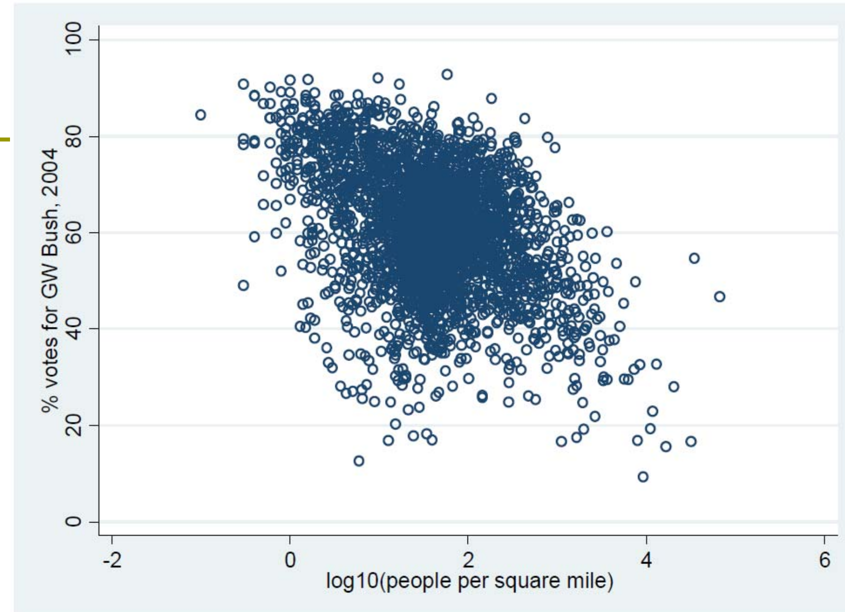
Data for Demonstration

- Percentage voted for Bush (*bush*)
- Logarithm of number of people living in a square mile (*logdens*)
- Population minority (*minority*)
- Proportion adults aged 25 or higher with at least 4 years of college education (*colled*)
- Census division—out of 9 geographic regions (*cendiv*)

```
. use c:\stata\data\presidential_elections_2004.dta, clear


. d


Contains data from c:\stata\data\presidential_elections_2004.dta
  obs:         3,054                    US counties -- 2004 election (Robinson,
2005)
 vars:            11                    6 May 2011 15:21
 size:       232,104 (97.8% of memory free)
-----------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------
fips            long   %9.0g                  FIPS code
state           str20  %20s                   State name
state2          str2   %9s                    State 2-letter abbreviation
region          byte   %9.0g       region     Region (4)
cendiv          byte   %15.0g      division   Census division (9)
county          str24  %24s                   County name
votes           float  %9.0g                  Total # of votes cast, 2004
bush            float  %9.0g                  % votes for GW Bush, 2004
logdens         float  %9.0g                  log10(people per square mile)
minority        float  %9.0g                  % population minority
colled          float  %9.0g                  % adults >25 w/4+ years college
-----------------------------------------------------------------------------
Sorted by:  fips
```

NOT Mixed Model

- Misleading: pools together politically, geographically, and economically diverse counties.
- There may be important differences across them with respect to the relationship between percentage voted for Bush and other characteristics (urban-rural differential)



- The traditional modeling approach does not capture the geographic differences in voting.
- The conventional approach assumes the same intercept and slope for all 3054 counties

Source	SS	df	MS
Model	122345.617	3	40781.8725
Residual	358593.826	3037	118.075017
Total	480939.443	3040	158.203764

Number of obs	=	3041
F(3, 3037)	=	345.39
Prob > F	=	0.0000
R-squared	=	0.2544
Adj R-squared	=	0.2537
Root MSE	=	10.866

bush	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logdens	-5.457462	.3031091	-18.00	0.000	-6.051781	-4.863142
minority	-.251151	.0125261	-20.05	0.000	-.2757115	-.2265905
colled	-.1811345	.0334151	-5.42	0.000	-.246653	-.115616
_cons	75.78636	.5739508	132.04	0.000	74.66099	76.91173

Conventional Regression -> Random Intercept

- The fitted model:

$$(2.1) \qquad y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_i \,, \ (\, i = 1, \ldots, n).$$

  - where X1=logdens, X2=minority, X3=colled
  - The subindex signals individual county
  - Right-hand side says nothing about possible geographic differences
  - None of the βs is county specific but assumed to be the same for all counties.
- Consider different geographic regions (9 census divisions: e.g., New England, Middle Atlantic, Mountain, Pacific).
  - Allow each region to have its own intercept: random intercept model
  - Each region is permitted to have its own intercept, but there's nothing more specific to it.
  - u_{0j} is the unique region effect in the model
  - The mean percent voting for Bush (mean of y) across counties within region is not the same in all regions at the average predictor values.

$$(2.2) \qquad y_{ij} = \beta_0 + \mathbf{u}_{0j} + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} + \varepsilon_{ij}$$

$$= \beta_{0j} + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} + \varepsilon_{ij} \ .$$

