# PSC 204B - Winter 2024 - Homework 6

Multilevel Modeling

Marwin Carmo

2024-02-23

```r
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(foreign)
library(haven)
library(performance)
library(lme4)
```

We will continue to use the *popularity* dataset we used in lab, as well as a new dataset for a repeated measures example. The second dataset comes from Chapter 10 of *Longitudinal Analysis* by Lesa Hoffman, and contains data from 207 older adults without a dementia diagnosis who were assessed roughly every two years for eight years on their memory (*prose recall*).

```r
popular = read_spss("popular2.sav")

memory = read_sas("SAS_Chapter10a.sas7bdat")
```

## Question 1

Use the *popularity* dataset to answer the following questions. In these questions, we are interested in predicting **teacher-rated** popularity of students (*popteach*) based on their gender.

### Part a) [1 point]

Calculate the intra-class correlation coefficient for how much variability in teacher-rated popularity there is across classrooms. Report the value, and state whether this is considered a large amount of variation (provide a reason, based on cut-offs provided in lecture).

```r
int_model <- lmer(popteach ~ 1 + (1|class), data= popular)
performance::icc(int_model)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.343
##   Unadjusted ICC: 0.343
```

**ICC Value:** 0.343 **Is this substantial?**: Since $\rho_{ic} > .05$, the errors can be considered substantial.

### Part b) [1.5 point]

Run 3 regression models, each predicting teacher-rated popularity using gender, and report the output:

- A OLS regression model

- A multilevel model with a random intercept only
- A multilevel model with a random intercept and random slope

```r
ols_mod <- lm(popteach ~ sex, data = popular)
ml_int_mod <- lmer(popteach ~ sex + (1|class), data = popular)
ml_sp_mod <- lmer(popteach ~ sex + (1 + sex|class), data = popular)

summary(ols_mod)
```

```
##
## Call:
## lm(formula = popteach ~ sex, data = popular)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8180 -0.8180  0.1820  0.7139  4.1820
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.28615    0.03743   114.5   <2e-16 ***
## sex          1.53185    0.05265    29.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.177 on 1998 degrees of freedom
## Multiple R-squared:  0.2976, Adjusted R-squared:  0.2973
## F-statistic: 846.6 on 1 and 1998 DF,  p-value: < 2.2e-16
```

```r
summary(ml_int_mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popteach ~ sex + (1 | class)
##    Data: popular
##
## REML criterion at convergence: 5777.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1587 -0.6728 -0.0121  0.6675  3.6230
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  class    (Intercept) 0.4731   0.6878
##  Residual             0.9303   0.9645
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  4.39327    0.07586   57.91
## sex          1.32022    0.04655   28.36
##
## Correlation of Fixed Effects:
##     (Intr)
## sex -0.311
```

```r
summary(ml_sp_mod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popteach ~ sex + (1 + sex | class)
##    Data: popular
##
## REML criterion at convergence: 5772.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1639 -0.6727 -0.0284  0.6788  3.5162
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  class    (Intercept) 0.54000  0.7348
##           sex         0.06164  0.2483   -0.45
##  Residual             0.91682  0.9575
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  4.39713    0.08046   54.65
## sex          1.32128    0.05322   24.83
##
## Correlation of Fixed Effects:
##     (Intr)
## sex -0.460
```

## Part c) [1 point]

How do the 3 models differ in terms of what is / is not allowed to differ across classrooms? In other words, what does it mean to include no random effects, a random intercept only, or a random intercept and random slope?

In the OLS model, the relationship between gender and popularity is assumed to be identical across all classrooms. That is, it gives us the average effect of gender on popularity. When we include a random intercept for classes, we allow them to have different average popularity levels, but thee effect of gender on popularity is still assumed to be the same. In the model with random intercepts and random slopes, classrooms can have different average popularity levels and the effect of gender on popularity is allowed to vary across classrooms.

## Part d) [1 point]

Report and interpret the fixed effects from the model with the random intercept and random slope.

**(Intercept)**: The intercept represents the average level of teacher-rated popularity across clusters when the student is a boy.

**Effect of sex**: The slope of sex represents the average effect of gender (or average difference between boys and girls) across clusters on teacher-rated popularity.

## Part e) [1 point]

How do the fixed effects and standard errors differ between the OLS model and the model with a random intercept and random slope?

