MLM Assignment 2

Christine Hedde-von-Westernhagen   Emilia Löscher   Kyuri Park

01 March, 2022

Data Description

In this assignment, we analyze the `curran_wide.csv` data, which contains the information about the age, antisocial behavior, reading skills, emotional support, cognitive stimulation, and mother's age of 221 sampled children. Antisocial behavior and reading skills are measured over 4 occasions. In this analysis, we do not use children's age and emotional support variables.

The (pre-processed) data specifics are as follows:

- `id`: children id
- `time`: measurement occasion ranging from 0 to 3
- `anti`: antisocial behavior (time-variant)
- `read`: reading recognition skills (time-variant & grand-mean centered)
- `momage`: mother's age measured at the first occasion (time-invariant & grand-mean centered)
- `homecog`: cognitive stimulation measured at the first occasion (time-invariant & grand-mean centered)

1. Convert the wide data file into a long format. Check the data and recode if necessary.

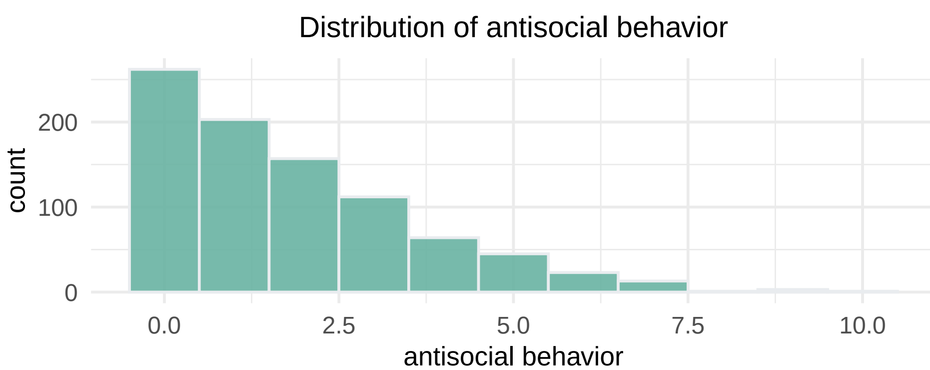
```
## # A tibble: 6 x 13
##      id anti1 anti2 anti3 anti4 read1 read2 read3 read4   sex momage homecog
##   <int> <int> <int> <int> <int> <dbl> <dbl> <dbl> <dbl> <int>  <int>   <int>
## 1    34     3     6     4     5   2.1   2.9   4.5   4.5     1     28       9
## 2    58     0     2     0     1   2.3   4.5   4.2   4.6     0     28       9
## 3   125     1     1     2     1   2.3   3.8   4.3   6.2     0     29      10
## 4   133     3     4     3     5   1.8   2.6   4.1   4       1     28       8
## 5   163     5     4     5     5   3.5   4.8   5.8   7.5     1     28      10
## 6   248     1     2     2     0   3.5   5.7   7     6.9     0     28       9
## # ... with 1 more variable: homeemo <int>
```