Random Intercept Model

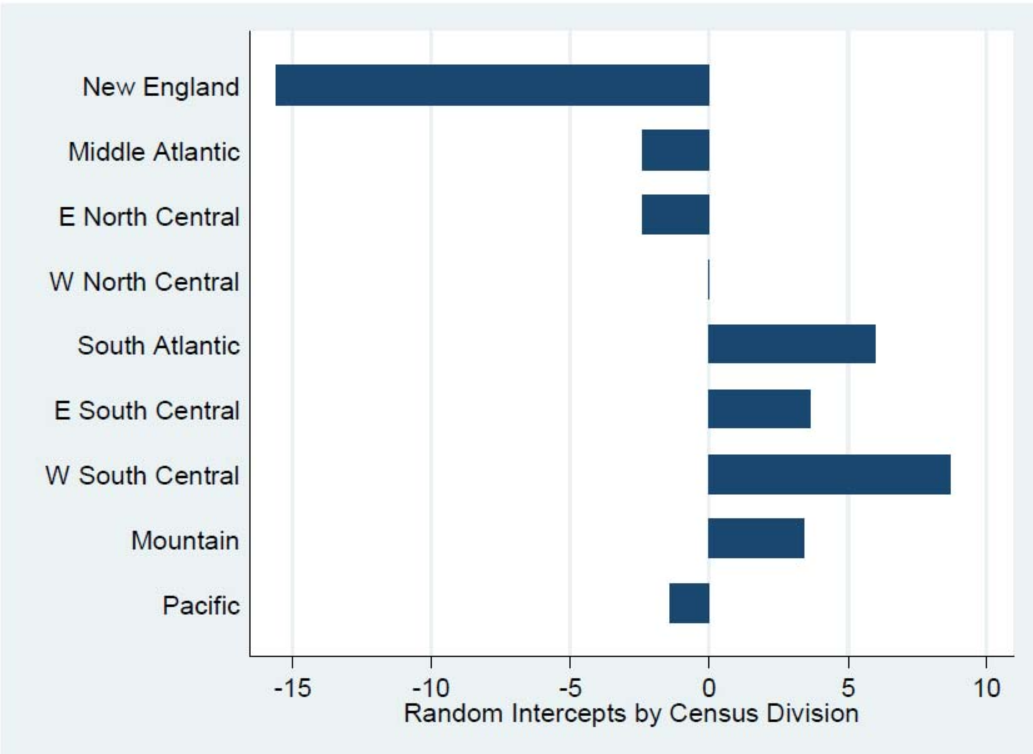
```
. xtmixed bush logdens minority colled || cendiv:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -11339.79
Iteration 1:   log likelihood =  -11339.79

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      3041
Group variable: cendiv                          Number of groups   =         9
                                                Obs per group: min =        67
                                                               avg =     337.9
                                                               max =       616

                                                Wald chi2(3)       =   1161.96
Log likelihood =  -11339.79                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logdens |   -4.52417   .3621775   -12.49   0.000    -5.234025   -3.814316
    minority |  -.3645394   .0129918   -28.06   0.000    -.3900029   -.3390758
      colled |  -.0583942   .0357717    -1.63   0.103    -.1285053     .011717
       _cons |   72.09305   2.294038    31.43   0.000     67.59682    76.58929
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Identity             |
                   sd(_cons) |   6.617132   1.600465      4.119005    10.63034
-----------------------------+------------------------------------------------
                sd(Residual) |   10.00339   .1284657      9.754742    10.25837
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   455.99 Prob >= chibar2 = 0.0000
```

- LR test: the RIM is a better model than the traditional regression
- The standard deviation of the random intercepts is found to be significant
- (2.4) gives the estimated model. Stars indicate that it is an estimated model

$$\text{(2.4)}\quad bush_{ij} = 72.13\text{-}4.70 logdens_{ij}\ \text{-}.37 minority_{ij}\text{-}.04 colled_{ij} + {}^{*}u_{0j} + {}^{*}\varepsilon_{ij}$$

Random Intercept Model

```
. predict randint0, reffects
. graph hbar (mean) randint0, over(cendiv)
```



- Evaluate the regional intercepts
- Produce best linear unbiased predictions (BLUPs) of the random effects
- randint0 is a new variable with values for 3054 counties from 9 regions, it has the same value within each region. The value is actually the mean of all its values for a given region
- Interpretation: at any given level of the 3 predictors (logdens, minority and colled), percentage of voters for Bush is on average 16 points lower in the New England (NE) region than in the West/North/Central region, and 22 percent lower in the NE region than in the South Atlantic region.

Random Intercept Model

- List all the BLUP values for the random effects: the same values were assigned to each county in the same census division.
- BLUP is used in linear mixed models for the estimation of random effects. It is similar to the best linear unbiased estimates (BLUEs) of fixed effects.
  - Robinson, G.K. (1991). "That BLUP is a Good Thing: The Estimation of Random Effects". *Statistical Science* **6**(1): 15–32
  - Henderson, C.R. (1975). "Best linear unbiased estimation and prediction under a selection model". *Biometrics* **31** (2): 423–447.

