Advanced multilevel models

In this lecture, we will consider how to assess significances of fixed and random effects via likelihood ratio tests, models with more than one predictor variable, model simplification strategies and 3-level models.

We will work predominantly in the framework of the random intercept and slope model from the previous lecture 4, and in the context of the student popularity data. Recall that the model is given by

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

where y_{ij} the outcome (i.e. popularity) for individual i in class j, x_{ij} the predictor (i.e. extraversion) for individual i in class j and u_j, v_j are the random effects (intercept and slope).

Assessing significances

Fixed effects

Here, we follow the procedure for the linear regression model scenario. We consider hypothesis tests

$$H_0 : a = 0 \quad \text{vs} \quad H_1 : a \neq 0$$

and

$$H_0 : b = 0 \quad \text{vs} \quad H_1 : b \neq 0$$

with typical focus on the latter. The t-values are given by $|\text{Estimate/Std.Error}|$, in this case 12.12 and 19.36, respectively. Note that the corresponding p-values give the probabilities of "obtaining by chance an even more extreme value of the test statistic t". Both p-values are approximately 0, so in both cases H_0 is clearly rejected, and we conclude that both fixed effects are needed in the model.

Random effects

For the random effects, we consider the hypothesis tests

$$H_0 : \sigma_u^2 = 0 \quad \text{vs} \quad H_1 : \sigma_u^2 \neq 0$$

and

$$H_0 : \sigma_v^2 = 0 \quad \text{vs} \quad H_1 : \sigma_v^2 \neq 0.$$

We can't simply proceed as before and compute a t-value of the form $|\text{Estimate/Std.Error}|$ as the standard errors of the variance estimators are not given to us from the model output and the sampling distribution of the test statistic is not a student t distribution. (In fact, the sampling distribution of the variance estimator is right skewed.)

For random effects, one should carry out chi-squared tests based on deviances. This is achieved by fitting the (full) model with and (reduced) model without the random effect term, and then comparing the difference in deviances with a χ^2 distribution with degrees of freedom given by the difference in numbers of parameters between the full and reduced models.

The deviance is defined as

$$D = -2\log L + c$$

where L is the likelihood (that is, the probability of seeing the data) and c is a constant that cancels out when taking differences in deviances between nested models. To this end, consider models M_0 and M_1 with M_0 nested inside M_1. Then, $L_0 < L_1$ and $D_0 > D_1$. We then compare M_0 and M_1 via the test statistic

$$D_0 - D_1 = -2\log L_0 + 2\log L_1 = -2\log\frac{L_0}{L_1}$$

and compare this to a χ^2 distribution (with the appropriate degrees of freedom).

Example 1: consider the student popularity data. Let's test for a random slope:

$$H_0 : \sigma_v^2 = 0 \quad \text{vs} \quad H_1 : \sigma_v^2 \neq 0.$$

Note that the absence of a random slope will also remove the covariance between the random slope and random intercept. Hence, there will be two fewer parameters in the reduced model. We use the following R code (uncomment as necessary):

```
# model0 <- lmer(formula = popularity ~  1+ extraversion +(1|class),
# data = pop.data)
# deviance(model0)
## [1] 5832.639
# model1 <- lmer(formula = popularity ~  1+ extraversion +
# (1+ extraversion|class), data = pop.data)
# deviance(model1)
## [1] 5779.395
# deviance(model0)-deviance(model1)
## [1] 53.24409
# qchisq(0.95,2)
## [1] 5.991465
```

The deviance difference is given by

$$D_0 - D_1 = 53.244.$$

At the 5% level of significance, this is compared to the 95% quantile of the χ_2^2 distribution, which is 5.99. So, H_0 is clearly rejected, and we decide that the random slope is necessary.

Remarks:

- R gives a deprecation warning, as a possible misuse of `deviance` is for testing for differences in fixed effects.

- An alternative and equivalent approach is to use `ranova(model1)`.
- We could now test for the inclusion of random intercepts although for this example we know that the variance of the random intercept variance is much larger than that of the slope (see last lecture). Moreover, a model with a random slope but no random intercept would be considered unusual without good modelling reasons (e.g. equal baseline response for all groups in the study).

Example 2: We include gender and teacher experience into the model for student popularity.

```
# model2 <-
# lmer(formula =
# popularity ~ 1 + gender + extraversion + experience
# + (1 + gender + extraversion | class),
# data    = pop.data)
```

Note that it is not possible to include a random slope for the class level variable `experience` in a meaningful way.

The usual model output can be obtained via `summary(model2)` for which we see a very small variance for the `gender` random slopes. We therefore test to see if a random effect for gender is required in the model:

```
# model3 <- lmer(formula = popularity ~ 1 + gender
# + extraversion + experience + (1 + extraversion | class),
#               data = pop.data)
# deviance(model3)-deviance(model2)
## [1]  1.513252
# qchisq(0.95,3)
## [1] 7.814728
```

for which we find insufficient evidence to reject the null hypothesis (that the gender effect variance is 0) and we do remove the random effect for gender on this occasion. We prefer `model3`, whose fit is displayed graphically below.

