**Lab session I Wednesday – Chp.13 – Crossed Random Effects**

-----------------------------------------------------------------------------------------------------

name: <unnamed>

log: J:\Multilevel TA\Chp13\_Crossed\_Random\_Effects.log

log type: text

opened on: 16 Apr 2012, 23:29:26

. \*\*chap13.do\*\*

. \*\*Crossed random effects\*\*

. \*The dataset from the book is not available, so we are using a dataset from

. \*Rabe-Hesketh and Skrondal (2008) Multilevel and Longitudinal Modeling Using Stata, Second Edition\*

. \*See also chapter 8.4 for this particular example\*

.

. clear all

. set more off

.

. \*We get the data online

.

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

.

. /\*The dependent var is attainment score at age 16 for pupils that attended various combinations

> of primary and secondary schools.

> pid - primary school identifier

> sid - secondary school identifier

> vrq - verbal reasoning score

> sex - sex \*/

.

. sum pid sid vrq sex

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

pid | 3435 70.7377 45.02572 1 148

sid | 3435 10.21951 5.55694 1 19

vrq | 3435 97.80437 13.29291 70 140

sex | 3435 .4937409 .5000336 0 1

.

. center vrq

.

.

. \*We explore the special nesting structure first.

. \*The tag() function lets us define unique primary/secondary school combinations:

.

. egen pick\_comb=tag(pid sid)

.

. \*We count the unique sid combinations for each primary school.

. \*To how many different secondary schools did a primary school sent the pupils?

.

. egen numsid=total(pick\_comb), by(pid)

.

. sort pid sid

. list pid sid numsid if pick\_comb & pid <10 in 1/100

+--------------------+

| pid sid numsid |

|--------------------|

2. | 1 1 3 |

42. | 1 9 3 |

54. | 1 18 3 |

56. | 2 7 1 |

63. | 3 5 1 |

|--------------------|

65. | 4 6 2 |

68. | 4 9 2 |

+--------------------+

.

. \*Kids who went to primary school 1 ended up in 3 different secondary schools: 1, 9, 18.

.

. \*Now we want to know for each primary school, up to how many secondary schools were sent to.

.

. egen pick\_pid=tag(pid)

.

. tab numsid if pick\_pid

numsid | Freq. Percent Cum.

------------+-----------------------------------

1 | 57 38.51 38.51

2 | 50 33.78 72.30

3 | 26 17.57 89.86

4 | 10 6.76 96.62

5 | 2 1.35 97.97

6 | 3 2.03 100.00

------------+-----------------------------------

Total | 148 100.00

.

. \*90% of primary schools send the kids to 3 different secondary schools.

.

. \*First we try a normal 2-level specification (nested in primary schools)

. \*The trick is to pretend that there exists a new level, in which all observations are nested. (\_all

> )

.

. xtmixed attain || pid: , mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -8585.9769

Iteration 1: log likelihood = -8585.9769

Computing standard errors:

Mixed-effects ML regression Number of obs = 3435

Group variable: pid Number of groups = 148

Obs per group: min = 1

avg = 23.2

max = 72

Wald chi2(0) = .

Log likelihood = -8585.9769 Prob > chi2 = .

------------------------------------------------------------------------------

attain | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

\_cons | 5.616488 .110596 50.78 0.000 5.399724 5.833252

------------------------------------------------------------------------------

------------------------------------------------------------------------------

Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]

-----------------------------+------------------------------------------------

pid: Identity |

var(\_cons) | 1.21634 .2039156 .8756968 1.689493

-----------------------------+------------------------------------------------

var(Residual) | 8.204201 .2020808 7.817539 8.609989

------------------------------------------------------------------------------

LR test vs. linear regression: chibar2(01) = 255.31 Prob >= chibar2 = 0.0000

. est store mod1

.

. \*Crossed effects

. xtmixed attain || \_all: R.sid || pid:, mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -8574.5655

Iteration 1: log likelihood = -8574.5655

Computing standard errors:

Mixed-effects ML regression Number of obs = 3435

-----------------------------------------------------------

| No. of Observations per Group

Group Variable | Groups Minimum Average Maximum

----------------+------------------------------------------

\_all | 1 3435 3435.0 3435

pid | 148 1 23.2 72

-----------------------------------------------------------

Wald chi2(0) = .

