R for DNEP Day Two

Andrew Stewart,
Division of Neuroscience and Experimental Psychology,
School of Biological Sciences,
University of Manchester.

Plan for Today

- Tutorial this morning looking at linear models and (generalised) linear mixed models shortened to (G)LMMs.
- LMMs allow for models with a combination of fixed and random effects (intercepts and slopes).
- Focus on designs of one factor with several levels, and 2 x 2 designs for continuous and dichotomous data.
- Examination of measures of model fit, and using *emmeans* to interpret interactions.
- You doing all of the above in this afternoon's lab.

So far…

- You saw the place of R in the context of open science.
- You learned the basics of programming in R including writing scripts.
- You learned some graphing basics, and how to do a number of statistical tests in R (e.g., correlation, t-tests, AN(C)OVA, regression).
- You learned how to use R Markdown to produce nice looking reports.

Workflow

Data plots using `ggplot()` **and** `pirateplot()`

Tidying and transforming using `dplyr` **and** `tidyr`

Visualise

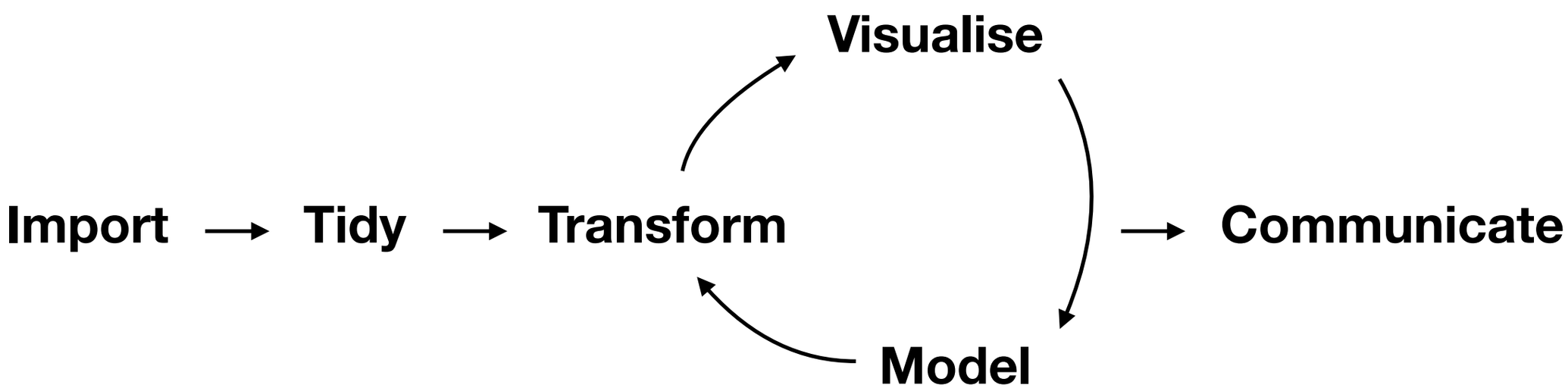
html, pdf and doc reports using `RMarkdown`

Import → **Tidy** → **Transform** → **Communicate**

Model

ANO(C)VA using `afex()` **and** `aov()`
Linear regression using `lm()` **and** `step()`
(Generalised) linear mixed models using `lmer()` **and** `glmer()`

Workflow



(Generalised) linear mixed models using `lmer()` and `glmer()`

Why Linear Mixed Models?

We are going to look at linear modelling and then (generalized) linear mixed modelling. (G)LMMs are taking experimental psychology by storm.

(G)LMMs are more powerful than ANOVA, allow for multiple random effects (typically subjects and items) simultaneously, subject and item covariates, nesting, unbalanced designs, normal and non-normal data distributions, cope with missing data, allow you to model both continuous and categorical IVs and DVs, operate over trial-level data, and allow you to determine the best statistical models to fit to your data that make the most theoretical sense…

Recap - Linear modelling in R

Imagine we have data corresponding to males and females and their height. This is the genderheightdata file.

Subject	Gender	Height
1	Male	180
2	Male	175
3	Male	179
4	Female	170
5	Female	165
6	Female	160

We might be interested in whether height is predicted by gender.

From Winter (2013)

So, we want to know whether Height is predicted by Gender.

Height ~ Gender

So, we fit a linear model like this:

```
> ourmodel <- lm(Height ~ Gender, genderheightdata)
```

The model is stored in the variable *ourmodel*.

```
> ourmodel <- lm(Height ~ Gender, genderheightdata)
> summary(ourmodel)

Call:
lm(formula = Height ~ Gender, data = genderheightdata)

Residuals:
 1  2  3  4  5  6
 2 -3  1  5  0 -5

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  165.000      2.309   71.45  2.3e-07 ***
GenderMale    13.000      3.266    3.98   0.0164 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4 on 4 degrees of freedom
Multiple R-squared:  0.7984,    Adjusted R-squared:  0.748
F-statistic: 15.84 on 1 and 4 DF,  p-value: 0.0164

>
```

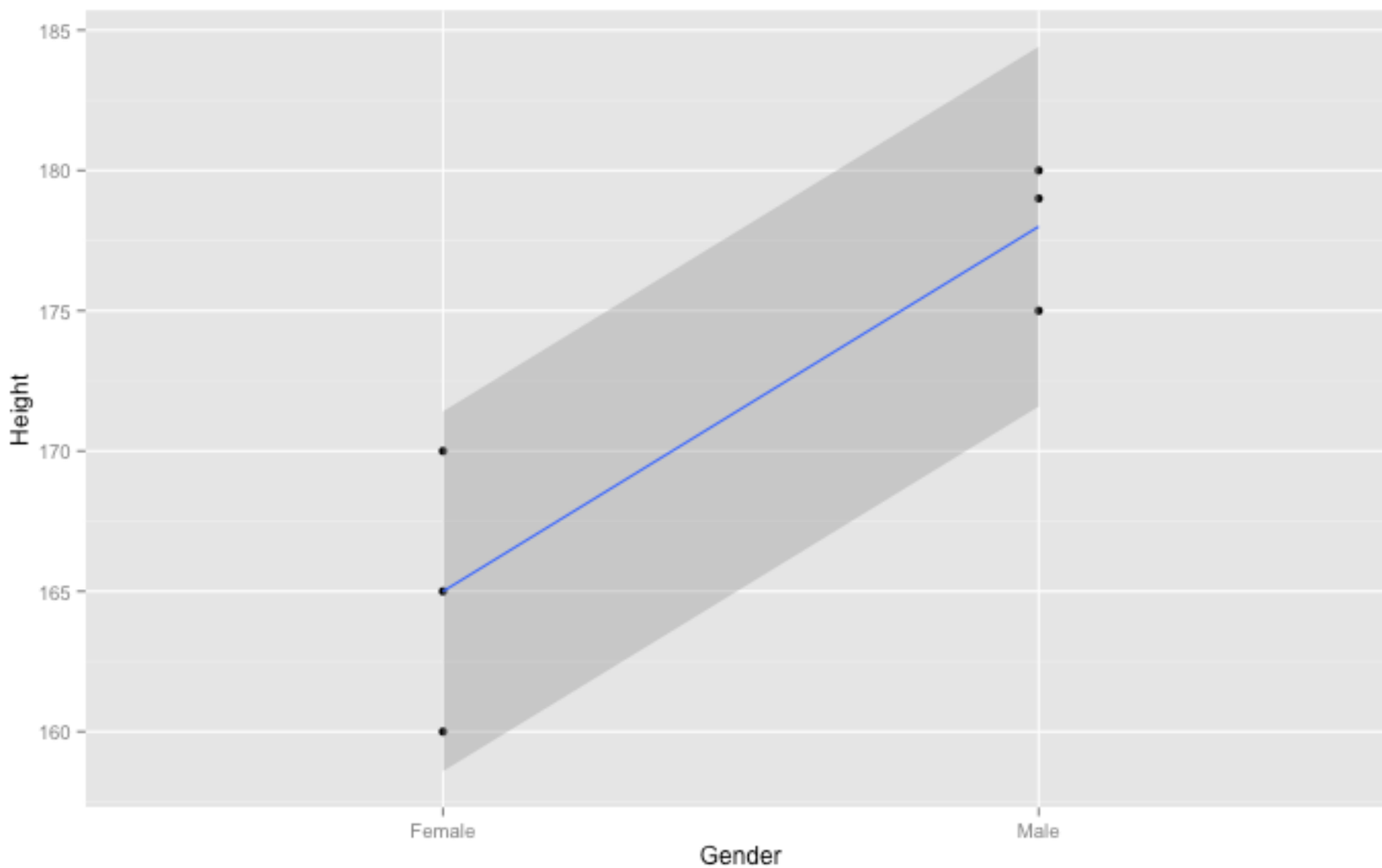
We have a significant predictor (Gender) and model (indicated by the F-ratio).

- For a model with one predictor, the p values associated with the model (i.e., the F value) and the predictor are the same. For models with more than one predictor, this won't be the case.

- The Intercept coefficient (165) corresponds to the mean Height of our reference category (Female). The estimate GenderMale (13) is the *difference* between our reference category and our Males. Females were taken as the reference category (i.e., the intercept) simply because R chooses this on an alphabetical basis (and *Female* comes before *Male*).

We can use ggplot to graph our data. Using the "lm" method, we can generate the linear model (or regression) line.

```
> scatter <- ggplot (genderheightdata, aes (group = 1, Gender, Height)) + geom_point()
> scatter + geom_smooth (method = "lm")
>
```



- Our predictor doesn't have to be categorical though. We're using the ageheightdata file here.

