2025LMMcourse Session 3 Practical

Fang Yang

2025-05-09

1 Part I. Outstanding questions from last time

1.1 What is ICC

First of all what is the residual of the random effects (sigma squared)? It is also called Residual Variance, which represents the within group variances.

When dealing with hierarchical modelling, the variances come from both within group (e.g., between different tests of a particular participant) and between group (e.g., between different participants).

To help us visualise this, recall our plot from last time

Predicted Effects of the Online Course on Vocabulary Test Score
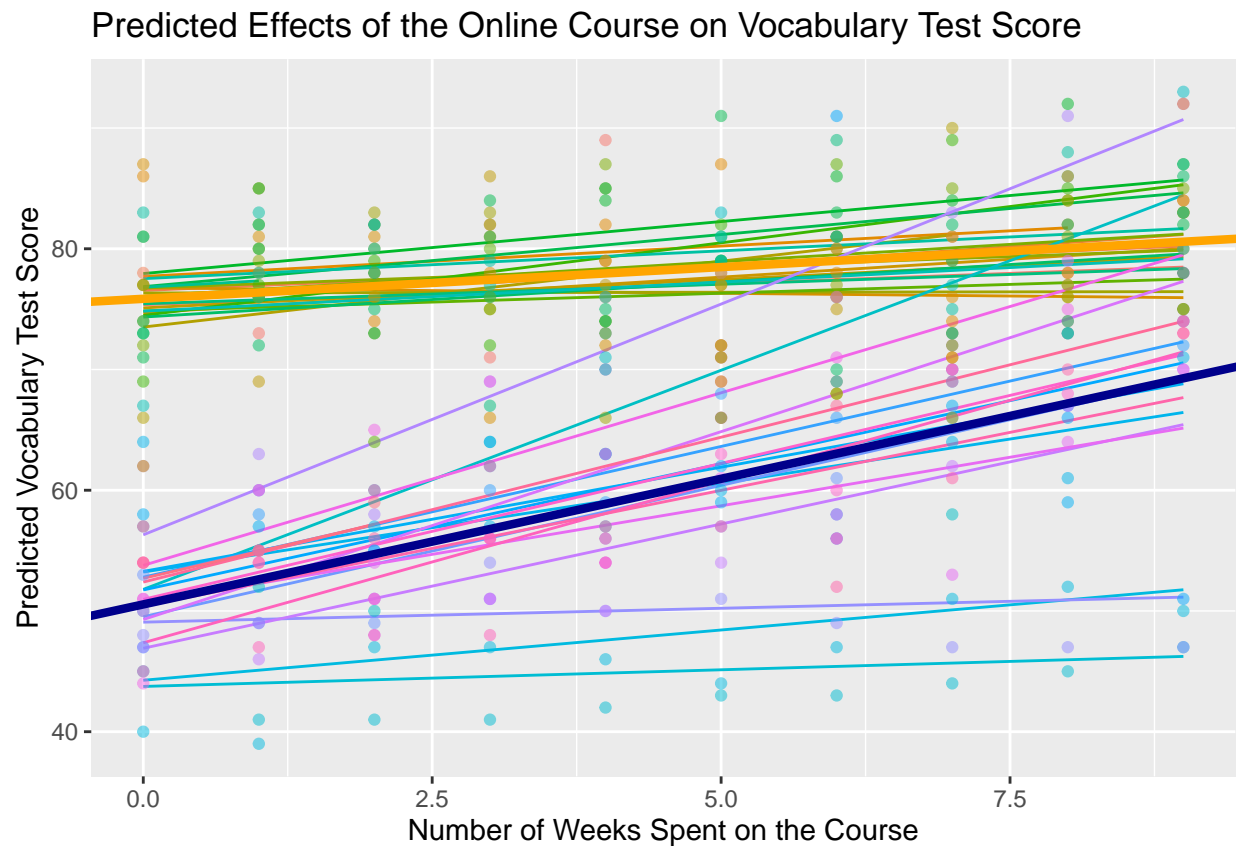
So the key is to understand between-group variance and within-group variance (i.e., residuals)

ICC stands for "IntraClass Correlation Coefficient".

It represents how much variation in a variable is attributable to the grouping factor.

ICC = Variance between the groups / Total variance in the variable = Between group variances / (Between Group Variances + Within Group Variances)

The ICC is calculated by dividing the random effect variance, sigma squaredi, by the total variance, i.e. the sum of the random effect variance (i.e., between-group variances) and the residual variance (i.e., within group variances), sigma squared or epsilon.

The higher the ICC is, the more justified for fitting a hierarchical model. It ranges from 0 to 1 (with no cutoffs).