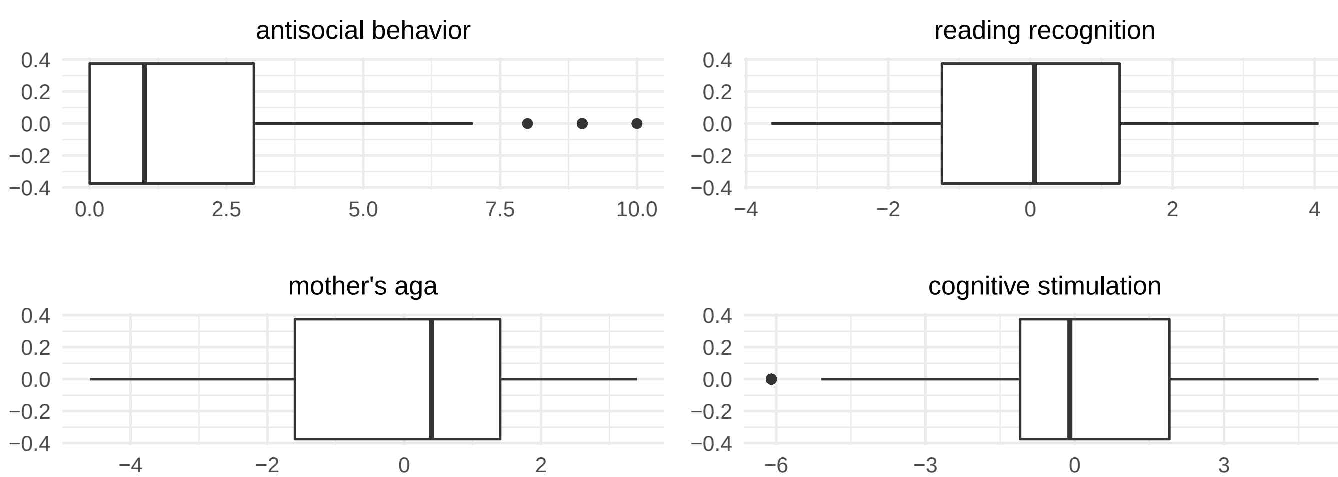
```
## # A tibble: 6 x 6
##      id  time  anti   read momage homecog
##   <int> <dbl> <int>  <dbl>  <dbl>   <dbl>
## 1    34     0     3 -2.25    2.40 -0.0995
## 2    34     1     6 -1.45    2.40 -0.0995
## 3    34     2     4  0.155   2.40 -0.0995
## 4    34     3     5  0.155   2.40 -0.0995
## 5    58     0     0 -2.05    2.40 -0.0995
## 6    58     1     2  0.155   2.40 -0.0995
```

