Using Linear mixed-effects models – why, when and how

Rob Davies

r.davies1@lancaster.ac.uk

May 2016

Aims for the class

1. Understand the motivation for *linear mixed-effects models* – the requirements of handling multilevel structured data
2. Introduce a multilevel structured dataset
3. Recognize alternative methods for analyzing multilevel structured data
4. Practise running linear mixed-effects models in R
5. Evaluating models using information criteria

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

Phenomena and data sets in the social sciences often have a multilevel structure

Repeated measures or *clustered data*

- 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
- Do multi-stage sampling
  - Find (sample) classes or schools – test (sample) children within classes or schools
  - Find (sample) clinics – test (sample) patients within clinics

The key insight: observations are clustered – correlated – *not independent*

Dependence of observations could be treated as a nuisance because an assumption of linear models is that observations are independent so failing to take dependence into account may result in incorrect inferences – the non-independence of observations means you have less information than their total number of suggests you have

Where we are going: *linear mixed-effects modelling*

Capture sources of variance due to *fixed effects* e.g. frequency and *random effects* e.g. differences between sampling units like people or words in intercepts or slopes

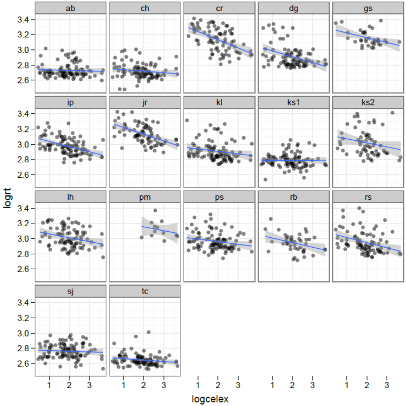


Figure : Effect of word frequency on word naming latencies of adult students

In psychological research, uniformity - the average participant - is a convenient simplification

We often average over individual differences to investigate experimental effects – or we study differences between participant groups averaging over responses to different stimuli

Both approaches cause problems but neither are necessary with linear mixed-effects models

Given variability among individuals or stimuli, focusing on the average appears risky

ML study – A concrete usage example

The data-set – experimental reading task – lexical decision

- All participants saw all 160 words and 160 matched non-words
- Effects of *TOWRE* measures of reading skill, age, ART measure of print exposure
- Effects of word attributes like item type (e.g. responses to words vs. non-words)
- Interactions between effects of *who* you are and effects of *what* kinds of stimuli you must respond to

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	item_numbe	subjectID	Test	Age	Years_in_edu	Gender	TOWRE_wor	TOWRE_non	ART_HRminu	RT	COT	Subject	Trial.order	item_name	Length
2	1	A16	ALT	21	16	F	97	55	10	491.3	851280.51	A16	303	went	4
3	1	B12	TLA	58	13	M	93	47	8	446.67	91425.8	fb11	27	went	4
4	1	A15	LTA	21	16	F	95	57	9	483.71	861801	A15	278	went	4
5	1	HR7	LTA	21	16	M	84	49	5	353.85	[illegible]	HR8	206	went	4

ML study – A concrete usage example

The data-set – experimental reading task – lexical decision

- Data are stacked in *long* format
- In each row, one observation relating to a response made by a participant to a stimulus
- There are columns to code for subject and stimulus identity
- Also variables coding for attributes that can be entered as *predictor* variables in the model of responses

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	item_numbe	subjectID	Test	Age	Years_in_edu	Gender	TOWRE_wor	TOWRE_non	ART_HRminu	RT	COT	Subject	Trial.order	item_name	Length
2	1	A16	ALT	21	16	F	97	55	10	491.3	851280.51	A16	303	went	4
3	1	B12	TLA	58	13	M	93	47	8	446.67	91425.8	fb11	27	went	4
4	1	A15	LTA	21	16	F	95	57	9	483.71	861801	A15	278	went	4

Figure : Notice the data are stacked in *long* format

Get the data for practice – download and read in the ML datset of responses to words and nonwords

