

Contextual Variables: Contextual and Compositional effects

MATH 4939

2026-03-01

Estimating within-group and between-group effects with mixed models

Let $\hat{\beta}_X^{MM}$ be the mixed-model estimate of the effect of X .

Then $\hat{\beta}_X^{MM}$ is an unbiased estimate of β_X^{Within} only if $\beta_X^{Within} = \beta_X^{Between}$ in the population. This is an assumption that is generally unlikely to be valid.

We will see:

- Using the group-mean¹ of X : $\bar{X}_j$ and the within-group deviation² of X : $X_{ij} - \bar{X}_j$ (also referred by other terms, such as CWG: centered-within-group variable)/
to estimate to estimate β_X^{Within} and $\beta_X^{Between}$.

¹which can generated using `spida2::cvar(X,group)`

²which can generated using `spida2::dvar(X,group)`

- The ‘parallelogram law’:

$$\beta_X^{Between} = \beta_X^{Within} + \beta_X^{Contextual}$$

- How the fixed-effects model can use $X_{ij} - \bar{X}_j$ or X_{ij} equivalently with $\bar{X}_j$, but which one you use in the random effects portion of the model can make a large difference.

Beware:

- The terms ‘contextual’ and ‘compositional’ are not used consistently by different authors – or even by the same author. I like ‘between’ as a synonym for ‘compositional’ because it isn’t ambiguous, but the world prefers words that sound more serious.
- Adding to the confusion (or creating it) is that the ‘effect’ of a variable (a predictor in a model) depends entirely on the other variables in the model. That is, the meaning of $\frac{\partial}{\partial X}E(Y)$ depends on the other variables because the choice of other variables determines the path along which predictors change as X changes. The choice determines the direction of $\frac{\partial}{\partial X}E(Y)$ viewed as a directional derivative.
- The *effect* of a *contextual variable* for a variable X is the *contextual effect* **if** the *raw* value of the variable is another variable in the model.
- The *effect* of a *contextual variable* is the *compositional effect* **if** the *centered within groups* value of the variable is another variable in the model.
- If you change a variable, X , in a model (e.g. centre it), you can change the meaning of the effects of all the other variables, unless X and its modified value remain within the orthogonal complement of the other variables.
- Try to retain the ideas, not just the terminology. It’s easier to apply the ideas in a new context.

This is an example using the Public schools in the `hs` data set in the **spida2** package.

```
library(spida2)
library(nlme)      # this is a version of nlme tailor-made for a specific purpose
```

This is a modified version of `nlme` that includes a `pdInd` class. See `?pdInd`

It can be installed with: `remotes::install_github('gmonette/nlme')`

To suppress this message load `nlme` with `'suppressPackageStartupMessages(library(nlme))'`

Attaching package: 'nlme'

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

`getData`

```
library(lattice)
library(latticeExtra)
library(car)
```

Loading required package: `carData`

```
hsp <- subset(hs, Sector == "Public")    # Public schools from the hs data set
```

```
caption <- 'Public schools with data ellipses and centroids'
```

```
p <- xyplot(mathach ~ ses, hsp, groups = school, pch = 1, cex = .2) +
```

```
layer(panel.ellipse(..., center.pch = 21, center.cex = 1, fill = '#444444'))
```