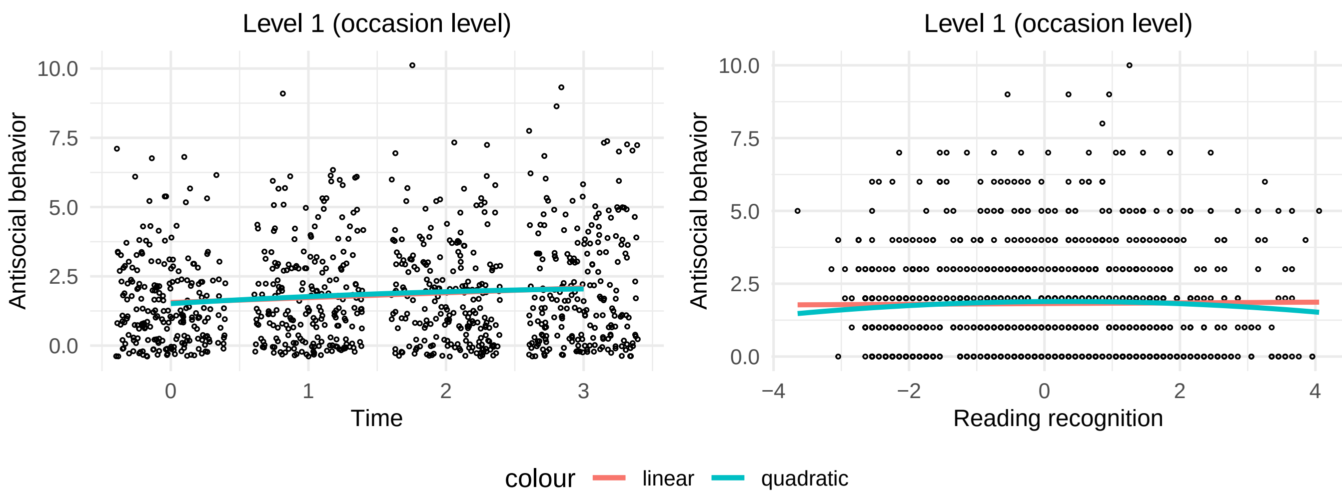
Table 1: Descriptive statistics

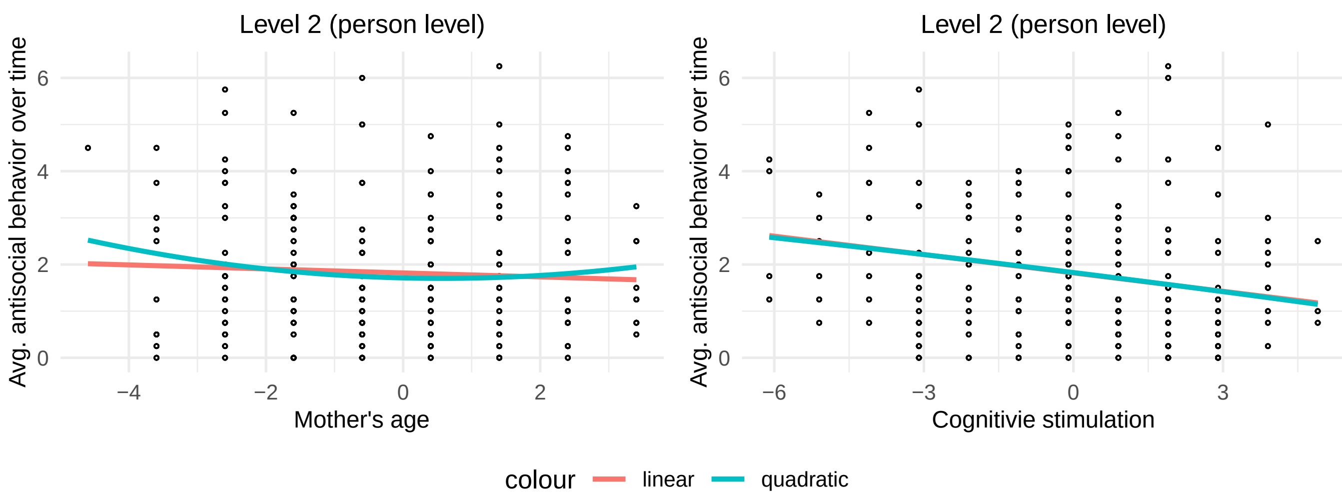
	n	mean	sd	median	min	max	skew	kurtosis	se
id	884	3679	2495	3410	34	8870	0.39	-1.05	83.92
time	884	1.5	1.12	1.5	0	3	0	-1.36	0.04
anti	884	1.82	1.82	1	0	10	1.12	1.05	0.06
read	884	0	1.62	0.05	-3.65	4.05	0.11	-0.77	0.05
momage	884	0	1.87	0.4	-4.6	3.4	-0.14	-0.85	0.06
homecog	884	0	2.45	-0.1	-6.1	4.9	-0.37	-0.42	0.08

- **Check the linearity assumption, report and include plots.**
- **Check for outliers (don't perform analyses, just look in the scatterplots), report.**









2. Answer the question: should you perform a multilevel analysis?

- **What is the mixed model equation?**

– Mixed Model Equation

$$y_{ti} = \beta_{00} + u_{0i} + e_{ti}$$

- y_{ti} refers to antisocial behavior of child i at time t.
- β_{00} refers to the overall intercept, which is the average antisocial behavior over all children.
- u_{0j} refers to the random residual error at the person level (level 2), which represents the deviation from the overall intercept (β_{00}) of child i.
- e_{ij} refers to the residual error at the occasion level (level 1).

- **Provide and interpret the relevant results.**

```
## model 1: random intercept model ((benchmark model to compute ICC))
model1 <- lmer(anti ~ 1 + (1|id), REML = FALSE, data= curran_long)
summary(model1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
```

```
## Formula: anti ~ 1 + (1 | id)
##    Data: curran_long
##
##      AIC      BIC   logLik deviance df.resid
##   3343.5   3357.9  -1668.8   3337.5      881
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4165 -0.5797 -0.2521  0.4752  4.1615
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 1.579    1.257
##  Residual             1.741    1.320
## Number of obs: 884, groups:  id, 221
##
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   1.81900    0.09547 221.00000   19.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **What is the intraclass correlation?**

The intraclass correlation (ρ) is calculated as follows:

$$\rho = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_e}$$

As shown below, the intraclass correlation equals to 0.476 in this case, which is deemed to be large.