Having read in *subjects.behaviour.items-310114.csv*, use *subset()* to remove errors

```
ML.all <- read.csv("subjects.behaviour.items-310114.csv",
header=T, na.strings = "-999")

ML.all.correct <- subset(ML.all, RT > 200)

summary(ML.all.correct)
```

```
> summary(ML.all.correct)
  item_number     subjectID      Test          Age         Years_in_education Gender   TOWRE_wordacc    TOWRE_nonwordacc ART_HRminusFR         RT
 Min.   :   1   B16     : 319   ALT:1545   Min.   :16.00   Min.   :11.00      F:4235   Min.   : 68.00   Min.   :16.0     Min.   : 1.00   Min.   : 262.2
 1st Qu.:  78   NH1     : 318   ATL:1779   1st Qu.:21.00   1st Qu.:13.00      M:6019   1st Qu.: 84.00   1st Qu.:50.0     1st Qu.: 7.00   1st Qu.: 550.2
 Median : 157   AA2     : 317   LAT:1509   Median :21.00   Median :16.00               Median : 93.00   Median :56.0     Median :12.00   Median : 640.8
 Mean   : 568   A19     : 314   LTA:1484   Mean   :37.14   Mean   :14.96               Mean   : 91.24   Mean   :52.5     Mean   :15.32   Mean   : 690.9
 3rd Qu.:1079   G20     : 314   TAL:2454   3rd Qu.:53.00   3rd Qu.:16.00               3rd Qu.: 98.00   3rd Qu.:57.0     3rd Qu.:21.00   3rd Qu.: 763.2
 Max.   :1160   A15     : 313   TLA:1483   Max.   :73.00   Max.   :19.00               Max.   :104.00   Max.   :63.0     Max.   :43.00   Max.   :1991.4
                (Other):8359
      COT              Subject       Trial.order      item_name         Length          Ortho_N           BG_Sum          BG_Mean       BG_Freq_By_Pos    item_type
 Min.   :  50094   A18     : 608   Min.   : 21.0   act     :   34   Min.   :3.000   Min.   : 0.000   Min.   :  1.0   Min.   :  1.0   Min.   :  1.0   nonword:4997
 1st Qu.: 295895   B16     : 319   1st Qu.:101.0   baz     :   34   1st Qu.:4.000   1st Qu.: 3.000   1st Qu.: 79.0   1st Qu.: 78.0   1st Qu.: 71.0   word   :5257
 Median : 553152   NH1     : 318   Median :181.0   both    :   34   Median :4.000   Median : 6.000   Median :158.0   Median :156.0   Median :150.0
 Mean   : 577406   AA2     : 317   Mean   :180.9   calf    :   34   Mean   :4.301   Mean   : 7.237   Mean   :157.4   Mean   :156.9   Mean   :147.7
 3rd Qu.: 812189   g20     : 314   3rd Qu.:261.0   cend    :   34   3rd Qu.:5.000   3rd Qu.:11.000   3rd Qu.:236.0   3rd Qu.:236.0   3rd Qu.:224.0
 Max.   :1583651   A15     : 313   Max.   :340.0   chred   :   34   Max.   :6.000   Max.   :24.000   Max.   :314.0   Max.   :316.0   Max.   :296.0
                   (Other):8065                    (Other):10050
     logrt            zAge.V1           zTOWRE_wordacc.V1   zTOWRE_nonwordacc.V1     zLength.V1          zOrtho_N.V1
 Min.   :2.419   Min.   :-1.1639665   Min.   :-2.5100285   Min.   :-4.132753   Min.   :-1.7989674   Min.   :-1.463842
 1st Qu.:2.740   1st Qu.:-0.8886561   1st Qu.:-0.7819251   1st Qu.:-0.283322   1st Qu.:-0.4164738   1st Qu.:-0.857033
 Median :2.807   Median :-0.8886561   Median : 0.1901331   Median : 0.395989   Median :-0.4164738   Median :-0.250223
 Mean   :2.821   Mean   : 0.0000000   Mean   : 0.0000000   Mean   : 0.000000   Mean   : 0.0000000   Mean   : 0.000000
 3rd Qu.:2.883   3rd Qu.: 0.8733306   3rd Qu.: 0.7301654   3rd Qu.: 0.509207   3rd Qu.: 0.9660197   3rd Qu.: 0.761125
```

A linear model (multiple regression) of an effect of some set of conditions

Linear models assume independence of observations – but if you take repeated measures then observations will be *dependent* – responses will cluster by person or stimulus

$$X_{ij} = \mu + (\mu_j - \mu) + \varepsilon_{ij} = \mu + \tau_j + \varepsilon_{ij} \tag{1}$$

- X_{ij} – the score of person *i* in condition *j*
- μ – the mean of all subjects who could be tested in the experiment
- μ_j – the mean score in condition *j*
- τ_j – the experimental effect – the extent to which the mean for condition *j* is different from the overall mean
- ε_{ij} – the amount to which person *i* in condition *j* differs from the mean for that group