Let's put this all in context. Recall our model results from last time.

```r
# recall our model from last time

# the intercept-only model
mMixed_simplest.model <- lmerTest::lmer(vocab_test_score ~ week * proficiency +
                            (1 | participant),
                      data = vocabdata,
                      control=lmerControl(optimizer="bobyqa",
                        optCtrl=list(maxfun=2e5)))
summary(mMixed_simplest.model)
tab_model(mMixed_simplest.model)

# Residual of the random effects (residual variance): sigma squared = 35.45
# this is the WITHIN-group (or residual) variance

# Random intercept variance: Tau00_participant = 25.47
# this is the BETWEEN-group variance for the intercept

# ICC = Between/(Between + Within) = 25.47/(25.47+35.45) = 0.42
```

```r
mMixed_reduced <- lmerTest::lmer(vocab_test_score ~ week * proficiency +
                            (1 + week | participant),
                      data = vocabdata,
                      control=lmerControl(optimizer="bobyqa",
                        optCtrl=list(maxfun=2e5)))
summary(mMixed_reduced)

tab_model(mMixed_reduced)

# Residual of the random effects (residual variance): sigma squared = 29.51.
# This is the WITHIN-group (or residual) variance

# Between-group random intercept variance: Tau00_participant = 10.09.
# This is the BETWEEN-group variance for the intercept.

# Between-group random slope variance: Tau11 = 0.69. This represents how much
# groups or subjects differ from each other in their slopes, i.e., in their
# responses to a particular fixed effect (e.g., week).

# Random slope-intercept correlation : Rho01 participant = 0.11.

# ICC = between-group variance/total variance
```

```
#        = (10.09 * 0.69)/(10.09 * 0.69 + 29.51)
#        = 0.19
```

Why on earth does the tab_model results report ICC = 0.52 then?

Because icc() calculates an adjusted and an unadjusted ICC, which both take all sources of uncertainty (i.e. of all random effects) into account. While the adjusted ICC only relates to the random effects, the unadjusted ICC also takes the fixed effects variances into account, more precisely, the fixed effects variance is added to the denominator of the formula to calculate the ICC (see Nakagawa et al. 2017). Typically, the adjusted ICC is of interest when the analysis of random effects is of interest. source: https://easystats. github.io/performance/reference/icc.html

```
# we can use the icc() function from the "performance" package to extract
# both ICC values directly

library("performance")
icc(mMixed_reduced) # unadjusted ICC = 0.19, matching our manual calculation.
```

2   Part II. Exercise - Continue exploring the vocab_learning data

Can you fit a model with a different structure of random effects and interpret the results?

Hint: consider including some or all of the following: - random intercept?
- random slope of one predictor? - random slopes of both predictors? - random slope of the interaction between the two predictors?

2.1   Example 1

Model with random intercept and random slope of "proficiency"

Tips: Deal with boundry fit issue.

Note the last line in the output; "boundary (singular) fit: see help('isSingular')", nearly suggesting overfitting. Basically, your model is over thinking - there is not as much variation in the data as the model tries to construct.

Solutions: 1. Remove the most complex part of the random effects structure (i.e. random slopes) 2. Maybe acceptable to remove a specific random effect term when its variance estimates are very low

2.2   Example 2

Model with intercept and slopes of both "week" and "proficiency"

Tips: Deal with convergence issue.

Note the last line in the output; Model failed to converge with max|grad| = 0.00665298 (tol = 0.002, component 1).The optimiser we are using can't find a suitable estimation for the maximum likelihood.

One solution is to adjust stopping (convergence) tolerances for the nonlinear optimizer, using the optCtrl() argument to lmerControl.

Now include optimizer control in the model and refit

It seems our best choice is (1 + week| participant) for this dataset.

3 Part III. A NEW DATASET

A tech company are exploring whether the daily time their customers spend on Instagram is dependent of customers' occupation: undergraduates, full-time employees, vs. the retired. They randomly sampled 15-20 customers from each population and collected other information (e.g., age, being a user of TikTok or not).They tracked the participants for 6 weeks.

