**SOCY7708: Hierarchical Linear Modeling**

**Instructor: Natasha Sarkisian**

**Class notes: HLM Model Building Strategies: Example**

Let’s follow the second (combined) model building strategy and evaluate a model where we consider two level 1 predictors – ses and female, and two level 2 predictors – sector and a set of dummies for school size. (Note that we could also consider two additional level 2 predictors that are aggregates of level 1 predictors – average school-level SES and gender composition of school; but we keep it simple for now.)

. mixed mathach c.ses##c.sector c.ses##i.sized i.female##i.sector i.female##i.sized || id: ses female, cov(unstr)

note: ses omitted because of collinearity.

note: 1.sector omitted because of collinearity.

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(11) = 677.31

Log likelihood = -23246.527 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 3.071058 .2823452 10.88 0.000 2.517671 3.624444

 sector | 2.36977 .4358575 5.44 0.000 1.515505 3.224035

 |

c.ses#c.sector | -1.283094 .2364085 -5.43 0.000 -1.746446 -.8197415

 |

 ses | 0 (omitted)

 |

 sized |

 2 | 1.23097 .5676327 2.17 0.030 .1184298 2.343509

 3 | 1.752608 .585621 2.99 0.003 .6048122 2.900404

 |

 sized#c.ses |

 2 | -.2519453 .2923376 -0.86 0.389 -.8249165 .3210258

 3 | -.1277427 .3097203 -0.41 0.680 -.7347833 .4792979

 |

 1.female | -.2816415 .4703985 -0.60 0.549 -1.203606 .6403227

 1.sector | 0 (omitted)

 |

 female#sector |

 1 1 | -.1739571 .4058262 -0.43 0.668 -.9693619 .6214477

 |

 female#sized |

 1 2 | -.712411 .5186187 -1.37 0.170 -1.728885 .3040631

 1 3 | -1.308136 .5277861 -2.48 0.013 -2.342578 -.2736945

 |

 \_cons | 11.03738 .5302974 20.81 0.000 9.998014 12.07674

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Unstructured |

 var(ses\_m) | .0864968 . . .

 var(female) | .7588742 . . .

 var(\_cons) | 4.090261 . . .

 cov(ses\_m,female) | -.1480118 . . .

 cov(ses\_m,\_cons) | .5930792 . . .

 cov(female,\_cons) | -.9053627 . . .

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

 var(Residual) | 36.36132 . . .

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

LR test vs. linear model: chi2(6) = 311.08 Prob > chi2 = 0.0000

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

Warning: Standard-error calculation failed.

This model has problems with variance components – let’s test each of them. First ses:

. mixed mathach c.ses##c.sector c.ses##i.sized i.female##i.sector i.female##i.sized || id: ses, cov(unstr)

note: ses omitted because of collinearity.

note: 1.sector omitted because of collinearity.

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(11) = 703.22

Log likelihood = -23248.671 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 3.048262 .2824976 10.79 0.000 2.494577 3.601947

 sector | 2.370088 .4077368 5.81 0.000 1.570938 3.169237

 |

c.ses#c.sector | -1.271897 .2364572 -5.38 0.000 -1.735345 -.8084498

 |

 ses | 0 (omitted)

 |

 sized |

 2 | 1.214981 .5279932 2.30 0.021 .1801335 2.249829

 3 | 1.718568 .5468995 3.14 0.002 .646665 2.790472

 |

 sized#c.ses |

 2 | -.2372528 .2926702 -0.81 0.418 -.8108759 .3363703

 3 | -.1103129 .3099481 -0.36 0.722 -.7178 .4971742

 |

 1.female | -.2954447 .4224582 -0.70 0.484 -1.123448 .5325581

 1.sector | 0 (omitted)

 |

 female#sector |

 1 1 | -.1500386 .368775 -0.41 0.684 -.8728244 .5727472

 |

 female#sized |

 1 2 | -.6956448 .4670362 -1.49 0.136 -1.611019 .2197294

 1 3 | -1.285465 .4759068 -2.70 0.007 -2.218225 -.3527048

 |

 \_cons | 11.05523 .4953209 22.32 0.000 10.08442 12.02604

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Unstructured |

 var(ses) | .0843152 .0868693 .0111923 .6351725

 var(\_cons) | 3.467264 .5180921 2.587004 4.647044

 cov(ses,\_cons) | .5406851 .2936755 -.0349084 1.116279

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

 var(Residual) | 36.48905 .6177096 35.29823 37.72005

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

LR test vs. linear model: chi2(3) = 306.79 Prob > chi2 = 0.0000

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

. estat ic

Akaike's information criterion and Bayesian information criterion

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

 Model | N ll(null) ll(model) df AIC BIC

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

 . | 7,185 . -23248.67 16 46529.34 **46639.42**

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

Note: BIC uses N = number of observations. See [R] BIC note.