A linear model (multiple regression) of an effect of some set of conditions

Linear models assume independence of observations – but if you take repeated measures then observations will be *dependent* – responses will cluster by person or stimulus

$$X_{ij} = \mu + (\mu_j - \mu) + \varepsilon_{ij} = \mu + \tau_j + \varepsilon_{ij} \tag{1}$$

- X_{ij} – the score of person i in condition j
- μ – the mean of all subjects who could be tested in the experiment
- μ_j – the mean score in condition j
- τ_j – the experimental effect – the extent to which the mean for condition j is different from the overall mean
- ε_{ij} – the amount to which person i in condition j differs from the mean for that group

We can account for between and within subject differences in our model

If you take repeated measures then observations will be *dependent*: for a slow responder, all their responses will be slower; for a difficult stimulus, all responses will be slower

$$X_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij} \quad (2)$$

- X_{ij} – the score of person i in condition j
- μ – the mean of all subjects who could be tested in the experiment
- π_i – add effect of being subject i – compared to average over all subjects
- τ_j – add effect of being in condition j – compared to average over all conditions
- ε_{ij} – the amount to which person i in condition j differs from the mean for that group

We can account for between and within subject differences in our model

If you take repeated measures then observations will be *dependent*: for a slow responder, all their responses will be slower; for a difficult stimulus, all responses will be slower

$$X_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij} \tag{2}$$

- X_{ij} – the score of person *i* in condition *j*
- μ – the mean of all subjects who could be tested in the experiment
- π_i – add effect of being subject *i* – compared to average over all subjects
- τ_j – add effect of being in condition *j* – compared to average over all conditions
- ε_{ij} – the amount to which person *i* in condition *j* differs from the mean for that group

A more realistic repeated measures model

Suppose that effects vary between subjects

$$X_{ij} = \mu + \tau_j + \pi_i + \pi_i\tau_j + \varepsilon_{ij} \tag{3}$$

- X_{ij} – the score of person *i* in condition *j* – grand mean
- μ – the mean of all subjects who could be tested in the experiment
- τ_j – add effect of being in condition *j* – compare average over all conditions (grand mean)
- π_i – add effect of being subject *i* – compare average over all subjects (grand mean)
- $\pi_i\tau_j$ – a subject by treatment interaction – different subjects (or words) react to conditions in different ways
- ε_{ij} – the amount to which person *i* in condition *j* differs from the mean for that group

The language as fixed effect fallacy Clark, 1973

- Historically, psychologists estimated effects taking into account error variance due to random differences between participants' responses
- But ignoring stimulus sampling could mean that effects are inferred to exist even when differences between conditions are due to random differences between responses to different stimuli

The language as fixed effect fallacy Clark, 1973

- Historically, psychologists estimated effects taking into account error variance due to random differences between participants' responses
- But ignoring stimulus sampling could mean that effects are inferred to exist even when differences between conditions are due to random differences between responses to different stimuli

A linear model taking into account *the random effects of items*

$$X_{ij} = \mu + \pi_i + \tau_j + \pi_i\tau_j + \beta_k + \pi_i\beta_k + \varepsilon_{ijk} \tag{4}$$

- β_k – effect of word k – unexplained differences in average response elicited by different stimuli
- $\pi_i\beta_k$ – the stimulus word by subject interaction – different people respond to different stimuli differently

A linear model taking into account *the random effects of items*

$$X_{ij} = \mu + \pi_i + \tau_j + \pi_i\tau_j + \beta_k + \pi_i\beta_k + \varepsilon_{ijk} \tag{4}$$

- β_k – effect of word k – unexplained differences in average response elicited by different stimuli
- $\pi_i\beta_k$ – the stimulus word by subject interaction – different people respond to different stimuli differently

Taking into account error variance due to subjects and items

Clark's (1973) *minF′* solution

$$minF' = \frac{MS_{\tau}}{MS_{\pi\tau} + MS_{\beta_k}} = \frac{F_1 F_2}{F_1 + F_2} \tag{5}$$

1. You start by *aggregating* your data
2. By-subjects data – for each subject, take the average of their responses to all the items
3. By-items data – for each item, take the average of all subjects' responses
4. You do separate ANOVAs, one for by-subjects (F1) data and one for by-items (F2) data
5. You put F1 and F2 together in the calculation of minF' *though psychologists tended to stop at step 4*

Taking into account error variance due to subjects and items

Clark's (1973) *minF′* solution

$$minF' = \frac{MS_{\tau}}{MS_{\pi\tau} + MS_{\beta_k}} = \frac{F_1 F_2}{F_1 + F_2} \tag{5}$$

