

Join us in building a kind, collaborative learning community via our updated [Code of Conduct](#).

Crossed vs nested random effects: how do they differ and how are they specified correctly in lme4?

Here is how I have understood nested vs. crossed random effects:

Nested random effects occur when a lower level factor appears only within a particular level of an upper level factor.

- For example, pupils within classes at a fixed point in time.
- In `lme4` I thought that we represent the random effects for nested data in either of two equivalent ways:

```
(1|class/pupil) # or
(1|class) + (1|class:pupil)
```

Crossed random effects means that a given factor appears in more than one level of the upper level factor.

- For example, there are pupils within classes measured over several years.
- In `lme4`, we would write:

```
(1|class) + (1|pupil)
```

However, when I was looking at a particular nested dataset, I noticed that both model formulas gave identical results (code and output below). However I have seen other datasets where the two formulas produced different results. So what is going on here?

```
mydata <- read.csv("https://web.archive.org/web/20160624172041if_/http://www-
personal.umich.edu/~bwest/classroom.csv")
# (the data is no longer at 'http://www-personal.umich.edu/~bwest/cClassroom.csv'
# hence the link to web.archive.org)
# Crossed version:
Linear mixed model fit by REML ['lmerMod']
Formula: mathgain ~ (1 | schoolid) + (1 | classid)
Data: mydata
```

REML criterion at convergence: 11768.8

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-4.6441	-0.5984	-0.0336	0.5334	5.6335

Random effects:

Groups	Name	Variance	Std.Dev.
classid	(Intercept)	99.23	9.961
schoolid	(Intercept)	77.49	8.803
Residual		1028.23	32.066

Number of obs: 1190, groups: classid, 312; schoolid, 107

Nested version:

```
Formula: mathgain ~ (1 | schoolid/classid)
```

REML criterion at convergence: 11768.8

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-4.6441	-0.5984	-0.0336	0.5334	5.6335

Random effects:

Groups	Name	Variance	Std.Dev.
classid:schoolid	(Intercept)	99.23	9.961
schoolid	(Intercept)	77.49	8.803
Residual		1028.23	32.066

Number of obs: 1190, groups: classid:schoolid, 312; schoolid, 107

[r](#) | [mixed-model](#) | [multilevel-analysis](#) | [lme4-nlme](#)

edited Mar 16 at 12:00



Eric Fail

77 3 17

asked Aug 8 '16 at 15:08



Joe King

1,213 2 23 43

1 This is probably due to the way the factors are coded. Do you have a link to the data ? – Robert Long Aug 8 '16 at 15:10

mydata <- read.csv("www-personal.umich.edu/~bwest/classroom.csv") – Joe King Aug 8 '16 at 15:18

1 Answer

This site uses cookies to deliver our services and to show you relevant ads and job listings. By using our site, you acknowledge that you have read and understand our [Cookie Policy](#), [Privacy Policy](#), and our [Terms of Service](#). Your use of Stack Overflow's Products and Services, including the Stack Overflow Network, is subject to these policies and terms.

You are not wrong in your understanding of what nested and crossed random effects are in the scenario that you describe. However, your definition of crossed random effects is a little narrow. A more general definition of crossed random effects is simply: **not nested**. We will look at this at the end of this answer, but the bulk of the answer will focus on the scenario you presented, of classrooms within schools.

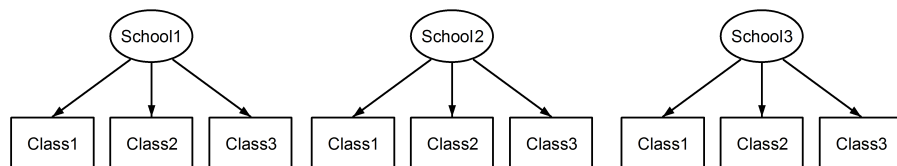
First note that:

Nesting is a property of the data, or rather the experimental design, not the model.

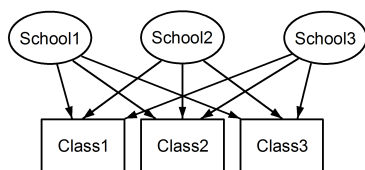
Also,

Nested data can be encoded in at least 2 different ways, and this is at the heart of the issue you found.

