Models With Three or More Levels and Cross-Classified Models

Introduction

A two level multilevel model imagines that *Level 1* units are nested in *Level 2* units. A three level multilevel model imagines that *Level 1* units are nested in *Level 2* units, which are in turn nested in *Level 3*.

A *cross classified* model imagines that the nesting is not hierarchical, but rather that there are two sets of clusters or nestings in which individuals may be nested.

Below, I describe the use of Stata, R, and Julia to estimate these models.

Three Or More Levels

The Data

I use the *longitudinal* data from *Multilevel Thinking* to which I have added an extra level of *United Nations Region* [@ArelBundock2018].

The Equation

$$\text{outcome}_{itj} = \beta_0 + \beta_1 \text{parental warmth}_{itj} + \beta_2 \text{physical punishment}_{itj} + \beta_3 \text{time}_{itj} + \quad (1)$$

$$\beta_4 \text{identity}_{itj} + \beta_5 \text{intervention}_{itj} + \beta_6 \text{HDI}_{itj} +$$

$$w_{0k} + u_{0j} + v_{0i} + e_{itjk}$$

Here we imagine w_{0k} (region), u_{0j} (country) and v_{0i} (family) are hierarchically nested effects.

Cross-Classified Models

The Data

I use the *cross-sectional* data from *Multilevel Thinking* to which I have added an extra level of a hypothetical language.

The Equation

$$\text{outcome}_{itj} = \beta_0 + \beta_1 \text{parental warmth}_{itj} + \beta_2 \text{physical punishment}_{itj} + \beta_3 \text{time}_{itj} + \quad (2)$$

$$\beta_4 \text{identity}_{itj} + \beta_5 \text{intervention}_{itj} + \beta_6 \text{HDI}_{itj} +$$

$$u_{0j} + m_{0m} + e_{ijm}$$

Here u_{0j} (country) and m_{0m} (language) are not nested hierarchically, but are *cross classified*.

Run The Models

Stata

```
use "fourlevel.dta", clear
```

```
mixed outcome || UNregion: || country: || family:
```

```
Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:  Log likelihood = -29061.686
Iteration 1:  Log likelihood = -29061.679
Iteration 2:  Log likelihood = -29061.679

Computing standard errors ...

Mixed-effects ML regression                             Number of obs = 9,000

        Grouping information
        ------------------------------------------------------------
                        |     No. of       Observations per group
         Group variable |     groups    Minimum    Average    Maximum
        ----------------+-------------------------------------------
               UNregion |          5        600    1,800.0      3,600
                country |         30        300      300.0        300
                 family |      3,000          3        3.0          3
        ------------------------------------------------------------
```

```
                                                      Wald chi2(0)  =     .
Log likelihood = -29061.679                           Prob > chi2   =     .

------------------------------------------------------------------------------
     outcome | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   54.05906    .987367    54.75   0.000     52.12385    55.99426
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
UNregion: Identity           |
                  var(_cons) |   4.172687   3.187885      .9334852    18.65194
-----------------------------+------------------------------------------------
country: Identity            |
                  var(_cons) |   2.849348   .8710225      1.565093    5.187414
-----------------------------+------------------------------------------------
family: Identity             |
                  var(_cons) |   11.72403     .57475      10.64997    12.90641
-----------------------------+------------------------------------------------
               var(Residual) |   28.23424   .5154842      27.24177    29.26286
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 1843.44               Prob > chi2 = 0.0000

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

```
mixed outcome t warmth physical_punishment i.identity i.intervention HDI || UNregion: || cou
```

```
Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:  Log likelihood = -28503.082
Iteration 1:  Log likelihood = -28503.039
Iteration 2:  Log likelihood = -28503.039

Computing standard errors ...