1. You start by *aggregating* your data
2. By-subjects data – for each subject, take the average of their responses to all the items
3. By-items data – for each item, take the average of all subjects' responses
4. You do separate ANOVAs, one for by-subjects (F1) data and one for by-items (F2) data
5. You put F1 and F2 together in the calculation of minF' *though psychologists tended to stop at step 4*

Taking into account error variance due to subjects and items

Clark's (1973) *minF′* solution

$$minF' = \frac{MS_{\tau}}{MS_{\pi\tau} + MS_{\beta_k}} = \frac{F_1 F_2}{F_1 + F_2} \quad (5)$$

1. You start by *aggregating* your data
2. By-subjects data – for each subject, take the average of their responses to all the items
3. By-items data – for each item, take the average of all subjects' responses
4. You do separate ANOVAs, one for by-subjects (F1) data and one for by-items (F2) data
5. You put F1 and F2 together in the calculation of minF' *though psychologists tended to stop at step 4*

Taking into account error variance due to subjects and items

Clark's (1973) *minF′* solution

$$minF' = \frac{MS_{\tau}}{MS_{\pi\tau} + MS_{\beta_k}} = \frac{F_1 F_2}{F_1 + F_2} \quad (5)$$

1. You start by *aggregating* your data
2. By-subjects data – for each subject, take the average of their responses to all the items
3. By-items data – for each item, take the average of all subjects' responses
4. You do separate ANOVAs, one for by-subjects (F1) data and one for by-items (F2) data
5. You put F1 and F2 together in the calculation of minF' *though psychologists tended to stop at step 4*

Taking into account error variance due to subjects and items

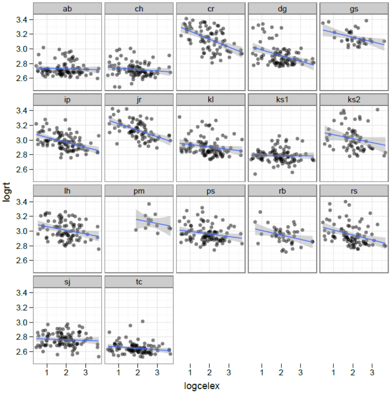
Clark's (1973) *minF′* solution

$$minF' = \frac{MS_{\tau}}{MS_{\pi\tau} + MS_{\beta_k}} = \frac{F_1 F_2}{F_1 + F_2} \tag{5}$$

1. You start by *aggregating* your data
2. By-subjects data – for each subject, take the average of their responses to all the items
3. By-items data – for each item, take the average of all subjects' responses
4. You do separate ANOVAs, one for by-subjects (F1) data and one for by-items (F2) data
5. You put F1 and F2 together in the calculation of minF' *though psychologists tended to stop at step 4*

The problem with minF' is that it is only good for ANOVA and ANOVA is only good for testing the effects of categorical variables – *factors*

Many dealt with the Clark problem, and allowed themselves to include predictors that were continuous variables, by performing regression analyses of by-items data



Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) argued there is a problem with multiple regression on by-items mean observations
- The approach reverses the language-as-fixed-effect problem
- Effects are assessed by comparison with an item-based error term
- Effects can be significant because of random variation between subjects in how they responded to items

Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) argued there is a problem with multiple regression on by-items mean observations
- The approach reverses the language-as-fixed-effect problem
- Effects are assessed by comparison with an item-based error term
- Effects can be significant because of random variation between subjects in how they responded to items

Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) argued there is a problem with multiple regression on by-items mean observations
- The approach reverses the language-as-fixed-effect problem
- Effects are assessed by comparison with an item-based error term
- Effects can be significant because of random variation between subjects in how they responded to items

Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) argued there is a problem with multiple regression on by-items mean observations
- The approach reverses the language-as-fixed-effect problem
- Effects are assessed by comparison with an item-based error term
- Effects can be significant because of random variation between subjects in how they responded to items

Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) suggested solutions to the problem with multiple regression on by-items mean observations
- A common approach is to perform a regression (linear model) on each subject and complete a t-test or ANOVA on the resulting per-subject coefficients
- Code for subject with $n - 1$ dummy variables and complete regression using subject, and subject by effect, predictors

Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) suggested solutions to the problem with multiple regression on by-items mean observations
- A common approach is to perform a regression (linear model) on each subject and complete a t-test or ANOVA on the resulting per-subject coefficients
- Code for subject with $n - 1$ dummy variables and complete regression using subject, and subject by effect, predictors

Repeated measures regression analysis

The problem with regression on by-items means data

- Lorch & Myers (1990) suggested solutions to the problem with multiple regression on by-items mean observations
- A common approach is to perform a regression (linear model) on each subject and complete a t-test or ANOVA on the resulting per-subject coefficients
- Code for subject with $n - 1$ dummy variables and complete regression using subject, and subject by effect, predictors