p

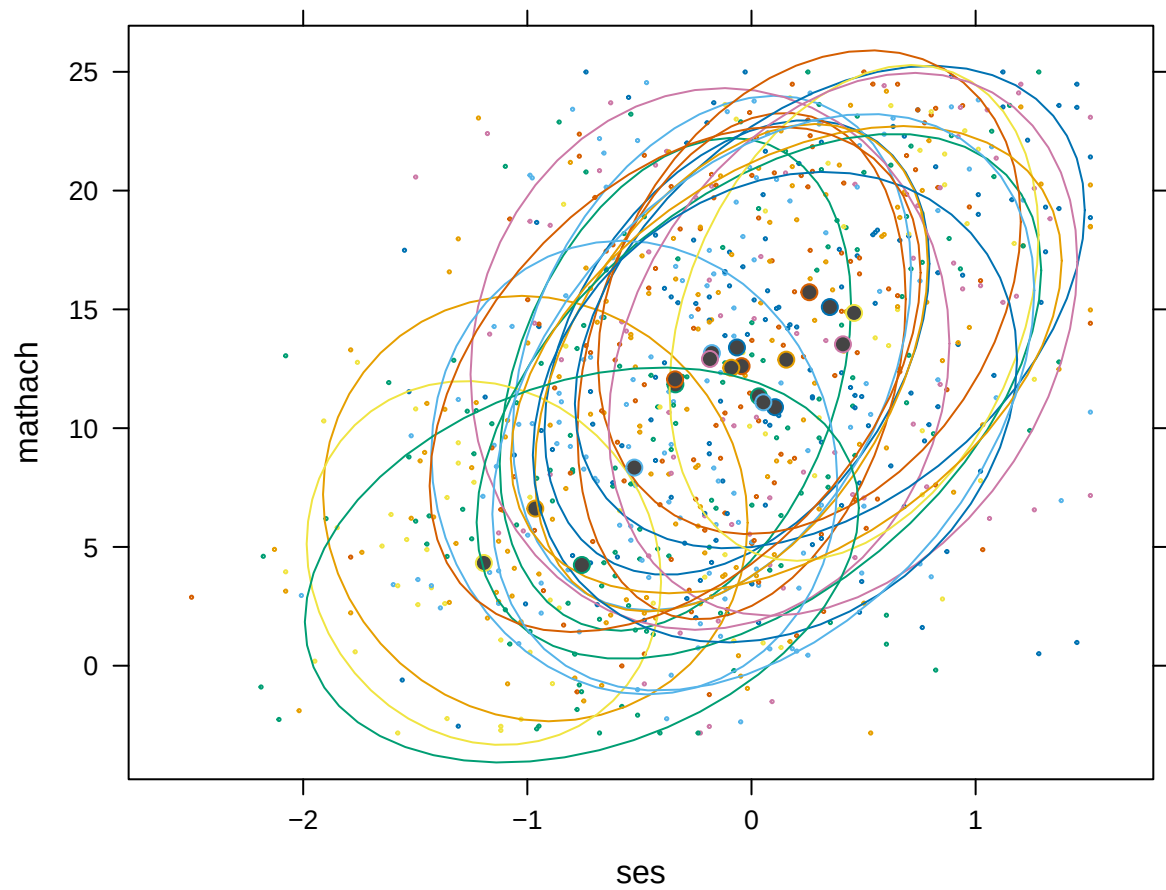


Figure 1: Public schools with data ellipses and centroids

Pooled, within and between models:

```
fit.pooled <- lm(mathach ~ ses, hsp)
fit.within <- lm(mathach ~ ses + factor(school), hsp) # What happens if we drop 'factor'?
fit.between <- lm(mathach ~ cvar(ses, school), hsp)

fit.list <- list(pooled = fit.pooled, within = fit.within, between = fit.between)
fit.list %>% lapply(coef) %>% lapply(`[, 2]` %>% cbind %>%
  {colnames(.) <- 'ses' ; .})
```

```
      ses
pooled  3.904946
within  2.902798
between 6.296687
```

```
hsp <- within(
  hsp,
  {
    pooled <- predict(fit.pooled)
    within <- predict(fit.within)
    between <- predict(fit.between)
  }
)
```

```
caption <- "Within school regression lines using OLS"
p + xyplot(within ~ ses, hsp, groups = school, type = 'l') +
  layer(panel.abline(v = 0))
```

