PSYC402-week-18-LME-2

Rob Davies (Lancaster University)

Targets for Week 18 – Ideas and skills

1. Practice how to tidy experimental data for mixed-effects analysis
2. Begin to develop understanding of crossed random effects of subjects and stimuli
3. Practice fitting linear mixed-effects models incorporating random effects of subjects and stimuli

To be modern psychological data analysts you will need to know the *what, why, when and how* of multilevel or mixed-effects models

This week, we make a *subtle* change and start talking more about Linear Mixed-effects models

Repeated measures data: we begin by *revising* our list of when we need mixed-effects models

- When we test the same people multiple times
  - Pre- and post-treatment
  - Multiple stimuli – everyone sees the same stimuli
  - Repeated testing – follow learning, development within individuals – in longitudinal designs
- When we do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Where we are going: linear mixed-effects models

- We need to learn how to estimate the effects of experimental variables
- *while also* taking into account sources of error variance like
  - the random differences between people we test
  - and the random differences between stimuli we present

The wider scientific impact – accepting diversity

- How do psychological effects *vary*?
- Uniformity is a common because convenient assumption
- We ask: How do people vary in their response?

The data we will work with: the CP study data

- As part of our lab work, we will practice steps often required to get data ready for mixed-effects model
- CP studied how 62 children read 160 words
- The data are in separate files and the files are *untidy*
  - CP study word naming rt 180211.dat reaction time for correct responses to word stimuli in reading
  - CP study word naming acc 180211.dat accuracy for all responses to word stimuli in reading
  - words.items.5 120714 150916.csv information about the 160 stimulus words presented in reading task
  - all.subjects 110614-050316-290518.csv information about the 62 participants

We will make data tidy

- What a horrible mess:
  - Psychological data collection often delivers *untidy* data
  - Here, we have data for different participants in separate columns
  - Each row holds the reaction times for the responses made by all participants to each stimulus word
  - Each cell holds the reaction time for the response made by a child to a word
  - We have missing values NA and reaction times

```
## # A tibble: 6 x 62
##   item_name AislingoC AlexB AllanaD  AmyR AndyD AnnaF Aoife
##   <chr>         <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl>  <dbl
## 1 act           595.   586       NA   693   597   627    649
## 2 ask           482.   864     1163   694.  616   631    538
## 3 both          458.   670     1114.  980  1019   796.   545
## 4 box           546    749.     975   678   589   604    574
## 5 broad         580   1474.      NA   789   684    NA    816
```

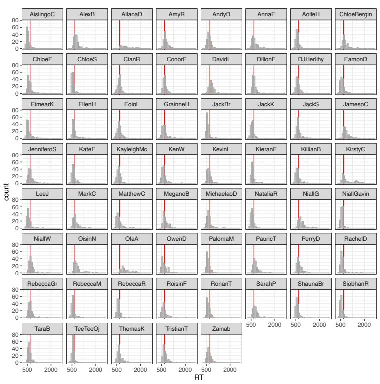

Next: When we do we need mixed-effects models?

When we do we need mixed-effects models? *When we have repeated measures data*

- In a reading study, we ask all individuals in a participant sample to read all words in a stimulus sample
- For each individual, we will have multiple observations and these observations will not be independent
  - One participant will tend to be slower or less accurate compared to another
  - Her responses may be more or less susceptible to the effects of the experimental variables
- The observed responses in different trials can be grouped by participants

Participants will vary for reasons we cannot explain

- Here you see a separate histogram plot for each participant
- Bars show the distribution of reaction time (RT)
- The red line shows the overall mean RT

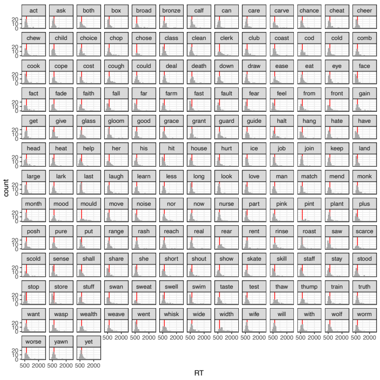


When we do we need mixed-effects models? When we have repeated measures data

- In a reading study, we ask all individuals in a participant sample to read all words in a stimulus sample
- For each stimulus, there are multiple observations and these observations will not be independent
  - One stimulus may prove to be more challenging to all participants compared to another, eliciting slower or less accurate responses
  - The effects of *within-items* experimental variables may be more or less prominent for responses to some stimuli than to others
- So the data can *also* be grouped by stimuli

Stimuli will vary for reasons we cannot explain

- Here you see a separate histogram plot for the responses to each word
- Bars show the distribution of reaction time (RT)
- The red line shows the overall mean RT



The language-as-fixed-effect fallacy

- If you are doing a repeated measures design study in which there are different participants
- And different tests or test items or stimuli
- And all participants respond to all stimuli
- Then you need to use mixed-effects models
- Because you need to deal with the random differences between people *and* the random differences between stimuli

The language as fixed effect fallacy

A very famous paper by Clark (1973)