**Fixed effect of intercept:** Similar in both models: OLS = 4.29, MLM = 4.40.

**SE of intercept:** Larger for the model with a random intercept and random slope: OLS = 0.037, MLM = 0.080.

**Fixed effect of gender:** Similar in both models: OLS = 1.53, MLM = 1.32.

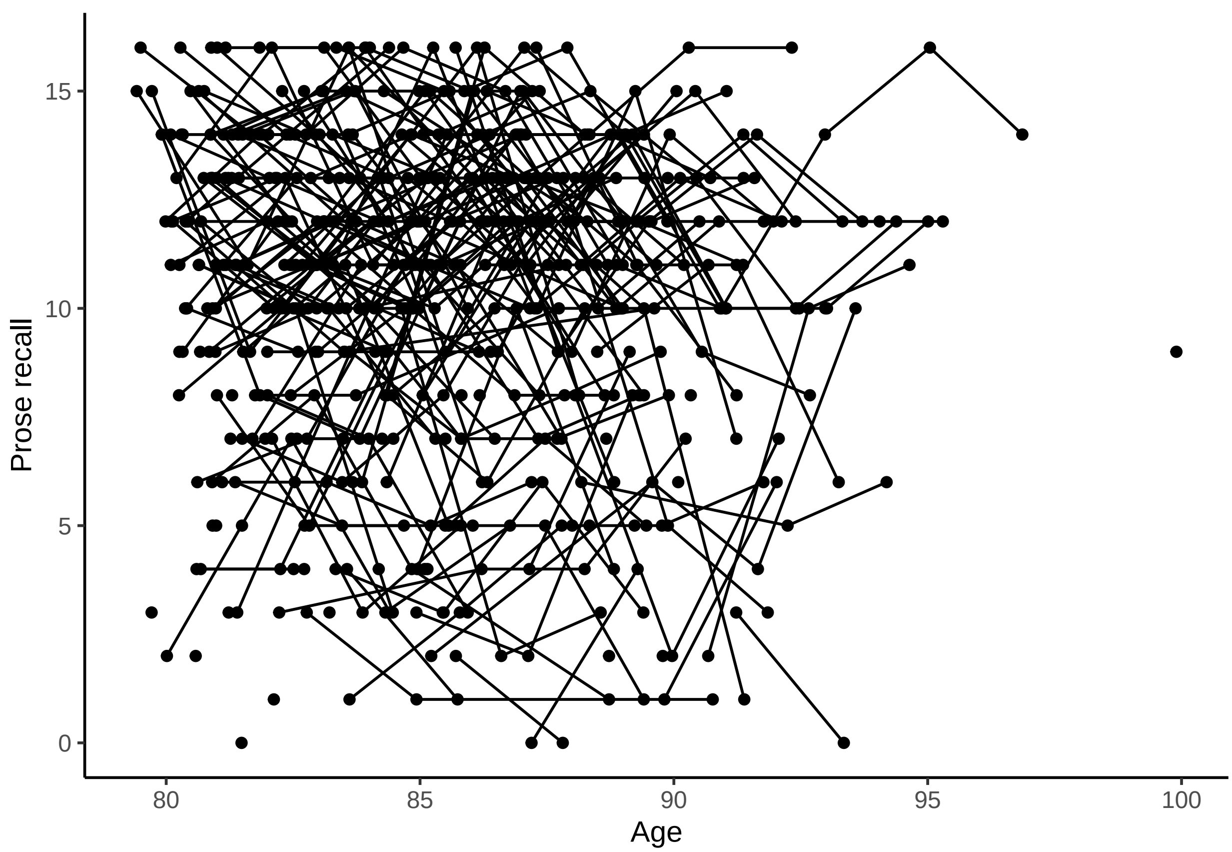
**SE of gender:** Very similar in both models: OLS = 0.053, MLM = 0.053.

# Question 2

## Part a) [1 point]

Create a plot to examine how prose recall (*recall*) changes with age (*tvage*), with age on the x-axis and prose recall on the y-axis. Be sure that observations from the same individual are connected with a line. Be sure to label your axes appropriately.

```
# code here
ggplot(data = memory, aes(x = tvage, y = recall, group = PersonID))+
  geom_point()+
  geom_line()+
  theme_classic()+
  labs(x = "Age", y = "Prose recall")
```




## Part b) [2 point]

Run a multilevel model to predict prose recall from age, including a random intercept and random slope. Be sure to include person-centered age (centered at age at the initial timepoint) and age at the initial timepoint as predictors. Note that age at the initial timepoint is already provided in the dataset (*ageT0*).

```r
# code here
memory_c <- memory |>
  dplyr::mutate(tvage_c = tvage - ageT0)

mod2b <- lmer(recall ~ ageT0 + tvage_c + (1 + tvage_c| PersonID), data = memory_c)

summary(mod2b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: recall ~ ageT0 + tvage_c + (1 + tvage_c | PersonID)
##    Data: memory_c
##
## REML criterion at convergence: 2841.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.89162 -0.47996  0.01651  0.51823  2.88728
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  PersonID (Intercept) 12.7181  3.5662
##           tvage_c      0.1266  0.3558   -0.48
##  Residual              4.1271  2.0315
## Number of obs: 557, groups:  PersonID, 207
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 25.05353    7.19744   3.481
## ageT0       -0.18260    0.08632  -2.115
## tvage_c     -0.04508    0.05066  -0.890
##
## Correlation of Fixed Effects:
##         (Intr) ageT0
## ageT0   -0.999
## tvage_c -0.077  0.060
```

## Part c) [1.5 points]

Report and interpret the fixed effects of the model.

**(Intercept):** The model's intercept is at 25.05, and corresponds to the expected prose recall for a person who is at the average initial age.

**Fixed effect of age at the initial timepoint:** The slope of initial age, which represents between-person differences in prose recall due to age. The negative slope of -0.18 indicates that people that are older at the start show lower prose recall.

**Fixed effect of age:** The slope of person-centered age is -0.05. It indicates that for every additional year a person gets older, their prose recall scores are expected to decrease by 0.05 points.

# Extra Credit

## EC 1 [0.5 point]

Using the ICC calculated in Question 1a, calculate the cluster effect. (*Note: each class has 100 students in it*)

```
# code

# does each class really have 100 students in it?
popular |>
  dplyr::with_groups(class,
                     summarise,
                     size = n()) |>
  dplyr::summarise(avg_size = mean(size))
```

```
## # A tibble: 1 x 1
##   avg_size
##      <dbl>
## 1       20
```

```
cluster_effect <- function(ni, icc) {
  1 + (ni - 1)*icc
}

cluster_effect(ni = 20, icc = 0.343)
```

```
## [1] 7.517
```

```
cluster_effect(ni = 100, icc = 0.343)
```

```
## [1] 34.957
```

## EC 2 [0.5 point]

Using the model run in Question 1b, with a random intercept and random slope - report and interpret the correlation between random effects.

The correlation between random effects of -0.45 indicates that classrooms that have larger intercepts are expected to have lower effects of gender on teacher-rated popularity.