Durham University
Research Methods Centre

MATH43515: Multilevel Modelling

Lecture 4: Basics of multilevel models

Module Convenor/Tutor:

Andy Golightly

Outline (Lecture 4)

- The two-level model with random intercept and slope
- Use of the R function lmer
- Interpretation of the fitted model
- Extracting and visualizing random effects
- Centering
- Intra-class correlations
- The intercept-only model
- The random intercept model

From individual regressions to a single model

Question: How we can we capture the inter-class variability of the extraversion-popularity relationship without fitting 100 separate regression models?

Answer: Build a distributional assumption on the intercept and slope directly into the model.

We assume now that intercepts and slopes are drawn from a normal distribution with some (unknown) means a and b.

A multilevel model with random intercept and random slope

Denote by y_{ij} the outcome (here: popularity) for individual i in class j, and x_{ij} the predictor (here: extraversion) for individual i in class j. This gives rise to the model

$$y_{ij} = a_j + b_j x_{ij} + \epsilon_{ij}$$

Lower level model (individuals)

where for class j,

$$a_j = a + u_j \text{ with } u_j \sim N(0, \sigma_u^2),$$

$$b_j = b + v_j \text{ with } v_j \sim N(0, \sigma_v^2),$$

and $\epsilon_{ij} \sim N(0, \sigma^2)$ as usual.

Upper level models (classes)

Random effects

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. However, the u_j and v_j may be (and usually are) correlated, i.e. there exists a variance matrix

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

where r is the correlation between u_j and v_j.

Fixed and random parts of the model

The preceding model can be rewritten as

$$y_{ij} = (a + u_j) + (b + v_j)x_{ij} + \epsilon_{ij} =$$

$$\underbrace{a + bx_{ij}}_{\text{fixed effects}} + \underbrace{\underbrace{u_j + v_j x_{ij}}_{\text{random effects}} + \epsilon_{ij}}_{\text{random part of the model}}$$

Models which contain fixed and random elements (such as this one) are also sometimes called mixed models or mixed effect models.

Marginal mean and variance

Taking expectation and variance over the random parts,

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

and

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

So,

- fixed effects specify the (marginal) mean.
- random effects specify the (marginal) variance.

Fitting this model in R

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

Fixed effects

Grouping factor

Random effects

Model output

σ_u^2

```
> summary(model1)
 Random effects:
  Groups   Name        Variance Std.Dev. Corr
```

σ_v^2

```
 class    (Intercept)  2.99680  1.7311
          extraversion 0.02595  0.1611   -0.97
          Residual     0.89495  0.9460
```

