THE UNIVERSITY of EDINBURGH
Centre for Data, Culture & Society

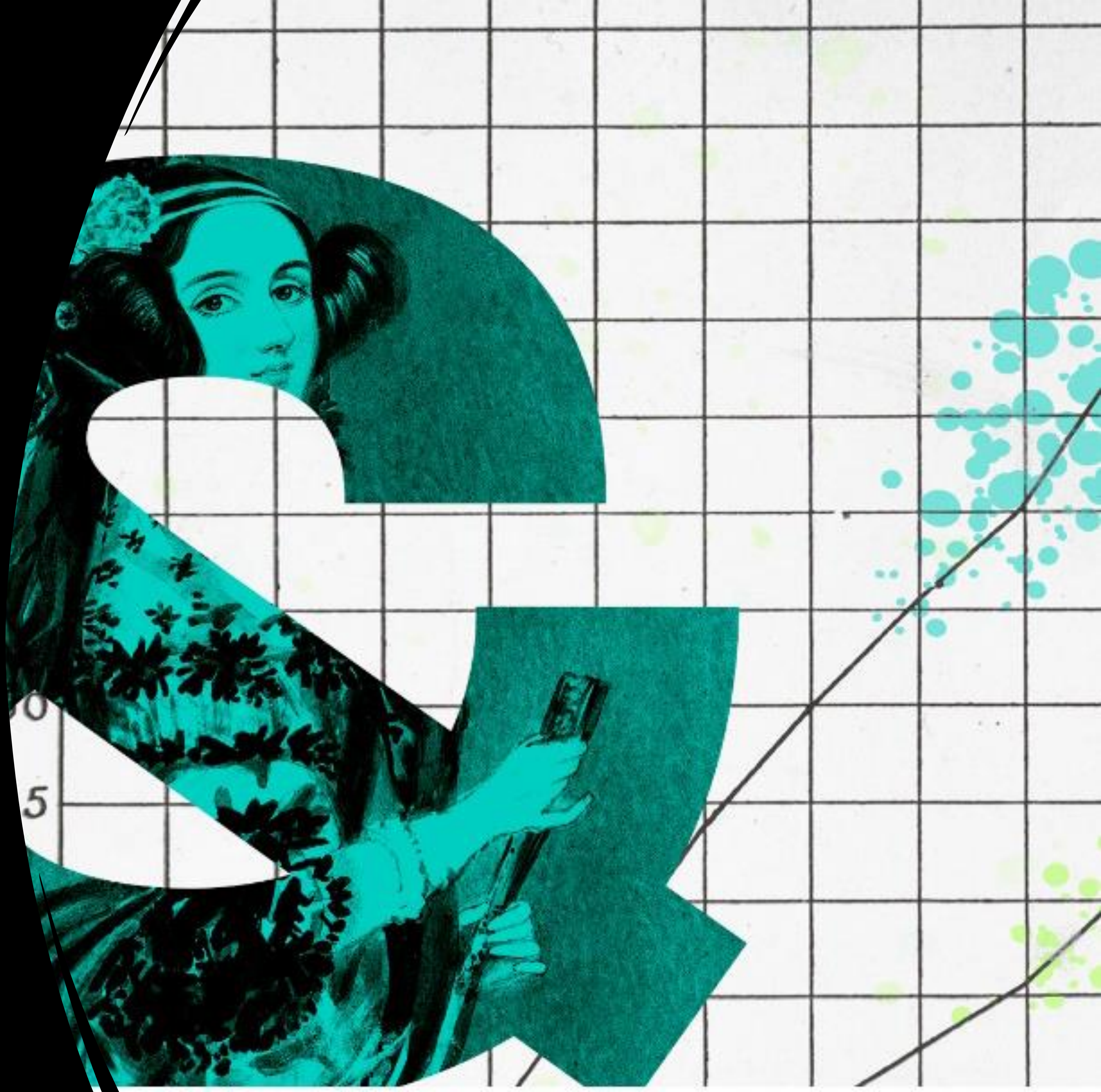
Regression & Mixed-Effects Modelling with R

