

Mixed-Effects Models in Practice II

2018-03-19

Preparation

We will use two datasets located in the *nbaLevel1.sav* file and the *nbaLevel2.sav* file. These data include player-level attributes for $n = 300$ NBA players, and team-level attributes for $N = 30$ different teams. The source of these data is: Woltman, Feldstein, MacKay, & Rocchi (2012). We will use these data to explore the question of whether how good a player is (Shots_on_five) predicts variation in life satisfaction.

The player-level attributes in the *nbaLevel1.sav* file include:

- Team_ID: The team ID number for each player
- Shots_on_five: A proxy for player quality/success. This indicates the number of successful shots (out of five taken). Higher values indicate a more successful player.
- Life_Satisfaction: Score on a survey of life satisfaction. Scores range from 5 to 25, with higher scores indicating more life satisfaction.

The team-level attributes in the *nbaLevel2.sav* file include:

- Team_ID: The team ID number
- Coach_Experience: Years of coaching experience in the NBA

```
# Load libraries
library(AICcmodavg)
library(broom)
library(dplyr)
library(ggplot2)
library(haven) #for reading in the .sav files
library(lme4) #for fitting mixed-effects models
library(sm)

# Read in player-level data
nbaL1 = read_sav(file = "~/Dropbox/epsy-8252/data/nbaLevel1.sav")

# Read in team-level data
nbaL2 = read_sav(file = "~/Dropbox/epsy-8252/data/nbaLevel2.sav")

# Join/merge the two files together
nba = left_join(nbaL1, nbaL2, by = "Team_ID")
head(nba)
```

Team_ID	Shots_on_five	Life_Satisfaction	Coach_Experience
01	3	18.8	2
01	3	18.0	2
01	4	21.0	2
01	4	20.5	2
01	3	19.0	2
01	2	12.1	2

Writing the Model as a Multilevel Model

One way we can express a mixed-effects model is by separating the model equation into a different equation for each level. For example, our mixed-effects models for the NBA data can be separated into a Level-1 (player-level) equation and a set of Level-2 (team-level) equations. Take the equation for the unconditional means model:

$$\text{Life Satisfaction}_{ij} = \beta_0 + b_{0j} + \epsilon_{ij},$$

We initially write the Level-1 equation. The level-1 equation here is:

$$\text{Life Satisfaction}_{ij} = \beta_{0j} + \epsilon_{ij},$$

Here, each player's outcome is a function of a team-level intercept and the player-specific error. Then we can write the level-2 (team-level) equation as:

$$\beta_{0j} = \beta_{00} + b_{0j}$$

This indicates that the team j 's intercept is a function of the average intercept (β_{00}) and a deviation for team j . Together these equations are referred to as the set of multilevel equations:

Level-1 :

$$\text{Life Satisfaction}_{ij} = \beta_{0j} + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_{00} + b_{0j}$$

If we have the multilevel equations, we can substitute the Level-2 equation(s) into the Level-1 equation to get the *composite* mixed-effects equation.

$$\text{Life Satisfaction}_{ij} = [\beta_{00} + b_{0j}] + \epsilon_{ij},$$

Here are some guidelines in helping you think about writing multilevel equations.

- Write the level-1 equation first. This will be an equation that expresses the outcome's relationship to a series of team-specific parameters and a player-specific residual.
- The number of team-specific parameters in the level-1 equation (aside from the residual) dictate the number of level-2 equations you will have.
- The team-specific parameters from the level-1 equation will be the outcomes in the level-2 equations.
- Random-effects are the residuals in the level-2 equations, and therefore are in the level-2 equations; one per equation.
- Variables from the data go to their appropriate level. For example player-level variables will be put in the level-1 equation, and team-level predictors will be put in the level-2 equations.

