

Longitudinal MLM Exercise with Code

Beth Grandfield

2/23/2022

Exercise 2b: General Instructions

This exercise is to be completed in lab and uploaded to blackboard. Complete the exercise in groups of 3.

While completing the exercise, fill in the excel file for each model. Upload the excel file at the end of the lab via Blackboard.

Longitudinal analysis

In this exercise, you will practice multilevel analysis on longitudinal data.

The file gpa2.csv contains the data for the GPA example discussed both in the lecture and in the book of Hox, chapter 5. The dependent variable is the GPA score of students, measured during 6 occasions. Besides GPA, we have the number of hours worked in off-campus jobs (Job) on each of these occasions, and the sex (1 = male, 2 = female) and high school GPA of the students.

Libraries used:

```
library(tidyverse)
library(ggplot2)
library(lme4)
library(lmerTest)
```

First inspect the data file GPA2.csv in R. As you can see, each row in the gpa2-file represents a case, and the variables are given in the columns.

```
GPA <- read.csv(file = "gpa2.csv", header=TRUE)
head(GPA)
```

1. At which level is each variable?
2. Which variable is the level 2 identification variable?
3. Convert the wide data file into a file that is suited for multilevel analysis using lme4 in R (that is, we need a long format).

```
library(tidyr)
GPA_long <- pivot_longer(data = GPA, cols = c(4:15),
                          names_to = c(".value", "time"),
                          names_pattern = "(gpa|job)(.)")
head(GPA_long)
```

4. Recode Time and Sex.

- Recode the variable Time (from 1-6, to 0-5).
- Recode the variable sex into dummy variables (i.e., using the values 0 and 1).

```
GPA_long$sex <- unclass(GPA_long$sex) - 1
GPA_long$time <- as.numeric(GPA_long$time) - 1
head(GPA_long)
```

5. Check the linearity assumption.

- Make a scatterplot of the variable GPA and Time (add a linear and quadratic fit line).

```
library(ggplot2)
ggplot(GPA_long,
  aes(x = time, y = gpa)) +
  geom_point() +
  geom_smooth(method = "lm",
    aes(color = "linear"),
    se = FALSE) +
  geom_smooth(method = "lm",
    formula = y ~ x + I(x^2),
    aes(color = "quadratic"),
    se = FALSE)
```

- Make a scatterplot of the variable GPA and Job (add a linear and quadratic fit line).

```
ggplot(GPA_long,
  aes(x = job, y = gpa)) +
  geom_point() +
  geom_smooth(method = "lm",
    aes(color = "linear"),
    se = FALSE) +
  geom_smooth(method = "lm",
    formula = y ~ x + I(x^2),
    aes(color = "quadratic"),
    se = FALSE)
```

6. Check for outliers (don't perform analyses, just look in the scatterplots)

Now we proceed with the multilevel analysis.

7. Specify the intercept only model.

- Inspect the output and compare the results with the model on the slides of the lecture. When they are the same, calculate and interpret the ICC.

```
library(lme4)
#recall from lab 1 that lmerTest will provide df, t, and p-values for fixed effects
library(lmerTest)

model_0 <- lm(gpa~1, data = GPA_long)
model_1 <- lmer(gpa~1 + (1|student) , REML = FALSE, data = GPA_long)
summary(model_0)
```

```
summary(model_1)

anova(model_1, model_0)
```

8. Specify the model with Time as a linear predictor of the trend over time.

- Should you include the variable time uncentered or centered?
- Inspect the output and compare the results with the model on the slides of lecture. When they are the same, you can go on to the next step.

```
model_2 <- lmer(gpa~1 + time + (1|student) , REML = FALSE, data = GPA_long)
summary(model_2)

# to check if time is a significant predictor:
anova(model_2, model_1)
```

9. Set up the model with the time-varying predictor Job.

- Inspect and compare the results with the model on the slides of the lecture. When they are the same, calculate and interpret the explained variance at level 1 and level 2.

```
mean_job <- mean(GPA_long$job)
GPA_long$job <- GPA_long$job - mean_job
model_3 <- lmer(gpa~1 + time + job + (1|student) , REML = FALSE, data = GPA_long)
summary(model_3)

# to check if job is a significant predictor:
anova(model_2, model_3)
```

10. Set up the model with the time-invariant predictors Gender and HighGPA.

- Inspect the output and compare the results with the model on the slides of the lecture. When they are the same, calculate and interpret the explained variance at level 2 and go on to the next step.

```
mean_highgpa <- mean(GPA_long$highgpa)
GPA_long$highgpa <- GPA_long$highgpa - mean_highgpa

model_4 <- lmer(gpa~1 + time + job + sex + highgpa + (1|student) , REML = FALSE,
               data = GPA_long)
summary(model_4)
anova(model_3, model_4)
```

11. For the time-varying predictors, check if the slopes are fixed or random.

- Start with the slope of Job: check if the variance of the slope for Job is significant. If the answer is no, you can remove u2 from the model and go on to the next step.
- Now turn to the variable Time: check if the variance of the slope for Time is significant. If the answer is yes, you can go on to the next step.

```

#random slope for job
model_5a <- lmer(gpa~1 + time + job + sex + highgpa + (1+job|student) ,
                REML = FALSE, data = GPA_long)
summary(model_5a)

anova(model_4, model_5a)

# random slope for time
model_5b <- lmer(gpa ~ 1 + time + job + sex + highgpa + (1 + time|student),
                REML = F, data = GPA_long)
summary(model_5b)
anova(model_5b, model_4)

```

12. Now, we'll check if Gender can (partly) explain why the trajectory over time differs between students. That is, we include Gender as a predictor of the slope for Time.

- Inspect the output and compare the results with the models on the slides of the lecture. When they are the same, calculate and interpret the explained slope variance.
- Make an interaction plot to show how the trajectory over time differs between boys and girls.
- For R one option is to draw an empty plot with time on the x-axis and GPA on the y-axis. Next, draw in the appropriate lines using `abline()` and the corresponding equations for the predicted outcomes over time for boys and girls.

```

model_6 <- lmer(gpa ~ 1 + time + job + sex + highgpa + sex * time + (1 + time|student),
                REML = F, data = GPA_long)
summary(model_6)
anova(model_6, model_5b)

# interaction plot
time <- 0:5
plot(x = time, ylim = c(1,4), xlim = c(0,5), type = "n", xlab = "time", ylab = "gpa")
col.sex <- c("cadetblue", "coral")
abline(a = model_6@beta[1], b = model_6@beta[2], col = col.sex[1], lwd = 2)

# girls start out higher and proceed faster over time:

abline(a = model_6@beta[1] + model_6@beta[4], b = model_6@beta[2] + model_6@beta[6],
       col = col.sex[2], lwd = 2)
legend("topleft", bty = "n", lwd = 2, col = rev(col.sex),
       legend = c("girls", "boys"))

```

13. Check the normality assumption for the level 1 residuals.

- The level 1 and 2 residuals can be accessed via the function `residuals(object, ...)`. Check the help file for more information and note the argument level and test what it does!
- Using the obtained residuals, (visually) inspect the normality assumption.
- Are the assumptions met?

```

qqnorm(residuals(model_6))
qqline(residuals(model_6))

```

14. Check the normality assumption for the level 2 residuals

- Are the assumptions met for the intercept and slope errors?

```
#intercept  
qqnorm(ranef(model_6)$student[,1])  
qqline(ranef(model_6)$student[,1])  
#slope  
qqnorm(ranef(model_6)$student[,2])  
qqline(ranef(model_6)$student[,2])
```