Basics of multilevel models

We ended the last lecture wondering how we can we capture the inter-class variability of the extraversion-popularity relationship without fitting 100 separate regression models. In this lecture we will consider the task of incorporating distributional assumptions on the intercept and slope directly into the model. We will assume now that intercepts and slopes are drawn from a normal distribution with (unknown) means a and b.

Two-level model with random intercept and slope

In what follows, we will use the student popularity data set from last lecture as a running example.

We denote by y_{ij} the outcome (i.e. popularity) for individual i in class j, and x_{ij} the predictor (i.e. extraversion) for individual i in class j. Consider the following model

$$\text{Level 1:} \quad y_{ij} = a_j + b_j x_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2)$$

where for class j

$$\begin{aligned} \text{Level 2:} \quad a_j &= a + u_j, \qquad u_j \sim N(0, \sigma_u^2) \\ b_j &= b + v_j, \qquad v_j \sim N(0, \sigma_v^2). \end{aligned}$$

Remarks:

- The variables u_j and v_j are called random effects.
- Random effects are unobserved (unlike the fixed effects, x_{ij}). They can be thought of as random draws from an infinitely large population of upper-level units (here classes).
- Note that the random errors, ϵ_{ij}, are assumed to be independent from the random effects.
- The u_j and v_j may be (and usually are) correlated, i.e. there exists a variance matrix

$$Var \begin{pmatrix} u_j \\ v_j \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & r\sigma_u\sigma_v \\ r\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}.$$

- Upon substitution of the level 2 equations into the level 1 equation we obtain

$$\begin{aligned} y_{ij} &= (a + u_j) + (b + v_j)x_{ij} + \epsilon_{ij} \\ &= a + bx_{ij} + u_j + v_j x_{ij} + \epsilon_{ij}. \end{aligned}$$

- Finally, note that the expectation of y_{ij} is

$$\begin{aligned} E(y_{ij}) &= a + bx_{ij} + E(u_j) + E(v_j)x_{ij} + E(\epsilon_{ij}) \\ &= a + bx_{ij}. \end{aligned}$$

The variance is

$$\begin{aligned} Var(y_{ij}) &= 0 + Var(u_j + v_j x_{ij}) + Var(\epsilon_{ij}) \\ &= \sigma_u^2 + x_{ij}^2 \sigma_v^2 + 2r\sigma_u\sigma_v x_{ij} + \sigma^2. \end{aligned}$$

So, the fixed effects specify the (marginal) mean and the random effects specify the (marginal) variance. We will revisit the usefulness of these results later in these notes.

Fitting in R

We fit with (uncomment to run)

```
# model1 <- lmer(formula = popularity ~  1+ extraversion +
#  (1+ extraversion|class), data = pop.data)
```

for which we obtain output:

```
# summary(model1)
# Random effects:
#   Groups   Name          Variance Std.Dev. Corr
#   class    (Intercept)   2.99680  1.7311
#            extraversion  0.02595  0.1611   -0.97
#            Residual      0.89495  0.9460
#  Number of obs: 2000, groups:  class, 100
#
# Fixed effects:
#                 Estimate Std. Error   df     t #value       Pr(>|t|)
# (Intercept)   2.46106    0.20309  96.71010   12.12     <2e-16 ***
# extraversion  0.49286    0.02545  89.75489   19.36     <2e-16 ***
```

From this output we see that

- The expected popularity rises by 0.49286 per extra point of extraversion.
- There is considerable variation between classes: The random effect variances are $\sigma_u^2 = 2.9968$ and $\sigma_v^2 = 0.02595$.
- The random effects are strongly negatively correlated (-0.97): For classes where the popularity is generally larger, increasing extraversion will have less additional effect.

Recall that the random effects are unobserved. Predicted values can be extracted via `ranef(model1)`. They can further be visualised via `plot_model(model1, type="re")` which requires the `sjPlot` package. The command `predict(model1)` is also useful for extracting predicted response values $\hat{y}_{ij}$. These can then be plotted for each group on the same plot.