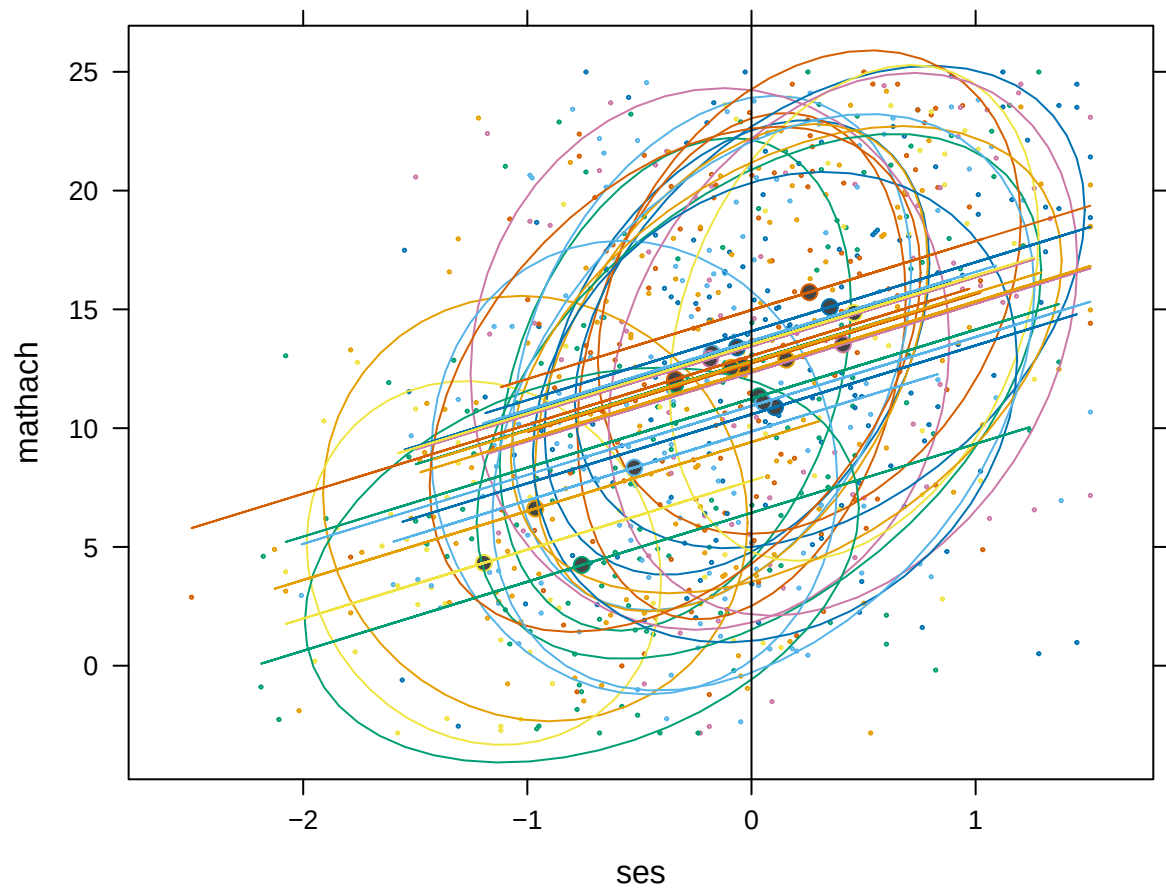


Figure 2: Within school regression lines using OLS

```
caption = "Pooled regression line using OLS"
p + xyplot(within ~ ses, hsp, groups = school, type = 'l') +
  xyplot(pooled ~ ses, hsp, type = 'l', lwd = 3) +
  layer(panel.abline(v = 0))
```

```
caption = "Between school regression line using OLS"
p + xyplot(within ~ ses, hsp, groups = school, type = 'l') +
  xyplot(pooled ~ ses, hsp, type = 'l', lwd = 3) +
  xyplot(between ~ cvar(ses,school), hsp, type = 'l', lwd = 4, col = 'red') +
  layer(panel.abline(v = 0))
```

What does a mixed model do?

```
fit.mm <- lme(mathach ~ ses, hsp, random = ~ 1 | school)
fit.mm
```

Linear mixed-effects model fit by REML

Data: hsp

Log-restricted-likelihood: -2744.814

Fixed: mathach ~ ses

(Intercept) ses

11.945822 3.226578

Random effects:

Formula: ~1 | school

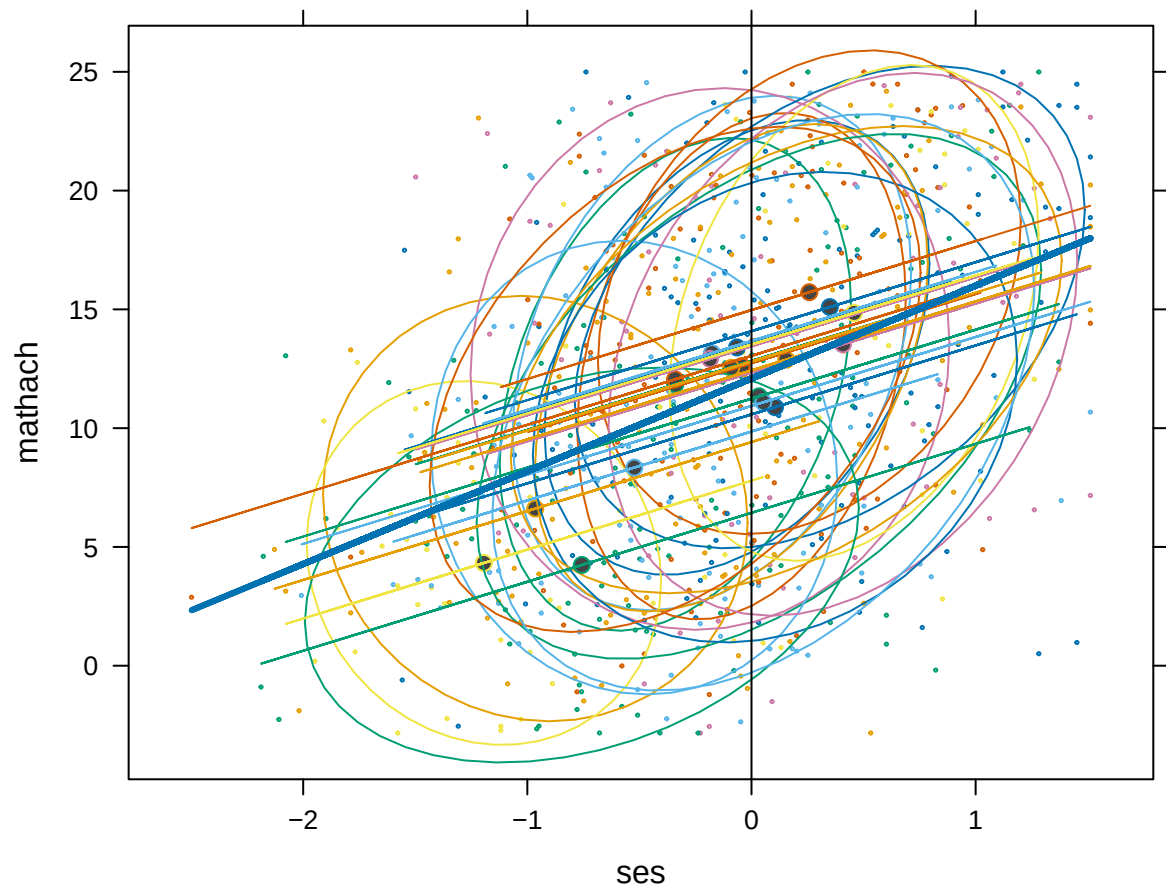


Figure 3: Pooled regression line using OLS

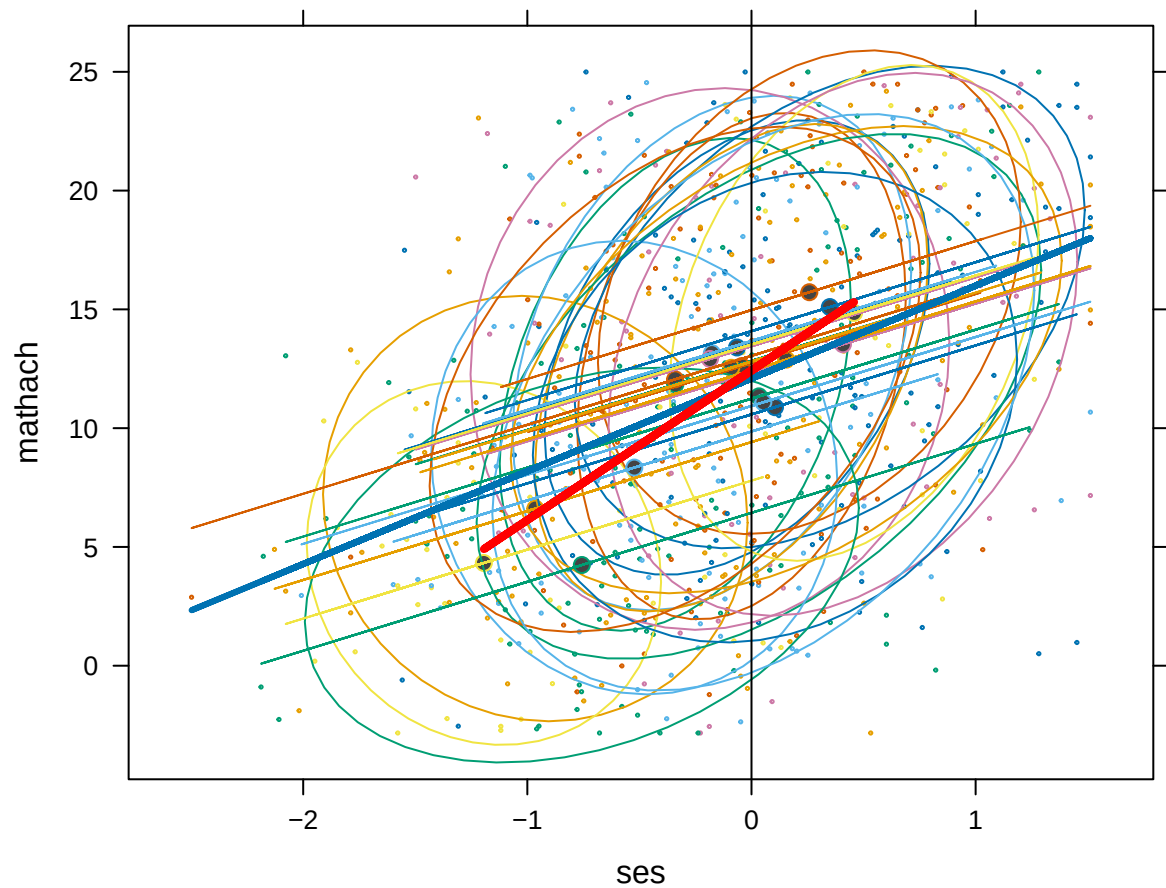


Figure 4: Between school regression line using OLS

(Intercept) Residual
StdDev: 1.81293 6.425983

Number of Observations: 833
Number of Groups: 19

```
fit.list <- c(fit.list, list(mixed = fit.mm))

(
  fit.list %>%
  lapply(
    function(fit) {
      if(class(fit)=='lm') coef(fit) else fixef(fit)
    }
  ) %>%
  sapply(`[, 2]` %>% cbind %>%
  {colnames(.) <- 'ses' ; .} %>%
  as.data.frame %>%
  round(3) %>%
  sortdf -> coef.table
)
```

	ses
within.ses	2.903
mixed.ses	3.227
pooled.ses	3.905

```
between.cvar(ses, school) 6.297
```

Using mixed models to estimate within, between and contextual effects

We get equivalent models by using

- raw ses and the within-group mean of ses, or
- centered-within-group ses and the within-group mean of ses

```
fit.contextual <- lme(mathach ~ ses + cvar(ses, school), hsp, random = ~ 1 | school)