```r
ICC <- 1.579/(1.579+1.741)
cat("ICC =", ICC)
```

```
## ICC = 0.4756024
```

- **What is your conclusion regarding the overall question regarding the necessity of performing a multilevel analysis?**

Yes we should perform the multilevel analysis in this case, because not only the data structure is nested (i.e., multiple measurements within each individual), but also the difference between individuals accounts for about *48%* of the total variance. In other words, the intraclass correlation – ICC: the proportion of the total variance explained by the between-individual differences – is *0.476*, which is high enough that the multilevel analysis is warranted.

3. Add the time-varying predictor(s).

- **Provide and interpret the relevant results and provide your overall conclusion.**

The time-varying predictor, `read` is not significant.

```r
## model2: add time predictor ((benchmark model for computing R2))
model2 <- lmer(anti ~ 1 + time + (1|id), REML = FALSE, data= curran_long)
summary(model2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: anti ~ 1 + time + (1 | id)
##    Data: curran_long
##
##      AIC      BIC   logLik deviance df.resid
##   3325.5   3344.6  -1658.7   3317.5      880
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2820 -0.5296 -0.1838  0.4780  4.1401
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 1.592    1.262
##  Residual             1.689    1.300
## Number of obs: 884, groups:  id, 221
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   1.5543     0.1120 400.3046  13.872  < 2e-16 ***
## time          0.1765     0.0391 663.0000   4.513 7.56e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## time -0.523
```

```
anova(model2, model1)
```

```
## Data: curran_long
## Models:
## model1: anti ~ 1 + (1 | id)
## model2: anti ~ 1 + time + (1 | id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model1    3 3343.5 3357.9 -1668.8   3337.5
## model2    4 3325.5 3344.6 -1658.7   3317.5 20.062  1  7.496e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## model3: add time-varying predictor, read
# center the predictor, read
curran_long$read <- curran_long$read - mean(curran_long$read)

model3 <- lmer(anti ~ 1 + time + read + (1|id), REML = FALSE, data= curran_long)
summary(model3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: anti ~ 1 + time + read + (1 | id)
##    Data: curran_long
##
##      AIC      BIC   logLik deviance df.resid
##   3327.2   3351.1  -1658.6   3317.2      879
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2985 -0.5234 -0.1704  0.4887  4.1580
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 1.576    1.255
##  Residual             1.693    1.301
## Number of obs: 884, groups:  id, 221
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   1.49940    0.15087 580.44942   9.938  < 2e-16 ***
## time          0.21307    0.07808 882.38989   2.729  0.00649 **
## read         -0.03376    0.06233 830.65151  -0.542  0.58819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) time
## time -0.776
## read  0.672 -0.865
```

```
anova(model3, model2)
```

```
## Data: curran_long
## Models:
## model2: anti ~ 1 + time + (1 | id)
## model3: anti ~ 1 + time + read + (1 | id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model2    4 3325.5 3344.6 -1658.7   3317.5
## model3    5 3327.2 3351.1 -1658.6   3317.2 0.2854  1     0.5932
```

4. On which level or levels can you expect explained variance?

- **Calculate and interpret the explained variances.**

We can expect the explained variances (R^2) at both occasion (level 1) and person level (level 2), as the level 1 predictor can explain the variances in both levels. The computed R^2 values for each level are:

- $R^2_{occasion}$ = -0.0024 ??????????
- R^2_{person} = 0.0101

```
m2var.lv1 <- 1.689
m2var.lv2 <- 1.592
m3var.lv1 <- 1.693
m3var.lv2 <- 1.576

## explained variance at level 1 (occasion level)
R2.lv1 <- (m2var.lv1 - m3var.lv1) / m2var.lv1
## explained variance at level 2 (person level)
R2.lv2 <- (m2var.lv2 - m3var.lv2) / m2var.lv2
cat("Explained variance at the occasion level =", round(R2.lv1,4) ,
    "\n", "Explained variance at the person level =", round(R2.lv2,4))
```

```
## Explained variance at the occasion level = -0.0024
##  Explained variance at the person level = 0.0101
```

5. Add the time invariant predictor(s) to the model.

- **Provide and interpret the relevant results and provide your overall conclusion.**