. est store sesrandom

. mixed mathach c.ses##c.sector c.ses##i.sized i.female##i.sector i.female##i.sized || id:

note: ses omitted because of collinearity.

note: 1.sector omitted because of collinearity.

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(11) = 699.99

Log likelihood = -23251.547 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 3.049839 .2763021 11.04 0.000 2.508297 3.591381

 sector | 2.371117 .4019846 5.90 0.000 1.583241 3.158992

 |

c.ses#c.sector | -1.265125 .2317514 -5.46 0.000 -1.71935 -.810901

 |

 ses | 0 (omitted)

 |

 sized |

 2 | 1.237825 .5204513 2.38 0.017 .2177591 2.257891

 3 | 1.7123 .538888 3.18 0.001 .6560985 2.768501

 |

 sized#c.ses |

 2 | -.2557244 .2866414 -0.89 0.372 -.8175312 .3060824

 3 | -.1155212 .303394 -0.38 0.703 -.7101625 .4791201

 |

 1.female | -.2856112 .4223186 -0.68 0.499 -1.11334 .5421179

 1.sector | 0 (omitted)

 |

 female#sector |

 1 1 | -.1414109 .3692235 -0.38 0.702 -.8650756 .5822539

 |

 female#sized |

 1 2 | -.7326045 .4664619 -1.57 0.116 -1.646853 .1816441

 1 3 | -1.283305 .4760334 -2.70 0.007 -2.216314 -.3502968

 |

 \_cons | 11.0998 .4879931 22.75 0.000 10.14335 12.05625

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Identity |

 var(\_cons) | 3.317526 .4858015 2.489824 4.420383

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

 var(Residual) | 36.55247 .6172766 35.36243 37.78255

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

LR test vs. linear model: chibar2(01) = 301.04 Prob >= chibar2 = 0.0000

. est store noslope

. lrtest sesrandom noslope

Likelihood-ratio test

Assumption: noslope nested within sesrandom

 LR chi2(2) = 5.75

Prob > chi2 = 0.0564

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.

. estat ic

Akaike's information criterion and Bayesian information criterion

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

 Model | N ll(null) ll(model) df AIC BIC

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

 . | 7,185 . -23251.55 14 46531.09 **46627.41**

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

Note: BIC uses N = number of observations. See [R] BIC note.

BIC value is 12 points lower in the model without SES variance so it’s preferred based on both BIC and LR test. Now let’s test the slope variance for female:

. mixed mathach c.ses##c.sector c.ses##i.sized i.female##i.sector i.female##i.sized || id: female, cov(unstr)

note: ses omitted because of collinearity.

note: 1.sector omitted because of collinearity.

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(11) = 696.49

Log likelihood = -23249.266 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 3.072735 .2756691 11.15 0.000 2.532434 3.613037

 sector | 2.367268 .432287 5.48 0.000 1.520001 3.214535

 |

c.ses#c.sector | -1.276275 .2314061 -5.52 0.000 -1.729823 -.8227277

 |

 ses | 0 (omitted)

 |

 sized |

 2 | 1.276582 .5631707 2.27 0.023 .1727874 2.380376

 3 | 1.766474 .5807221 3.04 0.002 .6282794 2.904668

 |

 sized#c.ses |

 2 | -.2675328 .28581 -0.94 0.349 -.8277101 .2926445

 3 | -.1308607 .3026749 -0.43 0.665 -.7240926 .4623711

 |

 1.female | -.2465819 .4707665 -0.52 0.600 -1.169267 .6761034

 1.sector | 0 (omitted)

