Linear mixed models in R
Day 2

JONAS WALTHER

Syntax of a crossed mixed-effects model

RT ~ Context + (1 | Subject)

Dependent variable

Independent variables/fixed effects

Random effects

Syntax of a crossed mixed-effects model

RT ~ Group * Context + Age + (1 + Context | Subject) + (1 | Item)

Dependent variable

Independent variables/fixed effects

Random effects (crossed between Subject and Item)

Syntax of a crossed mixed-effects model

RT ~ Group * Context + Age + (1 + Context | Subject) + (1 | Item)

Independent variables/fixed effects

: - interaction

* - short-form for single fixed effects and interaction between effects; equals to:

Group + Context + Group:Context

Interactions

Group*Context = Group + Context + Group:Context

	PL Context	UK Context
Experimental Group		
Control Group		

Interactions

No interaction – effects are additive

	PL Context	UK Context
Experimental Group	500ms	550ms
Control Group	480ms	530ms

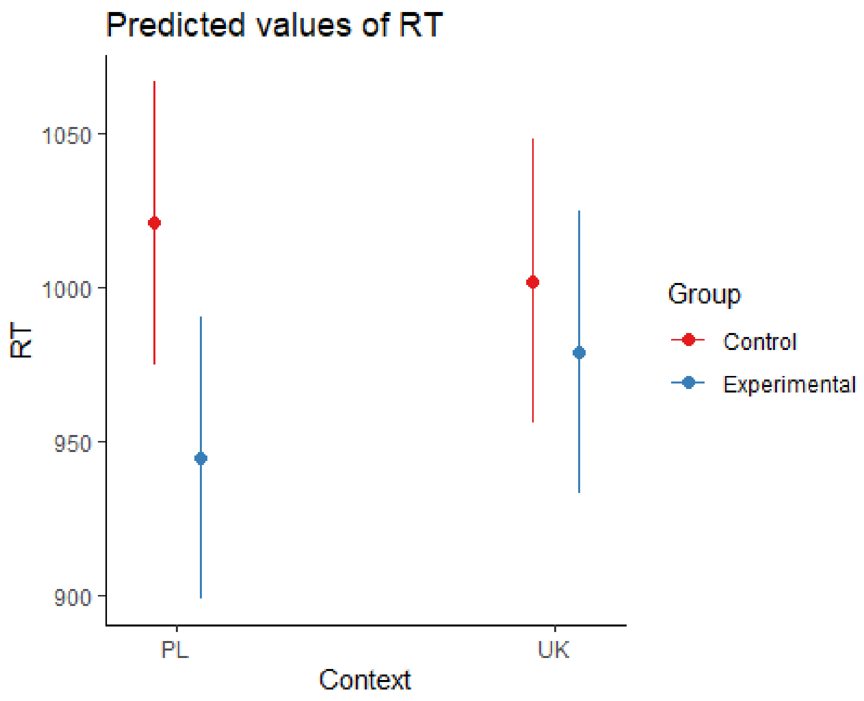
Sign. Interaction – effects are not additive

	PL Context	UK Context
Experimental Group	500ms	550ms
Control Group	480ms	480ms

Interactions

Significant interactions do not give sensible information on their own

- Additional tests are necessary to understand their effects



Syntax of a crossed mixed-effects model

RT ~ Group * Context + Age + (1 + Context | Subject) + (1 | Item)

Dependent variable

Independent variables/fixed effects

: - interaction

* - short-form for single fixed effects and interaction between effects; equals:

Group + Context + Group:Context

Random effects (crossed between Subject and Item)

Crossed vs. nested random effects

<u>Crossed</u>: each level of one random effect occurs at each level of the other random effect, and vice versa

- e.g., each subject sees every item, and each item is seen by every subject

<u>Nested</u>: There are some levels of one random effect (the nested one) that only occur within one level of the other random effect (the nesting one)

- E.g. tested several students in several schools. A given student only occurs in *one* school, not in every school

A model can include both (e.g. *Students* nested under *Schools*, but *Items* fully crossed with those)

Syntax of a crossed mixed-effects model

RT ~ Group * Context + Age + (1 + Context | Subject) + (1 | Item)

Nested random effect:

RT ~ Group * Context + Age + (1 | School/Student)

Identical to:

RT ~ Group * Context + Age + (1 | School) + (1|School:Student)

Summary-Output of an LME

Amount of variance for each random term

Coefficients for fixed effects

Correlation of fixed effects (NOT correlation of predictors or collinearity)

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: RT ~ Group + Context + Group:Context + Age + AoA + Trial + (1 +
##     Context | Subject) + (1 + Context | ItemNr)
##    Data: PN_Data
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ItemNr   (Intercept) 24021    154.99
##           ContextUK    3449     58.73   -0.15
##  Subject  (Intercept) 19522    139.72
##           ContextUK   16426    128.16   -0.41
##  Residual             55688    235.98
## Number of obs: 7227, groups:  ItemNr, 210; Subject, 74
##
## Fixed effects:
##                              Estimate Std. Error t value
## (Intercept)                 1019.6555    79.0877  12.893
## GroupExperimental            -59.6657    37.2576  -1.601
## ContextUK                    -11.2579    23.2922  -0.483
## Age                           -1.9501     2.4100  -0.809
## AoA                            3.2835     3.9845   0.824
## Trial                          0.4629     0.2875   1.610
## GroupExperimental:ContextUK   45.7348    32.1411   1.423
##
## Correlation of Fixed Effects:
##              (Intr) GrpExp CntxUK Age    AoA    Trial
## GrpExprmntl  0.196
## ContextUK   -0.127  0.290
## Age         -0.810 -0.309 -0.012
## AoA         -0.269 -0.217  0.001 -0.234
## Trial       -0.110 -0.001 -0.006  0.000  0.001
## GrpExpr:CUK  0.088 -0.404 -0.704  0.010 -0.001  0.000
```

Random effects-Output

For each random effect, how much each subject's effect differs from the overall fixed effect

```
ranef(model_GeneralModel)
## $Subject
##        (Intercept)   ContextUK
## AF1310   34.183826   71.901098
## AG0712  -26.795064  -10.985648
## AG0911 -243.449395  170.020494
## AJ1312  -58.494452   59.235755
## AJ1611 -239.331334  327.456270
## AJ3007 -148.513248  -77.695294
## AK0807  231.764526 -167.158225
## AK1612 -210.589850   13.285782
…
```

Syntax of a crossed mixed-effects model

RT ~ Group * Context + Age + (1 + Context | Subject) + (1 | Item)

Dependent variable

Independent variables/fixed effects

Random effects (crossed between Subject and Item)

- Random effects are grouping factors
- Always categorical
- Recommended to have at least five levels

Syntax of random effects

(1 + Context | Subject)

Random intercept

Random slopes

Variable name for coding levels of random factor

We added subject intercepts

$$y_{i\mathrm{Cond}} = \beta_0 + \beta_{Cond}x_{iCond} + \epsilon_{Cond} + u_{0i}$$

But we can also add subject slopes

$$y_{iCond} = \beta_0 + (\beta_{Cond} + u_{1i})x_{iCond} + \epsilon_{Cond} + u_{0i}$$

β_0 -average naming latency

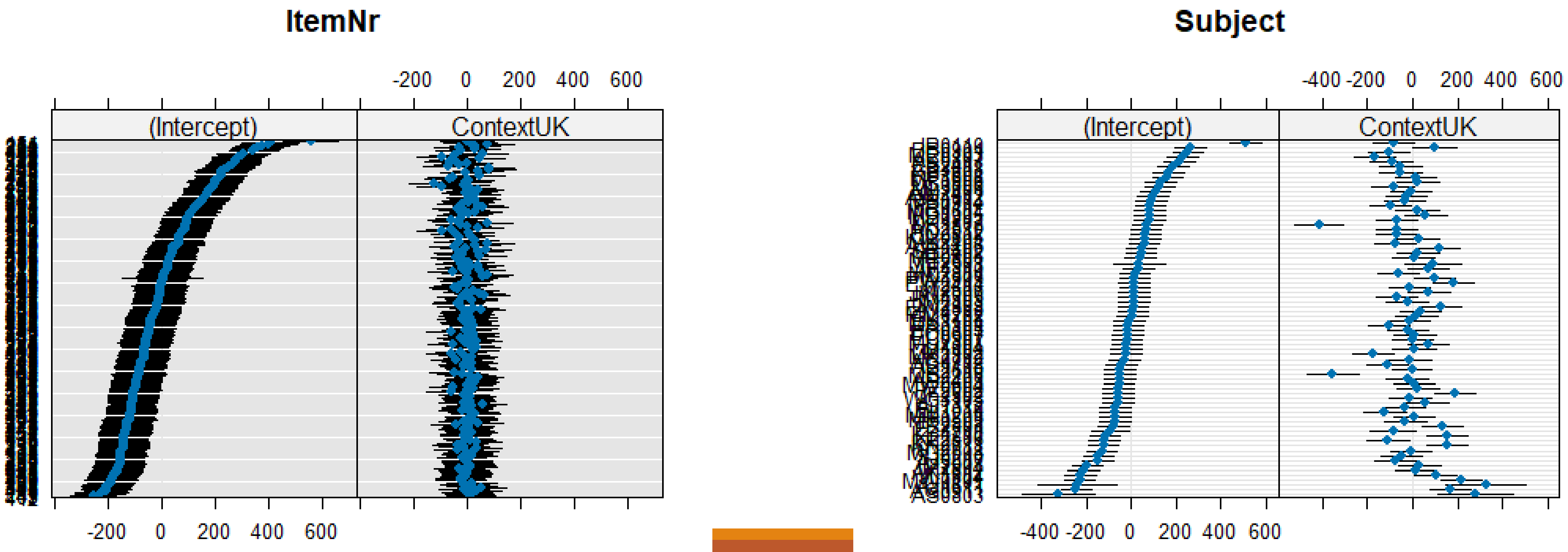
u_{0i} -subject dependant adjustment

ϵ_{Cond} -error/noise term

Subject adjustments for intercept and fixed effects

RT ~ Group + Context + Group:Context + Age + AoA + Trial + (1 + Context| Subject) + (1 + Context|ItemNr)

```r
library(lattice)
print(dotplot(ranef(model_GeneralModel, condVar = TRUE)))
```



Syntax of random effects

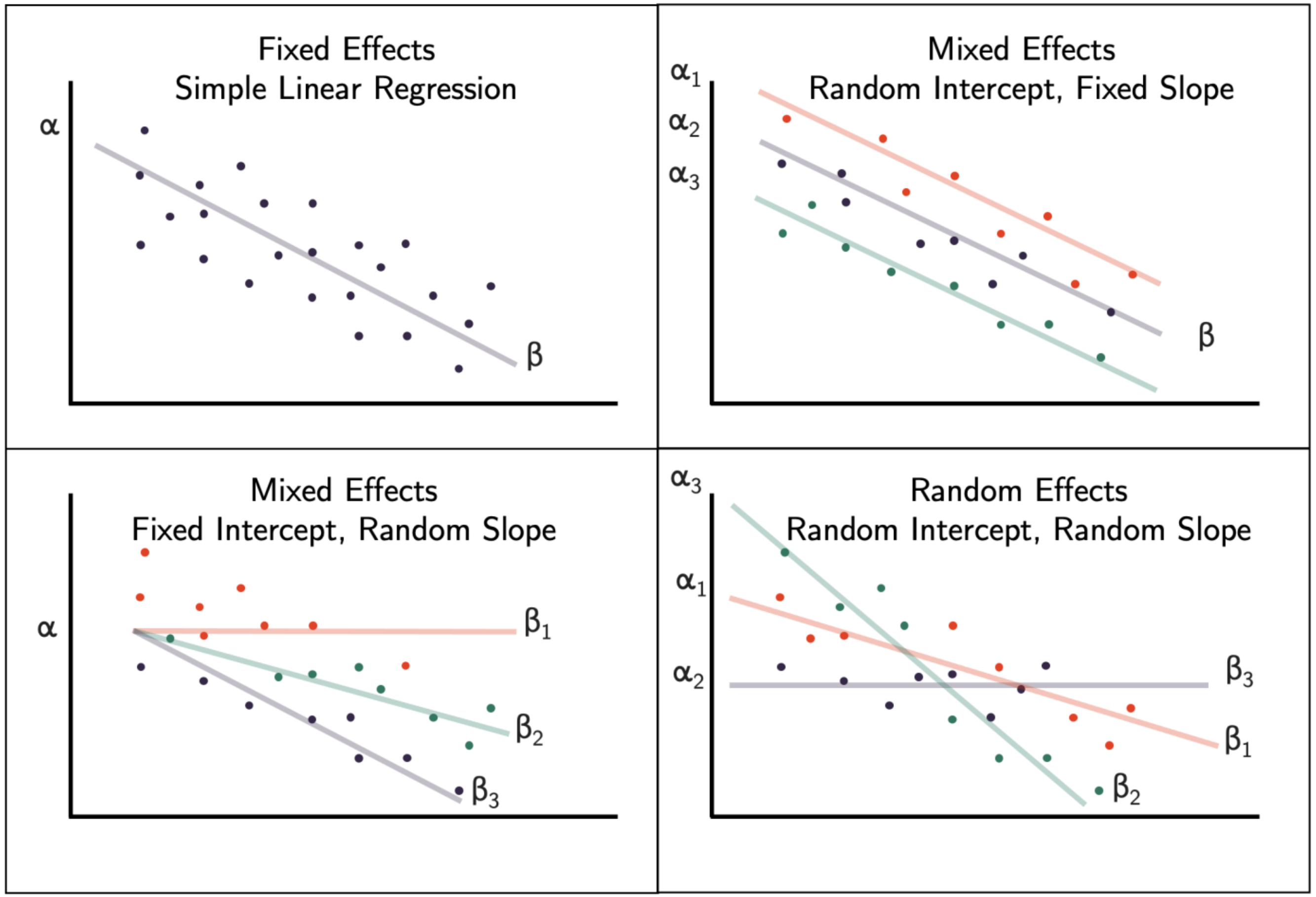
(1 + Context | Subject)

Random intercept

Random slopes

Variable name for coding levels of random factor

Random intercept and random slope



https://bookdown.org/steve_midway/DAR/random-effects.html

Random intercept and no slope

```
summary(lmer(RT ~ Context + (1 | Subject) + (1 | ItemNr),
data=PN_Data))


## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ItemNr   (Intercept) 23600    153.6
##  Subject  (Intercept) 15184    123.2
##  Residual             59859    244.7
```

Random intercept and random slope

```
summary(lmer(RT ~ Context + (1 + Context | Subject) + (1 |
ItemNr), data=PN_Data))
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ItemNr   (Intercept) 23827    154.4
##  Subject  (Intercept) 20662    143.7
##           ContextUK   17133    130.9    -0.46
##  Residual             56532    237.8
```

No intercept and random slope

```
summary(lmer(RT ~ Context + (0 + Context | Subject) + (1|
ItemNr), data=PN_Data))
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ItemNr   (Intercept) 23827    154.4
##  Subject  ContextPL   20663    143.7
##           ContextUK   20330    142.6    0.58
##  Residual             56532    237.8
```

Random intercept and random slope

```
summary(lmer(RT ~ Context + (1 + Context || Subject) + (1 + Context
| ItemNr), data=PN_Data))
```

```
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  ItemNr    (Intercept) 24424    156.28
##            ContextUK    3427     58.54   -0.17
##  Subject   ContextPL    8073     89.85
##            ContextUK    7906     88.91   -0.07
##  Subject.1 (Intercept) 12515    111.87
##  Residual              55673    235.95
```

Random structure: correlations between intercept and slope

(1 + Context | Subject)

Random intercept

Random slopes

Variable name for coding levels of random factor

| - random intercept and random slope are correlated with each other

|| - no correlations between random effects

(Context || Subject)

This expands to:

(0 + Context | Subject) + (1|Subject)

Random structure: correlations between intercept and slope

$y_{iCond} = \beta_0 + (\beta_{Cond} + u_{1i})x_{iCond} + \epsilon_{Cond} + u_{0i}$

Between-subject variability in the intercept:

$u_0 \sim Normal(0, \sigma_{u0})$

Between-subject variability in the slope:

$u_1 \sim Normal(0, \sigma_{u1})$

Within-subject variability:

$\epsilon \sim Normal(0, \sigma)$

```
model_Large8 = lmer(RT ~ Group+Context + (1
+Context|Subject) +(1| ItemNr), data=PN_Data)

## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ItemNr   (Intercept) 23832    154.4
##  Subject  (Intercept) 19854    140.9
##           ContextUK   16898    130.0    -0.44
##  Residual             56540    237.8
```

Random structure: correlations between intercept and slope

```
(1 +Context|Subject)

## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 19854    140.9
##           ContextUK   16898    130.0    -0.44
##  Residual             56540    237.8
```

```
(1 +Context||Subject)

## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  Subject   ContextPL    3843     61.99
##            ContextUK    4623     68.00   -1.00
##  Subject.1 (Intercept) 16011    126.53
##  Residual              56540    237.78
```

Random structure: correlations between intercept and slope

Correlations between intercepts and slopes are frequently assumed in research

- Group with higher intercept will also have a higher slope
- If they are present, they can be used to calculate better estimates
- Even low correlations can contribute to better estimates

But they make the model more complex

Correlations of 1 between slope and intercept can prevent the model from converging

- 1 or very close to one, usually means that the model failed to calculate the correlation
- Usually because there was not enough data
- Removing correlations by using || allows model to converge
- But it is recommended to also switch to a simpler model

Assumptions of linear mixed models

The explanatory variables are related linearly to the response and additive.

The residuals have constant variance.

The residuals are normally distributed.

The explanatory variables are related linearly to the response.

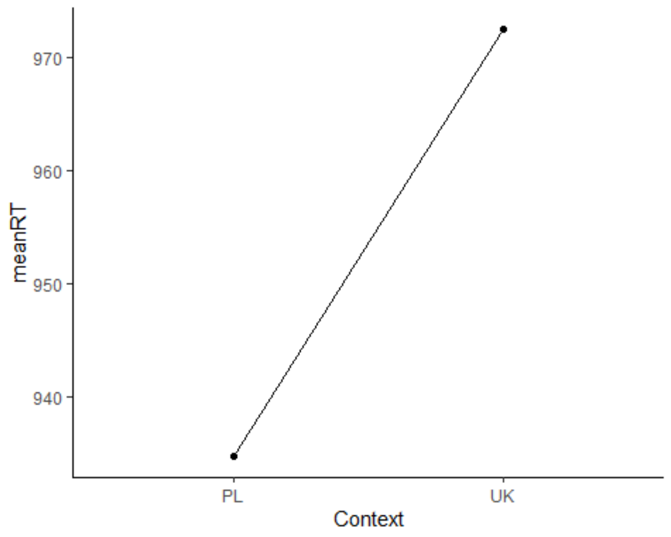
Linear mixed effects models rely on linear relationships between the variables

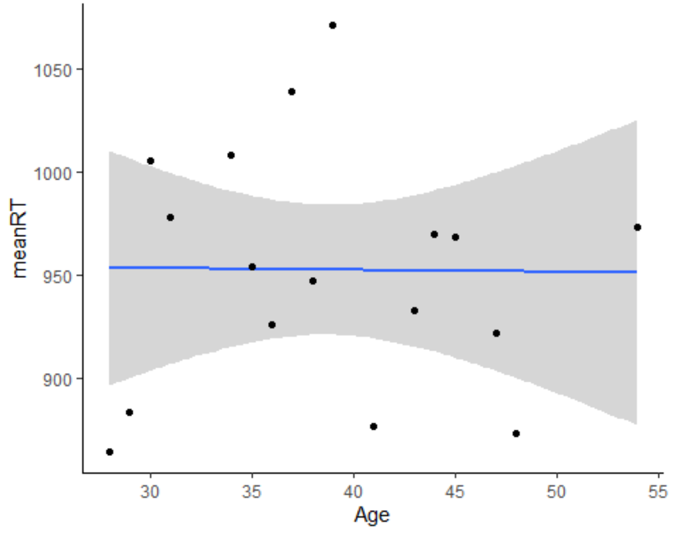
Categorical variables are treated as 2-level variables

Continuous variables are used in linear regression

But many effects are non-linear:

- Introduction of additional predictors and constructs to approximate non-linear effects
- Switch to Generalised-Mixed-Models and other tools





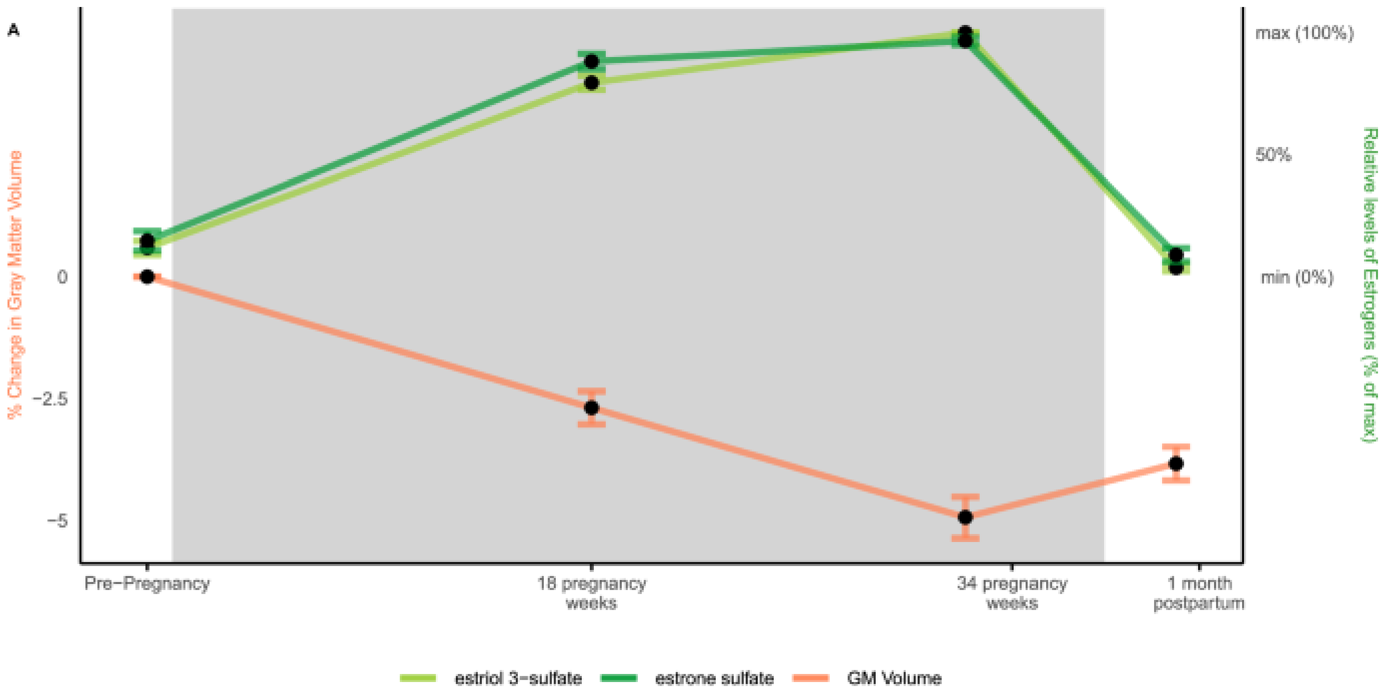
Analysing nonlinear effects with LMEs

Fixed effects in LMEs are always treated as linear

Many processes are non-linear

LMEs allow the abstraction of non-linear effects as linear ones

-> but you lose information and you could introduce erroneus effects



Servin-Barthet et al. (2025)

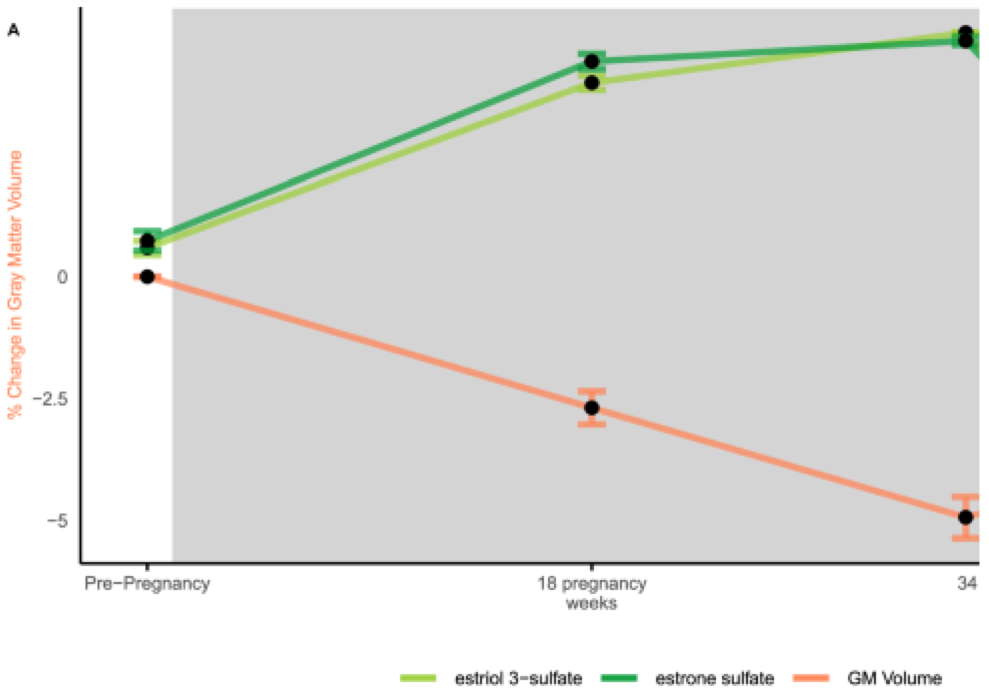
Analysing nonlinear effects with LMEs

Fixed effects in LMEs are always treated as linear

Many processes are non-linear

LMEs allow the abstraction of non-linear effects as linear ones

-> but you lose information and you could introduce erroneus effects



Servin-Barthet et al. (2025)

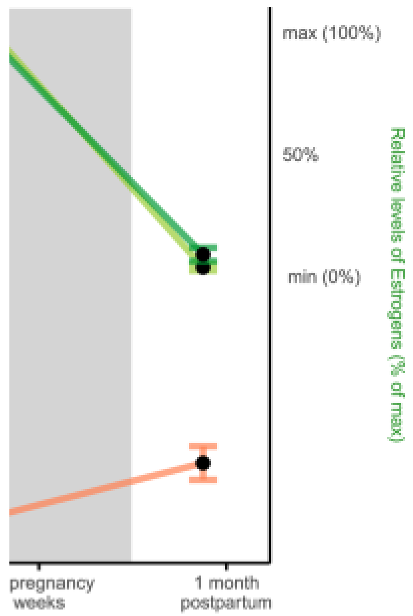
Analysing nonlinear effects with LMEs

Fixed effects in LMEs are always treated as linear

Many processes are non-linear

LMEs allow the abstraction of non-linear effects as linear ones

-> but you lose information and you could introduce erroneus effects



Servin-Barthet et al. (2025)

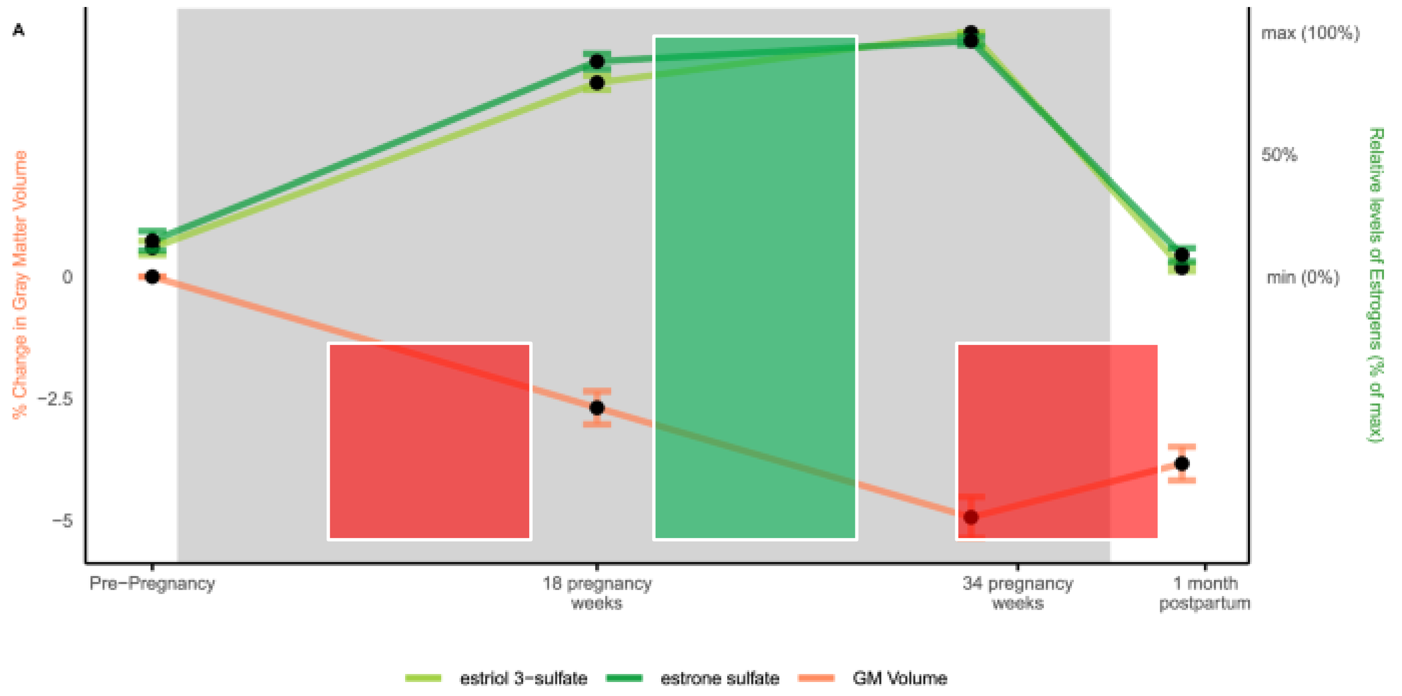
Analysing nonlinear effects with LMEs

Fixed effects in LMEs are always treated as linear

Many processes are non-linear

LMEs allow the abstraction of non-linear effects as linear ones

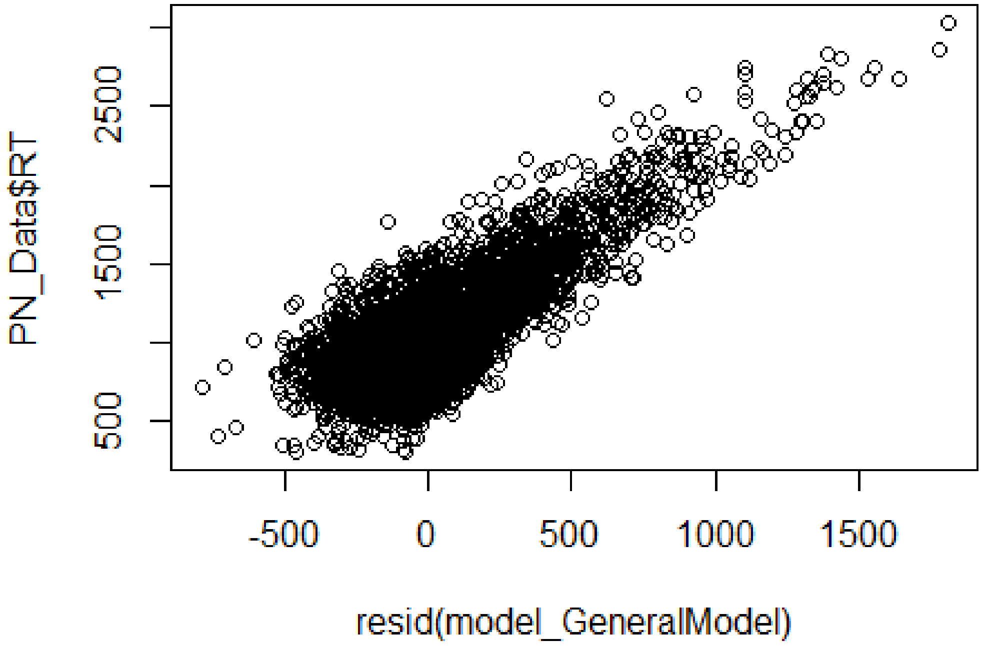
-> but you lose information and you could introduce erroneus effects



Servin-Barthet et al. (2025)

The explanatory variables are related linearly to the response.

```
plot(resid(model_GeneralModel), PN_Data$RT)
```

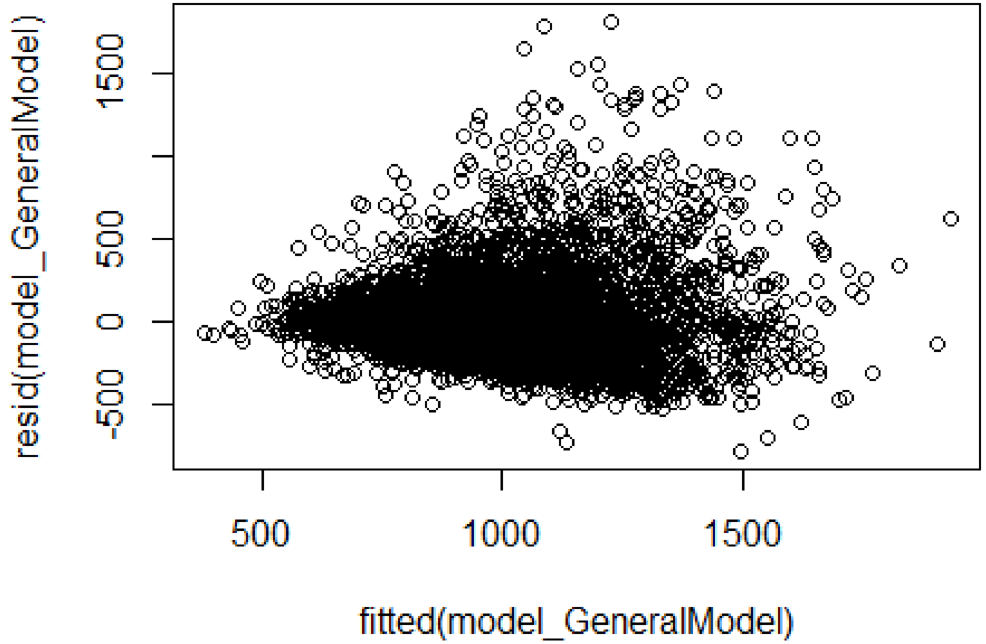


The residuals have constant variance.

$$y_{i\text{Cond}} = \beta_0 + \beta_{Cond} x_{iCond} + \epsilon_{Cond} + u_{0i}$$

Plotting residuals against fitted values will indicate if there is non-constant error variance.

```
plot(fitted(model_GeneralModel), resid(model_GeneralModel))
```



Homoscedasticity

We are looking for a consistent vertical spread

-> variability should stay the same across different levels of the dependent variable

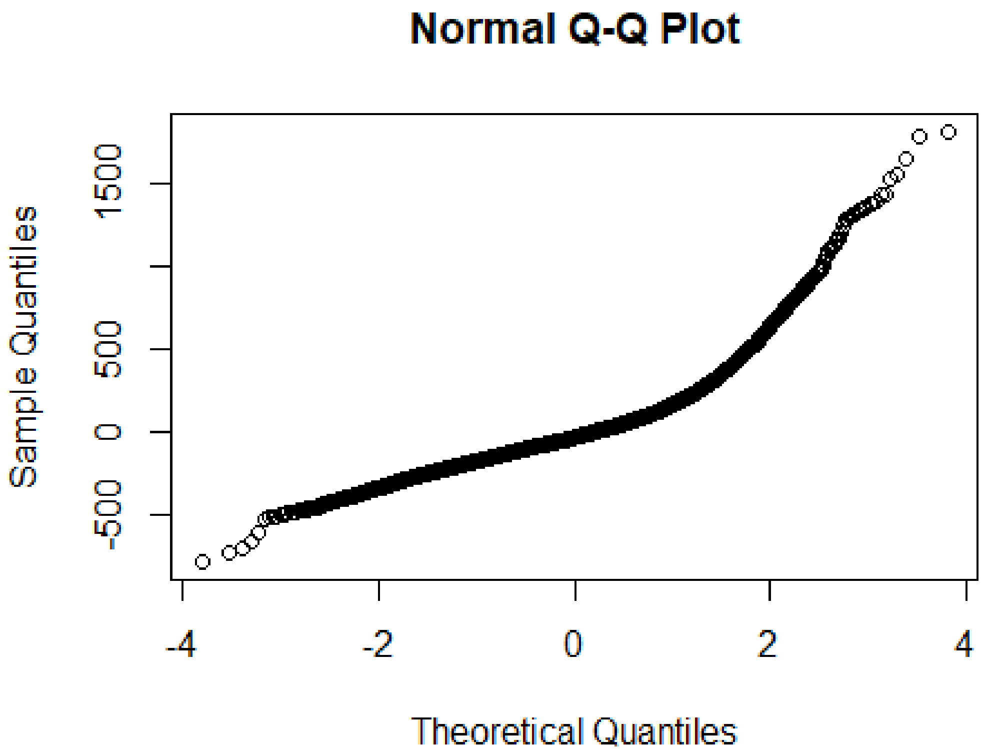
The errors are Normally distributed.

The errors are Normally distributed.

```
res_model <-
residuals(model_GeneralModel)
qqnorm(res_model)
```

If the line curves away from the diagonal, the normality assumption is not met.

The solution might be transformation of the data.



Transforming data

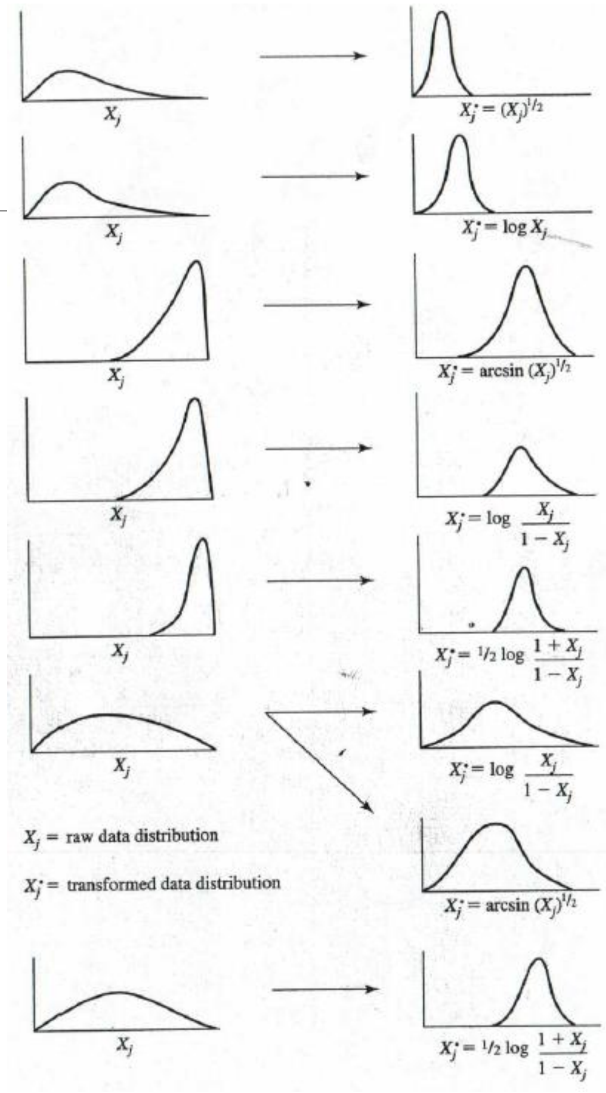
Many experimental results are not normally distributed (RT, questionnaire scales etc.)

Box-Cox procedure can be used to determine best transformation

However, transformation can also lead to bad results and a switch to GLM might be better (Lo and Andrews, 2015)

Transforming data

Data transformation is done by applying some function to the data, which changes the distribution shape, but maintains the overall data distribution



https://www.statisticssolutions.com/transforming-data-for-normality/

Box-Cox package

Box-Cox Transformation

- Find the exponent λ to approximate the normal distribution as close as possible

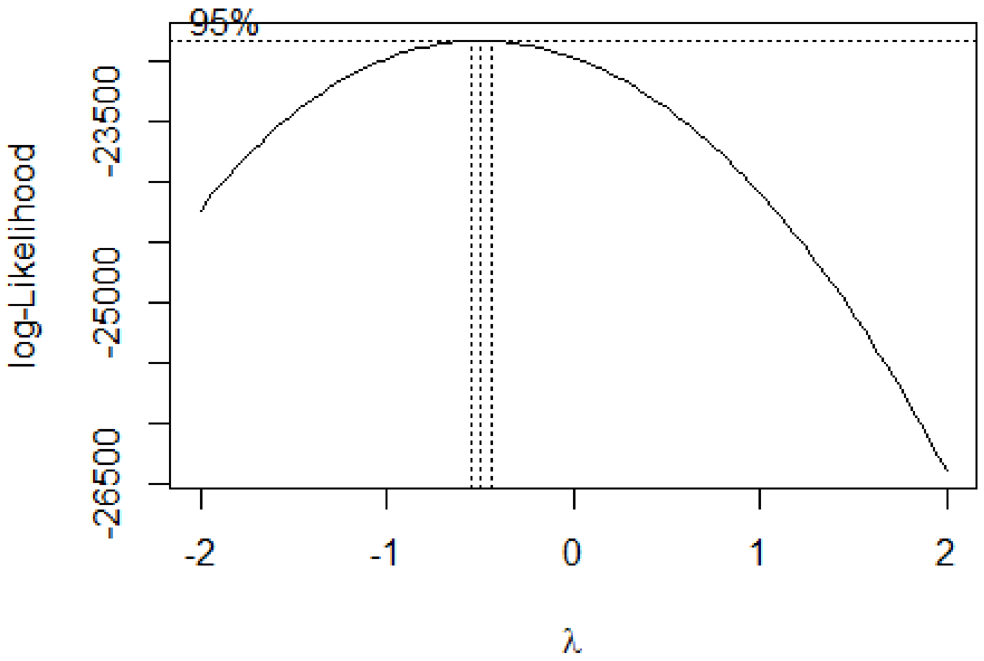
Not easily implemented for LMEs

But in linear regression

Lambda value (λ)	Transformed data (Y’)
-3	Y^-3 = 1/Y3
-2	Y^-2 = 1/Y2
-1	Y^-1 = 1/Y1
-0.5	Y^-0.5 = 1/(√(Y))
0	log(Y)
0.5	Y^0.5 = √(Y)
1	Y^1 = Y
2	Y^2
3	Y^3

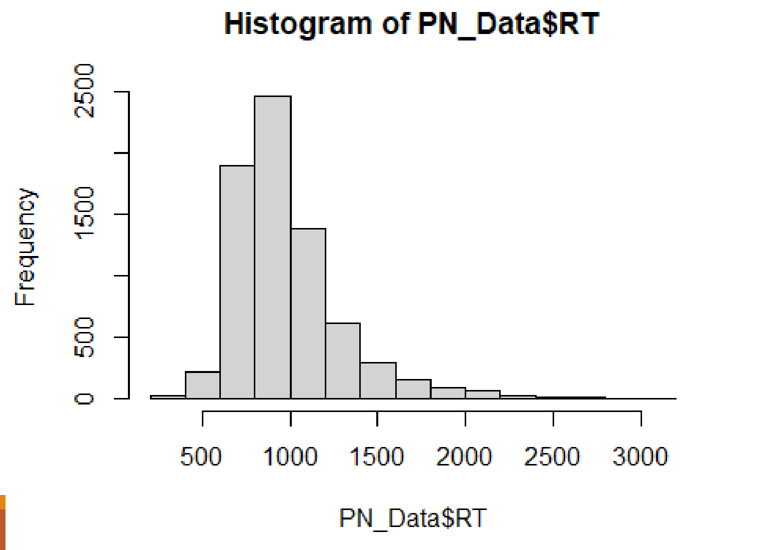
How to use Box-Cox package

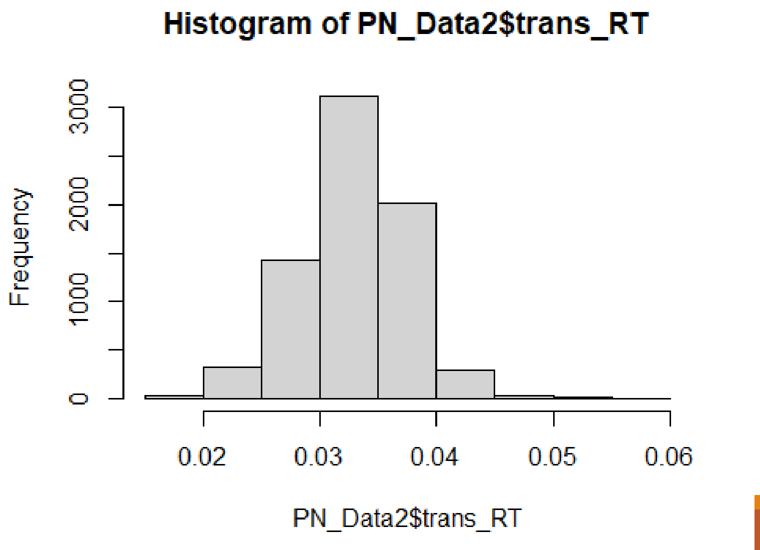
```
library(MASS)

m0 <- lm(RT~1+Context*Group, data
= PN_Data)

boxcox(m0)
```



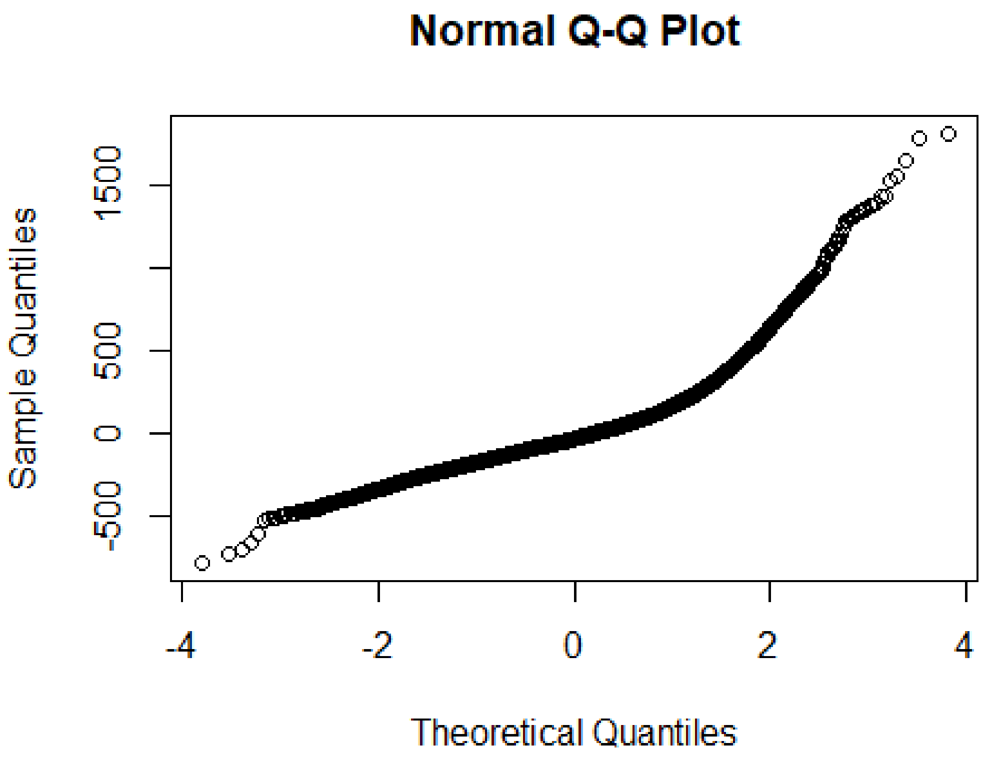
Transforming reaction times

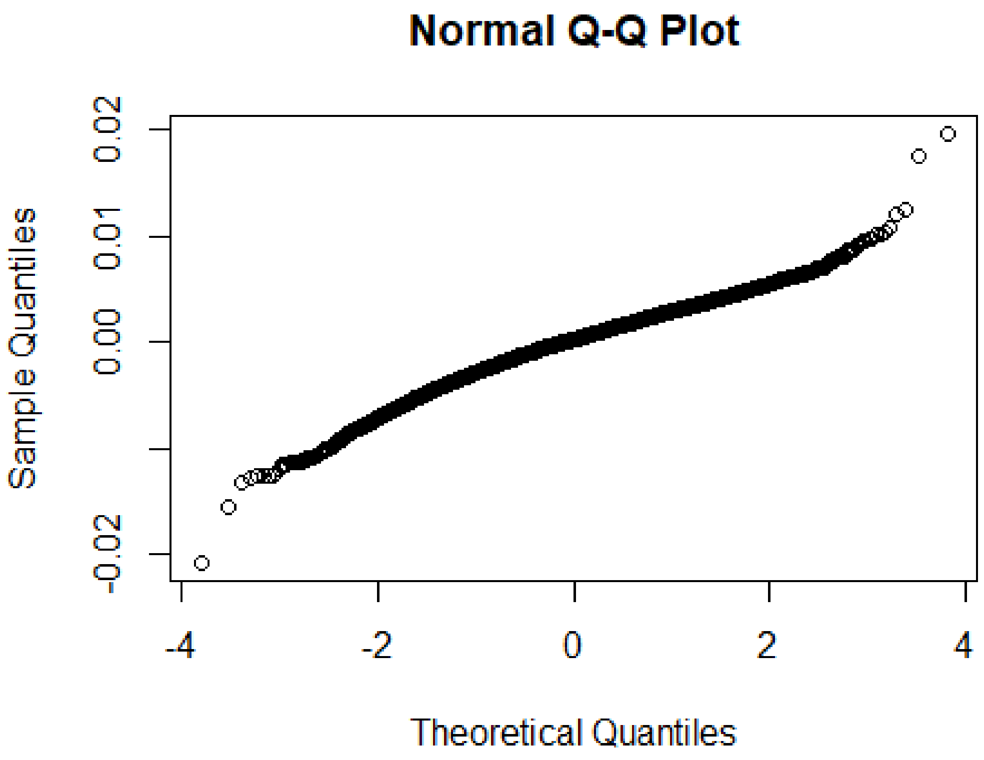
```
hist(PN_Data$RT)