e time spent on Instagram daily by different groups
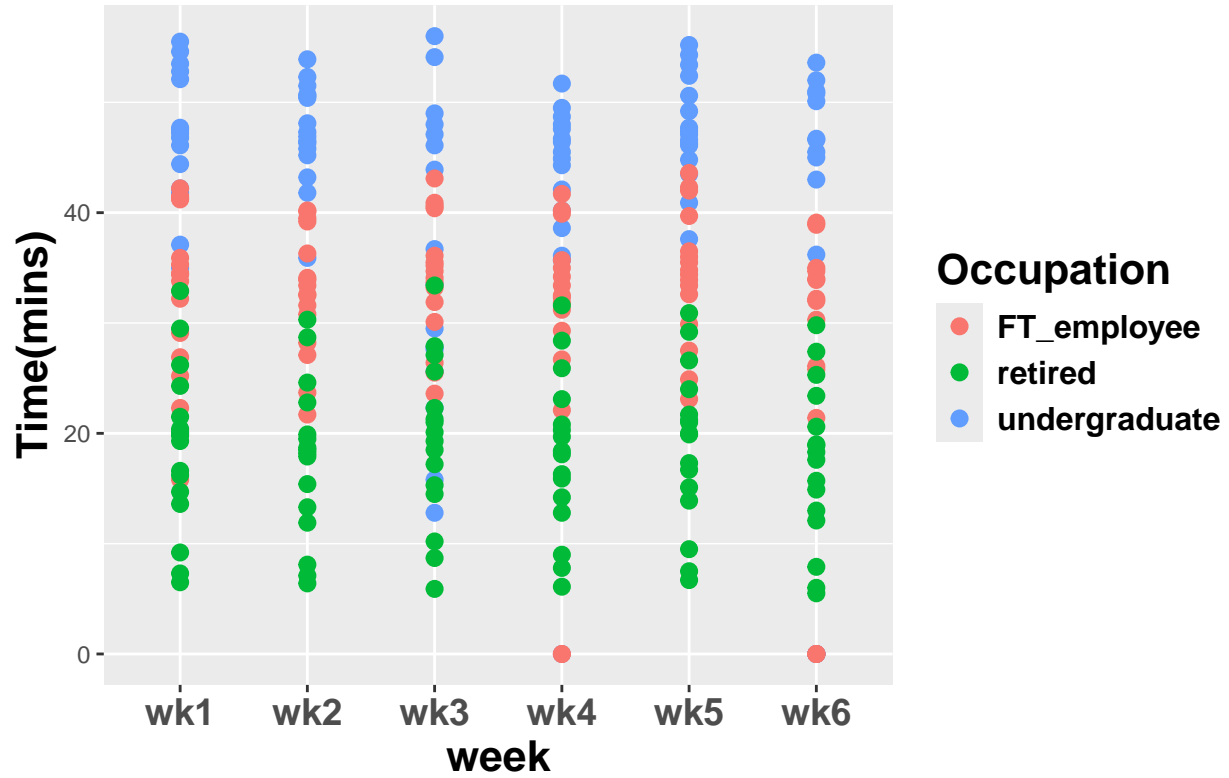
Average time spent on Instagram daily by different groups
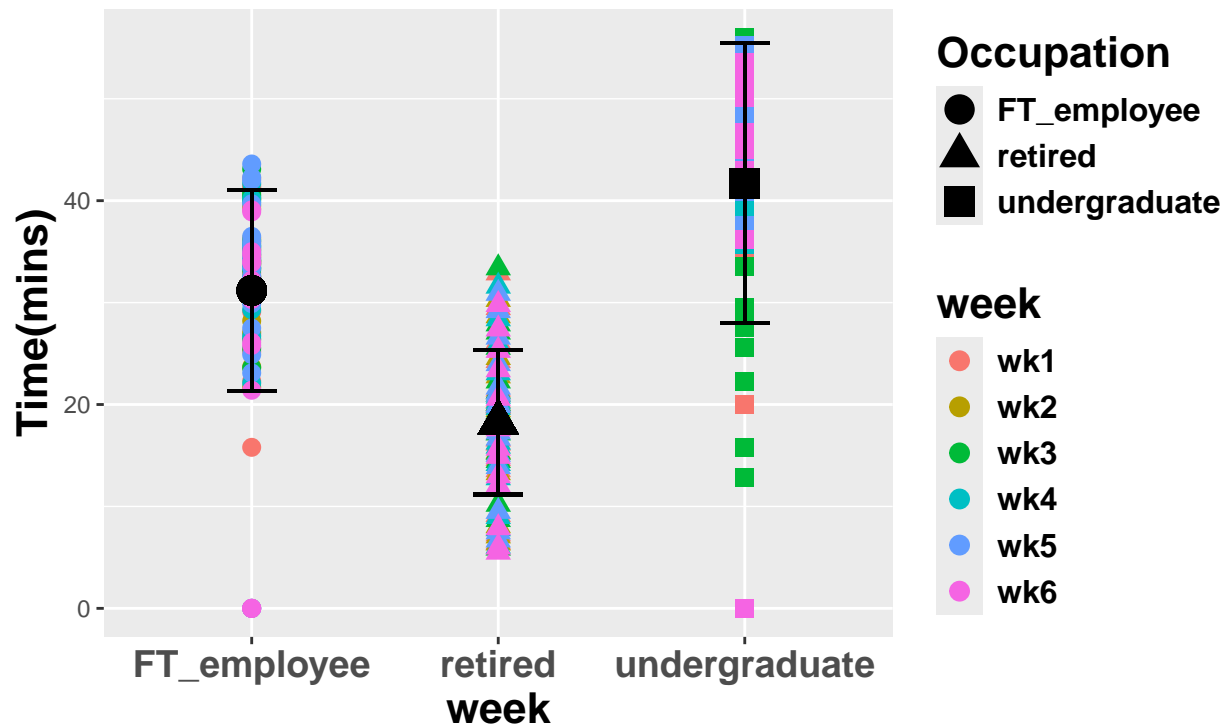
Next you can try to fit some models.

Start with a simple regression without random effects?

How about a mixed-effect model? What factors can be treated as random effects?

Anything else you may want to include as fixed effects? Do you have theoretical/practical justification for this?

4 Part IV. Self-reflection & Discussion

In what ways can linear mixed-effects models help you address the research questions of your own research ?