Fang Jackson-Yang

6 May 2024

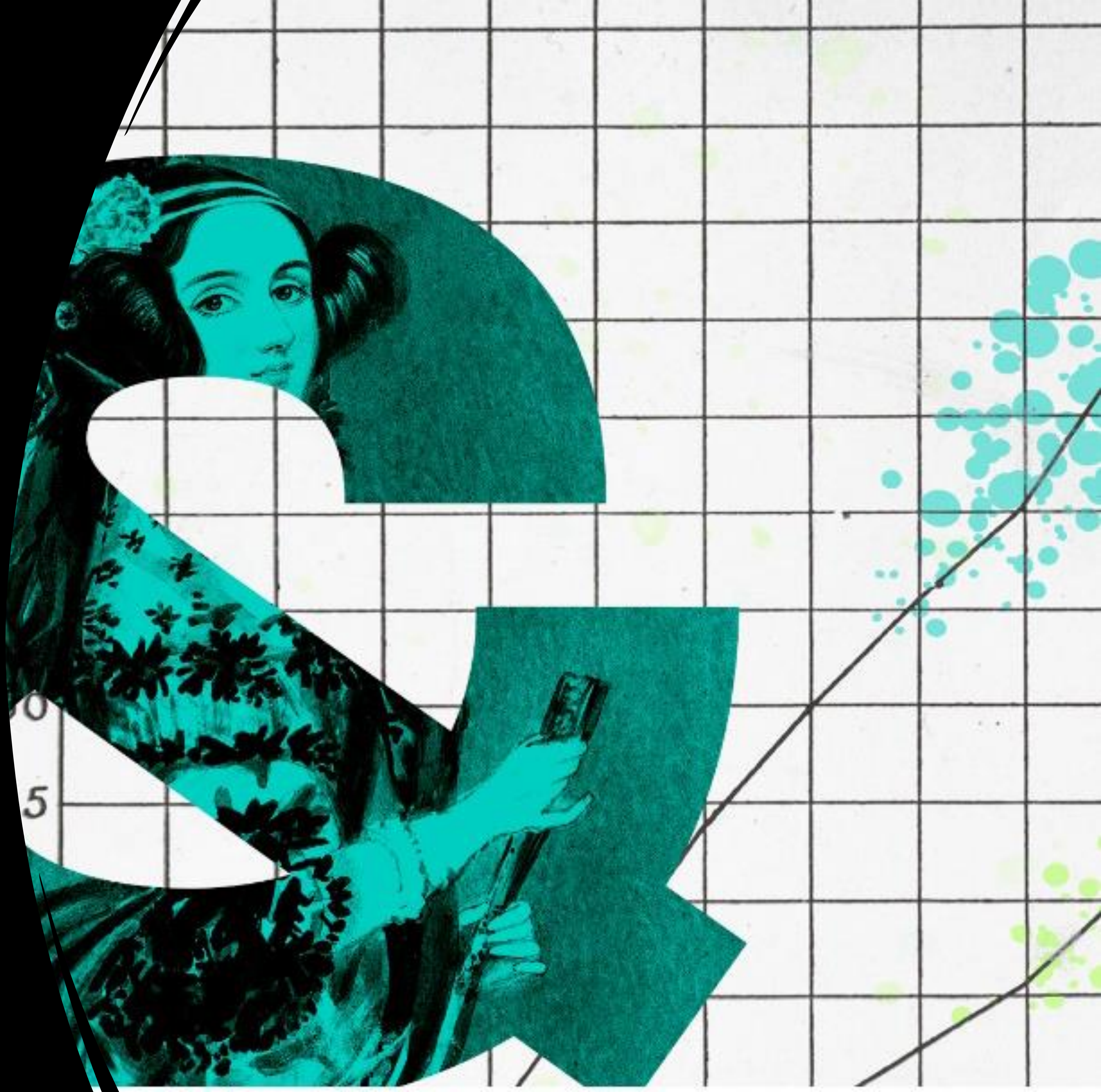
Course outline

- Session 1. Simple Regression; Individual Difference; Intro to Linear mixed-effect models (LMMs)
- Session 2. LMMs (lmer); Introduction to Generalised LMMs (glmer)
- Session 3. Ggeneralised LMMs (glmer); Model assumptions and diagnostics