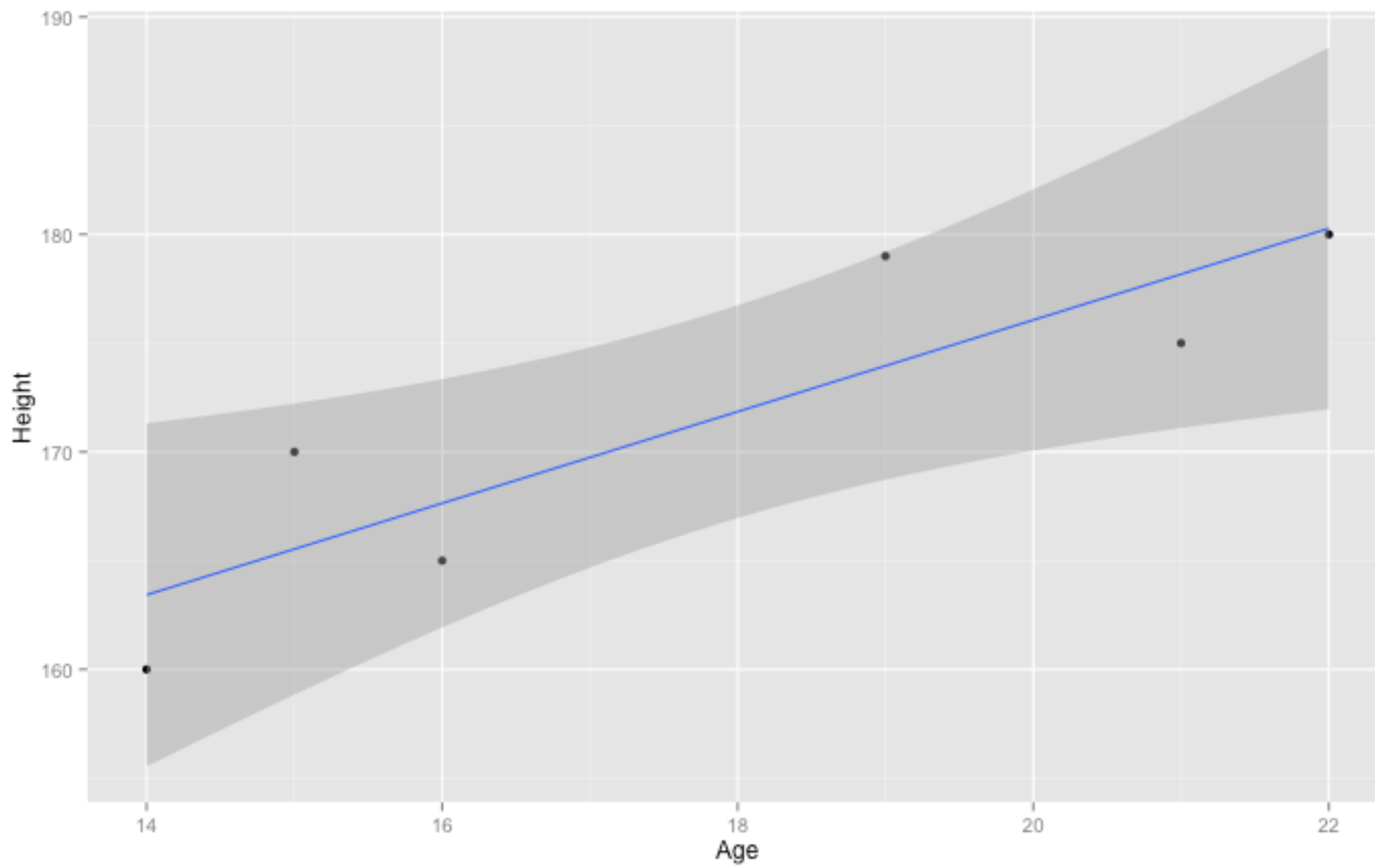
Subject	Age	Height
1	22	180
2	21	175
3	19	179
4	15	170
5	16	165
6	14	160

Is Height predicted by Age?

```
> ourmodel <- lm(Height ~ Age, ageheightdata)
> summary (ourmodel)

Call:
lm(formula = Height ~ Age, data = ageheightdata)

Residuals:
      1       2       3       4       5       6 
-0.2766 -3.1702  5.0426  4.4681 -2.6383 -3.4255 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 133.9362    10.5214  12.730 0.000219 ***
Age           2.1064     0.5817   3.621 0.022334 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.307 on 4 degrees of freedom
Multiple R-squared:  0.7663,	Adjusted R-squared:  0.7078 
F-statistic: 13.11 on 1 and 4 DF,  p-value: 0.02233

>
```

For every increase in Age by 1, Height increases by 2.1064. But of course, we know this relationship breaks down at a certain age - but for the data we have, we can fit a linear function.

```
> scatter <- ggplot (ageheightdata, aes (Age, Height)) + geom_point()
> scatter + geom_smooth (method = "lm")
>
```



Linear Mixed Models

What happens when we have multiple predictors?

Imagine we are interested in how a person's Gender predicts the Pitch of their voice. Let's imagine we're interested in how Polite they're trying to be *also* predicts the Pitch of their voice.

We observe the same 6 people in different situations (some requiring them to be extra polite while talking, others not.)

These observations are not independent of each other (which is an assumption of a linear model).

From Winter (2013)

- We can get around the lack of independence by treating participants as a random effect such that each participant has their own *individual* pitch baseline.

- This gives us a separate random intercept value for each participant - in other words, our model can account for individual variation.

- This is a *mixed effects linear model*:

pitch ~ politeness + gender + (1| subject) + error

This is our random effect and assumes a different intercept for each participant.

- Imagine also that we had different Politeness scenarios (e.g., 7 different scenarios that required participants to be polite, and 7 different scenarios that require them to be informal.)
- Each scenario in each category might have been a little different. One particular scenario might have had a higher Pitch mean than a different scenario - in other words, the scenarios will also have different baselines.

- We can capture the random effect of Item (called *scenario* in this dataset) in the same way we did for participants:

```
pitch ~ politeness + gender + (1|subject) + (1|scenario) + error
```

subject	gender	scenario	attitude	frequency
F1	F	1	pol	213.3
F1	F	1	inf	204.5
F1	F	2	pol	285.1
F1	F	2	inf	259.7
F1	F	3	pol	203.9
F1	F	3	inf	286.9
F1	F	4	pol	250.8
F1	F	4	inf	276.8
F1	F	5	pol	231.9
F1	F	5	inf	252.4
F1	F	6	pol	181.2

6 participants, and 7 items (scenarios). Each scenario appeared in two version - polite vs informal.

- We're using the politenessdata file here.

Note - for linear mixed models, data must be in Long format - every row is one time point per participant. This is really important! The following is from a 2 x 2 experiment.

Item no.

Participant no.

Our DV

Factor 1

Factor 2

	Subject	Item	RT	Sentence	Context
1	1	3	1270	Positive	Negative
2	1	7	739	Positive	Negative
3	1	11	982	Positive	Negative
4	1	15	1291	Positive	Negative
5	1	19	1734	Positive	Negative
6	1	23	1757	Positive	Negative
7	1	27	1052	Positive	Negative
8	2	4	1706	Positive	Negative
9	2	8	533	Positive	Negative
10	2	12	1009	Positive	Negative

Fixed vs. Random Effects

<u>Fixed effect</u> Data has been gathered from all the levels of the factor that are of interest. (Typically your experimental factors and maybe factors like gender).

<u>Random effect</u> The factor has many possible levels, interest is in all possible levels, but only a random sample of levels is included in the data. (Typically participants and items). Typically need > 5 levels.

For mixed effects linear modelling in R, we need to install the package *lme4*. This is the mixed effects model equivalent of *lm* which we used previously. We also want the *lmerTest* package and the *emmeans* package.

```
> install.packages (“lme4")

> install.packages (“lmerTest”)

> install.packages (“emmeans")
```

Gives us p-values for our model estimates.

Allows us to do pairwise comparisons.

Remember then to load them:

```
> library (lme4)

> library (lmerTest)

> library (emmeans)
```

```
> politeness.model <- lmer (frequency~attitude + (1|subject) + (1|scenario), data=politeness_data)
> summary (politeness.model)
Linear mixed model fit by REML ['lmerMod']
Formula: frequency ~ attitude + (1 | subject) + (1 | scenario)
   Data: politeness_data

REML criterion at convergence: 793.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2006 -0.5817 -0.0639  0.5625  3.4385

Random effects:
 Groups   Name        Variance Std.Dev.
 scenario (Intercept)  219     14.80
 subject  (Intercept) 4015     63.36
 Residual              646     25.42
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  202.588     26.754   7.572
attitudepol  -19.695      5.585  -3.527

Correlation of Fixed Effects:
            (Intr)
attitudepol -0.103
>
```

More variability in subjects than in scenarios.

Going from Informal to Polite contexts, people's voice pitch drops about 19.7Hz.

- So far we haven't accounted for Gender. We can add Gender as a fixed effect (because we know that Males have lower pitched voices than Females).

Our new fixed effect

```
> politeness.model <- lmer (frequency~attitude + gender + (1|subject) + (1|scenario), data=politeness_data)
> summary (politeness.model)
```

```
> politeness.model <- lmer (frequency~attitude + gender + (1|subject) + (1|scenario), data=politeness_data)
> summary (politeness.model)
Linear mixed model fit by REML ['lmerMod']
Formula: frequency ~ attitude + gender + (1 | subject) + (1 | scenario)
   Data: politeness_data

REML criterion at convergence: 775.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2591 -0.6236 -0.0772  0.5388  3.4795

Random effects:
 Groups   Name        Variance Std.Dev.
 scenario (Intercept) 219.5    14.81
 subject  (Intercept) 615.6    24.81
 Residual             645.9    25.41
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  256.846     16.116  15.938
attitudepol  -19.721      5.584  -3.532
genderM     -108.516     21.013  -5.164

Correlation of Fixed Effects:
            (Intr) atttdp
attitudepol -0.173
genderM     -0.652  0.004
>
```

Our variation due to subjects has dropped - we now explain quite a lot in terms of our factor Gender.

We have an effect of Situation, and also an effect of Gender. People have lower pitched voices in Polite Situations, and Males have lower pitched voices than females.

- To determine whether our mixed effects model is significant, we need to know whether it differs from what we'd expect if Politeness didn't influence voice pitch.

```
> politeness.null <- lmer (frequency~gender + (1|subject) + (1|scenario), data=politeness_data, REML=FALSE)
```

- This model which we call politeness.null removes our Attitude effect as a predictor (and includes a new term REML=FALSE which is needed to estimate the likelihood of the null model relative to our experimental model).

We re-create our politeness.model which includes our Attitude effect as a predictor (plus the likelihood estimator).

```
> politeness.model <- lmer (frequency~attitude + gender + (1|subject) + (1|scenario), data=politeness_data, REML=FALSE)
```

We can now compare the two models with each other using the anova function:

```
> anova (politeness.null, politeness.model)
```

This performs a likelihood ratio test on our 2 models and tells us whether they are significantly different from each other - this test only works with **nested** models.

```
> anova (politeness.null, politeness.model)
Data: politeness_data
Models:
politeness.null: frequency ~ gender + (1 | subject) + (1 | scenario)
politeness.model: frequency ~ attitude + gender + (1 | subject) + (1 | scenario)
                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
politeness.null   5 816.72 828.81 -403.36   806.72
politeness.model  6 807.10 821.61 -397.55   795.10 11.618      1  0.0006532 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

This is the important bit as the chi-squared test tells us whether our politeness model differs from the null. It does. We could write "...being polite lowers pitch ($\chi 2$ (1)=11.62, $p < .001$) by 19.7 Hz..."

Note, deviance equals the residual sum of squares in linear models.

- We can also compare the model without gender to the model with gender:

```
politeness.nogender <- lmer (frequency ~ attitude + (1|subject) + (1|scenario), data=politeness_data)
politeness.withgender <- lmer (frequency ~ attitude + gender + (1|subject) + (1|scenario), data=politeness_data)
anova (politeness.nogender, politeness.withgender)
```

```
> anova (politeness.nogender, politeness.withgender)
refitting model(s) with ML (instead of REML)
Data: politeness_data
Models:
object: frequency ~ attitude + (1 | subject) + (1 | scenario)
..1: frequency ~ attitude + gender + (1 | subject) + (1 | scenario)
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
object  5 817.04 829.13 -403.52   807.04
..1     6 807.10 821.61 -397.55   795.10 11.938      1    0.00055 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

The difference between the models is significant.

