Day 3: Random-intercept models

Introduction to Multilevel Models
EUI Short Course 22–27 May, 2011
Prof. Kenneth Benoit

May 25, 2011

Reliability

- Each overall error consists of the two error components ζ_j and ϵ_{ij}:

$$\xi_{ij} \equiv \zeta_j + \epsilon_{ij}$$

- The error components are independent, so it can be shown that the total variance is the sum of the between-subject and within-subject variances:

$$\begin{aligned} \mathrm{Var}(y_{ij}) &= \mathrm{Var}(\beta + \zeta_j + \epsilon_{ij}) \\ &= \mathrm{Var}(\beta) + \mathrm{Var}(\zeta_j + \epsilon_{ij}) \\ &= (0) + \psi + \theta \end{aligned}$$

- We can express the proportion of the total variance that is between subjects as:

$$\rho = \frac{\mathrm{Var}(\zeta_j)}{\mathrm{Var}(y_{ij})} = \frac{\psi}{\psi + \theta}$$

- ρ can also be thought of as reliability of measurements for the same subjects j. It is also analogous to R^2 in that it represents the proportion of the total variance that is "explained" by subjects

Intraclass correlation

- ρ can also be interpreted as the marginal correlation between measurements on two occasions i and i' for the same subject
- So ρ aso represents within-cluster correlation
- We estimate the ICC using parameter estimates for ψ and θ:

$$\hat{\rho} = \frac{\hat{\psi}}{\hat{\psi} + \hat{\theta}}$$

- Can contrast the ICC with Pearson's r as:

$$r = \frac{\frac{1}{J-1}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{i\cdot})(y_{i'j} - \bar{y}_{i'\cdot})}{s_{y_i} s_{y_{i'}}}$$

- Pearson's r provides a measure of *relative agreement*, based on deviations of each i from their respective means
- ICC provides a measure of absolute agreement – and is therefore affected by linear transformations of the measurements

Fixed v. random effects

- The model we have called "random intercepts" is also a one-way random-effects ANOVA model, written as:

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij} \quad \epsilon_{ij}|\zeta_j \sim N(0, \theta) \quad \zeta_j \sim N(0, \psi)$$

  where ζ_j is a random intercept

- An alternative is the one-way fixed-effects ANOVA model:

$$y_{ij} = \beta + \alpha_j + \epsilon_{ij} \quad \epsilon_{ij} \sim N(0, \theta) \quad \sum_{j=1}^{J} \alpha_j = 0$$

  where α_j is an unknown cluster-specific parameter

Fixed v. random effects: which to choose?

- Question: are we concerned about the *population* of clusters, or instead the particular clusters in the *sample*?
  - If we are interested in the variance ψ for the population of clusters, or inference for β when clusters and units are sampled from respective population, then use a random effects approach
  - If we are interested in the sample-specific "effects" α_j and inferences regarding β only when units (and not clusters) are considered randomly sampled, then use a fixed effects approach
- The choice mostly affects the standard error of $\hat{\beta}$ but also can affect $\hat{\beta}$ itself

Stata example using HSB data (xtreg)

```
. xtreg mathach, i(schoolid) mle nolog

Random-effects ML regression                    Number of obs      =      7185
Group variable: schoolid                        Number of groups   =       160

Random effects u_i ~ Gaussian                   Obs per group: min =        14
                                                               avg =      44.9
                                                               max =        67

                                                Wald chi2(0)       =      0.00
Log likelihood  = -23557.905                    Prob > chi2        =         .

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   12.63707   .2436216    51.87   0.000     12.15958    13.11456
-------------+----------------------------------------------------------------
    /sigma_u |   2.924631   .1826925                      2.587612    3.305544
    /sigma_e |   6.256868   .0527937                      6.154245    6.361202
         rho |   .1793109   .0185934                      .1452078    .2180551
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=  983.92 Prob>=chibar2 = 0.000
```

_cons overall population mean $\hat{\beta}$: 12.63707

/sigma_u between-subject standard deviation $\sqrt{\hat{\psi}}$ of ζ_j: 2.924631

/sigma_e within-subject SD $\sqrt{\hat{\theta}}$: 6.256868

rho intraclass correlation ρ, also computed as:

$$\hat{\rho} = \frac{\hat{\psi}}{\hat{\psi} + \hat{\theta}} = \frac{2.92^2}{6.26^2 + 2.92^2} = .18$$

Stata example using HSB data (xtmixed)

```
. xtmixed mathach || schoolid:, mle nolog

Mixed-effects ML regression                     Number of obs      =      7185
Group variable: schoolid                        Number of groups   =       160

                                                Obs per group: min =        14
                                                               avg =      44.9
                                                               max =        67


                                                Wald chi2(0)       =         .
Log likelihood = -23557.905                     Prob > chi2        =         .

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   12.63707   .2436173    51.87   0.000     12.15959    13.11455
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
schoolid: Identity           |
                   sd(_cons) |   2.924632   .1826955      2.587608    3.305552
-----------------------------+------------------------------------------------
                sd(Residual) |   6.256868   .0527937      6.154245    6.361202
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   983.92 Prob >= chibar2 = 0.0000
```