Log likelihood = -8574.5655 Prob > chi2 = .

------------------------------------------------------------------------------

attain | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

\_cons | 5.504009 .1749325 31.46 0.000 5.161148 5.846871

------------------------------------------------------------------------------

------------------------------------------------------------------------------

Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]

-----------------------------+------------------------------------------------

\_all: Identity |

var(R.sid) | .3481667 .1618119 .1400193 .8657381

-----------------------------+------------------------------------------------

pid: Identity |

var(\_cons) | 1.124362 .2059384 .7852353 1.60995

-----------------------------+------------------------------------------------

var(Residual) | 8.111477 .2004789 7.72791 8.514081

------------------------------------------------------------------------------

LR test vs. linear regression: chi2(2) = 278.13 Prob > chi2 = 0.0000

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

. est store mod2

.

. esttab mod1 mod2, se wide nostar ///

> transform(ln\*: exp(2\*@) exp(2\*@)) ///

> eqlabels(, none)

----------------------------------------------------------------

(1) (2)

attain attain

----------------------------------------------------------------

attain 5.616 (0.111) 5.504 (0.175)

lns1\_1\_1 1.216 (0.102) 0.348 (0.0809)

lnsig\_e 8.204 (0.101) 8.111 (0.100)

lns2\_1\_1 1.124 (0.103)

----------------------------------------------------------------

N 3435 3435

----------------------------------------------------------------

Standard errors in parentheses

.

. \*Very few changes. A little bit of variation is now distributed to secondary schools.

.

. \*Different ICC for different primary/secondary school combination.

.

. \*Same primary, but different secondary schools:

. dis 1.12/(.35+1.12+8.11)

.11691023

. \*.11691023

.

. \*Same secondary, but not the same primary school:

. dis .35/(.35+1.12+8.11)

.03653445

. \*.03653445

.

. \*Same primary AND secondary school:

. dis (.35+1.12)/(.35+1.12+8.11)

.15344468

. \*.15344468

.

.

.

. \*Crossed effects, fixed explanatory variables. Verbal aptitude and gender.

. xtmixed attain c\_vrq sex || \_all: R.sid || pid:, mle var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -7421.5244

Iteration 1: log likelihood = -7421.482

Iteration 2: log likelihood = -7421.482

Computing standard errors:

Mixed-effects ML regression Number of obs = 3435

-----------------------------------------------------------

| No. of Observations per Group

Group Variable | Groups Minimum Average Maximum

----------------+------------------------------------------

\_all | 1 3435 3435.0 3435

pid | 148 1 23.2 72

-----------------------------------------------------------

Wald chi2(2) = 3360.11

Log likelihood = -7421.482 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

attain | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

c\_vrq | .1596649 .0027756 57.52 0.000 .1542248 .1651051

sex | .1158734 .0714428 1.62 0.105 -.0241519 .2558988

\_cons | 5.570574 .0741593 75.12 0.000 5.425225 5.715924

------------------------------------------------------------------------------

------------------------------------------------------------------------------

Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]

-----------------------------+------------------------------------------------

\_all: Identity |

var(R.sid) | .0110688 .0222705 .0002145 .5711175

-----------------------------+------------------------------------------------

pid: Identity |

var(\_cons) | .2735176 .0610716 .1765742 .4236852

-----------------------------+------------------------------------------------

var(Residual) | 4.250266 .1048371 4.049677 4.460791

------------------------------------------------------------------------------

LR test vs. linear regression: chi2(2) = 87.72 Prob > chi2 = 0.0000

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

.

.

. \*Let's try to include a random slope (c\_vrq) on both levels.

. \*It's a bit complicated. For primary schools, we just include it after pid: .

. \* But for the secondary school level, we have to generate interaction terms of c\_vrq and all second

> ary schools.

.

. \*Create dummies for each secondary school

. tab sid, gen(id\_sid)

sid | Freq. Percent Cum.

