Multi-Level Modeling Centering Predictors

Andrew Zieffler

Educational Psychology

University of Minnesota

Driven to Discover℠

Read in and Prepare Data for these Notes

```
# Load foreign package to be able to read in SPSS data
> library(foreign)

# Read in the level-1 (player-level) data
> nbaL1 = read.spss(file = "http://www.tc.umn.edu/~zief0002/data/nbaLevel1.sav",
    to.data.frame = TRUE)

# Read in the level-2 (team-level) data
> nbaL2 = read.spss(file = "http://www.tc.umn.edu/~zief0002/data/nbaLevel2.sav",
    to.data.frame = TRUE)

# Merge nbaL2 into nbaL1 using the Team_ID variable
> nba = merge(nbaL1, nbaL2, by = "Team_ID")

# Load libraries
> library(ggplot2)
> library(lmerTest)
```

Centering the Predictors

When centering predictors in a multi-level model you have two choices...

- Grand mean centering
- Group mean centering

Grand mean centering: $X - \bar{X}$

Group mean centering: $X - \bar{X}_j$

The choice of centering will affect interpretation of the fixed-effects, and possibly, the random effects.

Grand Mean Centering

```
# Center predictor using grand mean
> nba$so5 = nba$Shots_on_five - mean(nba$Shots_on_five)

> lmer.b2 = lmer(Life_Satisfaction ~ 1 + so5 + (1 + so5 | Team_ID), data = nba)

> summary(lmer.b2)

REML criterion at convergence: 1379

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Team_ID  (Intercept) 1.18501  1.0886
          so5         0.09913  0.3148   1.00
 Residual             5.10616  2.2597
Number of obs: 300, groups: Team_ID, 30

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  14.6184     0.2402 13.9440   60.85   <2e-16 ***
so5           3.2887     0.1340 27.0580   24.55   <2e-16 ***
```

	Parameter	Raw	Centered Grand Mean
Fixed effects			
Intercept	Intercept	6.43 (0.32)	14.62 (0.24)
Shots_on_five	Intercept	3.29 (0.13)	3.29 (0.13)
Variance components			
Level-1	Within-persons	5.11	5.11
Level-2	Intercepts	0.09	1.19
	Slopes	0.10	0.10
Pseudo R^2 statistics and Goodness-of-fit			
	$R^2_{Y,\hat{Y}}$	0.84	0.84
	R^2_{ϵ}	0.650	0.650
	Deviance	1379.00	1379.00
	AIC	1391.02	1391.02
	BIC	1413.241	1413.241

Fixed and Random Effects

- Estimated intercepts are different (like OLS) between the raw score model and the grand mean centered model.
- Because of this, the variance component for intercepts will also be different

The raw score model and the grand mean centered model are **equivalent** linear models (they produce the same global measures of fit and the same predicted values).

Interpreting the fixed-effects

Level-1: $\hat{Y}_i = \hat{\beta}_0^* + \hat{\beta}_1^*(X - \bar{X})$

Level-2: $\hat{\beta}_0^* = \hat{\beta}_0 + b_{0j}$

$\hat{\beta}_1^* = \hat{\beta}_1 + b_{1j}$

Fixed-effects portion of the composite model

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1(X - \bar{X})$$

Interpretation of Intercept

The intercept coefficient is the average life satisfaction for all players who have a shooting success (`Shots_on_5`) equal to the grand mean (the average shooting success in the sample...across all teams).

Interpretation of the slope

The slope coefficient is the predicted difference in life satisfaction for players who are one unit different in their shooting success (`Shots_on_5`)...same interpretation as in the raw score model.

In our example...

Predicted Level-1 Model

$$\hat{\text{Life Satisfaction}} = 14.62 + 3.29(\text{Shooting success} - \overline{\text{Shooting success}})$$

Interpretation of Intercept

The mean life satisfaction for players who have an average shooting success in the sample is 14.62.

Interpretation of the slope

Each one-unit difference in shooting success (`Shots_on_five`) is associated with a 3.29 unit change in life satisfaction, on average.

Interpreting the random-effects

Level-1: $$\hat{Y}_i = \hat{\beta}_0^* + \hat{\beta}_1^*(X - \bar{X})$$

Level-2: $$\hat{\beta}_0^* = \hat{\beta}_0 + b_{0j}$$

$$\hat{\beta}_1^* = \hat{\beta}_1 + b_{1j}$$

Complete composite model

$$\hat{Y}_i = \hat{\beta}_0 + b_{0j} + \left(\hat{\beta}_1 + b_{1j}\right)(X - \bar{X})$$

After substituting the grand mean of X into the predicted model.

$$\hat{Y}_i = \hat{\beta}_0 + b_{0j}$$

This is the predicted life satisfaction for a player who plays on team *j* and whose shooting success is at the overall average. This is sometimes referred to as an **adjusted mean**.