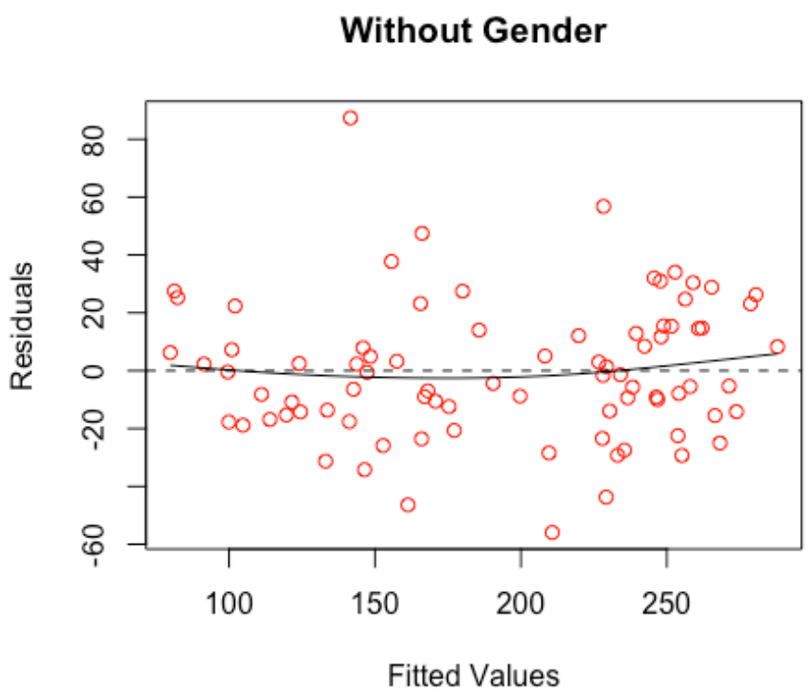
The second model has the lower AIC value (so is a better fit).

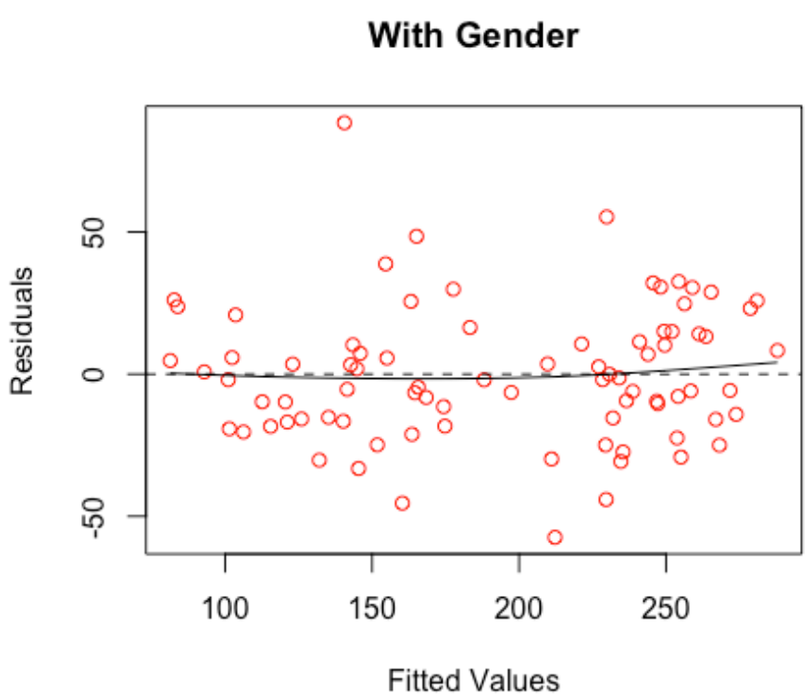
We can visualise the fit of the two models by plotting the residuals against the fitted data using:

```
#plot for the model without gender
plot(fitted(politeness.nogender), residuals(politeness.nogender),
       xlab = "Fitted Values", ylab = "Residuals" , main = "Without Gender",col="red")
abline(h = 0, lty = 2)
lines(smooth.spline(fitted(politeness.nogender), residuals(politeness.nogender)))

#plot for the model with gender
plot(fitted(politeness.withgender), residuals(politeness.withgender),
       xlab = "Fitted Values", ylab = "Residuals" , main = "With Gender", col="red")
abline(h = 0, lty = 2)
lines(smooth.spline(fitted(politeness.withgender), residuals(politeness.withgender)))
```

This adds a straight, dashed, line at y=0

This adds a line connecting the residuals and fitted data using a cubic smoothing spline.





Better models have better overlap between the solid line, and the dashed line. This is where there is minimisation in error between a model and the fitted data. You can see the model 'With Gender' fits the data better than 'Without Gender' as the solid line overlaps more with the dashed line. Consistent with what our AIC values tell us.

- So far we have accounted for the possibility that our participants and items might have different pitch baselines (which is why we introduced the separate random intercepts). But what if the effect of politeness is different for different participants, and also what if the effect of politeness is different for different scenarios?

- All this means is that the slopes of our lines might vary as a function of participant (so the difference between the two levels of politeness might be bigger for one person than for another) and as a function of scenario (so the difference between the two levels of politeness might also be bigger for one scenario than for another).

```
> coef (politeness.model)
$scenario
  (Intercept) attitudepol   genderM
1    243.4859   -19.72207 -108.5173
2    263.3592   -19.72207 -108.5173
3    268.1322   -19.72207 -108.5173
4    277.2546   -19.72207 -108.5173
5    254.9319   -19.72207 -108.5173
6    244.8015   -19.72207 -108.5173
7    245.9618   -19.72207 -108.5173

$subject
   (Intercept) attitudepol   genderM
F1    243.3684   -19.72207 -108.5173
F2    266.9443   -19.72207 -108.5173
F3    260.2276   -19.72207 -108.5173
M3    284.3536   -19.72207 -108.5173
M4    262.0575   -19.72207 -108.5173
M7    224.1292   -19.72207 -108.5173
```

The different intercepts for each item and for each participant take into account individual baseline differences. However, it doesn't take into account the fact our effect might be bigger for some participants than for others (and for some items than for others). In other words, the slopes are all currently the same.

```
> politeness.model <- lmer (frequency~attitude + gender + (1+attitude|subject) + (1+attitude|scenario), data=politeness_data, REML=FALSE)
```

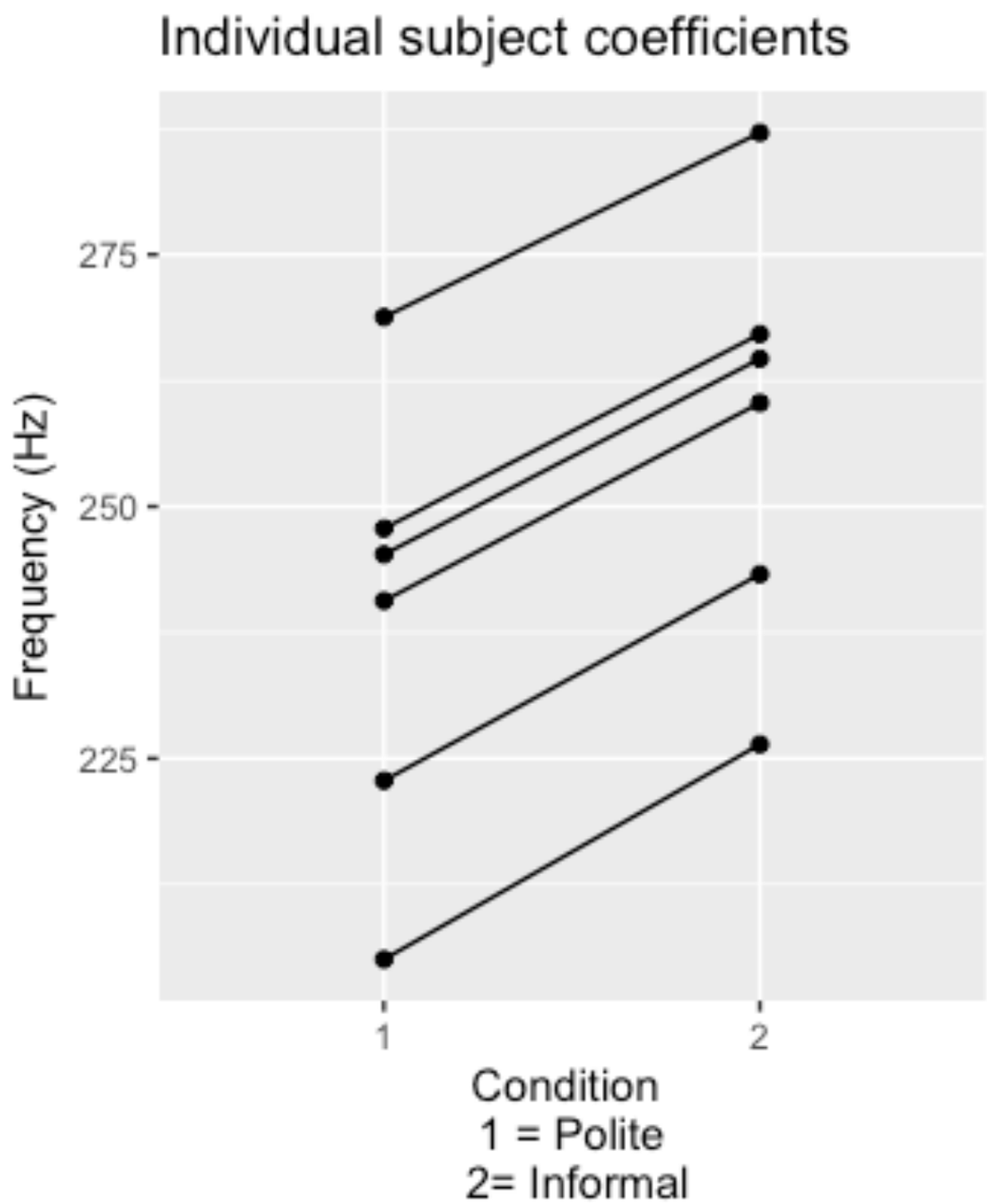
These modified terms tell the model to expect different intercepts for Frequency (which we had before) as well as differing slopes as a function of the factor Attitude. These are our random effects.