First consider an easy situation. Assumed model is a one-way random effects model with known intercept and variance components and normally distributed random effects:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \ldots, n_i; i = 1, \ldots, q$$

$$b_i \sim i.i.d.\ N\left(0, \sigma_b^2\right)$$

$$\varepsilon_{it} \sim i.i.d.\ N\left(0, \sigma_\varepsilon^2\right)$$

$$\varepsilon_{it} \perp b_i, \mu, \sigma_\varepsilon^2, and\ \sigma_b^2\ known$$

In which case the Best Linear Unbiased Predictor is given by

$$\tilde{b}_i = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i}(\bar{Y}_{i\cdot} - \mu)$$

```
. list cendiv randint0 in 1/100
```

	cendiv	randint0
1.	New England	-15.57527
2.	New England	-15.57527
3.	New England	-15.57527
4.	New England	-15.57527
5.	New England	-15.57527
6.	New England	-15.57527
7.	New England	-15.57527
8.	New England	-15.57527
9.	New England	-15.57527
10.	New England	-15.57527
11.	New England	-15.57527
12.	New England	-15.57527
13.	New England	-15.57527
14.	New England	-15.57527
15.	New England	-15.57527
16.	New England	-15.57527
17.	New England	-15.57527
18.	New England	-15.57527
19.	New England	-15.57527
20.	New England	-15.57527
21.	New England	-15.57527
22.	New England	-15.57527
23.	New England	-15.57527
24.	New England	-15.57527
25.	New England	-15.57527
67.	New England	-15.57527
68.	Middle Atlantic	-2.395582
69.	Middle Atlantic	-2.395582
70.	Middle Atlantic	-2.395582
71.	Middle Atlantic	-2.395582
72.	Middle Atlantic	-2.395582
73.	Middle Atlantic	-2.395582
74.	Middle Atlantic	-2.395582
75.	Middle Atlantic	-2.395582
76.	Middle Atlantic	-2.395582
77.	Middle Atlantic	-2.395582
78.	Middle Atlantic	-2.395582
79.	Middle Atlantic	-2.395582

Random Intercept Model

```
. xtmixed bush || cendiv:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -11895.132
Iteration 1:   log restricted-likelihood = -11895.132

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =      3054
Group variable: cendiv                          Number of groups   =         9

                                                Obs per group: min =        67
                                                               avg =     339.3
                                                               max =       618

                                                Wald chi2(0)       =         .
Log restricted-likelihood = -11895.132          Prob > chi2        =         .
```

bush	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	58.23352	2.383564	24.43	0.000	53.56182	62.90522

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cendiv: Identity				
sd(_cons)	7.103931	1.825418	4.293139	11.755
sd(Residual)	11.82181	.1514928	11.52859	12.12249

```
LR test vs. linear regression: chibar2(01) =   368.34 Prob >= chibar2 = 0.0000

. predict randint1, reffects

. table cendiv, contents (mean randint0 mean randint1 mean bush)
```

Census division (9)	mean(randint0)	mean(randint1)	mean(bush)
New England	-15.57527	-15.22133	42.38305
Middle Atlantic	-2.395582	-4.124419	54.03296
E North Central	-2.391467	-.4009925	57.82999
W North Central	.0169212	5.126793	63.38329
South Atlantic	5.96793	.8395588	59.07743
E South Central	3.617538	1.77551	60.02254
W South Central	8.68867	7.416742	65.69396
Mountain	3.456139	7.486413	65.79398
Pacific	-1.384879	-2.898272	55.2749

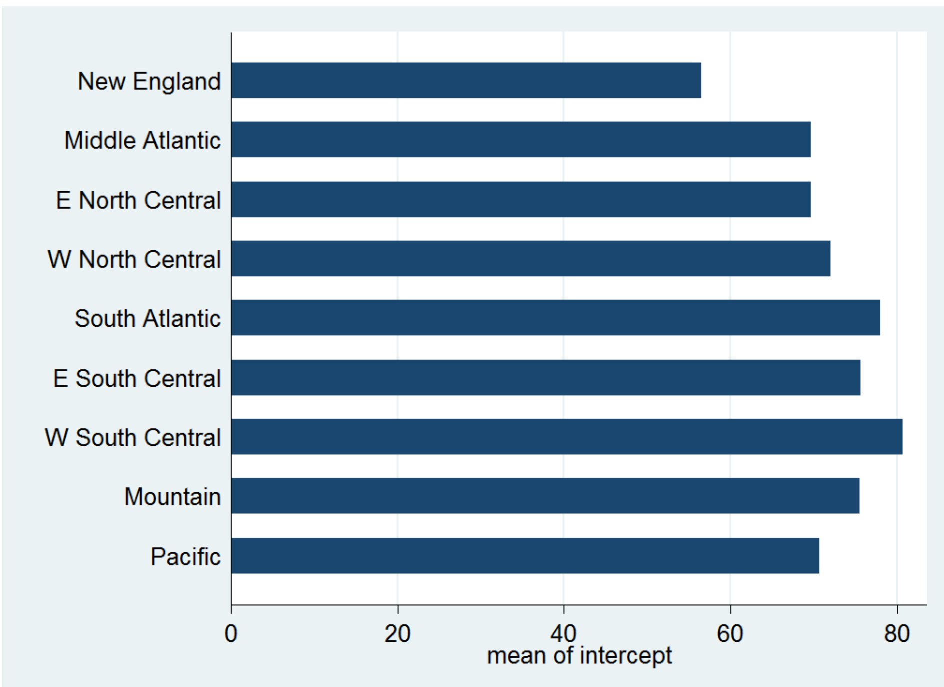
- Can we use the formula from last slide to calculate the BLUP?