To see this in action, let's consider the mixed-effects model that includes a random-effect for intercept and for the effect of shots-on-five. The composite mixed-effects model we previously wrote looks like this:

$$\text{Life Satisfaction}_{ij} = \beta_0 + b_{0j} + \beta_1(\text{Shots}_{ij}) + b_{1j}(\text{Shots}) + \epsilon_{ij}$$

Using our guidelines, we first write the level-1 equation which expresses life-satisfaction scores as a team-specific intercept and team-specific slope for shots-on-five. The residual term and the shots-on-five variable both have i subscripts indicating they also go in the level-1 equation. The level-1 equation is:

$$\text{Life Satisfaction}_{ij} = \beta_{0j} + \beta_{1j}(\text{Shots}_{ij}) + \epsilon_{ij}$$

Now we can write our level-2 equations. Since there are two team-specific parameters in the level-1 model, there will be two level-2 equations. The outcomes for these equations will be those team-specific parameters. Here the two level-2 equations are:

$$\begin{aligned}\beta_{0j} &= \beta_{00} + b_{0j} \\ \beta_{1j} &= \beta_{10} + b_{1j}\end{aligned}$$

You can check the substitution for yourself.

Predictor: Coach Experience in Level-2 Intercept Equation

The coach experience variable is a team-level variable. Thus, we would include it in the level-2 equations. Below we include the coach experience variable in the level-2 intercept equation (but not the slope equation):

Level-1 :

$$\text{Life Satisfaction}_{ij} = \beta_{0j} + \beta_{1j}(\text{Shots}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\begin{aligned}\beta_{0j} &= \beta_{00} + \beta_{01}(\text{CE}_j) + b_{0j} \\ \beta_{1j} &= \beta_{10} + b_{1j}\end{aligned}$$

If we substitute the level-2 equations into the level-1 equation and simplify,

$$\begin{aligned}\text{Life Satisfaction}_{ij} &= \left[\beta_{00} + \beta_{01}(\text{CE}_j) + b_{0j} \right] + \left[\beta_{10} + b_{1j} \right] (\text{Shots}_{ij}) + \epsilon_{ij} \\ &= \beta_{00} + \beta_{01}(\text{CE}_j) + b_{0j} + \beta_{10}(\text{Shots}_{ij}) + b_{1j}(\text{Shots}_{ij}) + \epsilon_{ij} \\ &= \beta_{00} + \beta_{10}(\text{Shots}_{ij}) + \beta_{01}(\text{CE}_j) + \left[b_{0j} + b_{1j}(\text{Shots}_{ij}) + \epsilon_{ij} \right]\end{aligned}$$

Including a team-level predictor in the level-2 intercept equation leads to the inclusion of a main-effect of that predictor in the composite (LMER) equation.

Predictor: Coach Experience in Level-2 Slope Equation

Below we include the coach experience variable in the level-2 slope equation (but not the intercept equation):

Level-1 :

$$\text{Life Satisfaction}_{ij} = \beta_{0j} + \beta_{1j}(\text{Shots}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\begin{aligned}\beta_{0j} &= \beta_{00} + b_{0j} \\ \beta_{1j} &= \beta_{10} + \beta_{11}(\text{CE}_j) + b_{1j}\end{aligned}$$

If we substitute the level-2 equations into the level-1 equation and simplify,

$$\begin{aligned}
\text{Life Satisfaction}_{ij} &= \left[\beta_{00} + b_{0j} \right] + \left[\beta_{10} + \beta_{11}(\text{CE}_j) + b_{1j} \right] (\text{Shots}_{ij}) + \epsilon_{ij} \\
&= \beta_{00} + b_{0j} + \beta_{10}(\text{Shots}_{ij}) + \beta_{11}(\text{CE}_j)(\text{Shots}_{ij}) + b_{1j}(\text{Shots}_{ij}) + \epsilon_{ij} \\
&= \beta_{00} + \beta_{10}(\text{Shots}_{ij}) + \beta_{11}(\text{CE}_j)(\text{Shots}_{ij}) + \left[b_{0j} + b_{1j}(\text{Shots}_{ij}) + \epsilon_{ij} \right]
\end{aligned}$$

Including a team-level predictor in the level-2 slope equation leads to the inclusion of an interaction in the composite (LMER) equation. If we want to include an interaction we also need to include all constituent main-effects in the model. Therefore we need a main-effect of coaching experience and a main-effect of shots-on-five in the model. The latter is there, but not the former. To get a main-effect of coaching experience in the composite equation it needs to be included in the level-2 intercept model.

ANY PREDICTORS INCLUDED IN THE LEVEL-2 SLOPE EQUATIONS NEED TO ALSO BE INCLUDED IN THE LEVEL-2 INTERCEPT EQUATIONS.