Interpretation of each RE for intercept

The b_{0j} estimate for each team is the difference in predicted life satisfaction between predicted life satisfaction for a player who plays on team *j* and whose shooting success is at the overall average. and a player whose hooting success is at the overall average

Interpretation of the Variance Component for the REs for intercept

The estimated variance in the b_{0j} estimates indicate the amount of variation in the team adjusted means.

Interpretation of the REs for slope and the variance for the REs for slope are the same as in the raw score model.

In our example...

```
# Get estimates of the random-effects
> ranef(lmer.b2)

$Team_ID
   (Intercept)         so5
01  0.27845168  0.08053478
02  1.29050189  0.37324353
03  1.37369664  0.39730541
04 -0.26348652 -0.07620651
05 -0.85077507 -0.24606418
06  0.61973897  0.17924310
```

There seems to be between-team variation in intercepts.

$$\hat{\sigma}_0^2 = 1.185$$

There seems to be between-team variation in slopes.

$$\hat{\sigma}_1^2 = 0.099$$

```
# Estimates of the variance-covariance matrix of the random effects
> varCorr(lmer.b2)$Team_ID

            (Intercept)        so5
(Intercept)    1.185014 0.34273402
so5            0.342734 0.09912675
attr(,"stddev")
(Intercept)         so5
   1.088584    0.314844
attr(,"correlation")
            (Intercept) so5
(Intercept)           1   1
so5                   1   1
```

$$\mathbf{G} = \begin{bmatrix} 1.185 & 0.343 \\ 0.343 & 0.099 \end{bmatrix}$$

Square roots of the variance estimates

The b_{0j} estimates and b_{1j} estimates are positively related. Teams that have a higher intercept also tend to have higher slopes.

Group Mean Centering

Compute the Group Means

```
> library(dplyr)

# Compute the mean for each team
> teams = nba %.%
    group_by(Team_ID) %.%
    summarise(meanShots = mean(Shots_on_five))

> head(teams)

  Team_ID meanShots
1      01       3.0
2      02       3.7
3      03       3.3
4      04       3.3
5      05       1.5
6      06       2.7
```

Merge the group means back into the level-1 data set

```
# Merge the team means with the nba data frame
> nba2 = merge(nba, teams, by = “Team_ID”)
```

Center the Predictor

```
# Center using the group means
> nba2$gcShots = nba2$Shots_on_five - nba2$meanShots

> head(nba2)

  Team_ID Shots_on_five Life_Satisfaction Coach_Experience meanShots gcShots
1      01             3            18.804                2         3       0
2      01             3            18.000                2         3       0
3      01             4            21.000                2         3       1
4      01             4            20.500                2         3       1
5      01             3            19.000                2         3       0
6      01             2            12.100                2         3      -1
```

Fit the Model

```
> lmer.b3 = lmer(Life_Satisfaction ~ 1 + gcShots + (1 + gcShots | Team_ID),
    data = nba2)

> summary(lmer.b3)

REML criterion at convergence: 1436.5

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Team_ID  (Intercept) 15.9589  3.9949
          gcShots      0.6645  0.8151   0.43
 Residual              4.6089  2.1468
Number of obs: 300, groups: Team_ID, 30

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  14.8067     0.7398 28.9980   20.01  < 2e-16 ***
gcShots       2.8892     0.2022 21.8030   14.29  1.5e-12 ***
```

	Parameter	Raw	Centered: Grand Mean	Centered: Group Mean
Fixed effects				
Intercept	Intercept	6.43 (0.32)	14.62 (0.24)	14.81 (0.74)
Shots_on_five	Intercept	3.29 (0.13)	3.29 (0.13)	2.89 (0.20)
Variance components				
Level-1	Within-persons	5.11	5.11	4.61
Level-2	Intercepts	0.09	1.19	15.96
	Slopes	0.10	0.10	0.66
Pseudo R^2 statistics and Goodness-of-fit				
	$R^2_{Y,\hat{Y}}$	0.84	0.84	0.865
	R^2_{ϵ}	0.650	0.650	0.684
	Deviance	1379.00	1379.00	1436.5
	AIC	1391.02	1391.02	1448.53
	BIC	1413.241	1413.241	1470.76

Fixed and Random Effects

- All of the estimated fixed-effects are different between the raw score model and the group mean centered model.
- The estimated random-effects are also very different.
- The pseudo R^2 and the goodness-of-fit measures are all different...

...The raw score model and the grand mean centered model are **not** equivalent linear models.

Interpreting the Fixed-Effects

Predicted Level-1 Model

$$\text{Life Sat}\hat{\text{i}}\text{sfaction}_{ij} = 14.81 + 2.89(\text{SO5}_i - \overline{\text{SO5}}_j)$$

Interpretation of Intercept

The intercept coefficient is now the average life satisfaction for all players who have a shooting success (`Shots_on_5`) equal to their team mean....this is just the average life satisfaction for all players of 14.81.

Interpretation of the slope

Each one-unit difference in shooting success (`Shots_on_five`) is associated with a 2.89 unit change in life satisfaction, on average, for *players on the same team*.

Interpretation of the residual variance component

The estimated residual variance provides a measure of the within-team (player-to-player) variation of life satisfaction scores after accounting for shooting success.

In our example...

There seems to be within-team variation in life satisfaction scores after accounting for shooting success.

$$\hat{\sigma}^2_\epsilon = 4.61$$

The residual variation (level-1) decreased by 68%.

$$r^2_\epsilon = \frac{14.61 - 4.61}{14.61} = 0.684$$

Remember...the change in level-1 residual variation should **always** be compared to the *unconditional random intercepts* model.

Interpreting the Random-Effects

Random Intercepts

Fitted Composite Model

$$\hat{Y}_{ij} = \hat{\beta}_{00} + b_{0j} + \left(\hat{\beta}_{10} + b_{1j}\right)(X_i - \bar{X}_j)$$

when $X_i - \bar{X}_j = 0$

$$\hat{Y}_{ij} = \hat{\beta}_{00} + b_{0j}$$

This is the predicted life satisfaction for a player who plays on team *j* and whose shooting success is at the **team average**.

When we compute this, it is the grand mean plus (or minus) the team deviation...this is just the estimate of each team's average life satisfaction.

These are sometimes referred to as **adjusted** means.

Interpreting the Random-Effects

Random Slopes

Fitted Composite Model

$$\hat{Y}_{ij} = \hat{\beta}_{00} + b_{0j} + \left(\hat{\beta}_{10} + b_{1j}\right)(X_i - \bar{X}_j)$$

Remember, the β_{10} estimate is the average effect of a one-unit difference in shooting success (`Shots_on_five`) for players on the same team.

The random effect, b_{1j}, estimate is the deviation from this average for team *j*

$$\hat{\beta}_{10} + b_{1j}$$

This is the effect of shooting success on predicted life satisfaction for team *j*.

```
# Get estimates of the random-effects
> ranef(lmer.b3)

$Team_ID
   (Intercept)    gcShots
01   1.8212020  0.4565972
02   5.9347180  0.2592425
03   4.6499038  0.3503464
04   1.9489869  0.3851830
05  -4.1516371 -0.4686615
06   1.1564187  0.8490884
```

Team 1: $\hat{\text{Life Satisfaction}} = 14.81 + 1.82 + (2.89 + 0.46)\text{gcShots}$

$$\hat{\text{Life Satisfaction}} = 16.63 + 3.35(\text{gcShots})$$

The estimated life satisfaction for a player who has an average shooting success on team 1 (`gcShots` = 0; `Shots_on_five` = 3.0) is, on average, 16.63.

On team 1, each one-unit difference in shooting success (`gcShots`) is associated with a 3.35 unit change in life satisfaction.

```
# Estimates of the variance-covariance matrix of the random effects
> varCorr(lmer.b3)$Team_ID

             (Intercept)   gcShots
(Intercept)    15.958874 1.3938218
gcShots         1.393822 0.6644517
attr(,"stddev")
(Intercept)     gcShots
  3.9948559   0.8151391
attr(,"correlation")
             (Intercept)   gcShots
(Intercept)    1.0000000 0.4280302
gcShots        0.4280302 1.0000000
```

Square roots of the variance estimates

There seems to be between-team variation in intercepts.

$$\hat{\sigma}_0^2 = 15.96$$

There seems to be between-team variation in slopes.

$$\hat{\sigma}_1^2 = 0.66$$

$$\mathbf{G} = \begin{bmatrix} 15.96 & 1.39 \\ 1.39 & 0.66 \end{bmatrix}$$

The b_{0j} estimates and b_{1j} estimates are positively related. Teams that have a higher intercept also tend to have higher slopes.

Pseudo-R^2 for the Random-Effects

The change in level-2 residual variation should **always** be compared to the *unconditional random intercepts and random slopes* model.

The pseudo-R^2 for intercept represents the decrease in the residual variation in the level-2 intercepts.

$$r_0^2 = \frac{0.093 - 15.96}{0.093} = -170.61$$

The pseudo-R^2 for slope represents the decrease in the residual variation in the level-2 slopes.

$$r_1^2 = \frac{0.099 - 0.66}{0.099} = -5.67$$

Remember, pseudo-R^2 values can be negative. They can also be greater than 1. In this case it is likely because the variation in life satisfaction is almost exclusively at level-2.