- Historically, psychologists tested effects against error variance due to differences between people
- They ignored differences due to stimuli
- This meant they were likely to find significant effects not because there were true differences between conditions
- But because there were random differences between stimuli presented in different conditions

Taking into account error variance due to subjects and items – Clark's (1973) *minF′* solution

$$minF' = \frac{MS_{effect}}{MS_{random-subject-effects} + MS_{random-word-differences}} = \frac{F_1 F_2}{F_1 + F_2} \quad (1)$$

1. You start by *aggregating* your data
   - By-subjects data – for each subject, take the average of their responses to all the items
   - By-items data – for each item, take the average of all subjects' responses to that item
2. You do separate ANOVAs, one for by-subjects (F1) data and one for by-items (F2) data
3. You put F1 and F2 together, calculating minF'

Using tidyverse functions, it is easy to calculate by-subjects and by-items RT averages

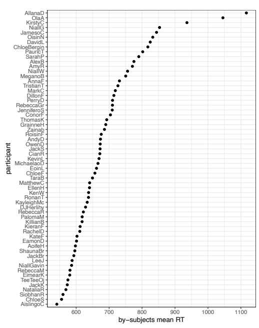
```
by.items.rt <- long.all.noNAs %>%
               group_by(item_name) %>%
               summarise(av_RT = mean(RT, na.rm = TRUE))
by.items.rt

by.subjects.rt <- long.all.noNAs %>%
                  group_by(subjectID) %>%
                  summarise(av_RT = mean(RT, na.rm = TRUE))
by.subjects.rt
```

- We can then join the by-items data with stimulus properties and analyze the effects of those properties (e.g. word frequency)
- *or* we can join the by-subjects data with participant attributes and analyze the effects of those attributes (e.g. participant group)
- We cannot look at *both* item and participant effects at the same

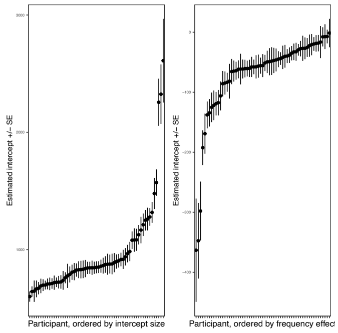
But analysing data only by-items means we lose track of participant differences

- *Lorch & Myers (1990)* warn: analyzing just by-items mean RTs assumes wrongly that *subjects are a fixed effect*
- We can see this is wrong because, for example, with the CP data, we can see that participant RT varies substantially



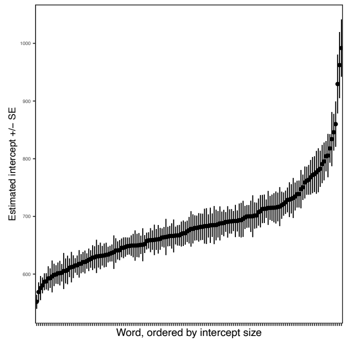
Participant differences in *both* average RT (or accuracy) *and* the impacts of effects

- These error bar plots show:
  - As points: the estimated intercept or the estimated effect of frequency on RT
  - Together with the standard errors of the estimates
  - For each participant analyzed separately
- We can see that participants vary greatly in both estimated intercept or slope *and* in uncertainty about estimates



Equally, analysing by-subjects data alone means we would lose track of random differences between stimuli

- These error bar plots show:
  - As points: the estimated intercept
  - Together with the standard errors of the estimate
  - For responses to each word analyzed separately
- We can see that responses to different words vary greatly in average speed – here, we ignore other effects



Next: So what do we do? We use mixed-effects models and we include random effects for both participants and stimuli

We account for differences between participants in intercept by modelling the intercept as two terms

$$\beta_{0i} = \gamma_0 + U_{0i} \tag{2}$$

- Where γ_0 is the average intercept
- And U_{0i} is the difference for each i child between *their* intercept and the average intercept

We account for differences between participants in slope by modelling the slope of effects as two terms

$$\beta_{1i} = \gamma_1 + U_{1i} \tag{3}$$

- Where γ_1 is the average slope
- And U_{1i} represents the difference for each i child between the slope of *their* frequency effect and the average slope

We account differences between items in intercepts by modelling the intercept as two terms

$$\beta_{0j} = \gamma_0 + W_{0j} \tag{4}$$

- Where γ_0 is the average intercept
- And W_{0j} represents the deviation, for each word, between the word intercept and the average intercept

Our model can now incorporate the random effects of *both* participants and words

$$Y_{ij} = \gamma_0 + \gamma_1 X_j + U_{0i} + U_{1i} X_j + W_{0j} + e_{ij} \quad (5)$$

- Where the outcome Y_{ij} is related to ...
- The average intercept γ_0 and differences between i children in the intercept U_{0i};
- The average effect of the explanatory variable frequency $\gamma_1 X_j$ and differences between i participants in the slope $U_{1i} X_j$;
- Plus the random differences between items in intercepts W_{0j}
- And the residual error variance e_{ij}.