- Consider New England:
- σ_b^2=7.103931^2
- σ_e^2=11.82181^2
- n_i=67
- Ave of $Y_{i.}$=42.38305
- u=58.23352

- BLUP of b_i=-15.22133

Random Intercept Model

```
. xtmixed bush logdens minority colled || cendiv:
. predict randint0, reffects
. gen intercept=_b[_cons]+randint0
. graph hbar (mean) intercept, over(cendiv)
```



$$(2.2)\qquad y_{ij} = \beta_0 + \mathbf{u}_{0j} + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} + \varepsilon_{ij}$$

$$= \beta_{0j} + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} + \varepsilon_{ij} \ .$$

- Total effects of intercepts=fixed effects + predicted random effects
- $E(y_{ij} | x_{1,ij}, x_{2,ij}, x_{3,ij}) = \beta_0 + E(u_{0j}) + \text{constant}$, for a given j.
- So the "randint0" gives the difference of the intercepts from the fixed effect intercept (population intercept) for each census division
- Interpretation: at any given level of the 3 predictors (logdens, minority and colled), percentage of voters for Bush is on average 16 points lower in the New England (NE) region than in the West/North/Central region, and 22 percent lower in the NE region than in the South Atlantic region.

Random Regression Model: Intercept+Slope

```
. xtmixed bush logdens minority colled || cendiv: logdens
Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -11298.734
Iteration 1:   log likelihood = -11298.734

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      3041
Group variable: cendiv                          Number of groups   =         9

                                                Obs per group: min =        67
                                                               avg =     337.9
                                                               max =       616

                                                Wald chi2(3)       =    806.25
Log likelihood = -11298.734                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logdens |  -3.310313   1.114965    -2.97   0.003    -5.495605   -1.125021
    minority |  -.3616886   .0130709   -27.67   0.000     -.387307   -.3360702
      colled |  -.1173469   .0360906    -3.25   0.001    -.1880833   -.0466105
       _cons |   70.12095   2.955209    23.73   0.000     64.32885    75.91305
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Independent          |
                sd(logdens)  |   3.113575   .8143897       1.86474    5.198768
                  sd(_cons)  |     8.5913   2.232214      5.162945    14.29619
-----------------------------+------------------------------------------------
                sd(Residual) |   9.825565   .1264176      9.580889    10.07649
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =   538.10   Prob > chi2 = 0.0000

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

- Compare this model with the RIM

$$(2.5)\qquad y_{ij} = \beta_0 + u_{0j} + \beta_1 X_{1,ij} + u_{1j} X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} + \varepsilon_{ij}$$
$$= \beta_{0j} + \beta_{1j} X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} + \varepsilon_{ij} \ .$$

Random Regression Model: Intercept+Slope

- Compare this model with the RIM

```
. quietly xtmixed bush logdens minority colled || cendiv:
. estimates store m1

. quietly xtmixed bush logdens minority colled || cendiv: logdens

. estimates store m2

. lrtest m1 m2

Likelihood-ratio test                                 LR chi2(1)  =     82.11
(Assumption: m1 nested in m2)                         Prob > chi2 =    0.0000

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

- Inclusion of random slope into the model brought significant improvement in it.

Random Regression Model: Intercept+Slope

- Are the random intercept and slope are correlated?

```
. xtmixed bush logdens minority colled || cendiv: logdens,
cov(unstructured)
```

- The random intercept and slope are correlated

- So we include the correlation.

```
Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  -11296.31
Iteration 1:   log likelihood =  -11296.31

Computing standard errors:
Mixed-effects ML regression                     Number of obs      =      3041
Group variable: cendiv                          Number of groups   =         9
                                                Obs per group: min =        67
                                                               avg =     337.9
                                                               max =       616
                                                Wald chi2(3)       =    799.68
Log likelihood =  -11296.31                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logdens |  -3.150009   1.169325    -2.69   0.007    -5.441844   -.8581751
    minority |  -.3611161   .0130977   -27.57   0.000    -.3867872   -.3354451
      colled |  -.1230445   .0361363    -3.41   0.001    -.1938704   -.0522186
       _cons |   69.85194   3.168479    22.05   0.000     63.64184    76.06204
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Unstructured         |
                sd(logdens)  |   3.282749   .8547255      1.970658    5.468447
                  sd(_cons)  |   9.240389   2.402183      5.551459     15.3806
        corr(logdens,_cons)  |   -.675152   .1958924     -.9096869   -.1140963
-----------------------------+------------------------------------------------
               sd(Residual)  |   9.823658   .1263468      9.579118    10.07444
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   542.95   Prob > chi2 = 0.0000

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


. lrtest m1 m2

Likelihood-ratio test                                 LR chi2(1)  =      4.44
(Assumption: m1 nested in m2)                         Prob > chi2 =    0.0352
```