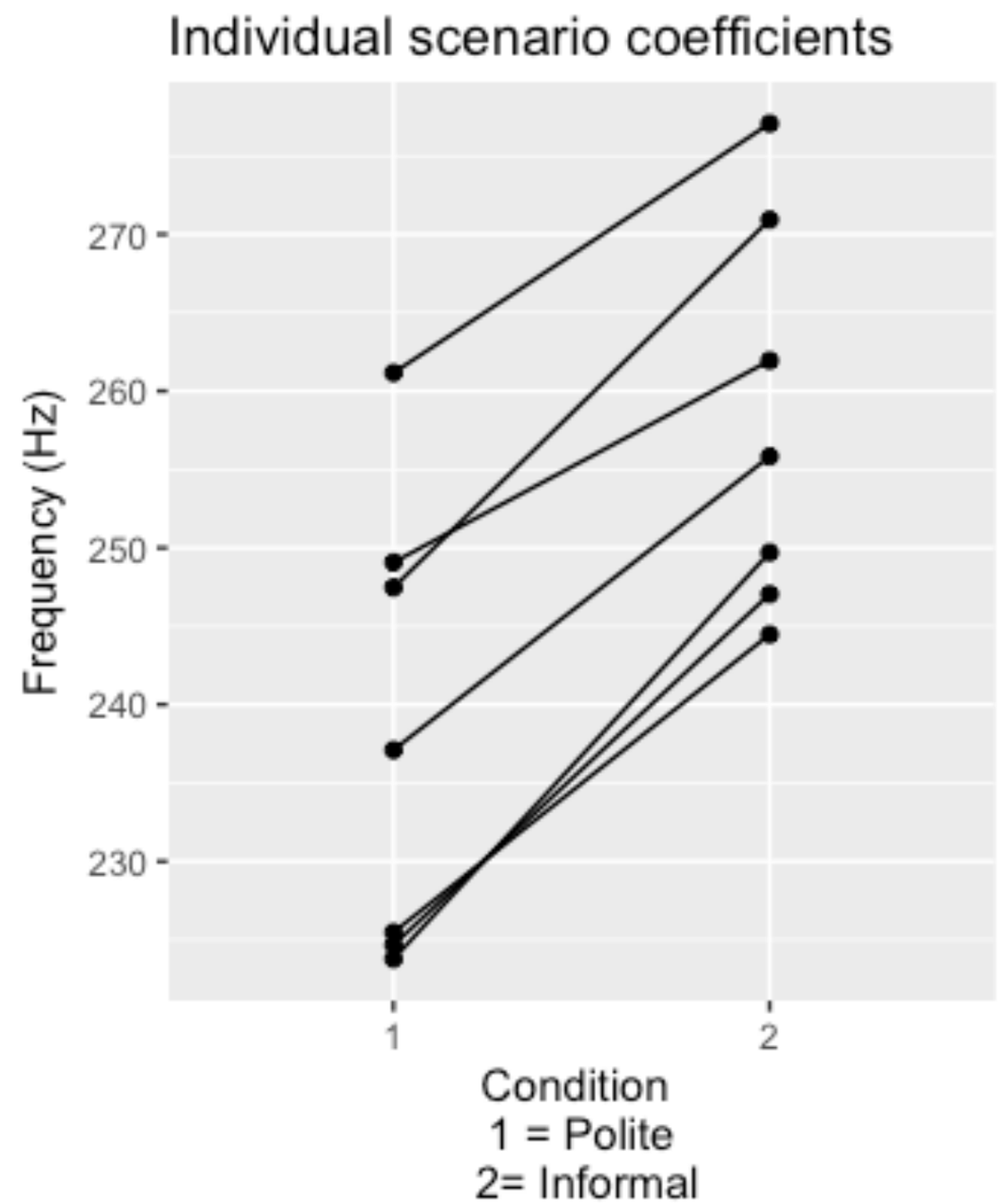
```
> coef (politeness.model)
$scenario
  (Intercept) attitudepol   genderM
1    245.2603   -20.43832 -110.8021
2    263.3012   -15.94386 -110.8021
3    269.1432   -20.63361 -110.8021
4    276.8309   -16.30132 -110.8021
5    256.0579   -19.40575 -110.8021
6    246.8605   -21.94816 -110.8021
7    248.4702   -23.55752 -110.8021

$subject
   (Intercept) attitudepol   genderM
F1    243.8053   -20.68245 -110.8021
F2    266.7321   -19.17028 -110.8021
F3    260.1484   -19.60452 -110.8021
M3    285.6958   -17.91951 -110.8021
M4    264.1982   -19.33741 -110.8021
M7    227.3551   -21.76744 -110.8021
```

The slopes between our two Attitude conditions differ for each item...

...and for each participant.

Plotting the slopes of our Politeness factor





- Again, we can compare our new model with the null (which needs to include the same random effects).

```
> politeness.null <- lmer (frequency~gender + (1+attitude|subject) + (1+attitude|scenario), data=politeness_data, REML=FALSE)
> anova (politeness.null, politeness.model)
Data: politeness_data
Models:
politeness.null: frequency ~ gender + (1 + attitude | subject) + (1 + attitude |
politeness.null:     scenario)
politeness.model: frequency ~ attitude + gender + (1 + attitude | subject) + (1 +
politeness.model:     attitude | scenario)
                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
politeness.null   9 819.61 841.37 -400.80   801.61
politeness.model 10 814.90 839.09 -397.45   794.90 6.7082      1   0.009597 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Our model is significant.

Examples of LMMs for Factorial Designs

- In the first case, we will look at a model where we have one factor with three levels. We have two sets of data we want to analyse - one is eye gaze duration data, the other is the number of times people re-read a section of text.
- In the second case, we will look at a model for a 2 x 2 repeated measures design - this time just with eye gaze duration data as people read a section of text.

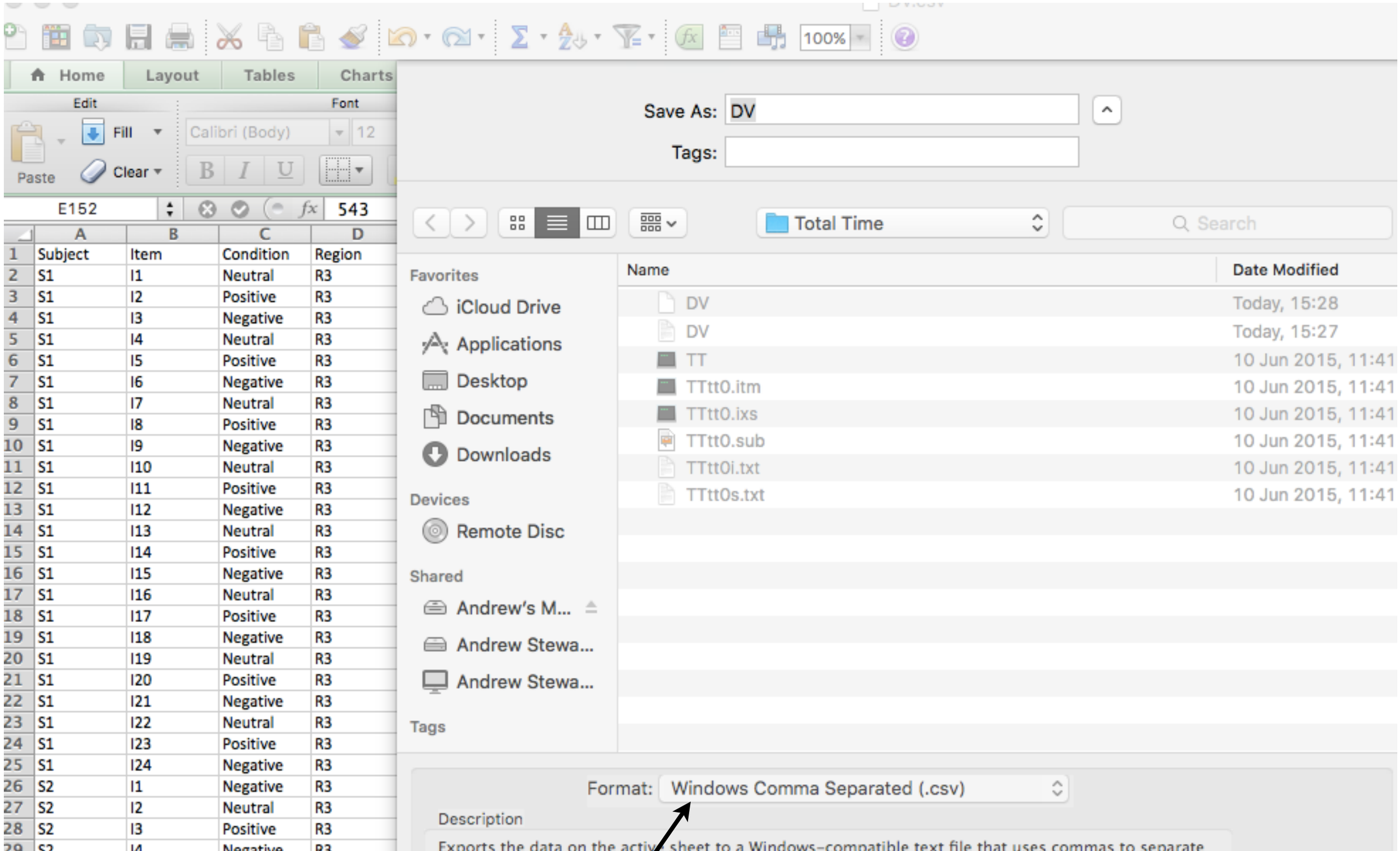
One factor with Three levels

- We are going to analyse eye movement data associated with reading a segment of text in one of three conditions - Positive, Negative, or Neutral.