σ^2

```
 Number of obs: 2000, groups:  class, 100
```

$\hat{a}$

```
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 ***
```

$\hat{b}$

Durham University

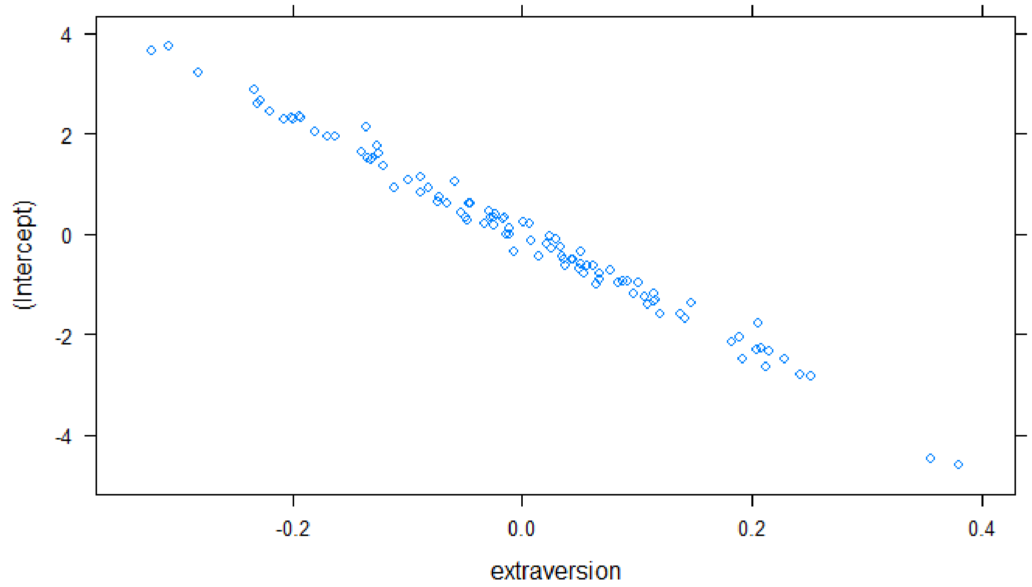
Some interpretation of fitted model

- 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.
- Note the increased standard errors of fixed effect parameters as compared to the simple linear model (the latter can be considered incorrect)

Extraction of predicted random effects

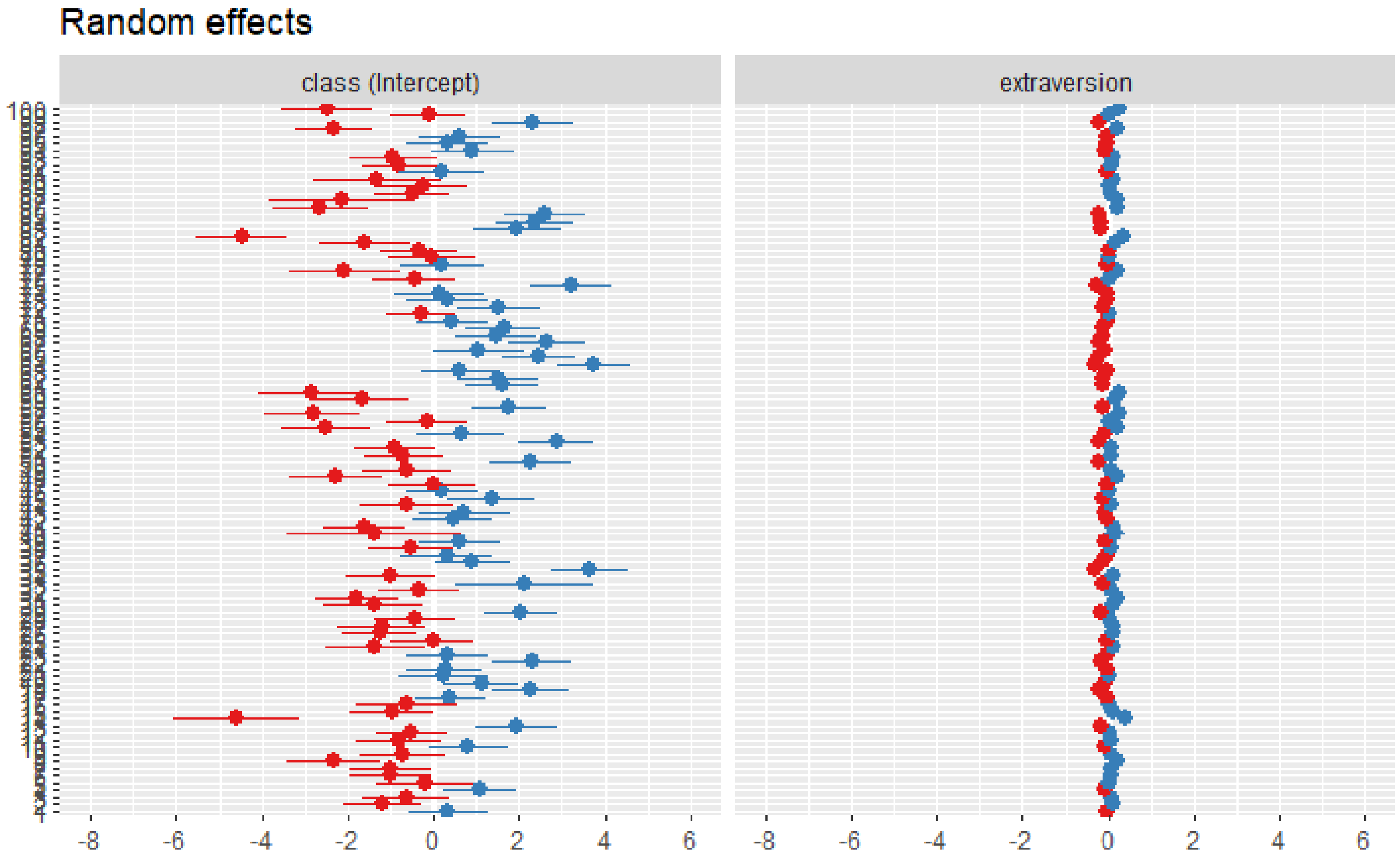
```
> ranef(model1)
$class
    (Intercept) extraversion
1    0.34088342 -0.027005565
2   -1.17789618  0.096951448
3   -0.62924340  0.057068990
4    1.08499615 -0.099897738
5   -0.19480592  0.021582089
6   -0.98327091  0.083734383
7   -1.00084117  0.064857187
...
100  -2.48262132  0.228942717
```