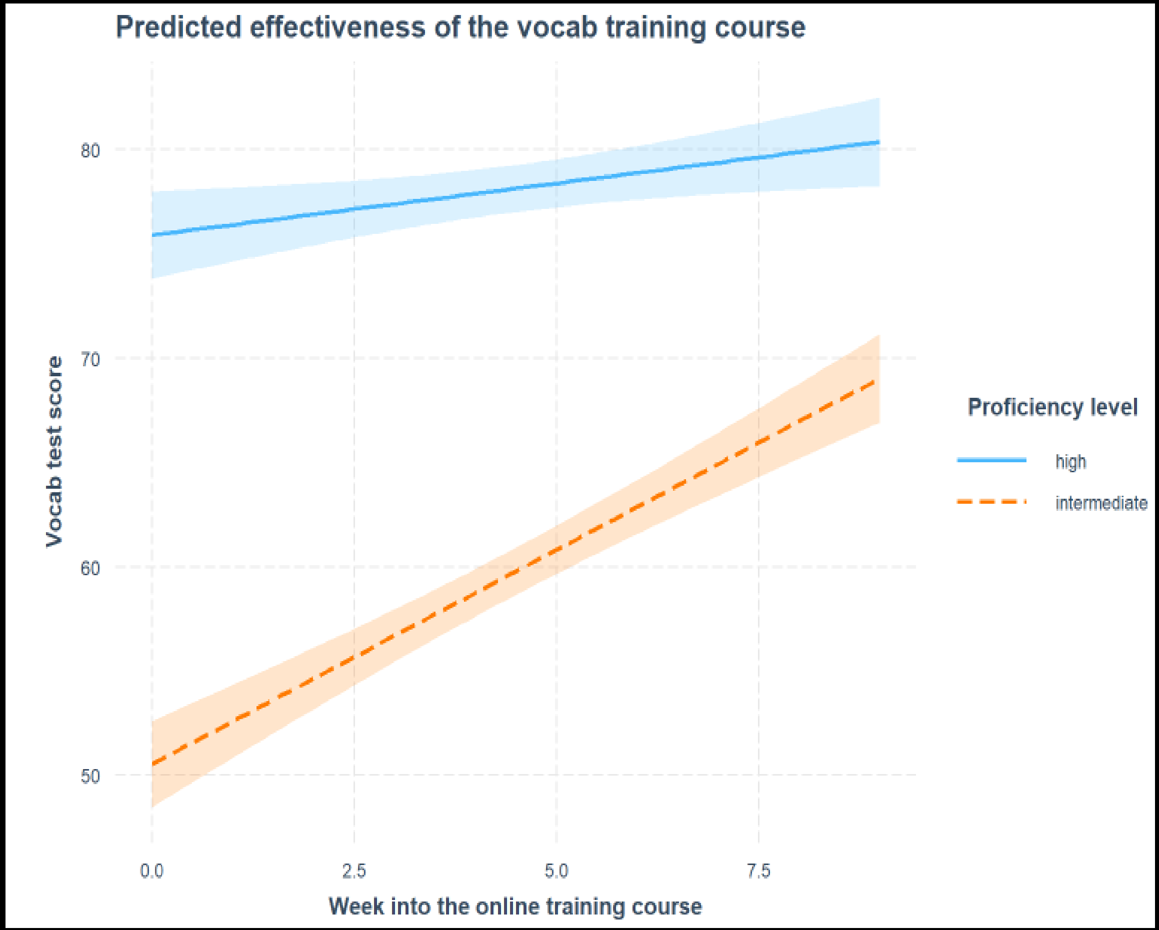
Session 2
Roadmap (today)

- Characteristics and objectives of LMMs
- Random Effects
- Model Comparison
- Applications
- Intro to Generalised LMMs



Linear Mixed Effects Models (LMMs) - lmer()

Recall problems with simple regression



Characteristics of LMMs

- Repeated-measures Factorial Design
  - a person or item being observed multiple times

- Measurements from a participant or an item are correlated
  - Thus cannot use simple regression due to violation of assumptions
  - more on this next week

- Random variability
  - More than one source
  - Across different groups of observations
  - Variance resulting from taking a random sample of observations

Objectives of LMMs

- Description
  - What occurs to subjects over time? etc.

- Inference
  - Is a treatment effective? Does substantial change occur over time?

- Prediction
  - What is the trend?