```
#install the lme4, lsmeans and lmertest packages first
install.packages ("lme4")
install.packages ("lmerTest")
install.packages ("emmeans")
library (lme4)
library (lmerTest)
library(emmeans)
```

```
#C1 = Neutral condition
#C2 = Negative condition
#C3 = Positive condition
```

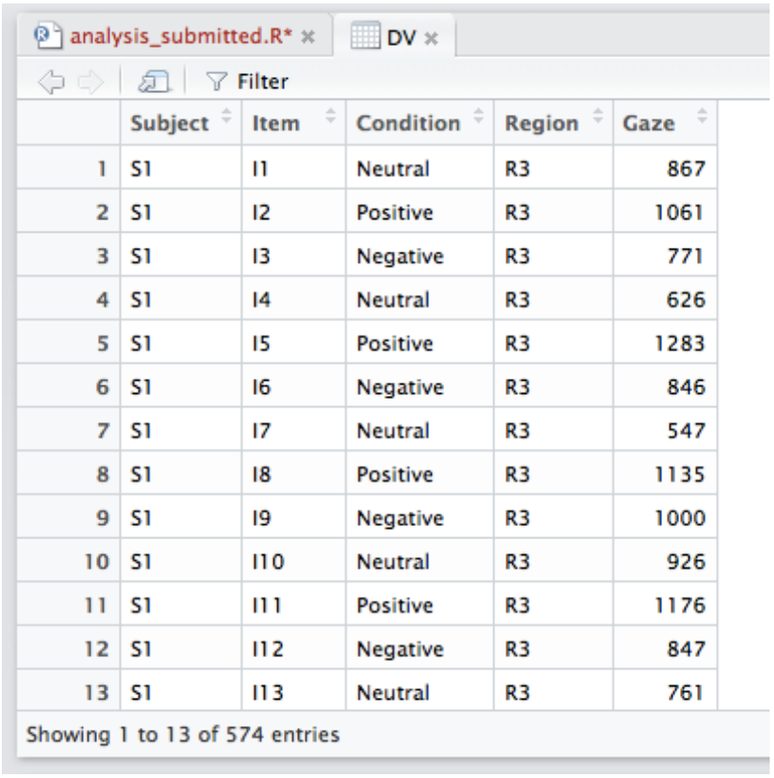
The *lmerTest* package gives us *p*-values for our fixed effects, while the *emmeans* allows us to conduct pairwise comparisons.



	A	B	C	D
1	Subject	Item	Condition	Region
2	S1	I1	Neutral	R3
3	S1	I2	Positive	R3
4	S1	I3	Negative	R3
5	S1	I4	Neutral	R3
6	S1	I5	Positive	R3
7	S1	I6	Negative	R3
8	S1	I7	Neutral	R3
9	S1	I8	Positive	R3
10	S1	I9	Negative	R3
11	S1	I10	Neutral	R3
12	S1	I11	Positive	R3
13	S1	I12	Negative	R3
14	S1	I13	Neutral	R3
15	S1	I14	Positive	R3
16	S1	I15	Negative	R3
17	S1	I16	Neutral	R3
18	S1	I17	Positive	R3
19	S1	I18	Negative	R3
20	S1	I19	Neutral	R3
21	S1	I20	Positive	R3
22	S1	I21	Negative	R3
23	S1	I22	Neutral	R3
24	S1	I23	Positive	R3
25	S1	I24	Negative	R3
26	S2	I1	Negative	R3
27	S2	I2	Neutral	R3
28	S2	I3	Positive	R3
29	S2	I4	Negative	R3

- First I need to re-save my data in Excel as a .csv file

- Our data file is called DV and looks like this:



analysis_submitted.R* × DV ×

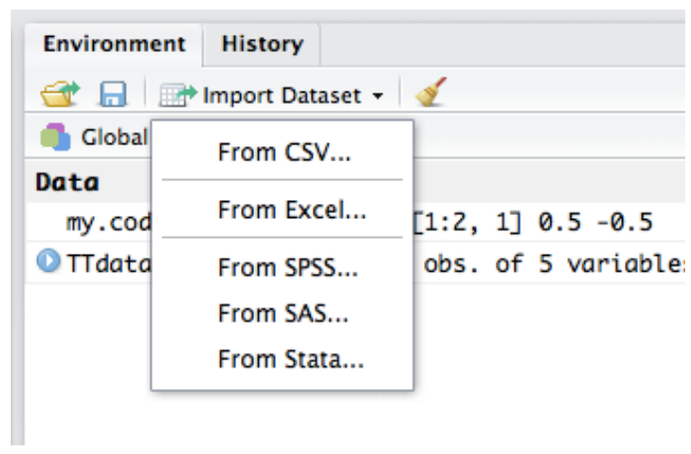
Filter

	Subject	Item	Condition	Region	Gaze
1	S1	I1	Neutral	R3	867
2	S1	I2	Positive	R3	1061
3	S1	I3	Negative	R3	771
4	S1	I4	Neutral	R3	626
5	S1	I5	Positive	R3	1283
6	S1	I6	Negative	R3	846
7	S1	I7	Neutral	R3	547
8	S1	I8	Positive	R3	1135
9	S1	I9	Negative	R3	1000
10	S1	I10	Neutral	R3	926
11	S1	I11	Positive	R3	1176
12	S1	I12	Negative	R3	847
13	S1	I13	Neutral	R3	761

Showing 1 to 13 of 574 entries

- The columns correspond to our Subject Number, our Item Number, our Condition, the Region of Text and the Gaze time (ms.)

- You then need to import the data file:



- Make sure you check that R correctly recognises your factors. In this case, it initially doesn't:

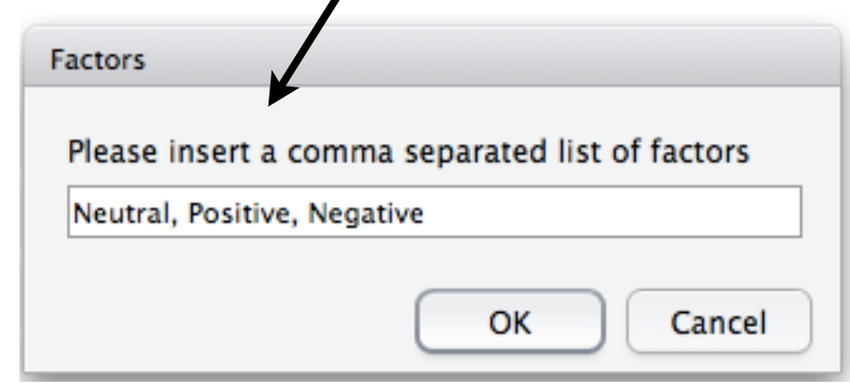
Data Preview:

Subject (character)	Item (character)	Condition (character)	Region (character)	Gaze (integer)
S1	I1	Neutral	R3	867
S1	I2	Positive	R3	1061
S1	I3	Negative	R3	771
S1	I4	Neutral	R3	626
S1	I5	Positive	R3	1283
S1	I6	Negative	R3	846
S1	I7	Neutral	R3	547
S1	I8	Positive	R3	1135
S1	I9	Negative	R3	1000
S1	I10	Neutral	R3	926
S1	I11	Positive	R3	1176
S1	I12	Negative	R3	847
S1	I13	Neutral	R3	761

Previewing first 50 entries.

The names of the columns you will use in your model (incl. the names of the random effects).

- For Condition, you need to select it as a Factor (not as a character string). Click on the down arrow, and then select Factor. Enter the levels separated by commas.



You can also use the function *as.factor* to turn your variable into a factor:

```
> DV$Condition <- as.factor(DV$Condition)
```

You can type the following to check the number of levels of the factor:

```
> levels (DV$Condition)
```

```
25 model.null <- lmer (Gaze ~ (1 + Condition| Subject) + (1 + Condition| Item), data=DV, REML=TRUE)
26 model.full <- lmer (Gaze ~ Condition + (1 + Condition| Subject) + (1 + Condition| Item), data=DV, REML=TRUE)
27 anova (model.null, model.full)
28 summary (model.full)
```

- Line 25 creates a variable called model.null associated with just random effects of Subjects and Items. Note there is no fixed effect.
- Line 26 create a variable called model.full which includes both the random and fixed effects.
- Line 27 tests where the model.full is a better fit to our data and model.null. If it is, it means adding the fixed effect means we are able to explain our data better than if we don't add it.
- Line 28 then asks for the model.full parameters to be displayed.

The Output

```
> anova(model.null, model.full)
refitting model(s) with ML (instead of REML)
Data: DV
Models:
object: Gaze ~ (1 + Condition | Subject) + (1 + Condition | Item)
..1: Gaze ~ Condition + (1 + Condition | Subject) + (1 + Condition |
..1:     Item)
       Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
object 14 8773.4 8834.4 -4372.7   8745.4
..1    16 8771.2 8840.8 -4369.6   8739.2 6.236      2    0.04425 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

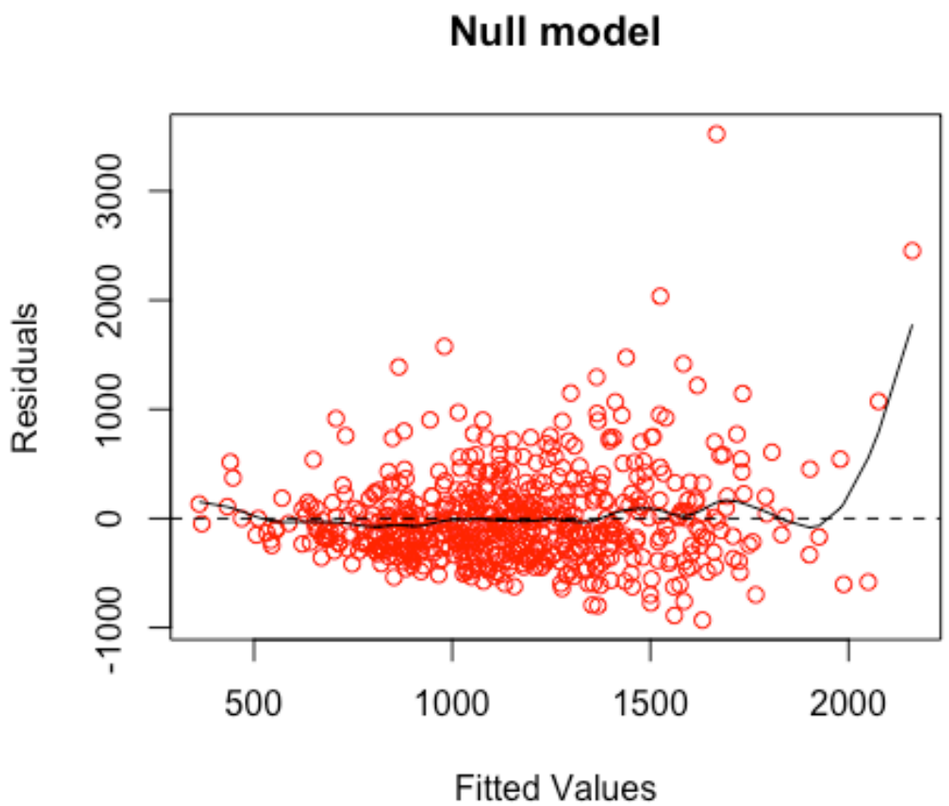
For model comparisons, a different parameter estimator must be used (R will do this for you). REML should be used to estimate parameters when you report them.