Day 3 focus: random intercept models

- One way to look at this is that we are reparameterizing β from the models introduced earlier, by adding explanatory variables X (covariates)
- This allows us to model directly the distinction between the effects of X that are *within-cluster* from those that are *between-cluster*
- Another way to look at it: we are extending the CLRM by adding random intercepts ζ_j
- We also discuss the measures of variation explained by X, and the coefficients of determination (R^2 equivalents) for random intercept models

Sample variances at different levels

- *Overall* standard deviation measured in deviations from overall mean $\bar{x}_{..}$:

$$s_{xO}^2 = \frac{1}{N-1} \sum_{j=1}^{J} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2$$

- *Between* standard deviation measured in deviations of cluster (level 2) means $\bar{x}_{.j}$ from overall means:

$$s_{xB}^2 = \frac{1}{J-1} \sum_{j=1}^{J} (\bar{x}_{.j} - \bar{x}_{..})^2$$

- *Within* standard deviation measured in deviations of observations (level 1) x_{ij} from the cluster means:

$$s_{xW}^2 = \frac{1}{N-1} \sum_{j=1}^{J} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2$$

Example using HSB dataset

```
. use http://www.stata-press.com/data/mlmus2/hsb.dta, clear

. xtsum mathach ses sector, i(schoolid)

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
mathach  overall |  12.74785   6.878246     -2.832     24.993 |     N =    7185
         between |             3.117651   4.239781   19.71914 |     n =     160
         within  |             6.186706  -6.926784   30.71674 | T-bar = 44.9063
                 |                                            |
ses      overall |  .0001434   .7793552     -3.758      2.692 |     N =    7185
         between |             .4139706  -1.193946   .8249825 |     n =     160
         within  |              .660588  -3.650597   2.856222 | T-bar = 44.9063
                 |                                            |
sector   overall |  .4931106   .4999873          0          1 |     N =    7185
         between |             .4976359          0          1 |     n =     160
         within  |                    0   .4931106   .4931106 | T-bar = 44.9063
```

- `T-bar` is the mean number of students per school (measured for each variable – same here because no missing)
- No within variance for `sector` because this is a level-2 only variable

Specification for random-intercept model

- Standard CLRM model with covariates:

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \xi_{ij}$$

- Error term: $\xi_{ij} \equiv \zeta_j + \epsilon_{ij}$
- Linear random-intercept model with covariates:

$$\begin{aligned} y_{ij} &= \beta_1 + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \zeta_j + \epsilon_{ij} \\ &= (\beta_1 + \zeta_j) + \beta_2 x_{2ij} + \cdots + \beta_p x_{pij} + \epsilon_{ij} \end{aligned}$$

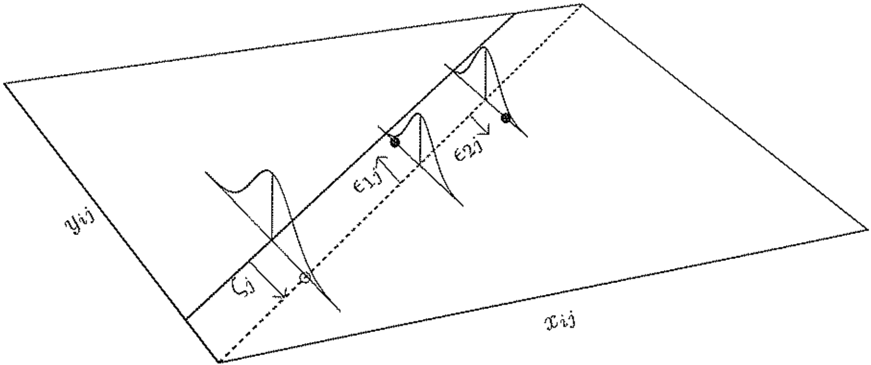
- Error assumptions (exogeneity):

$$\zeta_j | \mathbf{x}_{ij} \sim N(0, \psi)$$

and

$$\epsilon_{ij} | \mathbf{x}_{ij}, \zeta_j \sim N(0, \theta)$$

Illustration of random intercept model for one J group



Smoking and birth weight example from text: `xtreg`

```
. xtreg birwt smoke male mage hsgrad somecoll collgrad married black kessner2
> kessner3 novisit pretri2 pretri3, i(momid) mle

Random-effects ML regression                    Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3

                                                LR chi2(13)        =    659.47
Log likelihood  = -65145.752                    Prob > chi2        =    0.0000
```