Multilevel Equations for Fixed-Effects Models

Our conventional fixed-effects regression models (LM) can also be expressed as multilevel models. For example,

Level-1 :

$$\text{Life Satisfaction}_{ij} = \beta_{0j} + \beta_{1j}(\text{Shots}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_{00}$$

$$\beta_{1j} = \beta_{10}$$

Here the level-2 equations just include fixed-effects (no random-effects). Thus when we substitute them back into the level-1 model we only have fixed-effects in the model:

$$\text{Life Satisfaction}_{ij} = \beta_{00} + \beta_{10}(\text{Shots}_{ij}) + \epsilon_{ij}$$

Why are Multilevel Expressions Helpful?

Expressing the model as a set of multilevel equations can be helpful for readers. First, it explicitly separates the predictors into the two levels. Team-level predictors only appear in the level-2 models. Player-level predictors only appear in the level-1 model. Secondly, it helps us think about what the predictors at each level are actually doing. Level-1 predictors explain variation in the outcome at level-1. Level-2 predictors explain variation in the team-specific intercepts (or intercepts and slopes)—they explain level-2 variation. (Think about that partitioning of variation into levels we did.)

Checking Assumptions for the Mixed-Effects Model

To illustrate assumption checking in practice, we will evaluate the assumptions for a mixed-effects model that includes the fixed-effect of shots-on-five and random-effect for intercept and shots-on-five.

```
lmer.1 = lmer(Life_Satisfaction ~ 1 + Shots_on_five +
              (1 + Shots_on_five | Team_ID), data = nba, REML = FALSE)
```

In mathematical notation,

$$\text{Life Satisfaction}_{ij} = \beta_0 + \beta_1(\text{SO5}_{ij}) + b_{0j} + b_{1j}(\text{SO5}_{ij}) + \epsilon_{ij}$$

where,

$$\begin{aligned}\epsilon_i &\sim \mathcal{N}(0, \sigma_\epsilon^2) \\ b_{0j} &\sim \mathcal{N}(0, \sigma_{b_0}^2) \\ b_{1j} &\sim \mathcal{N}(0, \sigma_{b_1}^2)\end{aligned}$$

Just as in fixed-effects regression, we need to evaluate the assumptions; this time, however, we need to evaluate assumptions about the level-1 errors and for the random-effects.

Assumptions about the Level-1 Residuals

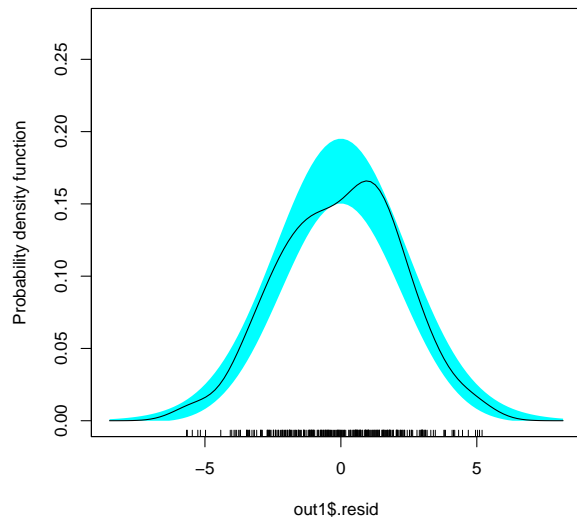
We will evaluate the level-1 residuals in the exact same way we evaluated the residuals from an LM analysis.

```
# Augment the model to get the level-1 residuals and fitted values
out1 = augment(lmer.1)
head(out1)
```