For each class, these are the *predicted* values of the random intercept and slope, given the data and the fitted model.



Durham University

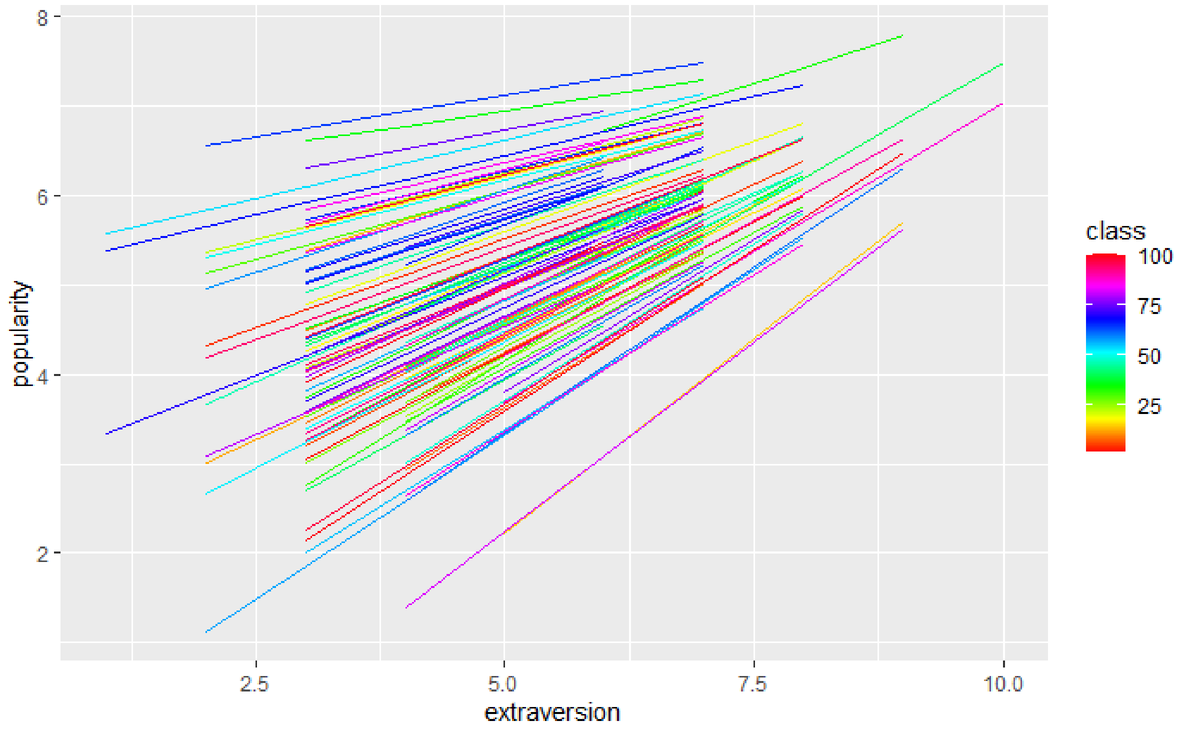
Visualizing random effects




```
> require(sjPlot)
> plot_model(model1, type="re")
```

Durham University

Predicted class-wise regression lines



Note the narrowing shape: Regression lines with large intercepts possess small slopes

Centering

Recall our intercept estimate of $\hat{a} = 2.46106$.