Random Regression Model: Intercept+Slope

- Predict intercepts and slopes within each of the 9 regions

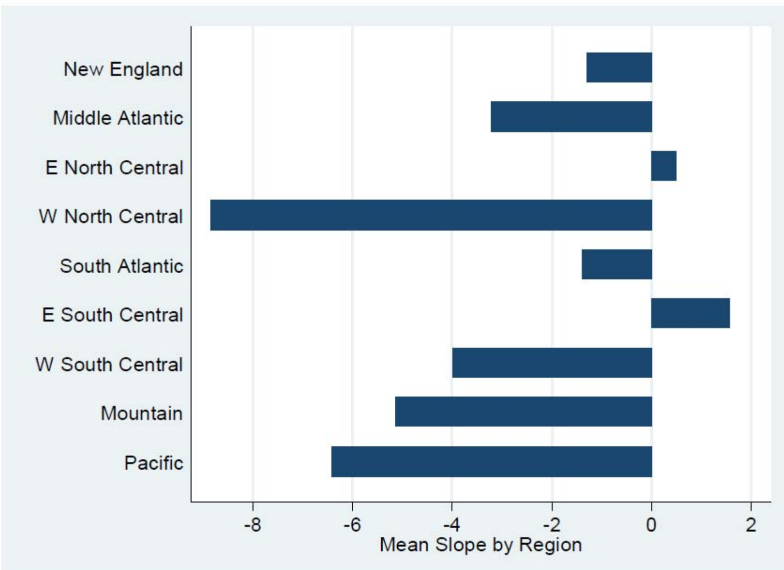
```
. predict randslo1 randint1, reffects

. table cendiv, contents (mean randslo1 mean randint1)
```

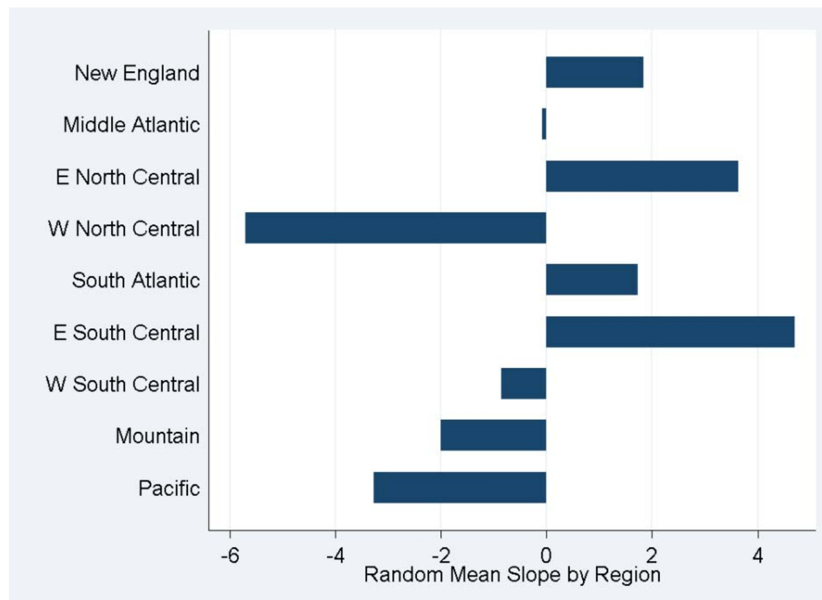
```
----------------------------------------------------
Census division |
(9)             | mean(randslo1)  mean(randint1)
----------------+-----------------------------------
    New England |       1.760172       -18.79535
Middle Atlantic |      -.0611391       -2.198096
E North Central |       3.618166       -9.007434
W North Central |       -5.65562        8.328082
 South Atlantic |       1.738479        2.875959
E South Central |       4.633666       -4.289312
W South Central |      -.8330051         10.8619
       Mountain |      -1.975749        7.232863
        Pacific |      -3.224968        4.991385
----------------------------------------------------
```

Random Regression Model: Intercept+Slope

- For each region, the mean slope is the fixed effect slope + the mean in that region of the pertinent random slopes: combined slope

```
. gen slope1 = randslo1 + _b[logdens]
. graph hbar (mean) slope1, over(cendiv) ytitle ("Mean Slope by
Region")
```





Random Regression Model: Multiple Random Slopes

- **. xtmixed bush logdens minority colled || cendiv: logdens minority colled**

```
Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -11184.804
Iteration 1:   log likelihood = -11184.804

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      3041
Group variable: cendiv                          Number of groups   =         9
                                                Obs per group: min =        67
                                                               avg =     337.9
                                                               max =       616
                                                Wald chi2(3)       =     52.49
Log likelihood = -11184.804                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logdens |  -2.717128   1.373684    -1.98   0.048    -5.409499   -.0247572
    minority |  -.3795605   .0560052    -6.78   0.000    -.4893286   -.2697924
      colled |  -.1707863   .1727742    -0.99   0.323    -.5094175     .167845
       _cons |   70.86653   3.435918    20.63   0.000     64.13225     77.6008
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Independent          |
                sd(logdens) |   3.868421   .9832899      2.350564    6.366421
               sd(minority) |    .153172   .0439569      .0872777    .2688161
                 sd(colled) |   .5032414   .1241234       .310334    .8160625
                  sd(_cons) |   10.01157   2.547813      6.079707    16.48625
-----------------------------+------------------------------------------------
               sd(Residual) |   9.375994   .1209753      9.141859    9.616124
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(4) =   765.96   Prob > chi2 = 0.0000

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

Random Regression Model: Multiple Random Slopes

```
. estimates store full

. quietly xtmixed bush logdens minority colled || cendiv: logdens minority

. estimates store nocolled

. lrtest nocolled full

Likelihood-ratio test                                 LR chi2(1)  =    197.33
(Assumption: nocolled nested in full)                 Prob > chi2 =    0.0000
```

- This shows that the full model fits significantly better
- And thus there's no evidence in the data set that would warrant excluding the random effect associated with the variable *colled*.

Fixed, Random and Total Effects

- Random effects help better fit models since they represent *heterogeneity* in the data
- Oftentimes, random effects are considered not of substantive interest
- In some empirical settings, however, one may be interested in the random effects themselves.
- The latter usually happens when of concern are the total effects, which are defined as the sum of random and fixed effects for a given predictor.