fit.compositional <- lme(mathach ~ dvar(ses, school) + cvar(ses, school),
                        hsp, random = ~ 1 | school)
wald(fit.contextual)
```

	numDF	denDF	F-value	p-value					
	3	17	404.3528	<.00001					
			Estimate	Std.Error	DF	t-value	p-value	Lower	0.95
(Intercept)			12.425644	0.369221	813	33.653679	<.00001	11.700906	
ses			2.902798	0.343606	813	8.448049	<.00001	2.228339	
cvar(ses, school)			3.511982	0.878414	17	3.998096	0.00093	1.658691	
			Upper	0.95					
(Intercept)			13.150383						
ses			3.577257						
cvar(ses, school)			5.365274						

```
wald(fit.compositional)
```

```

numDF denDF  F-value p-value
    3    17 404.3528 <.00001

              Estimate Std.Error DF  t-value  p-value Lower 0.95
(Intercept)   12.425644 0.369221 813 33.653679 <.00001 11.700906
dvar(ses, school) 2.902798 0.343606 813  8.448049 <.00001  2.228339
cvar(ses, school) 6.414780 0.808422  17  7.934941 <.00001  4.709159

              Upper 0.95
(Intercept)   13.150383
dvar(ses, school) 3.577257
cvar(ses, school) 8.120401