Structure of LMMS: Random effects

Pick up where we left from last week

- The structure of a simple regression

```
(1) m2bv2 <- lm(vocab_test_score ~ week * proficiency, data = vocabdata)
```

- The structure of linear mixed-effect models

```
(2) mixedm1 <- lmer(vocab_test_score ~ week * proficiency + (1 + week | participant), data = vocabdata)
```

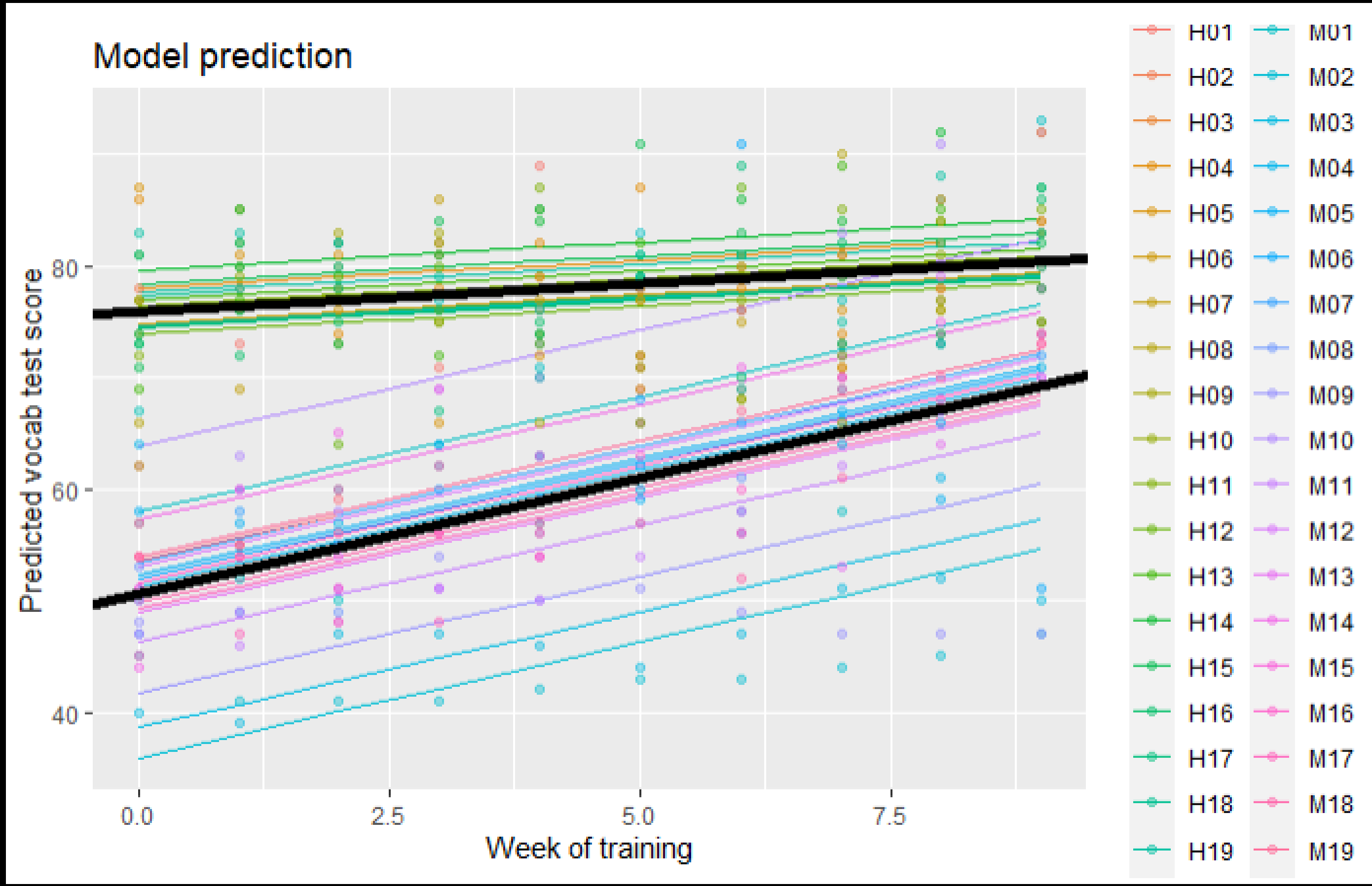
```
(3) mixedm2 <- lmer(vocab_test_score ~ week * proficiency + (1 | participant), data = vocabdata)
```