The dataset in your example is rather large, so I will use another schools example from the internet to explain the issues. But first, consider the following over-simplified example:



Here we have classes nested in schools, which is a familiar scenario. The important point here is that, between each school, the classes have the same identifier, **even though they are distinct if they are nested**. Class1 appears in School1, School2 and School3. However if the data are nested then Class1 in School1 is *not* the same unit of measurement as Class1 in School2 and School3. If they were the same, then we would have this situation:



which means that every class belongs to every school. The former is a nested design, and the latter is a crossed design (some might also call it multiple membership), and we would formulate these in lme4 using:

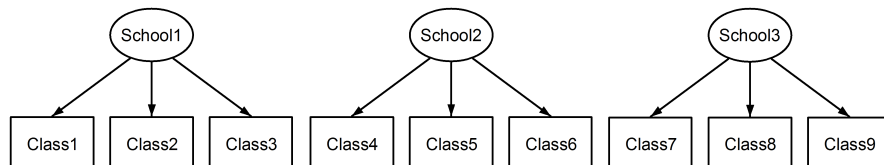
```
(1|School/Class)
```

and

```
(1|School) + (1|Class)
```

respectively. Due to the ambiguity of whether there is nesting or crossing of random effects, it is very important to specify the model correctly as these models will produce different results, as we shall show below. Moreover, it is not possible to know, just by inspecting the data, whether we have nested or crossed random effects. This can only be determined with knowledge of the data *and* the experimental design.

But first let us consider a case where the Class variable is coded uniquely across schools:



There is no longer any ambiguity concerning nesting or crossing. The nesting is explicit. Let us now see this with an example in R, where we have 6 schools (labelled `s1` to `s6`) and 4 classes within each school (labelled `a` to `d`):

```

      class
school a b c d
  I  50 50 50 50
  II  50 50 50 50
  III 50 50 50 50
  IV  50 50 50 50
  V   50 50 50 50
  VI  50 50 50 50

```

We can see from this cross tabulation that every class ID appears in every school, which satisfies your definition of crossed random effects (in this case we have *fully*, as opposed to *partially*, crossed random effects, because every class occurs in every school). So this is the same situation that we had in the first figure above. However, if the data are really nested and not crossed, then we need to explicitly tell lme4 :

```

> m0 <- lmer(extro ~ open + agree + social + (1 | school/class), data = dt)
> summary(m0)

Random effects:
Groups      Name      Variance Std.Dev.
class:school (Intercept)  8.2043  2.8643
school      (Intercept) 93.8421  9.6872
Residual                    0.9684  0.9841
Number of obs: 1200, groups: class:school, 24; school, 6

Fixed effects:
              Estimate Std. Error t value
(Intercept) 60.2378227  4.0117909  15.015
open         0.0061065  0.0049636   1.230
agree       -0.0076659  0.0056986  -1.345
social       0.0005404  0.0018524   0.292

> m1 <- lmer(extro ~ open + agree + social + (1 | school) + (1 | class), data = dt)
> summary(m1)

Random effects:
Groups      Name      Variance Std.Dev.
school      (Intercept) 95.887   9.792
class       (Intercept)  5.790   2.406
Residual                    2.787   1.669
Number of obs: 1200, groups: school, 6; class, 4

Fixed effects:
              Estimate Std. Error t value
(Intercept) 60.198841  4.212974  14.289
open         0.010834  0.008349   1.298
agree       -0.005420  0.009605  -0.564
social      -0.001762  0.003107  -0.567

```

As expected, the results differ because `m0` is a nested model while `m1` is a crossed model.

Now, if we introduce a new variable for the class identifier:

```

> dt$classID <- paste(dt$school, dt$class, sep=".")
> xtabs(~ school + classID, dt)

      classID
school I.a I.b I.c I.d II.a II.b II.c II.d III.a III.b III.c III.d IV.a IV.b
  I    50  50  50  50  0   0   0   0   0   0   0   0   0   0   0
  II   0   0   0   0  50  50  50  50  0   0   0   0   0   0   0
  III  0   0   0   0  0   0   0   0  50  50  50  50  0   0   0
  IV   0   0   0   0  0   0   0   0  0   0   0   0  50  50  0
  V    0   0   0   0  0   0   0   0  0   0   0   0   0   0   0
  VI   0   0   0   0  0   0   0   0  0   0   0   0   0   0   0

      classID
school IV.c IV.d V.a V.b V.c V.d VI.a VI.b VI.c VI.d
  I      0   0   0   0  0   0  0   0   0   0   0
  II     0   0   0   0  0   0  0   0   0   0   0
  III    0   0   0   0  0   0  0   0   0   0   0
  IV    50  50   0   0  0   0  0   0   0   0   0
  V     0   0  50  50  50  50  0   0   0   0   0
  VI     0   0   0   0  0   0  50  50  50  50  0

```

The cross tabulation shows that each level of class occurs only in one level of school, as per your definition of nesting. This is also the case with your data, however it is difficult to show that with your data because it is very sparse. Both model formulations will now produce the same output (that of the nested model `m0` above):

```

> m2 <- lmer(extro ~ open + agree + social + (1 | school/classID), data = dt)
> summary(m2)

Random effects:
Groups      Name      Variance Std.Dev.

```

```
      Estimate Std. Error t value
(Intercept) 60.2378227  4.0117882  15.015
open         0.0061065  0.0049636   1.230
agree        -0.0076659  0.0056986  -1.345
social        0.0005404  0.0018524   0.292

> m3 <- lmer(extro ~ open + agree + social + (1 | school) + (1 | classID), data = dt)
> summary(m3)

Random effects:
Groups      Name      Variance Std.Dev.
classID     (Intercept)  8.2043  2.8643
school      (Intercept) 93.8419  9.6872
Residual                    0.9684  0.9841
Number of obs: 1200, groups: classID, 24; school, 6

Fixed effects:
      Estimate Std. Error t value
(Intercept) 60.2378227  4.0117882  15.015
open         0.0061065  0.0049636   1.230
agree        -0.0076659  0.0056986  -1.345
social        0.0005404  0.0018524   0.292
```

It is worth noting that crossed random effects do not have to occur within the same factor - in the above the crossing was completely within school. However, this does not have to be the case, and very often it is not. For example, sticking with a school scenario, if instead of classes within schools we have pupils within schools, and we were also interested in the doctors that the pupils were registered with, then we would also have nesting of pupils within doctors. There is no nesting of schools within doctors, or vice versa, so this is also an example of crossed random effects, and we say that schools and doctors are crossed. A similar scenario where crossed random effects occur is when individual observations are nested within two factors simultaneously, which commonly occurs with so-called repeated measures *subject-item* data. Typically each subject is measured/tested multiple times with/on different items and these same items are measured/tested by different subjects. Thus, observations are clustered within subjects *and* within items, but items are not nested within subjects or vice-versa. Again, we say that subjects and items are *crossed*.

Summary: TL;DR

The difference between crossed and nested random effects is that nested random effects occur when one factor (grouping variable) appears only within a particular level of another factor (grouping variable). This is specified in `lme4` with:

```
(1|group1/group2)


where group2 is nested within group1 .
```

Crossed random effects are simply: *not nested*. This can occur with three or more grouping variables (factors) where one factor is separately nested in both of the others, or with two or more factors where individual observations are nested separately within the two factors. These are specified in `lme4` with:

```
(1|group1) + (1|group2)
```

edited May 30 at 7:06

answered Aug 8 '16 at 16:04



Robert Long
7,679 1 21 45

- 13 +6, this is a really nice answer. Remind me in a couple of days & I'll put a bounty on it. – gung ♦ Aug 8 '16 at 17:32
- 1 Excellent explanation. Is there a name for a possibly third type of experimental design where you create unique levels for two factors (possibly nested), so if I followed cohorts purchasing motorcycles from different dealers within different cities, how would I communicate to readers that I've basically created a single level factor `interaction(city, dealer)` ? – AdamO Aug 8 '16 at 19:12
- @AdamO I'm not sure about that. What do you mean by "followed" (are there repeated measures?) I'd need a bit more info and ideally some sample data. – Robert Long Aug 9 '16 at 5:51
- 1 +6, this is such a great answer that I felt I could top up @gung's bounty with another one. By the way, I am struggling with what probably is a very basic confusion ([here is my Q](#)) and would very much appreciate your help. – amoeba Aug 28 '16 at 14:32
- 1 Update to my previous comment. I have checked two old & authoritative books on experimental design: Maxwell and Delaney, 1990, *Designing Experiments and Analyzing Data*, and Montgomery, 1976, *Design and Analysis of Experiments*. Both talk about nested designs and both talk about repeated measures aka split-plots; both use the word "nested" *only* in the sense you are using it and never refer to repeated measures as nested. So I don't think there is any terminology difference after all. – amoeba Aug 30 '16 at 21:31