```
. quietly xtmixed bush logdens minority colled || cendiv: logdens minority colled

. predict relogdens reminority recolled re_cons, reffects

. d relogdens-re_cons
                 storage  display     value
variable name      type   format      label      variable label
-------------------------------------------------------------------------------
relogdens          float  %9.0g                  BLUP r.e. for cendiv: logdens
reminority         float  %9.0g                  BLUP r.e. for cendiv: minority
recolled           float  %9.0g                  BLUP r.e. for cendiv: colled
re_cons            float  %9.0g                  BLUP r.e. for cendiv: _cons
```

Fixed, Random and Total Effects

- Computing total effects, for example *colled*:

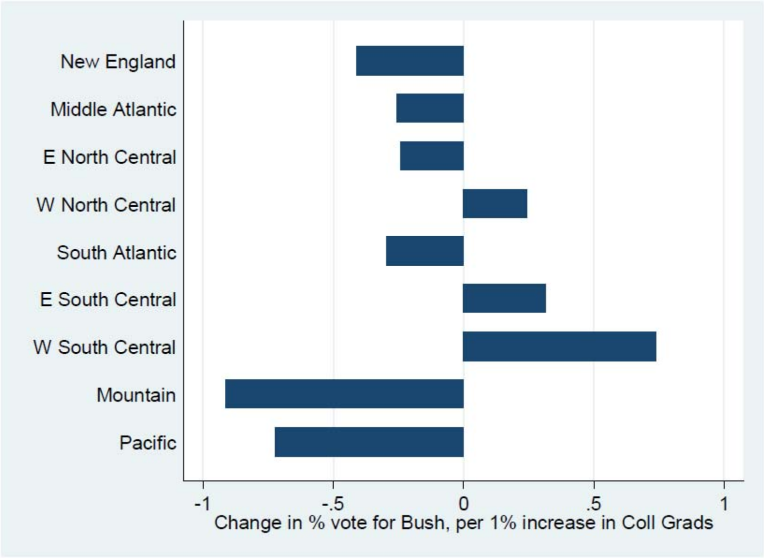
```
. gen tecolled = recolled + _b[colled]

. label variable tecolled "Random Plus Fixed Effect of Coll_Ed"

. d tecolled
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
tecolled        float  %9.0g                  Random Plus Fixed Effect of Coll_Ed
```

- Means for the random effects and total effects of *colled* per geographic region

```
. table cendiv, contents(mean recolled mean tecolled)
----------------------------------------------
Census division |
(9)             | mean(recolled)  mean(tecolled)
----------------+-----------------------------
    New England |      -.2406214      -.4122202
Middle Atlantic |      -.0856313      -.2572301
E North Central |      -.0695729      -.2411717
W North Central |       .4153244       .2437256
 South Atlantic |       -.125596      -.2971948
E South Central |       .4871398        .315541
W South Central |       .9101545       .7385557
       Mountain |      -.7414296      -.9130284
        Pacific |      -.5497676      -.7213664
----------------------------------------------
```

Fixed, Random and Total Effects

```
. graph hbar (mean) tecolled, over(cendiv) ytitle ("Change in %
vote for Bush, per 1% increase in Coll Grads")
```



- A model with a random effect for *colled* was better than the one without.
- In the fixed effect model, the various random effects across region will be averaged out into a single fixed effect.
- The total effects of *colled* range from substantially negative to substantially positive, will be lost.
- In fact, the fixed effect estimate for *colled* is -0.17 if using conventional regression, which would be quite misleading.

Nested Levels

- Is it meaning for to consider counties nested perhaps also in states?
- Consider variable *colled* nesting within states.

```
. xtmixed bush logdens minority colled || cendiv: logdens
minority colled || state: colled
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Independent          |
                sd(logdens)  |   2.703845   .7727275      1.544252    4.734186
               sd(minority)  |   .1465435   .0428326      .0826365    .2598728
                 sd(colled)  |   .3683903   .0962733       .220729    .6148326
                  sd(_cons)  |   8.416873   2.417524      4.793643    14.77869
-----------------------------+------------------------------------------------

state: Independent           |
                 sd(colled)  |   .1305727    .039009      .0727032    .2345047
                  sd(_cons)  |   5.883451   .7431715      4.593166    7.536196
-----------------------------+------------------------------------------------
               sd(Residual)  |   7.863302   .1027691      7.664436    8.067328
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(6) =  1695.92   Prob > chi2 = 0.0000

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

- All the random effects are significant

Nested Levels

```
. estimates store state
. lrtest full state
Likelihood-ratio test                                 LR chi2(2)  =    929.97
(Assumption: full nested in state)                    Prob > chi2 =    0.0000
```