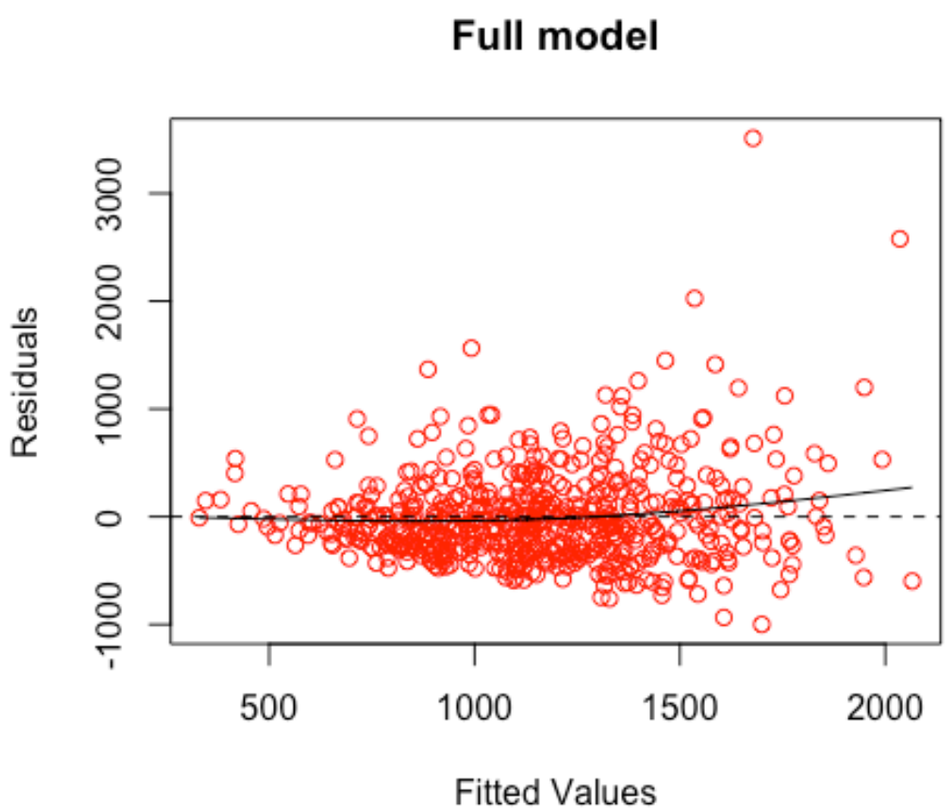
- Our two models differ significantly from each other. The one that fits our data the best has the lower AIC value. AIC is the Akaike Information Criterion and measures how much ‘information’ is not captured by our model (values that are relatively lower are better). NOTE - absolute AIC values cannot be interpreted - they have to be compared with the AIC value of another model.

```
Random effects:
 Groups   Name              Variance Std.Dev. Corr
 Subject  (Intercept)       108205   328.95
          ConditionNeutral    2589    50.88   -1.00
          ConditionPositive   6425    80.16   -1.00  1.00
 Item     (Intercept)        32985   181.62
          ConditionNeutral    1296    36.00    0.00
          ConditionPositive   3897    62.42   -0.54  0.84
 Residual                    204916   452.68
Number of obs: 574, groups:  Subject, 24; Item, 24

Fixed effects:
                 Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       1083.76      83.40   30.15  12.994 6.88e-14 ***
ConditionNeutral   101.04      48.05   52.01   2.103   0.0403 *
ConditionPositive  123.54      50.70   22.73   2.437   0.0231 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- This is what we're mainly interested in. We know the model itself is significantly better than the null model. These comparisons tells us what differences are driving the effect.

- Think of these like the contrasts that are used to interpret significant ANOVAs. In this case, the Neutral and Positive conditions are each being compared to the Negative condition (or the intercept of the regression line). The estimates tell us that the intercept is 1084 (which is the Negative condition mean). The Neutral mean is 1084+101, while the Positive mean is 1084+124.





Again we can examine model fit by plotting the residuals against the fitted values. The model with the fixed effect (the full model) is a better fit than the model which includes only the random effects.

A few points to note so far…

- Models can only be compared to each other using the ANOVA function if they are nested - in other words, if one model is a subset of the other. Models with different fixed and random effects structures cannot be compared in this way - use AIC or BIC comparisons.

- If using treatment coding for Contrasts, sometimes the Intercept (or reference level condition) chosen by R isn't the one you might want. You can change it using: `DV$Condition <- relevel (DV$Condition, ref = 3)` where ref corresponds to the level of the factor Condition you want as the intercept, DV corresponds to the datafile, and Condition corresponds to the factor you want to relevel.

What if our DV isn't a continuous variable?

- In eye movement work, we measure both gaze time (ms.) and also the number of times people re-read a region of text. For any one person reading a region of text, they either re-read it, or they don't. Thus, the data are binary (not continuous). In our data set, 1 corresponds to a region being re-read, 0 to not being re-read.

for workshop.R* | RO | DV

Filter

	Subject	Item	Condition	Region	DV
1	S1	I2	Positive	R3	0
2	S1	I3	Negative	R3	0
3	S1	I4	Neutral	R3	0
4	S1	I5	Positive	R3	0
5	S1	I6	Negative	R3	0
6	S1	I7	Neutral	R3	1
7	S1	I8	Positive	R3	0
8	S1	I9	Negative	R3	0
9	S1	I10	Neutral	R3	0
10	S1	I11	Positive	R3	0
11	S1	I12	Negative	R3	0
12	S1	I13	Neutral	R3	0

Showing 1 to 12 of 553 entries

- Here is our data file - our DV is categorical - either 1 or 0.

- For categorical data, we have to use the generalised linear model (*glmer*) and the binomial distribution. This is the syntax for such a model with both fixed and random effects:

```
model.full <- glmer (DV ~ Condition + (1 + Condition|Subject) + (1 + Condition|Item), data=RO, family=binomial)
```

- When we run it, we get an error (that you will get used to seeing again and again!)

```
> model.full <- glmer (DV ~ Condition + (1 + Condition|Subject) + (1 + Condition|Item), data=RO, family=binomial)
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.0171702 (tol = 0.001, component 1)
```

- So what can we do? We need to simplify the random effects structures. We can do this by dropping terms one by one until we find a model that can be fitted to our data. For example, we could drop the random slope from our items random effect first.

- For this particular example, the most complex model that works involves only random intercepts.

```
> model.interceptonly <- glmer (DV ~ Condition + (1|Subject) + (1|Item) , data=RO,
family=binomial)
> model.null <- glmer (DV ~  (1|Subject) + (1|Item), data=RO, family=binomial)
> anova (model.interceptonly, model.null)
```

```
> anova (model.interceptonly, model.null)
Data: RO
Models:
model.null: DV ~ (1 | Subject) + (1 | Item)
model.interceptonly: DV ~ Condition + (1 | Subject) + (1 | Item)
                    Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
model.null           3 601.97 614.91 -297.98   595.97
model.interceptonly  5 605.70 627.28 -297.85   595.70 0.2617      2     0.8773
```

- Our model with a fixed effect of Condition, and with random intercepts is no better than our model with just the random intercepts. In fact, it's worse - look at the AIC values. So we have no effect of Condition in our re-reading data.

Writing up LMM Results

- The analyses were carried out using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2017) to fit the linear mixed models for the reading time measure in R (R Development Core Team, 2017). The *glmer* function in the *lme4* package with Laplace approximation was used for the Re-reading measure. Below we report regression coefficients (b), standard errors, and *t*-values (for duration measures). Restricted maximum likelihood estimation was used for the reporting of linear mixed model parameters.

- Then you'll report the descriptive statistics and the parameter estimates…

LMMs for a 2 x 2 Repeated Measures Design

- Now let's take a 2 x 2 repeated measures design. We measured people's eye movements as they read either positive or negative information. Prior context set up expectations that the story was likely to continue with positive vs. negative information.

- Factor 1 is Context (Positive vs. Negative)

- Factor 2 is Sentence Type (Positive vs. Negative)

DV × | Workshop script2.R* ×

Filter

	Subject	Item	RT	Context	Sentence
1	1	3	1270	Positive	Negative
2	1	7	739	Positive	Negative
3	1	11	982	Positive	Negative
4	1	15	1291	Positive	Negative
5	1	19	1734	Positive	Negative
6	1	23	1757	Positive	Negative
7	1	27	1052	Positive	Negative
8	2	4	1706	Positive	Negative
9	2	8	533	Positive	Negative
10	2	12	1009	Positive	Negative
11	2	16	939	Positive	Negative
12	2	20	1848	Positive	Negative
13	2	24	1435	Positive	Negative
14	2	28	922	Positive	Negative

Showing 1 to 15 of 1,680 entries

- We have Subject number, Item number, RT (reading time), Context and Sentence.

- The first thing we need to do is to apply contrast weightings to our two factors. By default, the contrasts are dummy or treatment coded. We need to change them to deviation coded. This helps make the coefficients in the LMM make more sense as the intercept of the LMM will correspond to the Grand Mean (i.e., the mean of all four conditions).

```
contrasts (DV$Context)<-matrix (c(.5, -.5))
contrasts (DV$Sentence)<-matrix (c(.5, -.5))
```

- The first thing we are going to do is define our full model with our fixed effects and fully crossed Subject and Item random effects.

- Then we are going to define the null model with only the random effects.

```
model.full <- lmer (RT ~ Context*Sentence + (1+Context*Sentence |Subject) + (1+Context*Sentence |Item), data=DV, REML=TRUE)
model.null <- lmer (RT ~ (1+Context*Sentence |Subject) + (1+Context*Sentence |Item), data=DV, REML=TRUE)
```

- Note that we define our fixed effect using the notation Context*Sentence

- This is equivalent to (Context + Sentence + Context:Sentence) which corresponds to a main effect of Context, a main effect of Sentence and the interaction between the two (as represented by the colon symbol).

```
> anova (model.full, model.null)
refitting model(s) with ML (instead of REML)
Data: DV
Models:
..1: RT ~ (1 + Context * Sentence | Subject) + (1 + Context * Sentence |
..1:     Item)
object: RT ~ Context * Sentence + (1 + Context * Sentence | Subject) +
object:     (1 + Context * Sentence | Item)
       Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
..1    22 26720 26840 -13338    26676
object 25 26718 26853 -13334    26668 8.6625      3    0.03413 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Our model with the fixed effects (as well as the random effects) is a better fit for our data than is the model just with the random effects. Now we need to look at the model parameters using the ‘summary’ command…

```
> summary (model.full)
```

```
Fixed effects:
                    Estimate Std. Error      df t value Pr(>|t|)
(Intercept)          1568.75      76.24   50.07  20.577   <2e-16 ***
Context1              -36.20      86.01   29.77  -0.421   0.6768
Sentence1             -69.01      39.87   25.93  -1.731   0.0954 .
Context1:Sentence1   -168.73      80.36   25.51  -2.100   0.0458 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- We can see that the interaction is significant. But how do we know what difference(s) is/are driving this effect?
- Think back to ANOVA days - we need to now do something else…

- We can run pairwise comparisons. We can ask for a correction to be applied if we want to, but in this case we're doing to work out that correction by hand. There are only 2 theoretically meaningful pairwise comparisons, so we multiply the reported p value by 2 to manually apply Bonferroni correction.

- We use the *emmeans* function in the emmeans package.

```
> emmeans (model.full, pairwise ~ Context*Sentence, adjust="none")
$emmeans
 Context  Sentence   emmean       SE    df lower.CL upper.CL
 Positive Positive 1579.181 90.79178 48.12 1396.644 1761.718
 Negative Positive 1594.530 96.56705 36.99 1398.865 1790.194
 Positive Negative 1627.340 97.87425 41.97 1429.818 1824.862
 Negative Negative 1473.962 81.92644 39.71 1308.344 1639.580

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast                                estimate       SE    df t.ratio p.value
 Positive,Positive - Negative,Positive -15.34886 62.02007 27.31  -0.247  0.8064
 Positive,Positive - Positive,Negative -48.15900 97.23991 26.58  -0.495  0.6245
 Positive,Positive - Negative,Negative 105.21905 92.22797 29.04   1.141  0.2633
 Negative,Positive - Positive,Negative -32.81014 97.35194 31.48  -0.337  0.7383
 Negative,Positive - Negative,Negative 120.56791 92.61288 30.57   1.302  0.2027
 Positive,Negative - Negative,Negative 153.37805 50.68247 20.94   3.026  0.0064
```

Here we have the descriptive statistics associated with each of our 4 conditions.

Above are all the possible pairwise comparisons - only 2 are of theoretical interest to us:

1. A Negative meaning sentence following a Negative Context vs. the same Negative meaning following a Positive Context.

2. A Positive meaning sentence following a Negative Context vs. the same Positive meaning following a Positive Context.

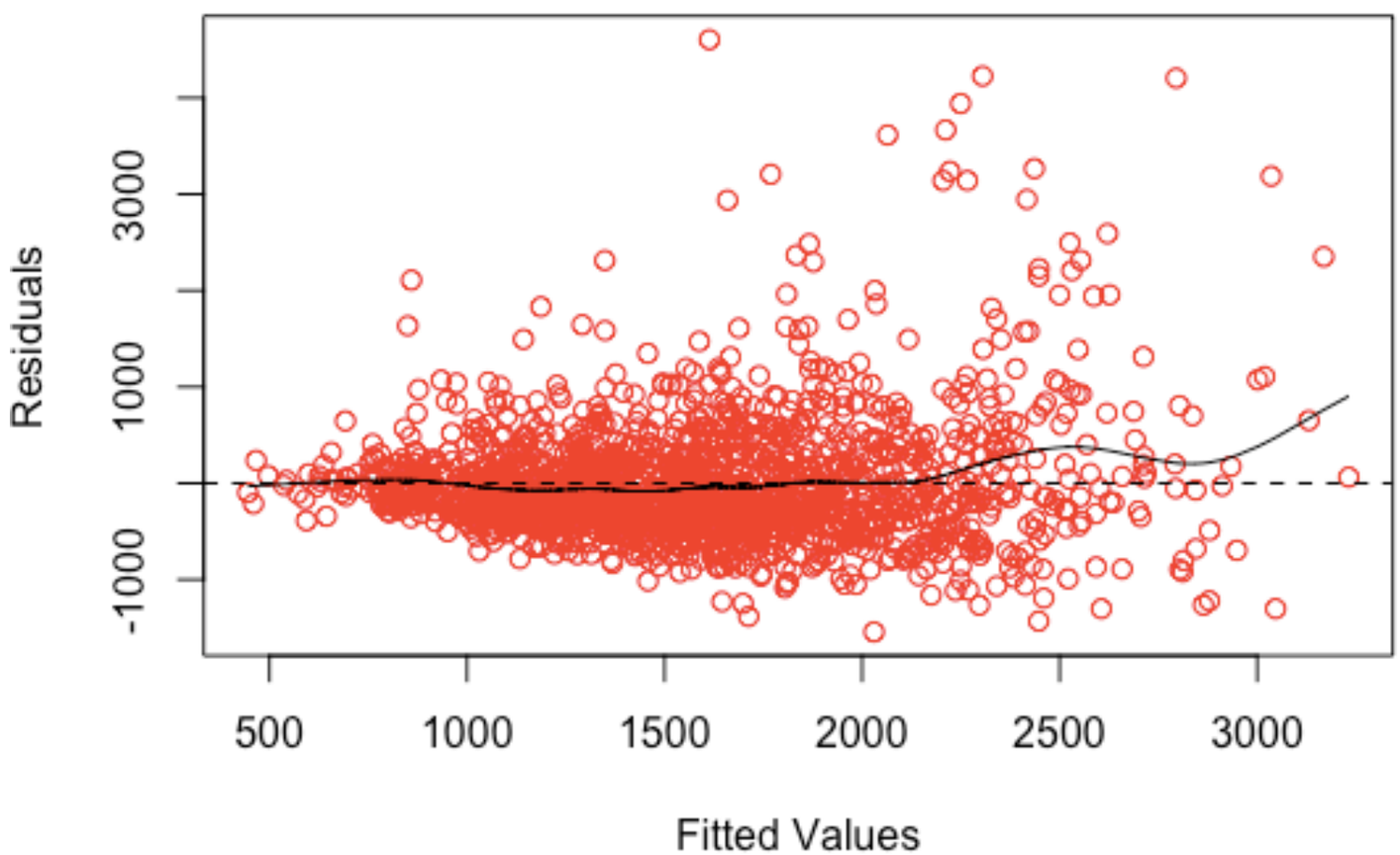
```
$contrasts
 contrast                                  estimate       SE    df t.ratio p.value
 Positive,Positive - Negative,Positive -15.34886 62.02007 27.31  -0.247  0.8064
 Positive,Positive - Positive,Negative -48.15900 97.23991 26.58  -0.495  0.6245
 Positive,Positive - Negative,Negative 105.21905 92.22797 29.04   1.141  0.2633
 Negative,Positive - Positive,Negative -32.81014 97.35194 31.48  -0.337  0.7383
 Negative,Positive - Negative,Negative 120.56791 92.61288 30.57   1.302  0.2027
 Positive,Negative - Negative,Negative 153.37805 50.68247 20.94   3.026  0.0064
```

- The two key comparisons reveal that Positive sentences are read no more quickly after Positive than after Negative context (1579 vs. 1595 ms.) while Negative Sentences are read more quickly after Negative than after Positive contexts (1474 vs. 1627 ms.)

- Note, the estimates in the contrasts correspond to the difference between comparisons in each pair.

We can visually assess the model fit by plotting the residuals against the observed data using:

```
#plot a graph of residuals - model fit is good the more the solid and dashed lines overlap
plot(fitted(model.full), residuals(model.full), xlab = "Fitted Values", ylab = "Residuals" , col="red")
abline(h = 0, lty = 2)
lines(smooth.spline(fitted(model.full), residuals(model.full)))
```



Looks like fit starts to drop off a bit for values > 2000.

- If we had re-reading (i.e., regression) data, we would also have to run an analysis using the *glmer* function on those data. The code would look like:

```
model.full <- glmer (Regressions ~ Context*Sentence + (1 + Context*Sentence|Subject)
+ (1 + Context*Sentence|Item), data=RO, family=binomial)
```

- To generate the pairwise comparisons (and to report the descriptives using the original measurement scale), we would use:

```
emmeans (model.full, pairwise~ Context*Sentence, adjust="none", type = "response")
```

- If we did not set the `type` parameter, then the descriptives would be on a log odds ratio scale (and harder to interpret).

Citing Packages

Remember to cite the packages you use, plus the version of R itself (with year info.) To find out how to cite a particular package, type:

```
> citation ("lme4")

To cite lme4 in publications use:

  Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models
  Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
```

And to find out which version of R you are using:

```
> version
                _
platform       x86_64-apple-darwin15.6.0
arch           x86_64
os             darwin15.6.0
system         x86_64, darwin15.6.0
status
major          3
minor          4.3
year           2017
month          11
day            30
svn rev        73796
language       R
version.string R version 3.4.3 (2017-11-30)
nickname       Kite-Eating Tree
```

Writing up These Results

The analyses were carried out using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) to fit the linear mixed models for the reading time measure in *R* (R Development Core Team, 2017). Pairwise comparisons conducted with the *emmeans* package (Lenth, 2018) were used to investigate the significant interaction for the reading time measure. Below we report regression coefficients (*b*), standard errors, and *t*-values. Restricted maximum likelihood estimation was used for the reporting of linear mixed model parameters. Deviation coding was used for each of the two experimental factors (Barr et al., 2013). Absolute values of the *t*-value greater than or equal to 1.96 indicate an effect that is significant at approximately the . 05 alpha level. For pairwise comparisons we report the *t*- values and *p*-values. Degrees of freedom are approximated using the Satterthwaite method.

	b	SE	t
Intercept	1569	76.24	20.577
Context	-36.20	86.01	-0.421
Sentence	-69.01	39.87	-1.731
Context x Sentence	-168.73	80.36	-2.100

You then report the two pairwise comparisons we conducted in the same way as you would do for ANOVA.

- When reporting the results of LMMs, it is important to provide all the information that someone would need to reproduce your analysis exactly. It's important to provide dates for the R packages you're using so that *exactly* the same version of R and associated packages can be used by someone else.

- We're moving toward a world where many of the top journals now ask for your analysis code to be uploaded as supplementary material. This could be your R script, or could be the R Markdown document. Either way, it should help with the need for reproducibility.

Addressing lack of convergence

```
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.0171702 (tol = 0.001, component 1)
```

If you see this message (which you will - again and again and again and again), it means you have to simplify your random effects structure so that a model can be identified.

Addressing lack of convergence

Simplify your random effects structure step by step. For an experiment with two factors (Factor 1 and Factor 2) we could simplify the participant and item random effects like this:

```
(1 + Factor 1*Factor 2| Participant) + (1 + Factor 1*Factor 2| Item)

(1 + Factor 1*Factor 2| Participant) + (1 + Factor 1+Factor 2| Item)

(1 + Factor 1+Factor 2| Participant) + (1 + Factor 1+Factor 2| Item)

(1 + Factor 1+Factor 2| Participant) + (1 + Factor 1| Item)
```

…

If you think your random effects looks too sparse when settling on a model that converges, you could try dropping one effect term entirely and then simplifying the other:

```
(1 + Factor 1*Factor 2| Participant) + (1 + Factor 1*Factor 2| Item)

(1 + Factor 1+Factor 2| Participant)

(1 + Factor 1| Participant)

(1 + Factor 2| Participant)

