

Stat 414 – Day 17
Imperfect Hierarchies (Cross-Classified, Multiple Membership) (Ch. 13)

Last Time: Logistic regression

- When do you use logistic regression?
- What is the logit transformation?
- How do we interpret the intercept? A slope?
- What are the variance components in a multilevel logistic regression?
- How do we compute an intraclass correlation coefficient for a random intercepts model?
- What are the “fitted values” in a multilevel logistic regression?
- What are the residuals in a multilevel logistic regression?

Example 1: The response variable is attainment score at age 16 for pupils that attended various combinations of primary and secondary schools. CrossAttainment.txt

pid = primary school identifier

sid = secondary school identifier

vrq = verbal reasoning score

sex = sex (Assuming sex = 0 = male, sex = 1 = female)

(a) Center the verbal reasoning score variable.

(b) How many unique primary and secondary school combinations are there? To how many different secondary schools did a primary school sent the pupils?

148 primary schools fed into 19 secondary schools; 57 primary schools went to only one secondary school, 50 fed into 2 different secondary schools, 26 fed into 3 and so on

(c) What happens if you try to fit a normal two-level random intercepts model with primary schools nested within secondary schools?

This runs ok because the schools are numbered uniquely. The results won't differ too much because the data is almost nested, most primary schools feed into just a few of the secondary schools.

Although R thinks there are 303 primary schools rather than realizing these are some of the same primary schools going into different secondary schools. Ends up deflating the “level 2” variance components.

(d) Fit a crossed model with random intercepts. How do the models compare? (*Hint:* Things don't change too much, why not?)

Compared to only nesting in primary schools, we do see that a little bit of the variance is now distributed to the secondary schools.

(e) Calculate the following intraclass correlation coefficients

- Same primary, but different secondary schools

$$1.13/9.61 = .1169$$

- Same secondary, but different primary schools

$$.372/9.61 = .0365$$

- Same primary and same secondary school

$$(1.13 + .372)/9.61 = .1534$$

(f) Now fit a model with crossed effects, and verbal aptitude and gender as fixed explanatory variables.

```
lmer(attain ~ 1 + cvrq + sex + (1 | sid) + (1| pid)
```

(g) Try to include a random slope for centered vrq for both classification variables. [Note: In other software packages, to create the random slopes for the second classification variable (pick the one with smaller number of categories), have to create some dummy variables for these categories and then “cross” these with vrq but restrict the covariance between all interaction dummies to be zero.]

```
> summary(lmer(attain ~ cvrq + sex + (cvrq|sid) + (cvrq|pid)))
```

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
pid	(Intercept)	0.2822112	0.53124	
	cvrq	0.0000219	0.00468	1.00
sid	(Intercept)	0.0192915	0.13889	
	cvrq	0.0000364	0.00603	0.66
Residual		4.2398698	2.05909	

Number of obs: 3435, groups: pid, 148; sid, 19

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.55684	0.07783	71.40
cvrq	0.15904	0.00317	50.19
sex	0.11100	0.07142	1.55

(h) Summarize what you learn from the model.

The random slopes don't seem to be significant... So there is a constant positive association between vrqc and attainment across the primary and secondary schools. After adjusting for vrq, the gender “effect” doesn't seem to be highly significant but we could check for random slopes (changes in the sex effect on attainment across the primary or secondary schools). Most of the variation is in the residuals – students going to the same primary and secondary schools.