PN_Data2 <-
  PN_Data %>%
  mutate(trans_RT = 1/sqrt(RT))


hist(PN_Data2$trans_RT)
```





Re-Test: normal distribution of residuals





Multicollinearity

Predictors can be highly correlated and cause problems with the model

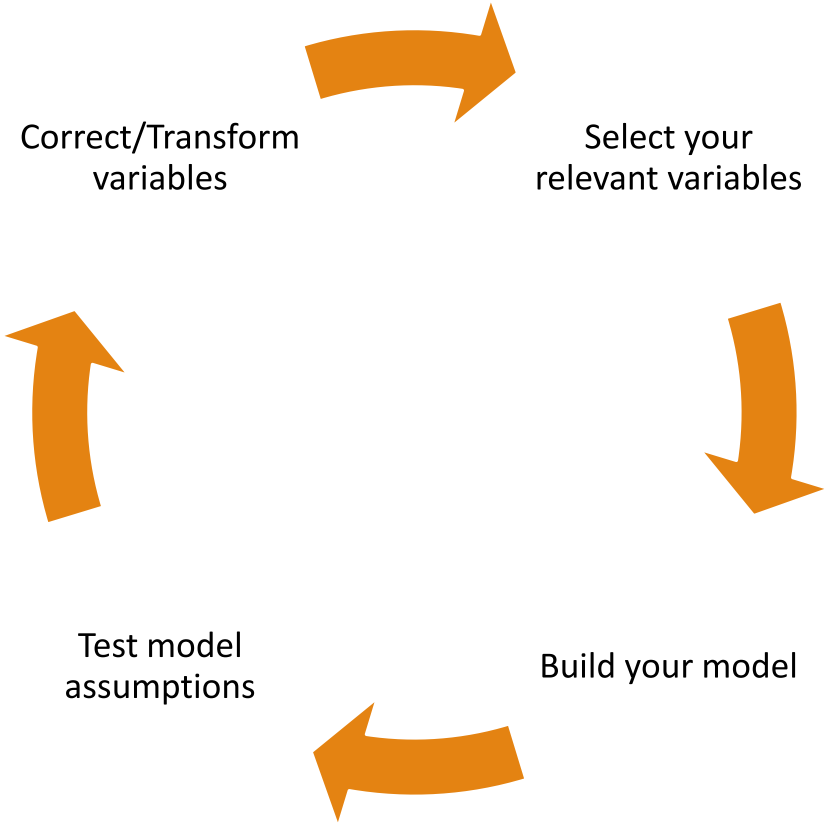
Tested with the Variance Inflation Factor (VIF)

```
library(car)
vif(model_GeneralModel)
##         Group       Context           Age           AoA         Trial
##      1.524975      1.981377      1.274221      1.199493      1.000071
## Group:Context
##      2.218775
```

VIF of 1 indicates no correlation between variables

VIF of 5 or higher might be problematic and coefficient estimates of the model might be highly unreliable

Creating your own models



Example experiment

Bilingual speaker name pictures in a Polish and an English Context

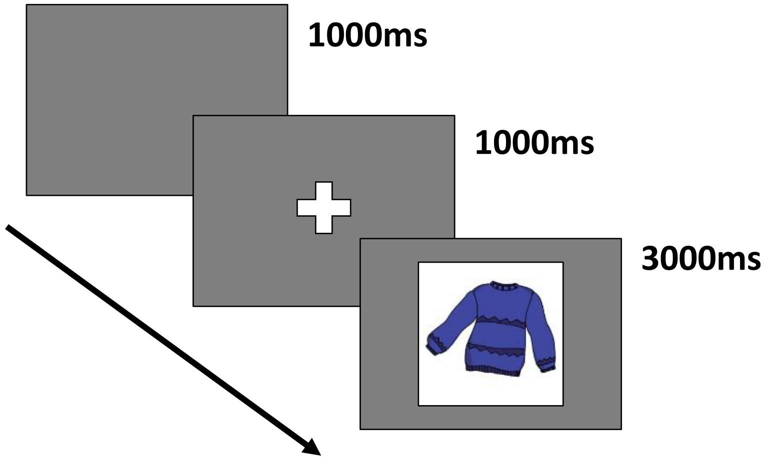
Every speaker saw every picture in both contexts

Experimental variables:

- Language context - UK vs PL
- Group – Experimental vs Control

Grouping factors: Subject and Item

Dependent variable: naming latency



Check the available data

```
head(PN_Data)
## # A tibble: 6 × 9
##   Subject     RT ItemNr Group        Context Trial lg.freq   Age   AoA
##   <chr>    <dbl> <chr>  <chr>        <chr>   <dbl>   <dbl> <dbl> <dbl>
## 1 AS3008    1049 127    Experimental UK          4    4.38    36     7
## 2 AW1912    1007 127    Experimental UK          4    4.38    31    15
## 3 JM2904     794 127    Experimental UK          4    4.38    39    12
## 4 LM1102     826 127    Experimental UK          4    4.38    43    15
## 5 MB0509     842 127    Experimental UK          4    4.38    28    16
## 6 MB2601    1131 127    Experimental UK          4    4.38    54    15
```

Check the available data

Dependent variable: RT

Grouping variables:

- Subject and Item

Experimental manipulations:

- Group
- Context

Background information:

- Age
- Age of L2 acquisition
- Trial
- Word frequency

RT ~ Context + (1 |Subject)

RT ~ Group*Context + (1 | Subject)

RT ~ Group* Context + (1 | Subject) + (1 |Item)

RT ~ Group* Context + Age + AoA + Trial + Freq + (1 | Subject) + (1 |Item)

RT ~ Group* Context + Age + AoA + Trial + Freq + (1 +Context +Trial+Freq| Subject) + (1 + Group + Age + AoA|Item)

Thank you for your attention!