Mixed-effects ML regression                           Number of obs =   9,000
```

Grouping information

Group variable	No. of groups	Observations per group Minimum	Average	Maximum
UNregion	5	600	1,800.0	3,600
country	30	300	300.0	300
id	3,000	3	3.0	3

Log likelihood = -28503.039

Wald chi2(6) = 1209.42
Prob > chi2 = 0.0000

outcome	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
t	.9433791	.0658667	14.32	0.000	.8142827	1.072476
warmth	.9140704	.0379156	24.11	0.000	.8397571	.9883837
physical_punishment	-1.008615	.0497772	-20.26	0.000	-1.106176	-.9110531
1.identity	-.1332133	.1516437	-0.88	0.380	-.4304294	.1640028
1.intervention	.8589263	.1519619	5.65	0.000	.5610865	1.156766
HDI	.0148561	.0196605	0.76	0.450	-.0236777	.0533899
_cons	50.16426	1.675219	29.94	0.000	46.88089	53.44763

Random-effects parameters	Estimate	Std. err.	[95% conf.	interval]
UNregion: Identity				
var(_cons)	4.722007	3.585939	1.065898	20.91884
country: Identity				
var(_cons)	2.863495	.8656459	1.583342	5.178668
id: Identity				
var(_cons)	8.421131	.4711947	7.546445	9.397199
var(Residual)	26.02919	.4752587	25.11417	26.97755

LR test vs. linear model: chi2(3) = 1844.00 Prob > chi2 = 0.0000

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

R

```
library(haven)

df4 <- read_dta("fourlevel.dta")
```

```
df4$identity <- factor(df4$identity)

df4$intervention <- factor(df4$intervention)
```

> **Caution**
>
> `lme4` does not directly provide p values in results, because of some disagreement over exactly how these p values should be calculated. Therefore, in this Appendix, I also call library `lmerTest` to provide p values for `lme4` results.

> **Tip**
>
> R prefers to use scientific notation when possible. I find that the use of scientific notation can be confusing in reading results. I turn off scientific notation by setting a penalty for its use: `options(scipen = 999)`.

```
library(lme4)

library(lmerTest)
```

```
Attaching package: 'lmerTest'

The following object is masked from 'package:lme4':

    lmer

The following object is masked from 'package:stats':

    step
```

```r
options(scipen = 999)

fit4A <- lmer(outcome ~  (1 | UNregion/country/id),
              data = df4)

summary(fit4A)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: outcome ~ (1 | UNregion/country/id)
   Data: df4

REML criterion at convergence: 58121.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7850 -0.6064 -0.0047  0.6020  3.4399

Random effects:
 Groups                 Name        Variance Std.Dev.
 id:(country:UNregion) (Intercept) 11.724   3.424
 country:UNregion      (Intercept)  2.842   1.686
 UNregion              (Intercept)  5.478   2.340
 Residual                          28.234   5.314
Number of obs: 9000, groups:
id:(country:UNregion), 3000; country:UNregion, 30; UNregion, 5

Fixed effects:
            Estimate Std. Error     df t value   Pr(>|t|)
(Intercept)   54.061      1.112  3.777    48.6 0.00000201 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit4B <- lmer(outcome ~ t + warmth + physical_punishment +
                identity + intervention + HDI +
                (1 | UNregion/country/id),
              data = df4)

summary(fit4B)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
lmerModLmerTest]
Formula:
outcome ~ t + warmth + physical_punishment + identity + intervention +
    HDI + (1 | UNregion/country/id)
   Data: df4

REML criterion at convergence: 57026.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6846 -0.6096 -0.0038  0.6138  3.6850

Random effects:
 Groups                 Name        Variance Std.Dev.
 id:(country:UNregion) (Intercept)  8.438   2.905
 country:UNregion      (Intercept)  2.979   1.726
 UNregion              (Intercept)  6.178   2.486
 Residual                           26.036   5.103
Number of obs: 9000, groups:
id:(country:UNregion), 3000; country:UNregion, 30; UNregion, 5