```

Note that variance estimates are the same, although the X matrix of the contextual model is more collinear and, in extreme cases, the fitting algorithm for the mixed model is more likely to produce numerically flawed results.

```
summary(fit.contextual)
```

```
Linear mixed-effects model fit by REML
```

```
Data: hsp
```

```
      AIC      BIC    logLik
5485.311 5508.918 -2737.655
```

```
Random effects:
```

```
Formula: ~1 | school
```

```
(Intercept) Residual
```

```
StdDev:    1.165459 6.422254
```

```
Fixed effects:  mathach ~ ses + cvar(ses, school)
                Value Std.Error   DF  t-value p-value
(Intercept)    12.425644 0.3692210  813 33.65368   0e+00
ses             2.902798 0.3436057  813  8.44805   0e+00
cvar(ses, school) 3.511982 0.8784138   17  3.99810   9e-04
```

```
Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.64983777 -0.77062095  0.03752833  0.73598158  2.69913720
```

```
Number of Observations: 833
Number of Groups: 19
```

```
summary(fit.compositional)
```

```
Linear mixed-effects model fit by REML
```

```
Data: hsp
      AIC      BIC    logLik
5485.311 5508.918 -2737.655
```

```
Random effects:
Formula: ~1 | school
      (Intercept) Residual
StdDev:    1.165459 6.422254
```

```
Fixed effects:  mathach ~ dvar(ses, school) + cvar(ses, school)
```

	Value	Std.Error	DF	t-value	p-value
(Intercept)	12.425644	0.3692210	813	33.65368	0
dvar(ses, school)	2.902798	0.3436057	813	8.44805	0
cvar(ses, school)	6.414780	0.8084219	17	7.93494	0

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.64983777	-0.77062095	0.03752833	0.73598158	2.69913720

Number of Observations: 833

Number of Groups: 19

Easier comparisons:

`wald(fit.contextual)`

	numDF	denDF	F-value	p-value
	3	17	404.3528	<.00001

	Estimate	Std.Error	DF	t-value	p-value	Lower 0.95
(Intercept)	12.425644	0.369221	813	33.653679	<.00001	11.700906
ses	2.902798	0.343606	813	8.448049	<.00001	2.228339
cvar(ses, school)	3.511982	0.878414	17	3.998096	0.00093	1.658691

Upper 0.95

(Intercept)	13.150383
ses	3.577257
cvar(ses, school)	5.365274

```
wald(fit.compositional)
```

```
      numDF denDF  F-value p-value
      3      17 404.3528 <.00001

      Estimate Std.Error DF  t-value  p-value Lower 0.95
(Intercept)   12.425644 0.369221  813 33.653679 <.00001 11.700906
dvar(ses, school) 2.902798 0.343606  813  8.448049 <.00001  2.228339
cvar(ses, school) 6.414780 0.808422   17  7.934941 <.00001  4.709159

      Upper 0.95
(Intercept)   13.150383
dvar(ses, school) 3.577257
cvar(ses, school) 8.120401
```

You can get the within and between estimates from both models

```
L.cntx <- rbind(
  'compositional (between) effect' = c(0, 1, 1),
  'within effect' = c(0,1,0),
  'contextual effect' = c(0,0,1)
)
L.comp <- rbind(
  'compositional (between) effect' = c(0, 0, 1),
  'within effect' = c(0,1,0),
  'contextual effect' = c(0,-1,1)
)
L.cntx
```

	[,1]	[,2]	[,3]
compositional (between) effect	0	1	1
within effect	0	1	0
contextual effect	0	0	1

L.comp

	[,1]	[,2]	[,3]
compositional (between) effect	0	0	1
within effect	0	1	0
contextual effect	0	-1	1

wald(fit.contextual, L.cntx)

	numDF	denDF	F-value	p-value
	1	2	17	67.16641

	Estimate	Std.Error	DF	t-value	p-value
compositional (between) effect	6.414780	0.808422	17	7.934941	<.00001
within effect	2.902798	0.343606	813	8.448049	<.00001
contextual effect	3.511982	0.878414	17	3.998096	0.00093

	Lower 0.95	Upper 0.95
compositional (between) effect	4.709159	8.120401
within effect	2.228339	3.577257
contextual effect	1.658691	5.365274

```
wald(fit.compositional, L.comp)
```

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

Comparing with the OLS estimates:

```
rbind(  
  wald(fit.within)[2, 1:2],  
  wald(fit.between)[2,1:2]  
)
```

	coef	se
ses	2.902798	0.3435176
cvar(ses, school)	6.296687	0.5601504

The contextual model and the compositional model ³ are **equivalent** since they have X matrices that span the same

³Even models can be contextual or compositional – from a computer science perspective this is an example of *overloading* or *ad hoc*

space. (**Exercise:** Prove it.)

Using the contextual model:

$$E(Y) = \psi_0 + \psi_1 X_{ij} + \psi_2 \bar{X}_j$$

With the compositional model:

$$E(Y) = \phi_0 + \phi_1 (X_{ij} - \bar{X}_j) + \phi_2 \bar{X}_j$$

Rearranging terms:

$$E(Y) = \phi_0 + \phi_1 X_{ij} + (\phi_2 - \phi_1) \bar{X}_j$$

Thus, the two expressions yield identical functions of X_{ij} and $\bar{X}_j$ iff

$$\psi_0 = \phi_0$$

$$\psi_1 = \phi_1$$

$$\psi_2 = \phi_2 - \phi_1$$

Equivalently:

$$\phi_2 = \psi_2 + \psi_1$$

polymorphism, closely related to the use of *generic functions* and *methods* in R

Note that:

$$\begin{aligned}\beta_0 &= \psi_0 = \phi_0 \\ \beta_{Within} &= \psi_1 = \phi_1 \\ \beta_{Contextual} &= \psi_2 = \phi_2 - \phi_1 \\ \beta_{Compositional} &= \beta_{Between} = \phi_2 = \psi_1 + \psi_2\end{aligned}$$

I think of this as the ‘parallelogram law’⁴. Imagine two schools with mean ses one unit apart, for example, 1 and 2. Suppose that $\beta_{Within} = 1/2$ and $\beta_{Between} = 3$. Let $\beta_0 = 1$.

⁴I just call it that because I think that giving ideas names that evoke a geometric mnemonic makes them harder to forget.

The compositional and the contextual models have equivalent fixed effects models:

$$\mathbf{X}_{Contextual} = \begin{bmatrix} 1 & X_{ij} & \bar{X}_{.j} \end{bmatrix}$$

$$\mathbf{X}_{Compositional} = \begin{bmatrix} 1 & (X_{ij} - \bar{X}_{.j}) & \bar{X}_{.j} \end{bmatrix}$$

Exercises:

- Show that $\mathcal{L}(\mathbf{X}_{Contextual}) = \mathcal{L}(\mathbf{X}_{Compositional})$
- Compare the within- and between-group estimates obtained with OLS and with mixed models. Compare the SEs of estimates. Are the differences small or large?

CWG vs raw variable for random effects

The models we used above did not include a random effect for X .

As far as possible mixed models should include random effects for level-1 variables. However, the number of parameters in the G matrix grows rapidly with additional random effects and the number of effects that can be included is limited unless one has an enormous amount of data.

Consider the following models where we use the verbose mode to see how the models converge:

```
fit1 <- lme(mathach ~ ses + cvar(ses,school), hsp, random = ~ 1 + ses | school,
           control = list(msVerbose = TRUE))
```

```
0:      4349.1922:   1.78170   2.13329 -5.25858
1:      4349.1747:   1.79280   2.47041 -5.27724
2:      4349.1641:   1.75189   2.45974 -5.28036
```

3:	4349.1592:	1.76856	2.32288	-5.31980
4:	4349.1582:	1.76364	2.35437	-5.33859
5:	4349.1574:	1.76273	2.36213	-5.37475
6:	4349.1487:	1.76561	2.42424	-5.92297
7:	4349.1416:	1.77706	2.45463	-6.47376
8:	4349.1267:	1.82164	2.49880	-8.07485
9:	4349.1196:	1.86750	2.51561	-9.39664
10:	4349.1168:	1.90783	2.52224	-10.4689
11:	4349.1164:	1.92948	2.52574	-11.0275
12:	4349.1163:	1.93635	2.52618	-11.2200
13:	4349.1163:	1.93849	2.52864	-11.2762
14:	4349.1163:	1.93814	2.52841	-11.2763
15:	4349.1163:	1.93791	2.52845	-11.2767
16:	4349.1163:	1.93790	2.52883	-11.2775
17:	4349.1163:	1.93820	2.52984	-11.2953
18:	4349.1163:	1.93863	2.53102	-11.3198

```
summary(fit1)
```

Linear mixed-effects model fit by REML

Data: hsp

AIC	BIC	logLik
5488.887	5521.937	-2737.444

Random effects:

Formula: ~1 + ses | school

```

Structure: General positive-definite, Log-Cholesky parametrization
              StdDev      Corr
(Intercept)  1.2417956 (Intr)
ses          0.5102312 0.669
Residual     6.4117298

```

```

Fixed effects: mathach ~ ses + cvar(ses, school)
              Value Std.Error DF t-value p-value
(Intercept)   12.374865 0.3881973 813 31.87777 0e+00
ses           2.921401 0.3642758 813  8.01975 0e+00
cvar(ses, school) 3.387174 0.8467063 17  4.00041 9e-04

```

```

Standardized Within-Group Residuals:
              Min              Q1              Med              Q3              Max
-2.65517053 -0.78396082  0.03691098  0.73850988  2.71116336

```

```

Number of Observations: 833
Number of Groups: 19

```

```

fit2 <- lme(mathach ~ ses + cvar(ses,school), hsp, random = ~ 1 + dvar(ses,school) | school,
            control = list(msVerbose = TRUE))

```

```

0:      4348.5861:  1.87756  1.87101 -4.38434
1:      4348.5831:  1.88780  1.90494 -4.38993
2:      4348.5827:  1.85703  1.91845 -4.40247
3:      4348.5803:  1.87137  1.92672 -4.43430

```

4:	4348.5793:	1.88097	1.91015	-4.46464
5:	4348.5759:	1.89488	1.91256	-4.57870
6:	4348.5643:	1.93322	1.92316	-5.03671
7:	4348.5461:	2.00126	1.94641	-6.09388
8:	4348.5356:	2.04819	1.96727	-7.04494
9:	4348.5309:	2.08050	1.99389	-7.85196
10:	4348.5298:	2.09812	2.01087	-8.32656
11:	4348.5297:	2.10644	2.01783	-8.53746
12:	4348.5297:	2.10940	2.01939	-8.59990
13:	4348.5297:	2.11063	2.01962	-8.61969
14:	4348.5297:	2.11131	2.01953	-8.62755

```
summary(fit2)
```

Linear mixed-effects model fit by REML

Data: hsp

	AIC	BIC	logLik
	5487.714	5520.764	-2736.857

Random effects:

Formula: ~1 + dvar(ses, school) | school

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	1.1783112	(Intr)
dvar(ses, school)	0.8495914	0.753
Residual	6.4014627	

```
Fixed effects:  mathach ~ ses + cvar(ses, school)
                Value Std.Error   DF  t-value p-value
(Intercept)    12.330098 0.3695787  813  33.36258  0.0000
ses             2.923381 0.3978351  813   7.34822  0.0000
cvar(ses, school) 2.885386 0.8703697   17   3.31513  0.0041
```

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.66426512	-0.78890506	0.04316219	0.75018313	2.70897716

Number of Observations: 833

Number of Groups: 19

There wasn't a large difference in the number of iterations in this case but with larger more complex models there can be a considerable difference. Sometimes, one or the other model will fail to converge.

Notice that there is a sensible difference in the estimate of the between effect.

We will see some reasons for this when we take up a deeper dive into the interpretation of the G matrix in the next section.