Analyses of results with simulated data and alternate procedures suggested that the per-subjects regression approach does not work well *Baayen et al. (2008)*

You can see how problems can arise – the approach cannot take into account variation in the reliability of estimates

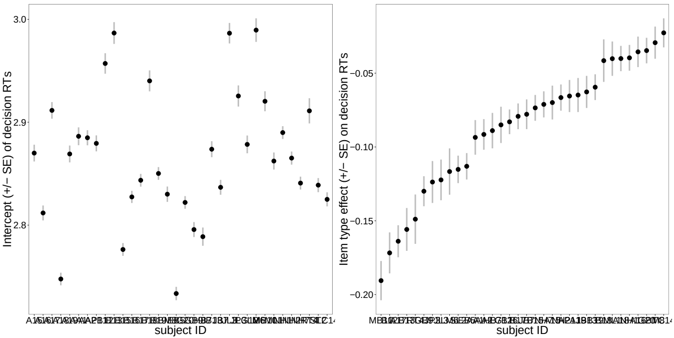


Figure : Per-subject estimates (points) with standard errors – ML data-set – intercepts and the effect of item type (words vs. non-words)

A more realistic repeated measures model of the item type and reading ability effects

The model is better because it takes into account random effects of subjects and items on intercepts, and the interactions between the effect of participant and the effect of stimulus item type

$$RT = \beta_0 + \beta_{wordreadingability} + \beta_{itemtype} + \beta_{ability*itemtype} + \beta_{subject} + \beta_{item} + \beta_{subject*itemtype} + \epsilon \quad (6)$$

- Allow *intercept to vary* – random effect of subject – some have slower average some have faster average than average overall
- Allow *slope of item type effect to vary* – random effect of subject – subjects can experience effects of (item type) with different directions or sizes

A more realistic repeated measures model of the item type and reading ability effects

The model is better because it takes into account random effects of subjects and items on intercepts, and the interactions between the effect of participant and the effect of stimulus item type

$$RT = \beta_0 + \beta_{wordreadingability} + \beta_{itemtype} + \beta_{ability*itemtype} + \beta_{subject} + \beta_{item} + \beta_{subject*itemtype} + \epsilon \quad (6)$$

- Allow *intercept to vary* – random effect of subject – some have slower average some have faster average than average overall
- Allow *slope of item type effect to vary* – random effect of subject – subjects can experience effects of (item type) with different directions or sizes

A more realistic repeated measures model of the item type and reading ability effects

Taking into account random effects of subjects and items

$$RT = \beta_0 + \beta_{wordreadingability} + \beta_{itemtype} + \beta_{ability*itemtype} + \beta_{subject} + \beta_{item} + \beta_{subject*itemtype} + \epsilon \quad (7)$$

- What about effect of random variation between stimuli?
  - Allow *intercept to vary* – random effect of items – some items harder and elicit slower average response and some easier and elicit faster responses on average
  - Allow *effect of reading ability to vary* – within-items effect of subject type can be different for different items

A more realistic repeated measures model of the item type and reading ability effects

Taking into account random effects of subjects and items

$$RT = \beta_0 + \beta_{wordreadingability} + \beta_{itemtype} + \beta_{ability*itemtype} + \beta_{subject} + \beta_{item} + \beta_{subject*itemtype} + \epsilon \quad (7)$$

- What about effect of random variation between stimuli?
  - Allow *intercept to vary* – random effect of items – some items harder and elicit slower average response and some easier and elicit faster responses on average
  - Allow *effect of reading ability to vary* – within-items effect of subject type can be different for different items

A more realistic repeated measures model of the item type and reading ability effects

Taking into account random effects of subjects and items

$$RT = \beta_0 + \beta_{wordreadingability} + \beta_{itemtype} + \beta_{ability*itemtype} + \beta_{subject} + \beta_{item} + \beta_{subject*itemtype} + \epsilon \quad (7)$$

- What about effect of random variation between stimuli?
  - Allow *intercept to vary* – random effect of items – some items harder and elicit slower average response and some easier and elicit faster responses on average
  - Allow *effect of reading ability to vary* – within-items effect of subject type can be different for different items

We do not model random effects directly – we just estimate the *spread* or variance of random differences

random intercepts – predicted differences (adjustments) between the overall average and the group average – and *random slopes* – predicted differences (adjustments) between the overall effect and the group average effect