Fixed effects:
                     Estimate Std. Error         df t value
(Intercept)          50.11857    1.78086   15.79112  28.143
t                     0.94338    0.06588 5998.37756  14.321
warmth                0.91406    0.03793 4745.28492  24.096
physical_punishment  -1.00876    0.04980 6483.46337 -20.257
identity1            -0.13324    0.15173 2969.00938  -0.878
intervention1         0.85872    0.15205 2971.85430   5.648
HDI                   0.01560    0.02006   24.39852   0.778
                              Pr(>|t|)
(Intercept)          0.00000000000000641 ***
t                   < 0.0000000000000002 ***
warmth              < 0.0000000000000002 ***
physical_punishment < 0.0000000000000002 ***
identity1                        0.380
intervention1        0.00000001780521096 ***
HDI                              0.444
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) t      warmth physc_ idntt1 intrv1
```

```
t           -0.073
warmth      -0.071 -0.002
physcl_pnsh -0.073 -0.007 -0.012
identity1   -0.040  0.000 -0.013 -0.003
interventn1 -0.045  0.000  0.039  0.019 -0.018
HDI         -0.738  0.000 -0.005  0.005 -0.001  0.001
```

Julia

```
using Tables, MixedModels, StatFiles, DataFrames, CategoricalArrays, DataFramesMeta

df4 = DataFrame(load("fourlevel.dta"))
```

```
@transform!(df4, :country = categorical(:country))

@transform!(df4, :UNregion = categorical(:UNregion))

@transform!(df4, :identity = categorical(:identity))

@transform!(df4, :intervention = categorical(:intervention))
```

```
m4A = fit(MixedModel, @formula(outcome ~ t + warmth +
                                  physical_punishment +
                                  identity + intervention +
                                  HDI +
                                  (1 | UNregion) +
                                  (1 | country) +
                                  (1 | id)), df4)
```

```
m4B = fit(MixedModel, @formula(outcome ~ t + warmth +
                                  physical_punishment +
                                  identity + intervention +
                                  HDI +
                                  (1 | UNregion) +
                                  (1 | country) +
                                  (1 | id)), df4)
```

Interpretation

Cross-Classified Models

The Data

The Equation

Run The Models

Stata

```
use "./simulate-and-analyze-multilevel-data/crossclassified.dta", clear
```

```
file ./simulate-and-analyze-multilevel-data/crossclassified.dta not found
r(601);

r(601);
```

```
mixed outcome || _all: R.country || _all: R.language
```

```
file ./simulate-and-analyze-multilevel-data/crossclassified.dta not found
r(601);


variable language not found
r(111);

r(111);
```

```
mixed outcome warmth physical_punishment i.identity i.intervention HDI || _all: R.country ||
```

```
file ./simulate-and-analyze-multilevel-data/crossclassified.dta not found
r(601);


variable language not found
r(111);

r(111);
```

R

```
library(haven)

dfCC <- read_dta("./simulate-and-analyze-multilevel-data/crossclassified.dta")
```

```
Error: './simulate-and-analyze-multilevel-data/crossclassified.dta' does not exist in current
```

```
dfCC$identity <- factor(dfCC$identity)
```

```
Error in eval(expr, envir, enclos): object 'dfCC' not found
```

```
dfCC$intervention <- factor(dfCC$intervention)
```

```
Error in eval(expr, envir, enclos): object 'dfCC' not found
```

```
library(lme4)

library(lmerTest)

options(scipen = 999)

fitCC_A <- lmer(outcome ~
                  (1 | country) +
                  (1 | language),
              data = dfCC)
```

```
Error: bad 'data': object 'dfCC' not found
```

```
summary(fitCC_A)
```

```
Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a
```

```
fitCC_B <- lmer(outcome ~ t + warmth + physical_punishment +
                  identity + intervention + HDI +
                  (1 | country) +
                  (1 | language),
                data = dfCC)
```

```
Error: bad 'data': object 'dfCC' not found
```

```
summary(fitCC_B)
```

```
Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a
```

Julia

Interpretation