Fixed effects

Random effects

Structure of LMMS: Random effects

```
(3) mixedm2 <- lmer(vocab_test_score ~ week * proficiency + (1 | participant), data = vocabdata)
```



Random intercept

Grouping

Structure of LMMS: Random effects

(3) mixedm2 <- lmer(vocab_test_score ~ week * proficiency + (1 | participant), data = vocabdata)

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest
]
Formula: vocab_test_score ~ week * proficiency + (1 | participant)
   Data: vocabdata

REML criterion at convergence: 2438.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3014 -0.5552  0.0434  0.6075  4.4641

Random effects:
 Groups      Name        Variance Std.Dev.
 participant (Intercept) 25.47    5.046
 Residual                35.45    5.954
Number of obs: 369, groups:  participant, 40

Fixed effects:
                             Estimate Std. Error       df t value Pr(>|t|)
(Intercept)                   75.9310     1.3932  65.6532  54.499  < 2e-16 ***
week                           0.5023     0.1558 334.4194   3.225  0.00138 **
proficiencyintermediate      -25.3754     1.9676  65.0011 -12.896  < 2e-16 ***
week:proficiencyintermediate   1.5755     0.2185 334.4965   7.210 3.76e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) week   prfcnc
week        -0.486
prfcncyntrm -0.708  0.344
wk:prfcncyn  0.346 -0.713 -0.477
```

Random intercept

Grouping

What happens to the overall quality of the model???

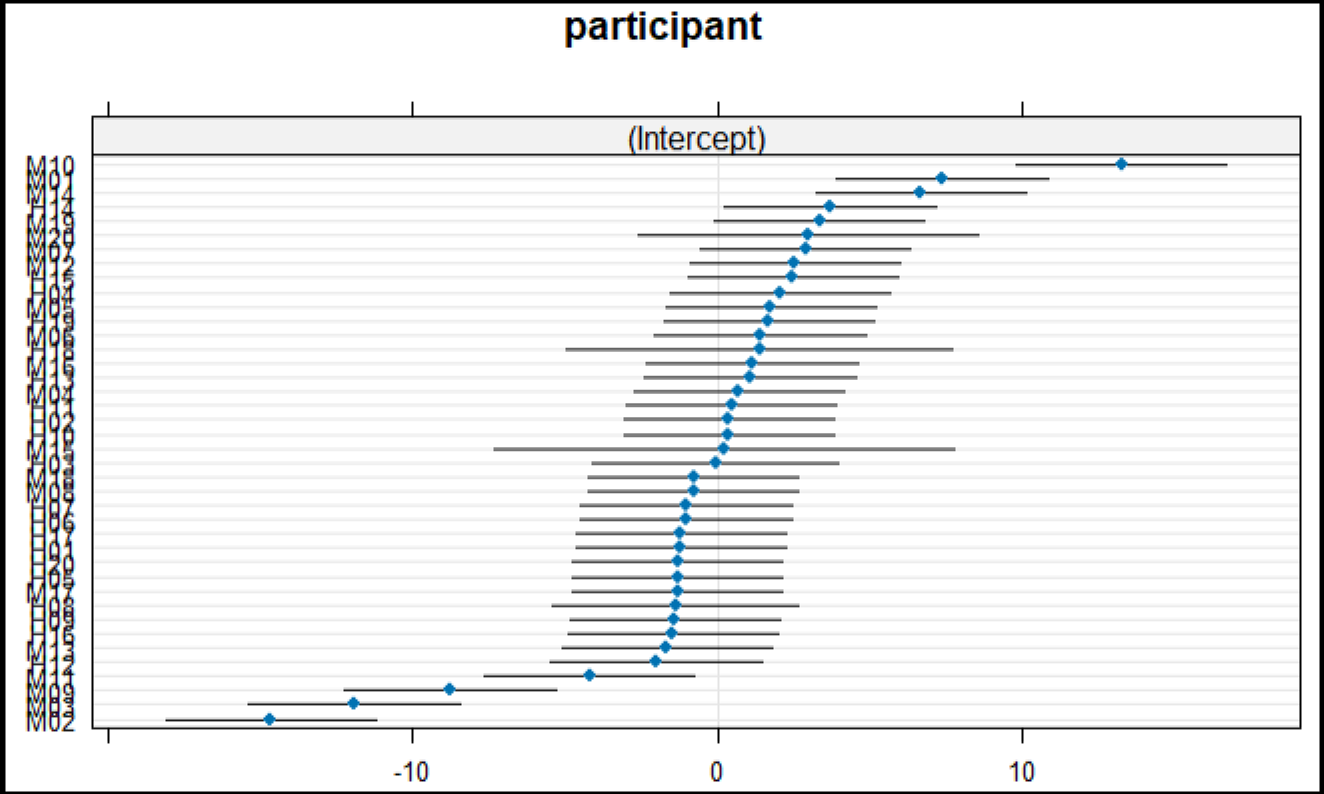
- well, in mixed models, we do not get a R^2 or other simple goodness-of-fit metrics

Structure of LMMS: Random effects