birwt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smoke	-218.3289	18.20988	-11.99	0.000	-254.0196	-182.6382
male	120.9375	9.558721	12.65	0.000	102.2027	139.6722
mage	8.100548	1.347266	6.01	0.000	5.459956	10.74114
hsgrad	56.84715	25.03538	2.27	0.023	7.778705	105.9156
somecoll	80.68607	27.30914	2.95	0.003	27.16115	134.211
collgrad	90.83273	27.99598	3.24	0.001	35.96162	145.7038
married	49.9202	25.50319	1.96	0.050	-.0651368	99.90554
black	-211.4138	28.27818	-7.48	0.000	-266.838	-155.9896
kessner2	-92.91883	19.92624	-4.66	0.000	-131.9736	-53.86411
kessner3	-150.8759	40.83414	-3.69	0.000	-230.9093	-70.84246
novisit	-30.03035	65.69213	-0.46	0.648	-158.7846	98.72387
pretri2	92.8579	23.19258	4.00	0.000	47.40127	138.3145
pretri3	178.7295	51.64145	3.46	0.001	77.51416	279.9449
_cons	3117.191	40.97597	76.07	0.000	3036.88	3197.503
/sigma_u	338.7674	6.296444			326.6487	351.3358
/sigma_e	370.6654	3.867707			363.1618	378.324
rho	.4551282	.0119411			.4318152	.4785967

```
Likelihood-ratio test of sigma_u=0: chibar2(01)= 1108.77 Prob>=chibar2 = 0.000
```

Smoking and birth weight example from text: xtmixed

```
. xtmixed birwt smoke male mage hsgrad somecoll collgrad married black
> kessner2 kessner3 novisit pretri2 pretri3, || momid:, mle

Mixed-effects ML regression                     Number of obs      =      8604
Group variable: momid                           Number of groups   =      3978

                                                Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3

                                                Wald chi2(13)      =    693.74
Log likelihood = -65145.752                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       birwt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -218.3286   18.15946   -12.02   0.000    -253.9205   -182.7368
        male |   120.9375   9.558003    12.65   0.000     102.2042    139.6708
        mage |   8.100566   1.344573     6.02   0.000     5.465251    10.73588
      hsgrad |   56.84716   25.03543     2.27   0.023     7.778611    105.9157
    somecoll |   80.68605   27.30906     2.95   0.003     27.16127    134.2108
    collgrad |   90.83268   27.99498     3.24   0.001     35.96354    145.7018
     married |   49.92022   25.50309     1.96   0.050    -.0649248    99.90537
       black |  -211.4138   28.27764    -7.48   0.000    -266.8369   -155.9906
    kessner2 |  -92.91882   19.92617    -4.66   0.000    -131.9734   -53.86424
    kessner3 |  -150.8758   40.83027    -3.70   0.000    -230.9017   -70.84992
     novisit |   -30.0303   65.69165    -0.46   0.648    -158.7836    98.72298
     pretri2 |   92.85784   23.19067     4.00   0.000     47.40497    138.3107
     pretri3 |   178.7294   51.63677     3.46   0.001      77.5232    279.9356
       _cons |   3117.191   40.88824    76.24   0.000     3037.051     3197.33
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
momid: Identity              |
                   sd(_cons) |   338.7686   6.296449      326.6499     351.337
-----------------------------+------------------------------------------------
                sd(Residual) |   370.6648   3.867695      363.1613    378.3234
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  1108.77 Prob >= chibar2 = 0.0000
```

Smoking and birth weight example from text: results compared

Table 3.1: Maximum likelihood estimates for smoking data (in grams)

	Full model		Null model		Level-2 cov.	
	Est	(SE)	Est	(SE)	Est	(SE)
Fixed part						
β_1 [_cons]	3,117	(41)	3,468	(7)	3,216	(26)
β_2 [smoke]	−218	(18)				
β_3 [male]	121	(10)				
β_4 [mage]	8	(1)				
β_5 [hsgrad]	57	(25)			131	(25)
β_6 [somecoll]	81	(27)			181	(27)
β_7 [collgrad]	91	(28)			233	(26)
β_8 [married]	50	(26)			115	(25)
β_9 [black]	−211	(28)			−201	(29)
β_{10} [kessner2]	−93	(20)				
β_{11} [kessner3]	−151	(41)				
β_{12} [novisit]	−30	(66)				
β_{13} [pretri2]	93	(23)				
β_{14} [pretri3]	179	(52)				
Random part						
$\sqrt{\psi}$	339		368		348	
$\sqrt{\theta}$	371		378		378	
Derived estimates						
R^2	0.09		0.00		0.05	
ρ	0.46		0.49		0.46	

Measures of fit for random intercept models

- Consider a null model without covariates, compared to a model with covariates
- The R^2 with OLS is the *proportional reduction in variance* from using the covariates model versus the null model:

$$R^2 = \frac{\hat{\sigma}_0^2 - \hat{\sigma}_1^2}{\hat{\sigma}_0^2}$$

- Snijders and Bosker (1999) propose a similar measures for the linear random-intercept model:

$$R^2 = \frac{\hat{\psi}_0 + \hat{\theta}_0 - (\hat{\psi}_0 + \hat{\theta}_0)}{\hat{\psi}_0 + \hat{\theta}_0}$$

- From the smoking and birthweight example (see earlier table):

$$\widehat{\psi}_1 + \widehat{\theta}_1 = 338.7686^2 + 370.6648^2 = 252156.56$$

It follows that

$$R^2 = \frac{278260.43 - 252156.56}{278260.43} = 0.09$$