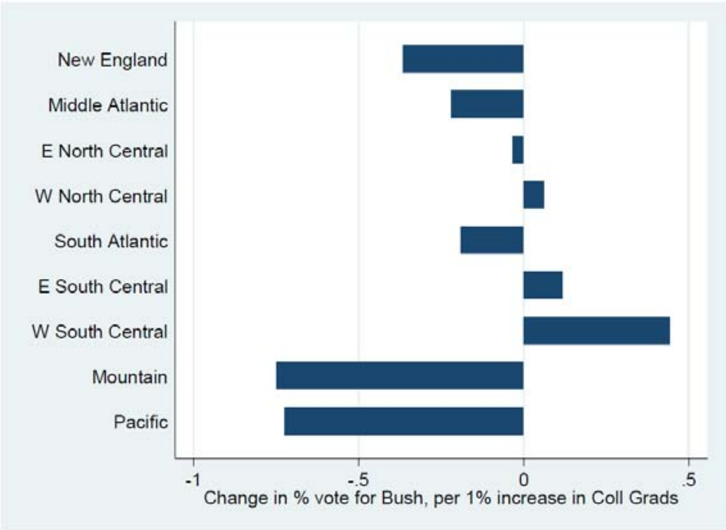
- The “state” model fits significantly better

```
. predict re*, reffects
re1             float   %9.0g               BLUP r.e. for cendiv: logdens
re2             float   %9.0g               BLUP r.e. for cendiv: minority
re3             float   %9.0g               BLUP r.e. for cendiv: colled
re4             float   %9.0g               BLUP r.e. for cendiv: _cons
re5             float   %9.0g               BLUP r.e. for state: colled
re6             float   %9.0g               BLUP r.e. for state: _cons
```

- Total effects for *colled*: tecolled2=re3+re5+_b[colled]

```
. graph hbar (mean) tecolled2, over(cendiv) ytitle ("Change in % vote for
Bush, per 1% increase in Coll Grads")
```

- The central regions have now somewhat closer to 0 (in absolute value)
- That for mountain regions marginally less negative
- Accounting for the intermediate level of nesting is associated with less pronounced effects of *colled* upon the response variable in those regions



Binary Responses

- In social and behavioral research, there is the case that examining relationships between discrete response and explanatory variables.

```
. use c:\stata\data\gss.dta, clear
(General Social Survey 2006)
```

```
Contains data from c:\stata\data\GSS_SwS1.dta
  obs:         1,595                          General Social Survey 2006
 vars:            17                          1 May 2011 13:45
 size:        54,230 (99.5% of memory free)
-----------------------------------------------------------------------------
storage  display     value
variable name    type   format      label      variable label
-------------------------------------------------------------------------------
id               int    %9.0g                  Case number (Statistics
w/Stata)
finalwt          float  %9.0g                  weight variable -- wtssall
cendiv           byte   %12.0g      cendiv     Census division (9)
age              byte   %8.0g       age        Age in years
educ             byte   %8.0g       educ       Highest year of schooling
completed
gender           byte   %9.0g       gender     Respondent gender
income           float  %9.0g       income     Family income in constant
dollars
loginc           float  %9.0g                  log10(family income, +1)
logsize          float  %9.0g                  log10(size of place in
thousands, +1)
married          byte   %9.0g       married    Married or unmarried?
minority         byte   %11.0g      minority   Minority status
politics         byte   %20.0g      politics   Think of self as liberal or
conservative?
bush             byte   %11.0g      bush       Voted for GW Bush in 2004?
grass            byte   %9.0g       grass      Should marijuana be made legal?
gunlaw           byte   %8.0g       gunlaw     Favor or oppose gun permits?
postlife         byte   %8.0g       postlife   Believe in life after death?
enviro           byte   %11.0g      enviro     Govt spending on environment?
-------------------------------------------------------------------------------
Sorted by:  id
```

Binary Responses

- Overall distributed votes

```
. tab bush
Voted for |
 GW Bush in |
      2004? |      Freq.     Percent        Cum.
------------+-----------------------------------
Kerry/Nader |        776       48.65       48.65
       Bush |        819       51.35      100.00
------------+-----------------------------------
      Total |      1,595      100.00
```

- We use *xtmelogit*
- Consider the log-odds of voting for Bush here.

Random Intercept Model

- Random intercept model

  (3.1) $ln[p_{ij}/(1- p_{ij})] = \beta_0 + \mathbf{u}_{0j} + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij}$.

- There is no error term: this is a model stated already in terms of the expectation of ‘event’.

```
. xtmelogit bush logsize minority educ || cendiv:
```

- The last line suggests that inclusion of a random intercept improves the model over the fixed effects only model (conventional logistic regression)
- We don’t have a residual random effect

```
Refining starting values:

Iteration 0:   log likelihood = -1008.5032  (not concave)
Iteration 1:   log likelihood = -1001.8839
Iteration 2:   log likelihood = -1001.4243

Performing gradient-based optimization:

Iteration 0:   log likelihood = -1001.4243
Iteration 1:   log likelihood = -1001.4226
Iteration 2:   log likelihood = -1001.4226

Mixed-effects logistic regression               Number of obs      =      1595
Group variable: cendiv                          Number of groups   =         9
                                                Obs per group: min =        62
                                                               avg =     177.2
                                                               max =       336

Integration points =   7                        Wald chi2(3)       =    156.51
Log likelihood = -1001.4226                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logsize |  -.2456232   .0723751    -3.39   0.001    -.3874758   -.1037707
    minority |  -1.988091   .1693794   -11.74   0.000    -2.320069   -1.656114
        educ |  -.0435621    .020059    -2.17   0.030     -.082877   -.0042472
       _cons |   1.331863   .3044779     4.37   0.000     .7350978    1.928629
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Identity             |
                   sd(_cons) |   .2631985   .0937197      .1309747    .5289069
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =     9.78 Prob>=chibar2 = 0.0009
```

Random Regression Model

- Will it be better to include a random slope for *logsize*?