```
(3) mixedm2 <- lmer(vocab_test_score ~ week * proficiency + (1 | participant), data = vocabdata)
```

Random intercept

grouping



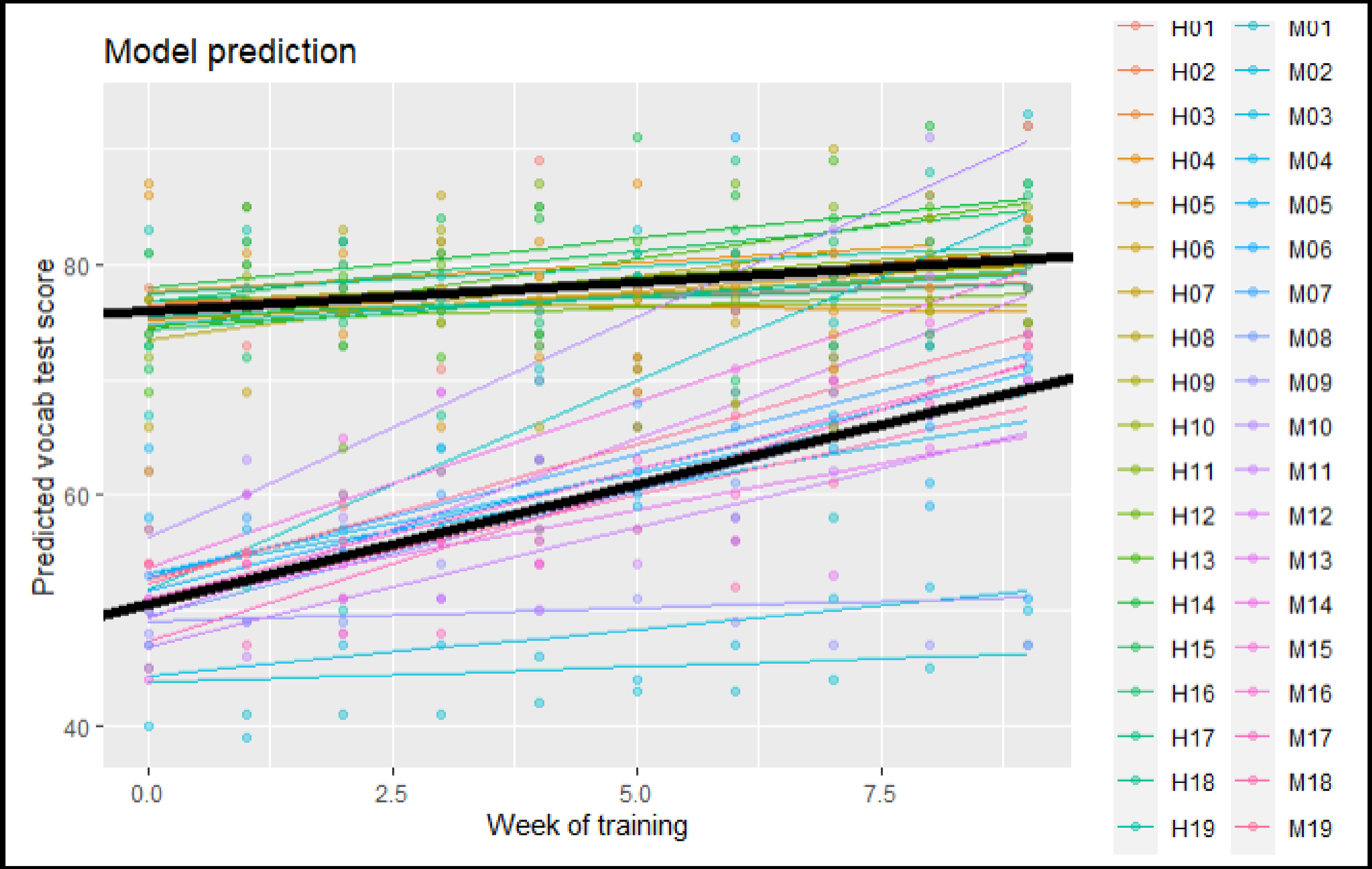
Structure of LMMS: Random effects