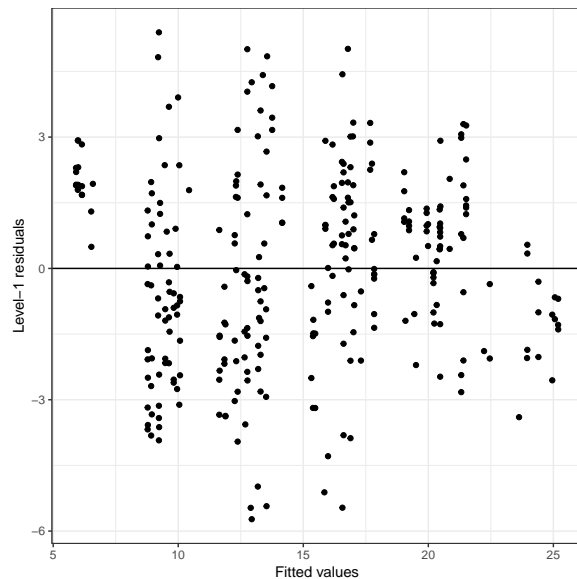
Life_Satisfaction	Shots_on_five	Team_ID	.fitted	.resid	.fixed	.mu	.offset	.sqrtXwt	.sqrtrwt	.weights	.v
18.8	3	01	16.6	2.194	16.3	16.6	0	1	1	1	2
18.0	3	01	16.6	1.390	16.3	16.6	0	1	1	1	1
21.0	4	01	20.0	1.011	19.6	20.0	0	1	1	1	1
20.5	4	01	20.0	0.511	19.6	20.0	0	1	1	1	0
19.0	3	01	16.6	2.390	16.3	16.6	0	1	1	1	2
12.1	2	01	13.2	-1.131	13.0	13.2	0	1	1	1	-1

The level-1 residuals are found in the `.resid` column, and the `.fitted` column contains the \hat{Y} values. As with LM residual analysis, we want to examine the normality of the residuals in a density plot (or some other plot that allows you to evaluate this), and the other assumptions by plotting the residuals against the fitted values in a scatterplot.

```
# Density plot of the level-1 residuals
sm.density(out1$.resid, model = "normal")
```



```
# Scatterplot of the level-1 residuals versus the fitted values
ggplot(data = out1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")
```



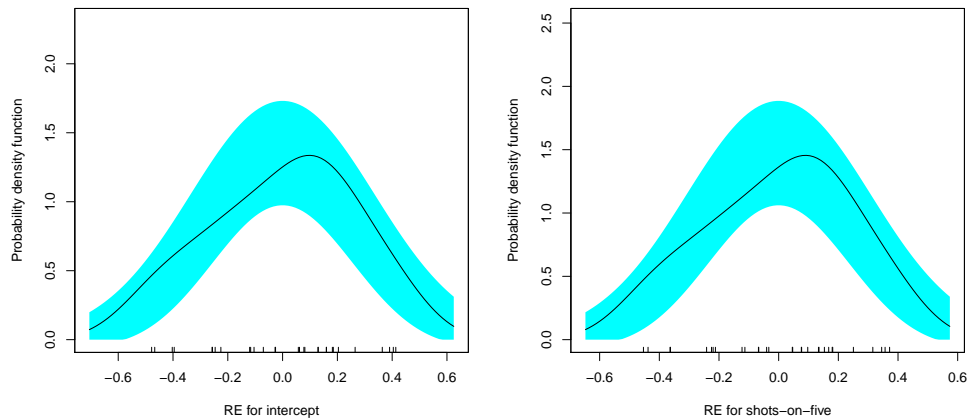
Based on the plots, the level-1 residual assumptions of the model seem reasonably met (maybe). The density plot suggests that the normality assumption is tenable. The scatterplot shows symmetry around the $Y = 0$ line (linearity) and fairly consistent variation in the residual throughout most of the fitted values (some lesser variation in the extremes).

Assumptions about the Random-Effects

We also need to examine the assumptions for the random-effects. The first assumption we will examine is the normality assumption for each set of REs.

```
# Obtain the RE for intercept and slope
re_int = ranef(lmer.1)$Team_ID[ , 1]
re_shots = ranef(lmer.1)$Team_ID[ , 2]

# Density plot of the RE for intercept
sm.density(re_int, model = "normal", xlab = "RE for intercept")
sm.density(re_shots, model = "normal", xlab = "RE for shots-on-five")
```



This looks good for both sets of REs. We will also plot each set of REs against the fitted values for the level-2 equations to check for linearity, and homogeneity of variance. The level-2 equations are:

$$\begin{aligned}\beta_{0j} &= \beta_{00} + b_{0j} \\ \beta_{1j} &= \beta_{10} + b_{1j}\end{aligned}$$

Thus the fitted-values for these are $\hat{\beta}_{00}$ and $\hat{\beta}_{10}$, respectively. In this example, the fitted values are just the fixed-effect estimates. If we had other level-2 predictors, that would not be the case. For example in the case of including coach experience at each level, the level-2 equations would be

$$\begin{aligned}\beta_{0j} &= \beta_{00} + \beta_{01}(\text{CE}_j) + b_{0j} \\ \beta_{1j} &= \beta_{10} + \beta_{11}(\text{CE}_j) + b_{1j}\end{aligned}$$

In this case the fitted values would be computed as $\hat{\beta}_{00} + \hat{\beta}_{01}(\text{CE}_j)$ and $\hat{\beta}_{10} + \hat{\beta}_{11}(\text{CE}_j)$.