------------+-----------------------------------

1 | 219 6.38 6.38

2 | 199 5.79 12.17

3 | 156 4.54 16.71

4 | 139 4.05 20.76

5 | 175 5.09 25.85

6 | 250 7.28 33.13

7 | 109 3.17 36.30

8 | 107 3.11 39.42

9 | 114 3.32 42.74

10 | 92 2.68 45.41

11 | 234 6.81 52.23

12 | 253 7.37 59.59

13 | 216 6.29 65.88

14 | 290 8.44 74.32

15 | 147 4.28 78.60

16 | 134 3.90 82.50

17 | 233 6.78 89.29

18 | 257 7.48 96.77

19 | 111 3.23 100.00

------------+-----------------------------------

Total | 3,435 100.00

.

. unab idvar: id\_sid\*

. foreach v of local idvar{

2. gen inter`v' = c\_vrq\*`v'

3. }

.

. \*Now we can include a random slope on both levels

. \*We restrict the covariance between all interaction dummies to 0.

. \*We also could specify random slopes for different variables on each level.

.

. xtmixed attain c\_vrq sex || \_all: R.sid || ///

> \_all:inter\*, cov(identity) nocons || pid: c\_vrq, var mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -7424.774

Iteration 1: log likelihood = -7421.3878

Iteration 2: log likelihood = -7421.3786

Iteration 3: log likelihood = -7421.3786

Computing standard errors:

Mixed-effects ML regression Number of obs = 3435

-----------------------------------------------------------

| No. of Observations per Group

Group Variable | Groups Minimum Average Maximum

----------------+------------------------------------------

\_all | 1 3435 3435.0 3435

pid | 148 1 23.2 72

-----------------------------------------------------------

Wald chi2(2) = 2902.56

Log likelihood = -7421.3786 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

attain | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

c\_vrq | .1597335 .0029859 53.50 0.000 .1538813 .1655856

sex | .1151836 .0714303 1.61 0.107 -.0248171 .2551843

\_cons | 5.569957 .07478 74.48 0.000 5.423391 5.716523

------------------------------------------------------------------------------

------------------------------------------------------------------------------

Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]

-----------------------------+------------------------------------------------

\_all: Identity |

var(R.sid) | .0124141 .0230308 .0003272 .4710613

-----------------------------+------------------------------------------------

\_all: Identity |

var(inter~d1..inter~19)(1) | .000021 .0000514 1.73e-07 .0025454

-----------------------------+------------------------------------------------

pid: Independent |

var(c\_vrq) | 1.68e-18 1.17e-15 0 .

var(\_cons) | .2750255 .0614916 .1774422 .4262742

-----------------------------+------------------------------------------------

var(Residual) | 4.245761 .1051668 4.044561 4.456969

------------------------------------------------------------------------------

LR test vs. linear regression: chi2(4) = 87.93 Prob > chi2 = 0.0000

(1) interid\_sid1 interid\_sid2 interid\_sid3 interid\_sid4 interid\_sid5 interid\_sid6 interid\_sid7

interid\_sid8 interid\_sid9 interid\_sid10 interid\_sid11 interid\_sid12 interid\_sid13

interid\_sid14 interid\_sid15 interid\_sid16 interid\_sid17 interid\_sid18 interid\_sid19

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

.

. /\*No significant random slopes for c\_vrq. Maybe there are gender differences?

>

> drop inter\*

> unab idvar: id\_sid\*

> foreach v of local idvar{

> gen inter`v' = sex\*`v'

> }

>

> xtmixed attain c\_vrq sex || \_all: R.sid || ///

> \_all:inter\*, cov(identity) nocons || pid: sex, var mle

>

> \*Mean Verbal apptitude score for primary schools

> bysort pid: egen mvrq=mean(c\_vrq)

>

> xtmixed attain c\_vrq c.mvrq##i.sex || \_all: R.sid || ///

> \_all:inter\*, cov(identity) nocons || pid: sex, var mle

> \*/

.

. \*Clearly, there are more things to discover, such as interactions of random intercepts.

. \*Specific combinations of primary/secondary schools might matter. But that would be too much now..

> .

.

. capture log close