```
(2) mixedm1 <- lmer(vocab_test_score ~ week * proficiency + (1 + week | participant), data = vocabdata)
```

Grouping

Random intercept

Random slope



Structure of LMMS: Random effects

(2) mixedm1 <- lmer(vocab_test_score ~ week * proficiency + (1 + week | participant), data = vocabdata)

Grouping

Random intercept

Random slope

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: vocab_test_score ~ week * proficiency + (1 + week | participant)
   Data: vocabdata

REML criterion at convergence: 2398.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5994 -0.5566  0.0445  0.6019  4.9182

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 10.0871  3.1760
             week         0.6937  0.8329   0.11
 Residual                29.5112  5.4324
Number of obs: 369, groups:  participant, 40

Fixed effects:
                            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)                  75.8547     1.0317  38.5521  73.526  < 2e-16 ***
week                          0.5255     0.2392  36.0100   2.197   0.0346 *
proficiencyintermediate     -25.3129     1.4532  38.2004 -17.419  < 2e-16 ***
week:proficiencyintermediate  1.5555     0.3395  35.1552   4.582 5.59e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) week   prfcnc
week        -0.306
prfcncyntrm -0.710  0.217
wk:prfcncyn  0.216 -0.705 -0.296
```

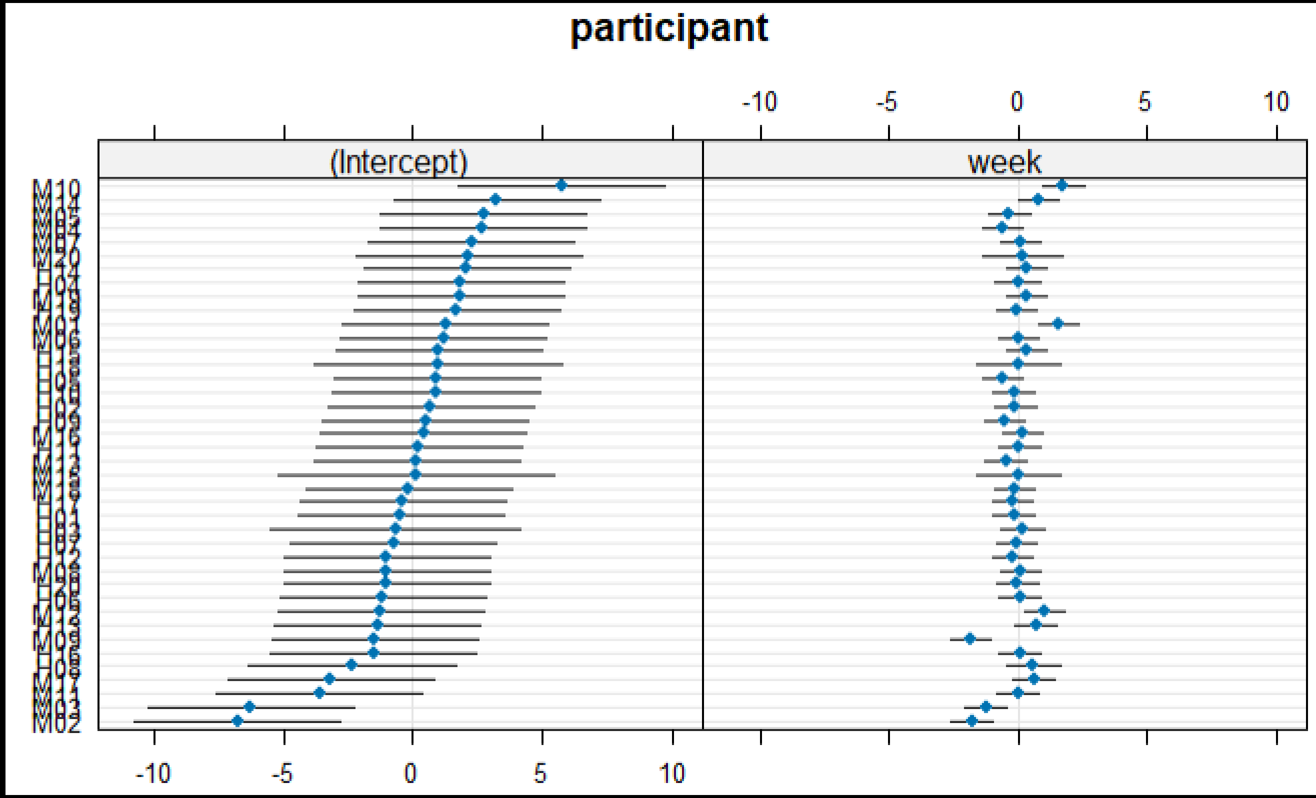
Structure of LMMS: Random effects

```
(2) mixedm1 <- lmer(vocab_test_score ~ week * proficiency + (1 + week | participant), data = vocabdata)
```