 |

 female#sector |

 1 1 | -.1725857 .4063304 -0.42 0.671 -.9689787 .6238073

 |

 female#sized |

 1 2 | -.7773901 .5187856 -1.50 0.134 -1.794191 .2394109

 1 3 | -1.331266 .5283803 -2.52 0.012 -2.366872 -.2956592

 |

 \_cons | 11.05971 .5258146 21.03 0.000 10.02914 12.09029

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Unstructured |

 var(female) | .768301 .5565779 .1857369 3.178079

 var(\_cons) | 3.992242 .6855878 2.851281 5.589767

 cov(female,\_cons) | -.959724 .5169883 -1.973003 .0535545

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

 var(Residual) | 36.43078 .6198063 35.23601 37.66606

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

LR test vs. linear model: chi2(3) = 305.60 Prob > chi2 = 0.0000

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

. est store femalerandom

. estat ic

Akaike's information criterion and Bayesian information criterion

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

 Model | N ll(null) ll(model) df AIC BIC

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

femalerandom | 7,185 . -23249.27 16 46530.53 **46640.61**

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

Note: BIC uses N = number of observations. See [R] BIC note.

. lrtest noslope femalerandom

Likelihood-ratio test

Assumption: noslope nested within femalerandom

 LR chi2(2) = 4.56

Prob > chi2 = 0.1022

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.

BIC value is 13 points lower in the model without FEMALE variance so it’s preferred based on both BIC and LR test. We conclude that neither level 1 predictor needs a random slope. However, we still want to investigate the cross-level interactions, so we estimate:

. mixed mathach c.ses##c.sector c.ses##i.sized i.female##i.sector i.female##i.sized || id:

note: ses omitted because of collinearity.

note: 1.sector omitted because of collinearity.

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(11) = 699.99

Log likelihood = -23251.547 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 3.049839 .2763021 11.04 0.000 2.508297 3.591381

 sector | 2.371117 .4019846 5.90 0.000 1.583241 3.158992

 |

c.ses#c.sector | -1.265125 .2317514 -5.46 0.000 -1.71935 -.810901

 |

 ses | 0 (omitted)

 |

 sized |

 2 | 1.237825 .5204513 2.38 0.017 .2177591 2.257891

 3 | 1.7123 .538888 3.18 0.001 .6560985 2.768501

 |

 sized#c.ses |

 2 | -.2557244 .2866414 -0.89 0.372 -.8175312 .3060824

 3 | -.1155212 .303394 -0.38 0.703 -.7101625 .4791201

 |

 1.female | -.2856112 .4223186 -0.68 0.499 -1.11334 .5421179

 1.sector | 0 (omitted)

 |

 female#sector |

 1 1 | -.1414109 .3692235 -0.38 0.702 -.8650756 .5822539

 |

 female#sized |

 1 2 | -.7326045 .4664619 -1.57 0.116 -1.646853 .1816441

 1 3 | -1.283305 .4760334 -2.70 0.007 -2.216314 -.3502968

 |

 \_cons | 11.0998 .4879931 22.75 0.000 10.14335 12.05625

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Identity |

 var(\_cons) | 3.317526 .4858015 2.489824 4.420383

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

 var(Residual) | 36.55247 .6172766 35.36243 37.78255

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

LR test vs. linear model: chibar2(01) = 301.04 Prob >= chibar2 = 0.0000

Female main effect is not significant, but because there are significant interactions with it, we’d keep it. Let’s test for joint significance of other non-significant terms (doesn’t include non-significant dummies in sets of dummies, sets should be evaluated as a whole):

. test 1.female#1.sector=0

 ( 1) [mathach]1.female#1.sector = 0

 chi2( 1) = 0.15

 Prob > chi2 = 0.7017

. test 2.sized#c.ses=0, acc

 ( 1) [mathach]1.female#1.sector = 0

 ( 2) [mathach]2.sized#c.ses = 0

 chi2( 2) = 0.94

 Prob > chi2 = 0.6240

. test 3.sized#c.ses=0, acc

 ( 1) [mathach]1.female#1.sector = 0

 ( 2) [mathach]2.sized#c.ses = 0

 ( 3) [mathach]3.sized#c.ses = 0

 chi2( 3) = 1.01

 Prob > chi2 = 0.7990

We can jointly omit all those:

. mixed mathach c.ses##c.sector i.female##i.sized || id:

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(8) = 698.49

Log likelihood = -23252.051 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 2.906627 .1400363 20.76 0.000 2.632161 3.181093

 sector | 2.292641 .3582754 6.40 0.000 1.590434 2.994848

 |

c.ses#c.sector | -1.27794 .2109064 -6.06 0.000 -1.691309 -.8645712

 |

 1.female | -.3719948 .385254 -0.97 0.334 -1.127079 .3830891

 |

 sized |

 2 | 1.234907 .5191091 2.38 0.017 .217472 2.252342

 3 | 1.692167 .5386236 3.14 0.002 .6364839 2.74785

 |

 female#sized |

 1 2 | -.7090223 .4646167 -1.53 0.127 -1.619654 .2016097

 1 3 | -1.219811 .4604496 -2.65 0.008 -2.122276 -.3173464

 |

 \_cons | 11.13458 .4845513 22.98 0.000 10.18488 12.08428

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Identity |

 var(\_cons) | 3.326955 .4861819 2.498373 4.430335

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

 var(Residual) | 36.55583 .6173111 35.36572 37.78598

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

LR test vs. linear model: chibar2(01) = 302.48 Prob >= chibar2 = 0.0000

Finally, we should consider whether we want to simplify those coding of size – could we combine groups 2 and 3?

. test 2.sized=3.sized

 ( 1) [mathach]2.sized - [mathach]3.sized = 0

 chi2( 1) = 1.14

 Prob > chi2 = 0.2865

. test 1.female#2.sized=1.female#3.sized, acc

 ( 1) [mathach]2.sized - [mathach]3.sized = 0

 ( 2) [mathach]1.female#2.sized - [mathach]1.female#3.sized = 0

 chi2( 2) = 2.27

 Prob > chi2 = 0.3212

Indeed, they don’ produce anything distinct, so we can combine them and use a single dummy for size – small vs large.

. tab sized, gen(sized\_)

 RECODE of |

 size | Freq. Percent Cum.

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

 1 | 1,524 21.21 21.21

 2 | 2,984 41.53 62.74

 3 | 2,677 37.26 100.00

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

 Total | 7,185 100.00

. tab sized\_1

 sized== |

 1.0000 | Freq. Percent Cum.

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

 0 | 5,661 78.79 78.79

 1 | 1,524 21.21 100.00

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

 Total | 7,185 100.00

It is coded 1=small, 0=large. We could also reverse code it if we’d like. So here is our final parsimonious model. I will also compare it to the original one:

. mixed mathach c.ses##c.sector i.female##i.sized\_1 || id:

Mixed-effects ML regression Number of obs = 7,185

Group variable: id Number of groups = 160

 Obs per group:

 min = 14

 avg = 44.9

 max = 67

 Wald chi2(6) = 696.34

Log likelihood = -23253.187 Prob > chi2 = 0.0000

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

 mathach | Coefficient Std. err. z P>|z| [95% conf. interval]

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

 ses | 2.908528 .1400094 20.77 0.000 2.634115 3.182942

 sector | 2.227118 .3332109 6.68 0.000 1.574037 2.8802

 |

c.ses#c.sector | -1.284285 .2107547 -6.09 0.000 -1.697357 -.8712136

 |

 1.female | -1.345095 .1811197 -7.43 0.000 -1.700083 -.990107

 1.sized\_1 | -1.451899 .4824572 -3.01 0.003 -2.397498 -.5063004

 |

female#sized\_1 |

 1 1 | .9753494 .4256332 2.29 0.022 .1411236 1.809575

 |

 \_cons | 12.62697 .2444691 51.65 0.000 12.14782 13.10612

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

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

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

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

id: Identity |

 var(\_cons) | 3.320192 .4849929 2.493589 4.420805

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

 var(Residual) | 36.56893 .6175188 35.37842 37.79949

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

LR test vs. linear model: chibar2(01) = 303.16 Prob >= chibar2 = 0.0000

. estat ic

Akaike's information criterion and Bayesian information criterion

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

 Model | N ll(null) ll(model) df AIC BIC

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

 . | 7,185 . -23253.19 9 46524.37 **46586.29**

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

Note: BIC uses N = number of observations. See [R] BIC note.

. lrtest . femalerandom

Likelihood-ratio test

Assumption: . nested within femalerandom

 LR chi2(7) = 7.84

Prob > chi2 = 0.3467

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.

. lrtest . sesrandom

Likelihood-ratio test

Assumption: . nested within sesrandom

 LR chi2(7) = 9.03

Prob > chi2 = 0.2504

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.

BIC is more than 40 points lower for this model than for the initial models with SES or FEMALE random slope, and LR test confirms that this parsimonious model fits just as well as those more complex ones, so we stick to this one.