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.

This is unfortunate. If we have an explicit parameter in the model, we would like it to mean something!

Centering (cont'd)

Therefore, in the context of multilevel model, it is common practice to centre 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)
```

Centering (cont'd)

```
> summary(model1a)
Random effects:
 Groups   Name          Variance Std.Dev. Corr
 class    (Intercept)   0.89178  0.9443
          centre.extrav 0.02599  0.1612   -0.88
 Residual               0.89492  0.9460
Number of obs: 2000, groups:  class, 100

Fixed effects:
              Estimate   Std. Error   df         t value   Pr(>|t|)
(Intercept)   5.03127    0.09702      97.07723   51.86     <2e-16 ***
centre.extrav 0.49286    0.02546      89.69832   19.36     <2e-16 ***
```

Centering (cont'd)

```
> summary(model1a)
Random effects:
 Groups   Name          Variance Std.Dev. Corr
 class    (Intercept)   0.89178  0.9443
          centre.extrav 0.02599  0.1612   -0.88
 Residual               0.89492  0.9460
Number of obs: 2000, groups:  class, 100

Fixed effects:
              Estimate Std. Error       df t value Pr(>|t|)
(Intercept)    5.03127    0.09702 97.07723   51.86   <2e-16 ***
centre.extrav  0.49286    0.02546 89.69832   19.36   <2e-16 ***
```

- 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 predictor take(s) their mean value.

How much variance is explained by the grouping structure?

Easiest way of approaching this question: Consider the \`empty', or \`intercept-only' model:

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

where γ_0 is a (fixed effect) intercept, u_j is a group-specific random effect with variance σ_u^2, and ϵ_{ij} is random error with variance σ^2.

This model does not explain any variance in the response. It just decomposes it

$$Var(y_{ij}) = 0 + Var(u_j) + Var(\epsilon_{ij})$$
$$= \sigma_u^2 + \sigma^2$$

into the variance components σ_u^2 and σ^2.

Decomposing variances (cont'd)

The proportion of the total variance explained by the grouping structure is then given by the intra-class-correlation (ICC)

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

The ICC can also be interpreted as the expected correlation between two randomly drawn units that are in the same group.

In the social and educational sciences, ICC values as small as 0.25 can be considered "large".

ICC for pupil popularity data

```
> 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
```

36% of the variance of the popularity scores resides on the group (class) level.

ICC for pupil popularity data (cont'd)

The ICC can be calculated 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
```

The words "adjusted" and "Conditional" are misleading here. This ICC is in fact unconditional (the model is not conditional on any covariates!)