Grouping

Random intercept

Random slope



Structure of LMMS: Summary

- lmer (y ~ $\boxed{X_1 * X_2 +}$
  $(\boxed{1 + X_1 * X_2} \mid \boxed{Grouping_1}) +$
  $(\boxed{1 + X_1} \mid \boxed{Grouping_2})$,
  data = datafilename)

You can have more than one source of individual differences (e.g., group by students; by schools)

Exercise

Can you fit a model with a different structure of random effects?

- 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?

Deal with common issues: Convergence

```
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.0206805 (tol = 0.002, component 1)
```

Solutions:

- Adjust stopping (convergence) tolerances for the nonlinear optimizer, using the optCtrl() argument to lmerControl.

- Center and scale continuous predictor variables (e.g., vocabdata<-scale(vocabdata$week, center=TRUE, scale=FALSE)

Deal with common issues: Boundary singular fit

```
boundary (singular) fit: see ?isSingular
```

Solutions:

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

Model Comparison & Selection

Which model should you choose?

- Hint: remember the anova() function we used last time?
    - anova(simpler model, complex model)

```
anova(mMixed1_pval, mMixed_reduced)


## Data: vocabdata
## Models:
## mMixed1_pval: vocab_test_score ~ week * proficiency + (1 | participant)
## mMixed_reduced: vocab_test_score ~ week * proficiency + (1 + week | participant)
##                npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mMixed1_pval      6 2451.4 2474.8 -1219.7   2439.4
## mMixed_reduced    8 2416.2 2447.5 -1200.1   2400.2 39.151  2  3.151e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Comparison & Selection

Which model should you choose?

- **Barr, Levy, Scheepers, & Tily, (2013) :** keep it maximal
- **Matuschek, Kliegl, Vasishth, Baayen & Bates (2017) :** make it parsimonious to balance Type1 error and power

Different approaches been promoted by different researchers; either is fine as long as you explain your rationale.

Applications: Self-reflection & Discussion

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

Generalised Linear Mixed-effects Models - glmer()

Binary outcome

- Simulated data based on real psycholinguistic findings.

- How would you describe the image? (A) or (B)?

(A) Gromit gave Wallace some cheese.
(B) Gromit gave some cheese to Wallace.

- People have a tendency to reuse a recently encountered sentence structure (priming effect).

Next time

- glmer () for binary outcome

```
M = glmer( y ~ x + (1 + x | grouping), data, family = 'binomial')
```

Further reading

- Paper

Brown, VA. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science.* 4(1). doi:10.1177/2515245920960351

Advances in Methods and Practices in Psychological Science

aps | ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Open access | Research article | First published online March 25, 2021

An Introduction to Linear Mixed-Effects Modeling in R

Violet A. Brown View all authors and affiliations

All Articles | https://doi.org/10.1177/2515245920960351

- E-Book

https://vasishth.github.io/Freq_CogSci/

Shravan Vasishth, Daniel Schad, Audrey Bürki, Reinhold Kliegl

Linear Mixed Models in Linguistics and Psychology: A Comprehensive Introduction

THANK YOU