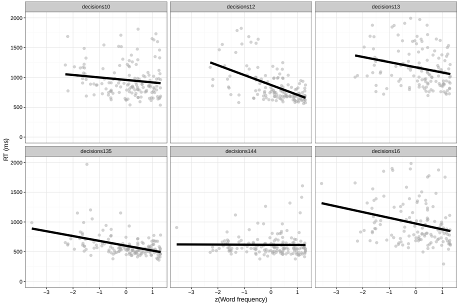


Figure : Individual differences in intercepts and in the effect of word frequency on lexical decision RTs

The advantages of mixed-effects models

- Approach is not restrictive about predictors or data structure
  - ANOVA is OK for experimental designs, categorical factors, data sets without missing values
- Can test effects at different levels of hierarchy
- We can allow random effects of both subjects and items – solving the ‘language-as-fixed-effect’ problem
- Estimation robust to imbalances in data

The advantages of mixed-effects models

- Approach is not restrictive about predictors or data structure
  - ANOVA is OK for experimental designs, categorical factors, data sets without missing values
- Can test effects at different levels of hierarchy
- We can allow random effects of both subjects and items – solving the 'language-as-fixed-effect' problem
- Estimation robust to imbalances in data

The advantages of mixed-effects models

- Approach is not restrictive about predictors or data structure
  - ANOVA is OK for experimental designs, categorical factors, data sets without missing values
- Can test effects at different levels of hierarchy
- We can allow random effects of both subjects and items – solving the ‘language-as-fixed-effect’ problem
- Estimation robust to imbalances in data

The advantages of mixed-effects models

- Approach is not restrictive about predictors or data structure
  - ANOVA is OK for experimental designs, categorical factors, data sets without missing values
- Can test effects at different levels of hierarchy
- We can allow random effects of both subjects and items – solving the ‘language-as-fixed-effect’ problem
- Estimation robust to imbalances in data

We focus on building a series of models up to the most complex model supported by the data

- A minimal (empty) model of the data might assume that the data we observe can be predicted only given the average value of observations – intercept
- And random effects of grouping variables like subjects or stimulus items on the average outcome

We focus on building a series of models up to the most complex model supported by the data

- A minimal (empty) model of the data might assume that the data we observe can be predicted only given the average value of observations – intercept
- And random effects of grouping variables like subjects or stimulus items on the average outcome

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                     (1|subjectID) + (1|item_name),
 data = ML.all.correct)
```

- `lmer()` run a *Linear Mixed-effects model* using the *lmer()* function
- `full.lmer0 <- lmer()` the model creates an object *full.lmer0* we can work with e.g. to get a summary of estimates
- `logrt ~ (1|subjectID) + (1|item_name)` the *lmer()* function interprets the model formula we specify
- `data = ML.all.correct` applying that model to the data we name *ML.all.correct*

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                    (1|subjectID) + (1|item_name),
 data = ML.all.correct)
```

- `lmer()` run a *Linear Mixed-effects model* using the *lmer()* function
- `full.lmer0 <- lmer()` the model creates an object *full.lmer0* we can work with e.g. to get a summary of estimates
- `logrt ~ (1|subjectID) + (1|item_name)` the *lmer()* function interprets the model formula we specify
- `data = ML.all.correct` applying that model to the data we name *ML.all.correct*

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                    (1|subjectID) + (1|item_name),
 data = ML.all.correct)
```

- `lmer()` run a *Linear Mixed-effects model* using the *lmer()* function
- `full.lmer0 <- lmer()` the model creates an object *full.lmer0* we can work with e.g. to get a summary of estimates
- `logrt ~ (1|subjectID) + (1|item_name)` the *lmer()* function interprets the model formula we specify
- `data = ML.all.correct` applying that model to the data we name *ML.all.correct*

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                    (1|subjectID) + (1|item_name),
 data = ML.all.correct)
```

- `lmer()` run a *Linear Mixed-effects model* using the *lmer()* function
- `full.lmer0 <- lmer()` the model creates an object *full.lmer0* we can work with e.g. to get a summary of estimates
- `logrt ~ (1|subjectID) + (1|item_name)` the *lmer()* function interprets the model formula we specify
- `data = ML.all.correct` applying that model to the data we name *ML.all.correct*

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                    (1|subjectID) + (1|item_name),
data = ML.all.correct)
```

- `(1|subjectID) + (1|item_name)` specify random effects
  - `(1|...)` – random effect on intercepts
  - `subjectID` or `item_name` – effects of subjects or items – specified by subject identity code (ID) or item name in coding variables
- The intercept `logrt ~ 1` is taken as required automatically, given the random intercepts terms

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                    (1|subjectID) + (1|item_name),
data = ML.all.correct)
```