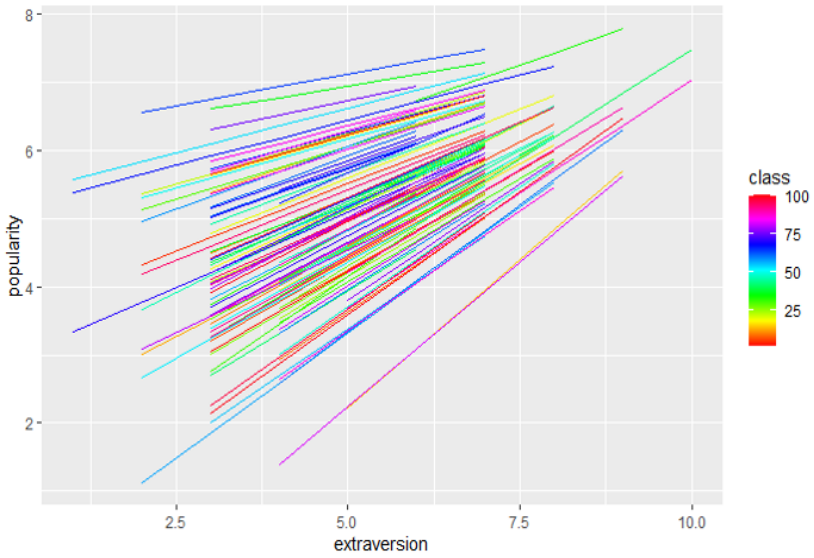

Figure 1: Predicted regression lines by class. Note that lines with large intercepts possess small slopes.

Centering

Recall our intercept estimate of $\hat{a} = 2.46$. Is this interpretable? In our case, x_{ij} is the value of extraversion of pupil i in class j. This was measured on a self-rating 10-point scale from 1 (min) to 10 (max). Hence, the intercept describes the expected response value at a non-defined predictor value. If x_{ij} were to include the value 0 as a possibility, then we'd interpret the intercept as the average response when $x_{ij} = 0$.

We can remedy this issue by centering the explanatory variables:

```
# m.Ex <- mean(pop.data£extraversion)
# m.ex
# pop.data£centre.extrav <- pop.data£extraversion-m.ex
# model1a <- lmer(formula = popularity ~
# 1+ centre.extrav +
# (1+ centre.extrav|class), data = pop.data)
```

Remarks

- Centering leaves slope estimates, standard errors, and random effects, unchanged.
- The intercept estimates, standard errors, and random effects, do change.
- Overall the fitted models are equivalent. This is just a reparametrization.
- The gain is the better interpretability. The intercept estimate is now the expected value of the outcome when the predictors take their mean value.

Intra-class correlation (ICC)

Consider the 'empty', or 'intercept-only' model:

$$y_{ij} = \gamma_0 + u_j + \epsilon_{ij}.$$

Note that

$$Var(y_{ij}) = \sigma_u^2 + \sigma^2.$$

This model decomposes the variance in the response into σ_u^2 and σ^2.

We now define the proportion of total variance explained by the grouping structure as the **intra-class correlation (ICC)**:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}$$

This can also be be interpreted as the expected correlation between two randomly drawn units that are in the same group. Note that in the social and educational sciences, ICC values as small as 0.25 can be considered "large".

For the pupil popularity data we have the following output

```
# intercept.only.model <- lmer(formula = popularity ~ 1
# + (1|class), data = pop.data)
# summary(intercept.only.model)
# Random effects:
# Groups       Name         Variance   Std.Dev.
# class       (Intercept)    0.7021    0.8379
# Residual                   1.2218    1.1053
# Number of obs: 2000, groups:  class, 100
#
# rho = 0.7021/(0.7021+1.2218)
# rho
# [1] 0.3649358
```

giving an ICC of 0.365.

We can also calculate ICC directly using certain R packages, such as `performance`:

```
# require(performance)
# icc(intercept.only.model)

# Intraclass correlation coefficient
#        Adjusted ICC: 0.365
#        Conditional ICC: 0.365
```

For the purposes of this course, we will (in general) ignore conditional ICC as it uses a different null model (when there are covariates).

Other models

So far, we have considered a **random intercept and slope model**:

$$y_{ij} = a + bx_{ij} + u_j + v_j x_{ij} + \epsilon_{ij}$$

and the **intercept only / empty model**:

$$y_{ij} = \gamma_0 + u_j + \epsilon_{ij}$$

If we take the empty model and add a covariate, we obtain the **random intercept model**:

$$y_{ij} = a + bx_{ij} + u_j + \epsilon_{ij}.$$

Remarks

- In such a model, each class gets assigned its own intercept, but the impact (i.e. slope) of the fixed effect (e.g. extraversion) on the response is equal for all classes.
- We still have $Var(y_{ij}) = \sigma_u^2 + \sigma^2$ and so the ICC can be calculated in the exactly same way as was done for the empty model.

Caveat: As soon as random slopes are involved, the previous expression for ICC becomes incorrect, since now $Var(y_{ij}) \neq \sigma_u^2 + \sigma^2$. Although a naive ICC can still be calculated, it is best avoided!