Below we set up a data set that puts the fitted values and random-effects for the level-2 equations to be used to create a scatterplot.

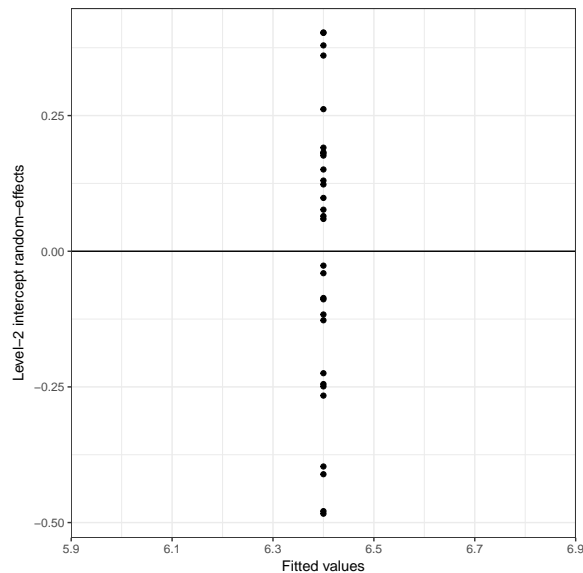
```
# Set up level-2 intercept data set
level_2_int = data.frame(
  fitted = 6.40,
  random_effect = ranef(lmer.1)$Team_ID[ , 1]
)

head(level_2_int)
```

fitted	random_effect
6.4	0.077
6.4	0.379
6.4	0.403

fitted	random_effect
6.4	-0.086
6.4	-0.245
6.4	0.176

```
ggplot(data = level_2_int, aes(x = fitted, y = random_effect)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-2 intercept random-effects")
```



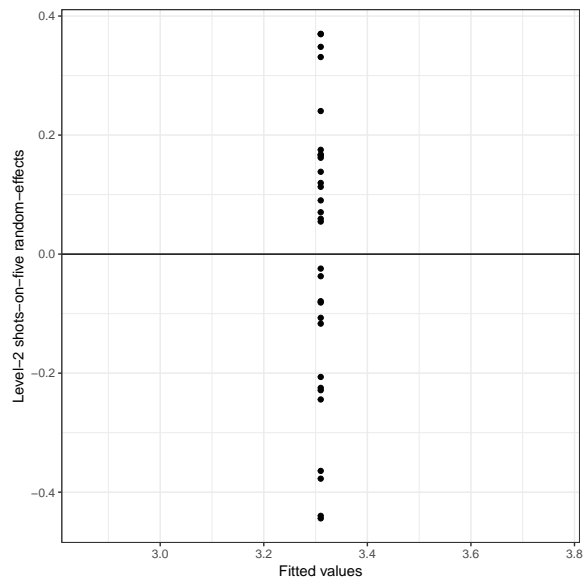
There is only one fitted value, so only a single vertical strip of dots. (With other level-2 predictors in the model this would look more like residual plots you are used to seeing.) Because of this, the assumption of homogeneity of variance is met and the assumption of linearity also seems tenable. Now we will do the same analysis for the second level-2 equation.

```
# Set up level-2 shots-on-five data set
level_2_shots = data.frame(
  fitted = 3.31,
  random_effect = ranef(lmer.1)$Team_ID[ , 2]
)

head(level_2_shots)
```

fitted	random_effect
3.31	0.070
3.31	0.348
3.31	0.370
3.31	-0.079
3.31	-0.225
3.31	0.162


```
ggplot(data = level_2_shots, aes(x = fitted, y = random_effect)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-2 shots-on-five random-effects")
```



The assumptions also look good for this set of REs.

References

Woltman, H., Feldstein, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52–69.