```
# model4: add time-invariant predictors, momage & homecog
model4 <- lmer(anti ~ 1 + time + read + momage + homecog + (1|id), REML = FALSE, data= curran_long)
summary(model4)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: anti ~ 1 + time + read + momage + homecog + (1 | id)
##    Data: curran_long
##
##      AIC      BIC   logLik deviance df.resid
##   3319.8   3353.3  -1652.9   3305.8      877
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3498 -0.5380 -0.1664  0.4945  4.0858
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 1.483    1.218
##  Residual             1.691    1.300
## Number of obs: 884, groups:  id, 221
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  1.533e+00  1.500e-01  6.013e+02  10.219  < 2e-16 ***
## time         1.905e-01  7.856e-02  8.834e+02   2.425  0.01549 *
## read        -1.296e-02  6.284e-02  8.412e+02  -0.206  0.83661
## momage       5.458e-04  5.180e-02  2.236e+02   0.011  0.99160
## homecog     -1.304e-01  3.923e-02  2.196e+02  -3.324  0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) time   read   momage
## time    -0.785
## read     0.681 -0.867
## momage  -0.097  0.123 -0.142
## homecog -0.056  0.071 -0.082 -0.229
```

```
anova(model4, model3)
```

```
## Data: curran_long
## Models:
## model3: anti ~ 1 + time + read + (1 | id)
## model4: anti ~ 1 + time + read + momage + homecog + (1 | id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model3    5 3327.2 3351.1 -1658.6   3317.2
## model4    7 3319.8 3353.3 -1652.9   3305.8 11.398  2    0.00335 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## do we REMOVE 'read' since it is not sig?
model4a <- lmer(anti ~ 1 + time + momage + homecog + (1|id), REML = FALSE, data= curran_long)
summary(model4a)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: anti ~ 1 + time + momage + homecog + (1 | id)
##    Data: curran_long
##
##      AIC      BIC   logLik deviance df.resid
##   3317.8   3346.5  -1652.9   3305.8      878
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3437 -0.5467 -0.1676  0.4893  4.0784
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 1.488    1.22
##  Residual             1.689    1.30
## Number of obs: 884, groups:  id, 221
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  1.554e+00  1.099e-01  4.102e+02  14.138  < 2e-16 ***
## time         1.765e-01  3.910e-02  6.630e+02   4.513 7.56e-06 ***
## momage      -9.752e-04  5.133e-02  2.210e+02  -0.019 0.984859
## homecog     -1.311e-01  3.915e-02  2.210e+02  -3.348 0.000956 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) time   momage
## time    -0.533
## momage   0.000  0.000
## homecog  0.000  0.000 -0.244
```

```
anova(model4a, model3)
```

```
## Data: curran_long
## Models:
## model3: anti ~ 1 + time + read + (1 | id)
## model4a: anti ~ 1 + time + momage + homecog + (1 | id)
##         npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model3     5 3327.2 3351.1 -1658.6   3317.2
## model4a    6 3317.8 3346.5 -1652.9   3305.8 11.356  1   0.000752 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6. On which level or levels can you expect explained variance?

- **Calculate and interpret the explained variances.**

We can expect the explained variances (R^2) at the person level (level 2), as the level 2 predictor can only explain the variance in level 2. The computed R^2 value for the person level is:

- $R^2_{person} = 0.0653$

```
m2var.lv1 <- 1.689
m2var.lv2 <- 1.592
m4var.lv1 <- 1.689     # depends on which model we use, 4 or 4a? I am going with 4a the one without 'read'
m4var.lv2 <- 1.488

## explained variance at level 2 (person level)
R2.lv2 <- (m2var.lv2 - m4var.lv2) / m2var.lv2
cat("Explained variance at the person level =", round(R2.lv2,4))
```

```
## Explained variance at the person level = 0.0653
```

7. For the time-varying predictor(s), check if the slope is fixed or random.

- **What are the null- and alternative hypotheses?**

- H_0:
- H_1:

- **Provide and interpret the relevant results.**

The variance of time and read are significant when added separately but the model does not converge when they are put together in a model.. ### • Provide an overall conclusion.

8. If there is a random slope, set up a model that predicts the slope variation.

- **Provide and interpret the relevant results and provide your overall conclusion.**

9. Decide on a final model.

- **provide the separate level 1 and 2 model equations, as well as the mixed model equation.**

- **Check the normality assumption for both the level-1 and level-2 errors, report**

Contribution

- Christine :

- Emilia :

- Kyuri: