Interpreting Contextual and Compositional Effects

2022-03-22

Contents

1 *Introduction*

This is an example using only the public school data from the 'hs' data set in 'spida2'.

We will see that:

> Including the contextual mean of 'ses' in each school in the model with `cvar(ses, id)` along with 'ses' itself allows you to estimate both the within-school and the between-school 'effects' of 'ses'.

Consider three fixed effects models along with a random intercept:

- `mathach ~ 1 + ses + cvar(ses, school)`
- `mathach ~ 1 + dvar(ses, school) + cvar(ses, school)`
- `mathach ~ 1 + ses`

2 Setup

```
library(spida2)
library(nlme)
```

```

Attaching package: 'nlme'

The following object is masked from 'package:spida2':

    getData
```

```
library(car)
```

```
Loading required package: carData
```

3 4 models

```
fit_contextual <-
  hs %>%
  subset(Sector == 'Public') %>%
  lme(mathach ~ 1 + ses + cvar(ses, school), ., random = ~ 1 | school)

fit_compositional <-
  hs %>%
  subset(Sector == 'Public') %>%
  lme(mathach ~ 1 + dvar(ses,school) + cvar(ses, school), ., random = ~ 1 | school)

fit_single_predictor <-
  hs %>%
  subset(Sector == 'Public') %>%
  lme(mathach ~ ses, ., random = ~ 1 | school)

fit_pooled <-
  hs %>%
  subset(Sector == 'Public') %>%
  lm(mathach ~ ses, .)
```

4 Contextual model

```
summary(fit_contextual)$tTable
```

```
                      Value Std.Error  DF   t-value       p-value
(Intercept)       12.425644 0.3692210 813 33.653679 3.296983e-156
ses                2.902798 0.3436057 813  8.448049  1.357154e-16
cvar(ses, school)  3.511982 0.8784138  17  3.998096  9.310398e-04
```

5 *Compositional model*

```
summary(fit_compositional)$tTable
```

```
                      Value Std.Error  DF   t-value       p-value
(Intercept)       12.425644 0.3692210 813 33.653679 3.296983e-156
dvar(ses, school)  2.902798 0.3436057 813  8.448049  1.357154e-16
cvar(ses, school)  6.414780 0.8084219  17  7.934941  4.078418e-07
```

6 *Single predictor model*

```
summary(fit_single_predictor)$tTable
```

```
                 Value Std.Error  DF   t-value      p-value
(Intercept) 11.945822 0.4760807 813 25.092008 2.42112e-103
ses          3.226578 0.3277559 813  9.844453  1.11434e-21
```

7 OLS pooled model

```
summary(fit_pooled) $ coefficients  # for lm fits
```

```
              Estimate Std. Error  t value      Pr(>|t|)
(Intercept) 12.091298  0.2331223 51.86675 8.843077e-263
ses          3.904946  0.2973014 13.13464  6.057271e-36
```

8 *Comparison of coefficients*

```
compareCoefs(fit_contextual, fit_compositional, fit_single_predictor, fit_pooled)
```

```
Warning in compareCoefs(fit_contextual, fit_compositional,
fit_single_predictor, : models to be compared are of different classes

Calls:
1: lme.formula(fixed = mathach ~ 1 + ses + cvar(ses, school), data = .,
  random = ~1 | school)
2: lme.formula(fixed = mathach ~ 1 + dvar(ses, school) + cvar(ses, school),
  data = ., random = ~1 | school)
3: lme.formula(fixed = mathach ~ ses, data = ., random = ~1 | school)
4: lm(formula = mathach ~ ses, data = .)

                  Model 1 Model 2 Model 3 Model 4
(Intercept)        12.426  12.426  11.946  12.091
SE                  0.369   0.369   0.476   0.233

ses                 2.903           3.227   3.905
SE                  0.344           0.328   0.297

cvar(ses, school)   3.512   6.415
SE                  0.878   0.808

dvar(ses, school)           2.903
SE                          0.344
```

9 *Estimating compositional effect from the contextual model and vice-versa*

```
wald(fit_contextual,
     rbind(
       "within effect"        = c(0,1, 0),
       "contextual effect"    = c(0,0, 1),
       "compositional effect" = c(0,1, 1)))
```

```
     numDF denDF  F-value p-value
   1     2    17 67.16641 <.00001
                        Estimate Std.Error DF  t-value  p-value Lower 0.95
   within effect        2.902798 0.343606  813 8.448049 <.00001 2.228339
   contextual effect    3.511982 0.878414   17 3.998096 0.00093 1.658691
   compositional effect 6.414780 0.808422   17 7.934941 <.00001 4.709159
                        Upper 0.95
   within effect        3.577257
   contextual effect    5.365274
   compositional effect 8.120401
```

```
wald(fit_compositional,
     rbind(
       "within effect"        = c(0,1, 0),
       "contextual effect"    = c(0,-1,1),
       "compositional effect" = c(0,0, 1)))
```

```
     numDF denDF  F-value p-value
   1     2    17 67.16641 <.00001
                        Estimate Std.Error DF  t-value  p-value Lower 0.95
   within effect        2.902798 0.343606  813 8.448049 <.00001 2.228339
   contextual effect    3.511982 0.878414   17 3.998096 0.00093 1.658691
   compositional effect 6.414780 0.808422   17 7.934941 <.00001 4.709159
                        Upper 0.95
   within effect        3.577257
   contextual effect    5.365274
   compositional effect 8.120401
```

10 Practical Implications for Statistical Analyses

Notes:

- Thinking about *between effects* by using the contextual variable `cvar(X, cluster)` is only meaningful if the cluster means of X vary systematically between groups. With a **balanced variable** where the the values of X are the same in each group, there is no *between effect* to estimate since you can't estimate the 'effect' of a constant (unless you can justify dropping the intercept term). So don't bother with 'cvar' for a balanced variable.
- Introducing `cvar(X, cluster)` has two main purposes:
  1. to be able to estimate the between-cluster relationship, and
  2. to be able to estimate the within-cluster relationship in a way that is unbiased by the between-cluster relationship.
- For model parsimony you might want to consider dropping the contextual variable. You can drop the *contextual variable* if if the true *contextual effect* is zero. You can test this hypothesis with the coefficient of `cvar(X, cluster)` in the model
  `Y ~ X + cvar(X, cluster)`
  Note that this is **NOT** the same as testing the coefficient of `cvar(X, cluster)` in the model
  `Y ~ dvar(X, cluster) + cvar(X, cluster)`
- Some analysts will first fit:
  `Y ~ X + cvar(X, cluster)`
  then consider whether the coefficient of
  `cvar(X, cluster)` is small enough to warrant dropping it.
  If they don't drop it, they would switch to the equivalent model
  `Y ~ dvar(X, cluster) + cvar(X, cluster)`
  which, usually, has better numerical properties due to lower collinearity.
- Often, a motivation to get an unbiased estimate of the within-effect is that it estimates the 'effect of X' controlling for potential confounders, measured or not, known or not, that are constant within level-1 units. Thus, including a contextual mean may provide an unbiased estimate of the within-cluster causal effect of X.