Diagnostics

There is significant overlap here with the checking procedure we met for linear regression models. We can plot residuals against fitted values and QQ plots of residuals via

```
# par(mfrow=c(1,2))
# plot(model3)
# qqnorm(resid(model3))
# qqline(resid(model3), col = "red")
```

We can also check the normality assumption of the random intercepts and slopes via

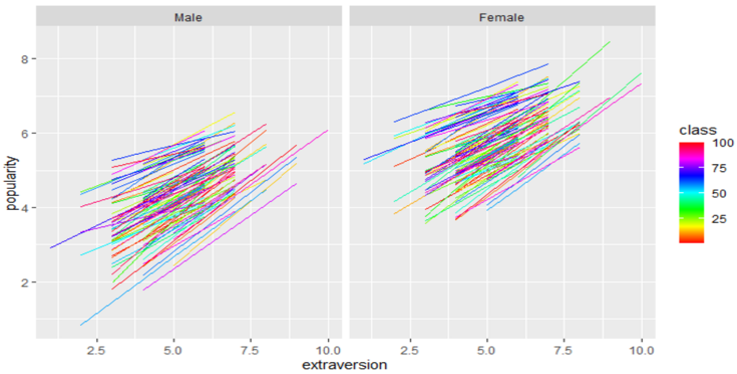

Figure 1: Predicted regression lines by class and gender (`model3`).

```
# par(mfrow=c(1,2))
# qqnorm(ranef(model3)£class[,1] )
# qqline(ranef(model3)£class[,1], col = "red")
# qqnorm(ranef(model3)£class[,2])
# qqline(ranef(model3)£class[,2], col = "red")
```

Three level models

Adding an additional level is straightforward, at least in principle. The following figure illustrates a three level data structure with pupils nested within classes nested within schools.

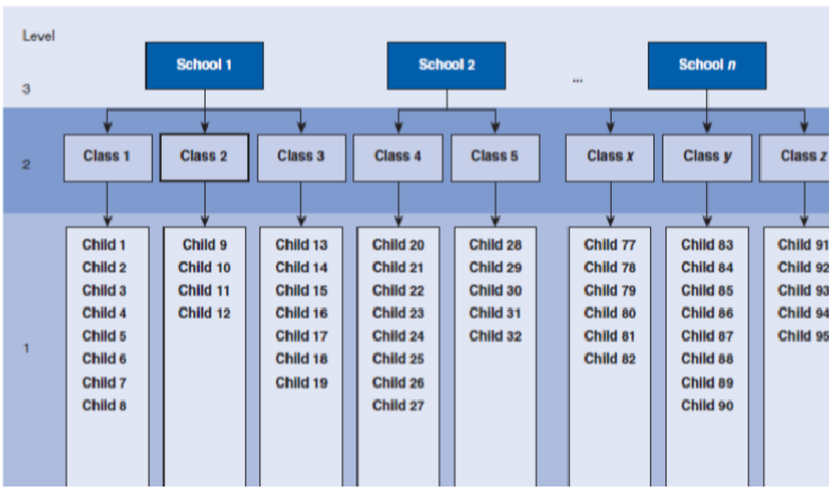

Figure 2: Three level data structure illustration.

We begin with the empty model given by

$$y_{ijk} = \gamma_0 + u_{jk} + v_k + \epsilon_{ijk}$$

where

- y_{ijk} is the response variable for the ith person in the jth class in the kth school,
- γ_0 is the overall intercept (average),
- u_{jk} is the random effect for the class and $u_{jk} \sim N(0, \sigma_u^2)$,
- v_{jk} is the random effect for the school and $v_k \sim N(0, \sigma_v^2)$,
- ϵ_{ijk} is the (student level) residucal error term and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$.

The school, class effects and the student level residual errors are assumed independent.

Variance partition components (VPCs)

For the empty model, it should be clear that

$$Var(y_{ijk}) = \sigma_u^2 + \sigma_v^2 + \sigma_e^2$$

where

- σ_v^2 is the variance component at the school level,
- σ_u^2 is the variance component at the class level,
- σ_e^2 is the individual (student) variance.

Variance partition coefficients (VPCs) report the proportion of the observed response variation that lies at each level of the model hierarchy. The school level VPC is calculated as

$$\frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}.$$

The class level VPC is calculated as

$$\frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}.$$

The individual level VPC is calculated as

$$\frac{\sigma_e^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}.$$

Intraclass correlation coefficients (ICCs)

Intraclass correlation coefficients (ICCs) measure the model implied correlation (i.e. similarity or homogeneity) of the observed responses within a given cluster

The school level ICC is calculated as the correlation between two individuals i and i' within the same school k, but different classrooms j and j' as

$$\frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}.$$

The classroom level ICC is calculated as the correlation between two students i and i' within the same classroom j and therefore the same school k as

$$\frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}.$$

We see that the classroom level ICC does not coincide with the classroom level VPC.

R code for fitting three level models

Consider an application with the following variables:

- `Math` – maths score (response variable)
- `ActiveTime` – time spent studying (student level covariate)
- `ClassSize` – class level covariate
- `Classroom` – class identifier
- `School` – school identifier
- `StudentID` – student identifier

We fit the empty model and summarise with

```
# Model.0 <- lmer(Math ~ 1+(1|School)+(1|School:Classroom),
#                         data=Sim3level)
# summary(Model.0)
```

We can include the two covariates `ActiveTime` and `ClassSize` to give the model:

$$y_{ijk} = \gamma_0 + \beta_1 \texttt{ActiveTime}_{ijk} + \beta_2 \texttt{ClassSize}_{jk} + u_{jk} + v_k + \epsilon_{ijk}.$$

Fitting this model can be achieved with

```
# Model.1 <- lmer(Math ~ ActiveTime+ClassSize
#                    +(1|School)
#                    +(1|School:Classroom),
#                    data=Sim3level)
```

or equivalently (and more succinctly)

```
# Model.1 <- lmer(Math ~ ActiveTime+ClassSize
#                    +(1|School/Classroom),
#                    data=Sim3level)
```

A discussion of the R code output and whether or not the covariates improve fit can be found in the lecture slides.