(3.2) $$ln[p_{ij}/(1-p_{ij})] = (\beta_0+u_{0j}) + (\beta_1+u_{1j})X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} .$$

```
. xtmelogit bush logsize minority educ
    || cendiv: logsize, covariance(unstructured)
. estimates store a
```

- The random intercepts are perhaps not significant as an effect
- They are not really related to random slopes

```
Refining starting values:

Iteration 0:   log likelihood = -1006.9894  (not concave)
Iteration 1:   log likelihood =   -995.258
Iteration 2:   log likelihood = -993.39684

Performing gradient-based optimization:

Iteration 0:   log likelihood = -993.39684  (not concave)
Iteration 1:   log likelihood = -993.29552
Iteration 2:   log likelihood = -993.19391
Iteration 3:   log likelihood = -993.18826
Iteration 4:   log likelihood = -993.18808
Iteration 5:   log likelihood = -993.18808

Mixed-effects logistic regression               Number of obs      =      1595
Group variable: cendiv                          Number of groups   =         9
                                                Obs per group: min =        62
                                                               avg =     177.2
                                                               max =       336
Integration points =   7                        Wald chi2(3)       =    136.94
Log likelihood = -993.18808                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logsize |  -.3067214   .1351804    -2.27   0.023      -.57167   -.0417728
    minority |  -1.929399   .1705009   -11.32   0.000    -2.263574   -1.595223
        educ |   -.041411   .0201692    -2.05   0.040    -.0809419     -.00188
       _cons |   1.361903   .3014102     4.52   0.000     .7711501    1.952656
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Unstructured         |
                sd(logsize)  |   .3249856   .1153993      .1620356    .6518051
                  sd(_cons)  |   .2166197   .1880638      .0395099    1.187653
        corr(logsize,_cons)  |  -.8708498   .2677054     -.9982073    .6831364
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(3) =    26.25   Prob > chi2 = 0.0000
```

Random Slope Only Model

- Dispense with the random intercepts

$$(3.3) \qquad ln[p_{ij}/(1-p_{ij})] = \beta_0 + (\beta_1+u_{1j})X_{1,ij} + \beta_2 X_{2,ij} + \beta_3 X_{3,ij} \,.$$

```
. xtmelogit bush logsize minority educ
    || cendiv: logsize, nocons
. estimates store b
. lrtest b a
```

- The model (3.3), the random-slope-only (RSO) is not significantly worse than the random regression (RR) Model (3.2)
- The RSO model is preferable

```
Mixed-effects logistic regression               Number of obs      =      1595
Group variable: cendiv                          Number of groups   =         9
                                                Obs per group: min =        62
                                                               avg =     177.2
                                                               max =       336

Integration points =   7                        Wald chi2(3)       =    150.81
Log likelihood = -994.02453                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        bush |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logsize |  -.2988801   .1097208    -2.72   0.006    -.5139289   -.0838312
    minority |  -1.969727   .1666701   -11.82   0.000    -2.296395    -1.64306
        educ |  -.0414891   .0201459    -2.06   0.039    -.0809745   -.0020038
       _cons |   1.361989   .2909969     4.68   0.000     .7916459    1.932333
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cendiv: Identity             |
                sd(logsize) |   .2340086    .071626       .128438    .4263536
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =    24.58 Prob>=chibar2 = 0.0000
```

```
Likelihood-ratio test                                 LR chi2(2)  =      1.67
(Assumption: b nested in a)                           Prob > chi2 =    0.4332
```

Model Choice

- Compare RSO (random-slope-only) with RIM (random intercept model): they are not nested, can not use likelihood ratio test
- Akaike's Information Criterion (AIC)

(3.4) AIC = -2*loglikelihood + 2*#parameters.

- Log-likelihood values are from the Stata pertinent model outputs
- Number of parameters can be counted:
  - RIM model (3.1) has 5 parameters (4 fixed and 1 random)
  - RSO model (3.3) has 5 parameters (4 fixed and 1 random)
- gen AIC_RIM=-2*(-1001.4226)+2*5
- gen AIC_RSO=-2*(-994.02453)+2*5
- AIC_RIM=2012.845
- AIC_RSO=1998.049

- RSO has a lower AIC, so it is preferable to the RIM
- The decision suggests that the regional variability in the pattern of voting for Bush is best modeled as variations in the effect of logsize (an indicator of urbanization or unban-ness) upon the odds of such vote, after accounting for minority and educational levels

Day 2 Conclusion

- This part was devoted to intermediate and some advanced topics within multilevel modeling
- Issues pertaining to the choice between single level statistical models and two level models, as well as between two-level and three-level models, are of special relevance in empirical behavioral, social and biomedical research.
- These issues were attended to from a model fitting perspective, in the context of mixed models
- Mixed modeling issues pertaining to random effect and total effect estimation were also discussed

The End