We can do all this in one move using lmer()

```
lmer.all <- lmer(RT ~  Lg.UK.CDcount +
                          (Lg.UK.CDcount + 1||subjectID) +
                          (1|item_name),

                 data = long.all.noNAs)

summary(lmer.all)
```


We can do all this in one move using lmer()

```
lmer.all <- lmer(RT ~  Lg.UK.CDcount +
                          (Lg.UK.CDcount + 1||subjectID) +
                          (1|item_name),

                 data = long.all.noNAs)

summary(lmer.all)
```

- lmer.all <- lmer(...) create a linear mixed-effects model object using the lmer() function
- RT ∼ Lg.UK.CDcount the fixed effect in the model is expressed as a formula in which the outcome RT is predicted ∼ by word frequency, given by Lg.UK.CDcount in the dataset
- We use data = long.all.noNAs to specify the data we are analyzing

We can do all this in one move using lmer()

```
lmer.all <- lmer(RT ~  Lg.UK.CDcount +
                          (Lg.UK.CDcount + 1||subjectID) +
                          (1|item_name),

                 data = long.all.noNAs)

summary(lmer.all)
```

- We add (...|subjectID) to specify random differences between sample groups (here, participants), specified using the dataset subjectID coding variable name
- We add (...1 |subjectID) to account for random differences between participants in intercepts, coded 1
- and (Lg.UK.CDcount ... |subjectID) to account for random differences between participants in the slope of the frequency effect, specified using the dataset Lg.UK.CDcount variable name

We can do all this in one move using lmer()

```
lmer.all <- lmer(RT ~  Lg.UK.CDcount +
                          (Lg.UK.CDcount + 1||subjectID) +
                          (1|item_name),

                 data = long.all.noNAs)

summary(lmer.all)
```

- We add the term (...|itemname) to specify random effects corresponding to random differences between sample groups (here, items) coded using the itemname variable name
- We add (1 |itemname) to account for random differences between sample groups (words) in intercepts, coded 1

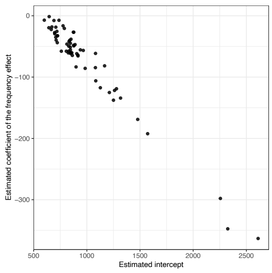
We usually do not aim to examine the specific deviations

We estimate just the *spread of deviations* by-participants or by-words: the **variance**

- $var(U_{0i})$ variance of deviations by-participants from the average intercept;
- $var(U_{1i} X_j)$ variance of deviations by-participants from the average slope of the frequency effect;
- $var(W_{0j})$ variance of deviations by-items from the average intercept;
- $var(e_{ij})$ residuals, at the response level, after taking into account all other terms.

Expect random effects will *covary*

- Participants who are slower to respond also show the frequency effect more strongly
- The scatterplot shows the relationship between per-participant estimates of
- The intercept and the slope
- The strong relationship is clear



How do you report a mixed-effects model?

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ Lg.UK.CDcount + (Lg.UK.CDcount + 1 || subjectID) + (1 |
##     item_name)
##    Data: long.all.noNAs
##
## REML criterion at convergence: 116976.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1794 -0.5474 -0.1646  0.3058 12.9485
##
## Random effects:
##  Groups      Name          Variance Std.Dev.
##  item_name   (Intercept)     3397    58.29
##  subjectID   Lg.UK.CDcount   3623    60.20
##  subjectID.1 (Intercept)   112307   335.12
##  Residual                   20704   143.89
## Number of obs: 9085, groups:  item_name, 159; subjectID, 61
##
## Fixed effects:
##               Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)     971.07      51.86  94.62  18.723  < 2e-16 ***
## Lg.UK.CDcount   -72.33      10.79 125.27  -6.703 6.23e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## Lg.UK.CDcnt -0.388
```


How do you report a mixed-effects model?

- Explain what variables went into the analysis: say what the outcome and predictor variables were
- Report the model equation RT $\sim$ frequency + (frequency + 1 || participant) + (1 | word)
- Report a table of coefficients: variable, estimate of coefficient of effect; SE; t (or z); and p
- Add to that table a report of the random effects terms: variances
- You should comment on the coefficient estimates; you may (or may not) comment on the random effects variances

Next week: we need to be ready to trouble shoot

- I stopped the model from estimating the covariance between random effects of participants on items and on slopes
- using (frequency + 1 || participant) not (frequency + 1 | participant)
- next week I explain why: convergence

```
library(lmerTest)
lmer.all <- lmer(RT ~  Lg.UK.CDcount +
                       (Lg.UK.CDcount + 1||subjectID) +
                       (1|item_name),

          data = long.all.noNAs)

summary(lmer.all)
```


Summary – Week 18: crossed random effects

1. Psychological studies often have repeated measures designs
   - When there are multiple observations for each person or stimulus
   - Because each person has to respond to multiple stimuli
   - And each stimulus is shown to multiple people
2. Mixed-effects models can be specified by the researcher
   - to account for random differences between participants or between stimuli
   - in the intercepts or the slopes of explanatory variables

Human diversity and how people vary: the challenge, the promise

- Variation is a fact and mixed-effects models enable us to take into account random differences between people
- *But* these models also allow us – this is new – to examine the nature of the variation directly