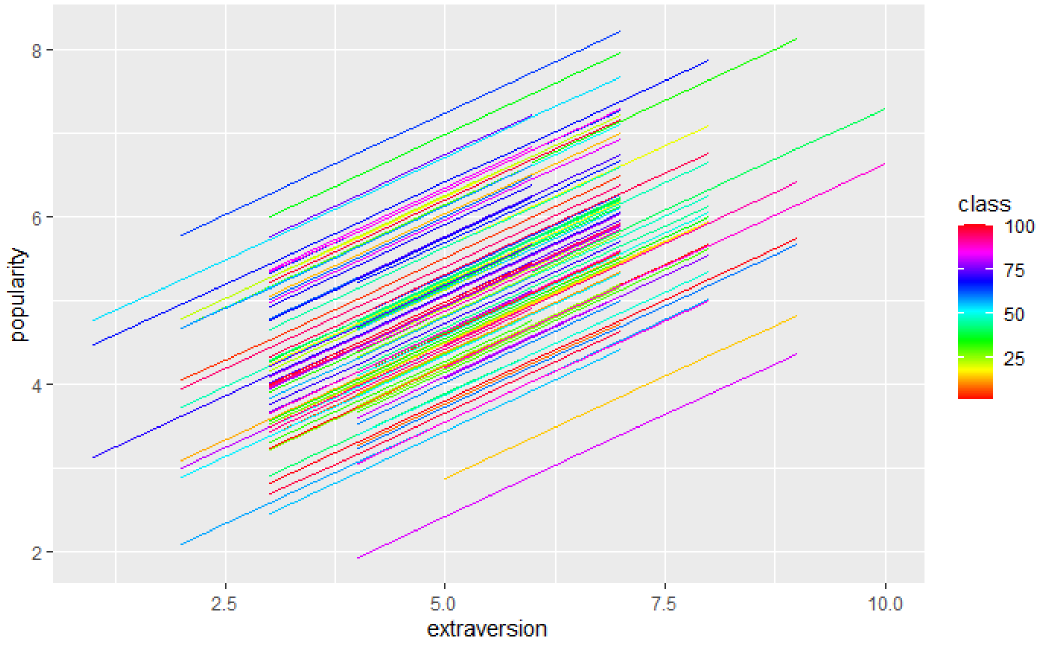
The random intercept model

If there are no random slopes, that is $\sigma_v^2 = 0$, we talk of a random intercept model.

```
model0 <- lmer(formula = popularity ~ 1+ extraversion
+(1|class), data    = pop.data)

Random effects:
 Groups   Name        Variance Std.Dev.
 class    (Intercept) 0.8406   0.9168
 Residual             0.9304   0.9646
Number of obs: 2000, groups:  class, 100

Fixed effects:
              Estimate Std. Error       df     t value Pr(>|t|)
(Intercept)  2.542e+00  1.411e-01 4.380e+02   18.01   <2e-16
extraversion 4.863e-01  2.015e-02 1.965e+03   24.13   <2e-16
```



In such a model, each class gets assigned its own intercept, but the impact (=slope) of the fixed effect (here extraversion) on the response is equal for all classes.

Durham University

ICC for the random intercept model

In terms of complexity, the random intercept model sits between the intercept-only-model and the random-intercept-and-random-slope-model.

Since we have still

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

the formula for the ICC stays the same:

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

Sometimes called adjusted ICC in this context.

Manual computation (from variances given in model output):

```
> 0.8406/(0.9304+0.8406)
0.4746471
```

Automated computation:

```
> icc(model0)
# Intraclass Correlation Coefficient
 Adjusted ICC: 0.475
Conditional ICC: 0.391
```

See for details:
https://easystats.github.io/performance/reference/icc.html

ICC in the presence of random slopes

As soon as random slopes are involved, the previous expression for ICC becomes incorrect, since now $Var(y_{ij}) \neq \sigma_u^2 + \sigma^2$. However, it is still sometimes used, “naïve ICC”:

```
> summary(model1)$varcor
Groups  Name        Std.Dev. Corr
 class   (Intercept) 1.73113
         extraversion 0.16110 -0.966
 Residual            0.94602
> 1.73113^2/(1.73113^2+0.94602^2)
[1] 0.7700391
```

Very likely an overestimate; could be checked via function icc….

Better avoid computing ICC under random slopes!

lmer (lme4)

Formula notation according to manual

Formula	Alternative	Meaning
`(1 | g)`	`1 + (1 | g)`	Random intercept with fixed mean.
`0 + offset(o) + (1 | g)`	`-1 + offset(o) + (1 | g)`	Random intercept with *a priori* means.
`(1 | g1/g2)`	`(1 | g1)+(1 | g1:g2)`	Intercept varying among g1 and g2 within g1.
`(1 | g1) + (1 | g2)`	`1 + (1 | g1) + (1 | g2).`	Intercept varying among g1 and g2.
`x + (x | g)`	`1 + x + (1 + x | g)`	Correlated random intercept and slope.
`x + (x || g)`	`1 + x + (1 | g) + (0 + x | g)`	Uncorrelated random intercept and slope.

Table 2: Examples of the right-hand-sides of mixed-effects model formulas. The names of grouping factors are denoted g, g1, and g2, and covariates and *a priori* known offsets as x and o.

To be further illustrated in this week's practical!

https://cran.r-project.org/web/packages/lme4/

Thank you!!!!!

Durham University
Research Methods Centre

andrew.golightly@durham.ac.uk