- `(1|subjectID) + (1|item_name)` specify random effects
  - `(1|...)` – random effect on intercepts
  - `subjectID` or `item_name` – effects of subjects or items – specified by subject identity code (ID) or item name in coding variables
- The intercept `logrt ~ 1` is taken as required automatically, given the random intercepts terms

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                    (1|subjectID) + (1|item_name),
data = ML.all.correct)
```

- `(1|subjectID) + (1|item_name)` specify random effects
  - `(1|...)` – random effect on intercepts
  - `subjectID` or `item_name` – effects of subjects or items – specified by subject identity code (ID) or item name in coding variables
- The intercept `logrt ~ 1` is taken as required automatically, given the random intercepts terms

R code for running a linear mixed-effects model

Start with an empty model specifying just the fixed effect of the intercept (overall average outcome) and the random effects of subjects and of items on intercepts (random differences in average outcomes)

```
full.lmer0 <- lmer(logrt ~
                         (1|subjectID) + (1|item_name),
data = ML.all.correct)
```

- `(1|subjectID) + (1|item_name)` specify random effects
  - `(1|...)` – random effect on intercepts
  - `subjectID` or `item_name` – effects of subjects or items – specified by subject identity code (ID) or item name in coding variables
- The intercept `logrt ~ 1` is taken as required automatically, given the random intercepts terms

R code for running a linear mixed-effects model

An empty model specifying just the fixed effect of the intercept and random effects of subjects and of items on intercepts – *summary(full.lmer0)* will deliver model estimates

```
full.lmer0 <- lmer(logrt ~
                     (1|subjectID) + (1|item_name),
data = ML.all.correct)
summary(full.lmer0)
```

```
> summary(ML.all.correct.lmer.0)
Linear mixed model fit by maximum likelihood
t-tests use  Satterthwaite approximations to degrees of freedom ['lmerMod']
Formula: logrt ~ (1 | subjectID) + (1 | item_name)
   Data: ML.all.correct

     AIC      BIC   logLik deviance df.resid
-17980.8 -17951.8   8994.4 -17988.8    10250

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5446 -0.6442 -0.1498  0.4667  5.2431

Random effects:
 Groups    Name        Variance Std.Dev.
 item_name (Intercept) 0.002681 0.05178
 subjectID (Intercept) 0.003044 0.05517
 Residual              0.009281 0.09634
Number of obs: 10254, groups:  item_name, 320; subjectID, 34

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  2.823356   0.009941 40.520000     284   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

R code for running a linear mixed-effects model – *notice the moving parts*

$$full.lmer0 < -lmer(logrt \sim (1|subjectID) + (1|itemname), data = ML.all.c \quad (8)$$

- You specify the model name full.lmer0, the dependent variable logrt, the subject and item coding variables subjectID, itemname, and the name of the data frame ML.all.correct
- You need to have previously used *read.csv()* to enter the data frame in R's workspace, and you need to know what the variables in the dataframe are called

R code for running a linear mixed-effects model – *notice the moving parts*

$$full.lmer0 < -lmer(logrt \sim (1|subjectID) + (1|itemname), data = ML.all.c \tag{8}$$

- You specify the model name full.lmer0, the dependent variable logrt, the subject and item coding variables subjectID, itemname, and the name of the data frame ML.all.correct
- You need to have previously used *read.csv()* to enter the data frame in R's workspace, and you need to know what the variables in the dataframe are called

LMEs – build-up the *fixed* effects while holding the *random* effects constant

```
full.lmer1 <- lmer(logrt ~

zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

(1|subjectID) + (1|item_name),
data = ML.all.correct, REML = F)
summary(full.lmer1)
```

- `zAge + zTOWRE_wordacc` add fixed effects – z- because they were standardized
- *Fixed effects* reproducible effects – manipulated, selected – of theoretical or practical interest
- `summary(full.lmer1)` – print a model summary

LMEs – build-up the *fixed* effects while holding the *random* effects constant

```
full.lmer1 <- lmer(logrt ~

zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

(1|subjectID) + (1|item_name),
data = ML.all.correct, REML = F)
summary(full.lmer1)
```

- `zAge + zTOWRE_wordacc` add fixed effects – z- because they were standardized
- *Fixed effects* reproducible effects – manipulated, selected – of theoretical or practical interest
- `summary(full.lmer1)` – print a model summary

LMEs – build-up the *fixed* effects while holding the *random* effects constant

```
full.lmer1 <- lmer(logrt ~

zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

(1|subjectID) + (1|item_name),
data = ML.all.correct, REML = F)
summary(full.lmer1)
```

- `zAge + zTOWRE_wordacc` add fixed effects – z- because they were standardized
- *Fixed effects* reproducible effects – manipulated, selected – of theoretical or practical interest
- `summary(full.lmer1)` – print a model summary

LMEs – build-up the *fixed* effects while holding the *random* effects constant

```
full.lmer1 <- lmer(logrt ~
zAge + zTOWRE_wordacc + zTOWRE_nonwordacc + (1|subjectID) + (1|ite
data = ML.all.correct, REML = F)
summary(full.lmer1)
```

```
> summary(ML.all.correct.lmer.1)
Linear mixed model fit by maximum likelihood
t-tests use  Satterthwaite approximations to degrees of freedom ['lmerMod']
Formula: logrt ~ zAge + zTOWRE_wordacc + zTOWRE_nonwordacc + (1 | subjectID) +      (1 | item_name)
   Data: ML.all.correct

     AIC      BIC   logLik deviance df.resid
-17981.9 -17931.3   8998.0 -17995.9    10247

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5397 -0.6439 -0.1489  0.4675  5.2441

Random effects:
 Groups    Name        Variance Std.Dev.
 item_name (Intercept) 0.002681 0.05178
 subjectID (Intercept) 0.002458 0.04958
 Residual              0.009281 0.09634
Number of obs: 10254, groups:  item_name, 320; subjectID, 34

Fixed effects:
                    Estimate Std. Error         df t value Pr(>|t|)
(Intercept)        2.8233188  0.0090345 42.1200000 312.505   <2e-16 ***
zAge               0.0130997  0.0088762 33.9600000   1.476    0.149
zTOWRE_wordacc     0.0001089  0.0114635 33.9700000   0.010    0.992
zTOWRE_nonwordacc -0.0174850  0.0113883 33.9700000  -1.535    0.134
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

How do we know if increasing *model complexity* by adding predictors actually helps us to account for variation in outcome values?

Simplicity and parsimony – we look at evaluating models, next

- Trade-off between too much and too little simplicity in model selection – variable selection
  - Models with too many parameters may tend to identify effects that are spurious
  - Effects may be *unintuitive* and hard to explain *and* not reproduced in future samples

How do we know if increasing *model complexity* by adding predictors actually helps us to account for variation in outcome values?

Simplicity and parsimony – we look at evaluating models, next

- Trade-off between too much and too little simplicity in model selection – variable selection
  - Models with too many parameters may tend to identify effects that are spurious
  - Effects may be *unintuitive* and hard to explain *and* not reproduced in future samples

How do we know if increasing *model complexity* by adding predictors actually helps us to account for variation in outcome values?

Simplicity and parsimony – we look at evaluating models, next

- Trade-off between too much and too little simplicity in model selection – variable selection
    - Models with too many parameters may tend to identify effects that are spurious
    - Effects may be *unintuitive* and hard to explain *and* not reproduced in future samples

Summary

The advantages of mixed-effects models in R

- Mixed-effects models are required where clusters of observations are analyzed
  - Repeated measures designs – multiple observations per participant or stimulus
- Traditional methods e.g. slopes-as-outcomes are approximations to linear mixed-effects models
- The *lmer()* R code allows you to specify your model $logrt \sim itemtype * ability + (1|subject$ in a simple, flexible language

Summary

The advantages of mixed-effects models in R

- Mixed-effects models are required where clusters of observations are analyzed
  - Repeated measures designs – multiple observations per participant or stimulus
- Traditional methods e.g. slopes-as-outcomes are approximations to linear mixed-effects models
- The *lmer()* R code allows you to specify your model $logrt \sim itemtype * ability + (1|subject$ in a simple, flexible language

Summary

The advantages of mixed-effects models in R

- Mixed-effects models are required where clusters of observations are analyzed
  - Repeated measures designs – multiple observations per participant or stimulus
- Traditional methods e.g. slopes-as-outcomes are approximations to linear mixed-effects models
- The *lmer()* R code allows you to specify your model $logrt \sim itemtype * ability + (1|subject$ in a simple, flexible language

Summary

The advantages of mixed-effects models in R

- Mixed-effects models are required where clusters of observations are analyzed
  - Repeated measures designs – multiple observations per participant or stimulus
- Traditional methods e.g. slopes-as-outcomes are approximations to linear mixed-effects models
- The *lmer()* R code allows you to specify your model $logrt \sim itemtype * ability + (1|subject$ in a simple, flexible language