…
```

You want to avoid random effects with just random intercepts (i.e., no slopes) as that can inflate the Type 1 error rate (Barr et al., 2013).

A few other LMM things…

- You can add participant and item covariates as fixed effects, and you can have a variety of continuous and categorical variables in your LMM. LMMs are *very* flexible.

- You'll find that sometimes several models fit your data - always run likelihood comparison tests to determine which is the best fit. If you have a selection where not one is statistically better than the others, choose the model that makes most *theoretical* sense.

Important Point

- Add as many random slopes (not just random intercepts) as your experimental design allows - for most cases, we expect variation between participants in terms of how they'll respond to different levels of an experimental condition (which is why we add participants as a random slope) and also variation between our experimental items to different levels of an experimental condition (which is why we also add items as a random slope).
- If the full model with random slopes and intercepts does not converge, then gradually simplify your random effects structures (e.g., drop an interaction term first, then drop a main effect etc.) until you find a model that does converge.

LMMs for Ordinal DVs

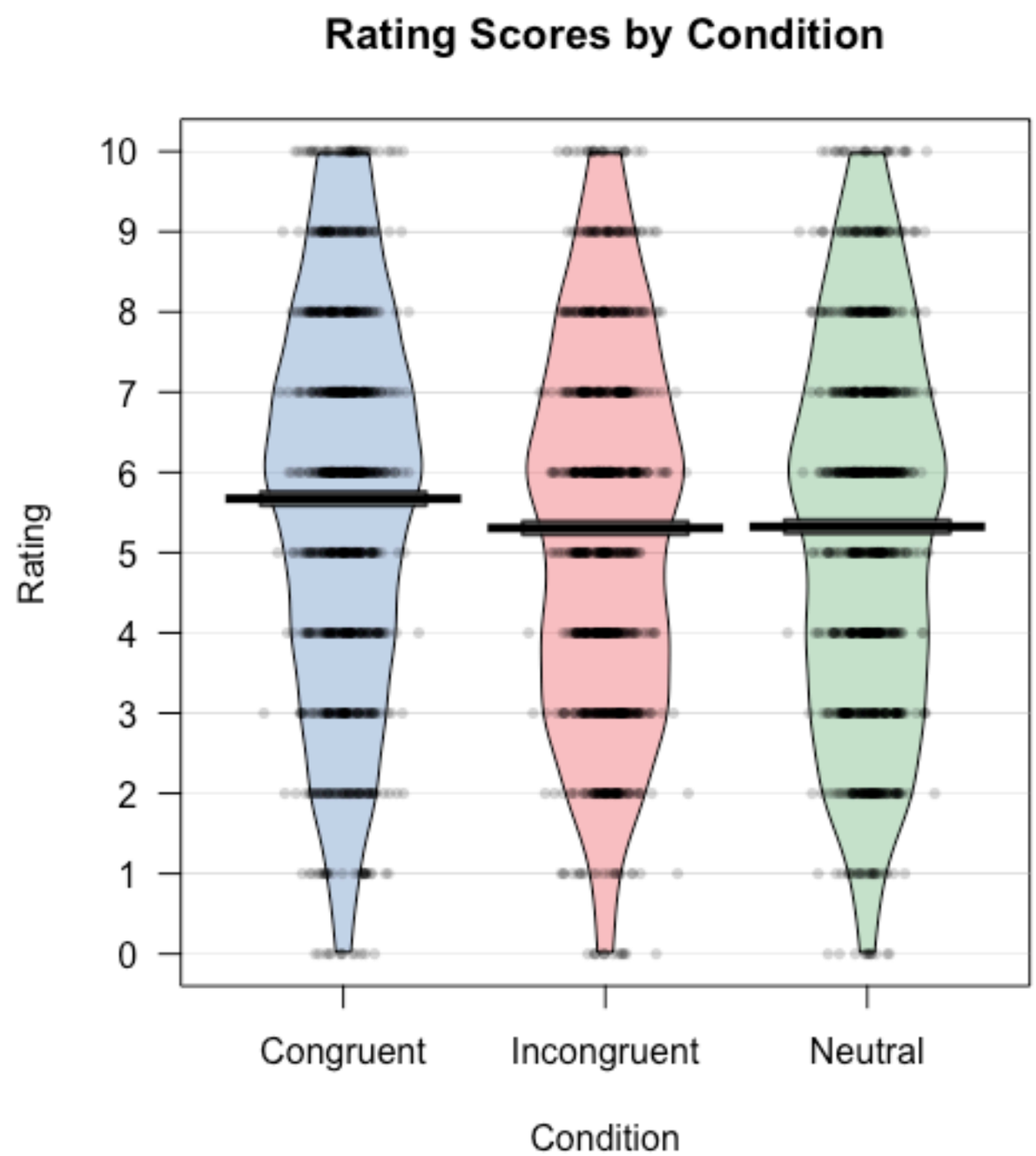
- Often we might collect data using a Likert scale. These data are ordinal and so we should used the cumulative-link mixed model function (CLMM) in the package called `ordinal`. Works similarly to LMMs in `lme4` but with one or two minor syntax changes…
- An example: we had 42 participants rate images of sports on a scale of 0-10 corresponding to how much they liked each one. Before each rating measure, they saw a video of a sport that matched or mismatched the one they then had to rate (with a neutral video as baseline).

- We want to know whether people's ratings were influenced by whether or not the sport they rated matched the one they had just seen.
- We have Subject, Image, and SportType as our random effects.
- VideoCondition corresponds to our condition - 2 is match, 3 is mismatch, and 4 is neutral.
- Our DV is the column 'ratings'.

Filter

	Subject	order	Image	SportType	VideoCondition	RTs	Questionnaire	SportExperienceTest1	SportExperienceTest2	ratings
1	1	1	5	5	2	2644	21	0	0	7
2	1	2	2	2	2	1606	21	0	0	4
3	1	3	10	10	2	1512	21	0	0	6
4	1	4	1	1	2	2217	21	0	0	8
5	1	5	3	3	2	1988	21	0	0	2
6	1	6	9	9	2	2876	21	0	0	9

- Plotting our data suggests the Congruent condition is producing higher scores than our other two conditions.



- Before we build our null and experimental models, we need to ensure our DV is coded as an ordinal variable.

```
> Main$ratings <- as.ordered (Main$ratings)

> model.clm.null <- clmm (ratings ~ 1 + (1 + VideoCondition|Subject) + (1 +
VideoCondition|SportType) + (1 + VideoCondition|Image), data=Main)

> model.clm4 <- clmm (ratings ~ VideoCondition + (1 + VideoCondition|Subject) +
(1 + VideoCondition|SportType) + (1 + VideoCondition|Image), data=Main)
```

- The syntax for our null model requires we have an explicit intercept (represented by a 1 in the fixed effects structure) similar to when we built regression models (this is different to how we specify a null model in `lme4` syntax).

- First, let's test whether our experimental model and null models differ:

```
> anova (model.clm.null, model.clm4)
```

```
## Likelihood ratio tests of cumulative link models:
##
##                formula:
## model.clm.null ratings ~ 1 + (1 + VideoCondition | Subject) + (1 + VideoCondition | SportType) + (1 +
VideoCondition | Image)
## model.clm4     ratings ~ VideoCondition + (1 + VideoCondition | Subject) + (1 + VideoCondition |
SportType) + (1 + VideoCondition | Image)
##                link: threshold:
## model.clm.null logit flexible
## model.clm4     logit flexible
##
##                no.par   AIC  logLik LR.stat df Pr(>Chisq)
## model.clm.null     28 10841 -5392.6
## model.clm4         30 10837 -5388.4  8.5295  2    0.01406 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We can see that they do - and our experimental model has the lower AIC value.

- Let's explore the effect of our Condition factor using *emmeans*:

```
> emmeans (model.clm4, pairwise ~ VideoCondition, adjust="none")
```

```
## $emmeans
##  VideoCondition    emmean        SE  df  asymp.LCL asymp.UCL
##  Congruent      0.6084163 0.2597503 Inf  0.0993151 1.1175176
##  Incongruent    0.2917028 0.2449736 Inf -0.1884367 0.7718422
##  Neutral        0.3153088 0.2436285 Inf -0.1621942 0.7928119
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast                   estimate         SE  df z.ratio p.value
##  Congruent - Incongruent  0.31671360 0.09146945 Inf   3.463  0.0005
##  Congruent - Neutral      0.29310751 0.09391144 Inf   3.121  0.0018
##  Incongruent - Neutral   -0.02360608 0.08587502 Inf  -0.275  0.7834
```

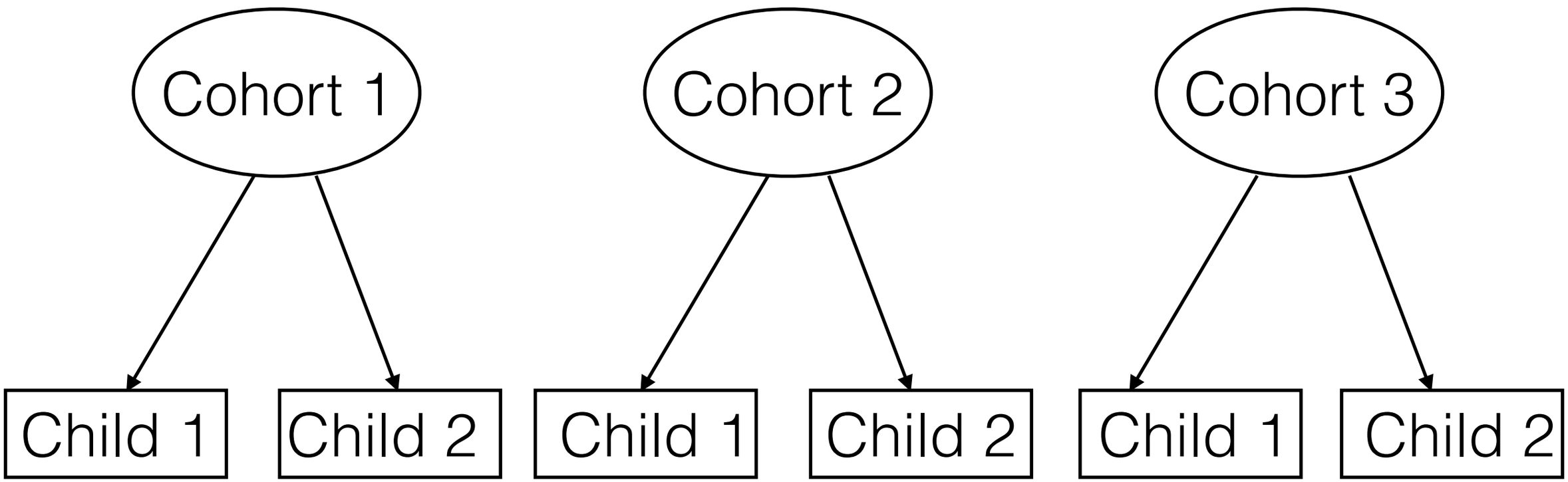
- The pairwise comparisons tell us that the Congruent condition differs from the Incongruent and Neutral conditions, but that the Incongruent and Neutral conditions do not differ.

- We can conclude that people's ratings for how much they liked particular sports were influenced by whether they had just seen a video depicting the sport. When the video and sport matched, they give the sport a higher rating when when the video and sport mismatched.

Crossed vs. Nested Random Effects

- In most experimental designs, your participant and item random factors are likely to be crossed - so random effects notation for a one factor experiment is `(1+Factor| Subjects) + (1+Factor| Items)`
- In some cases though, your factors might be *nested*. Nesting is a property of your data.
- To illustrate:

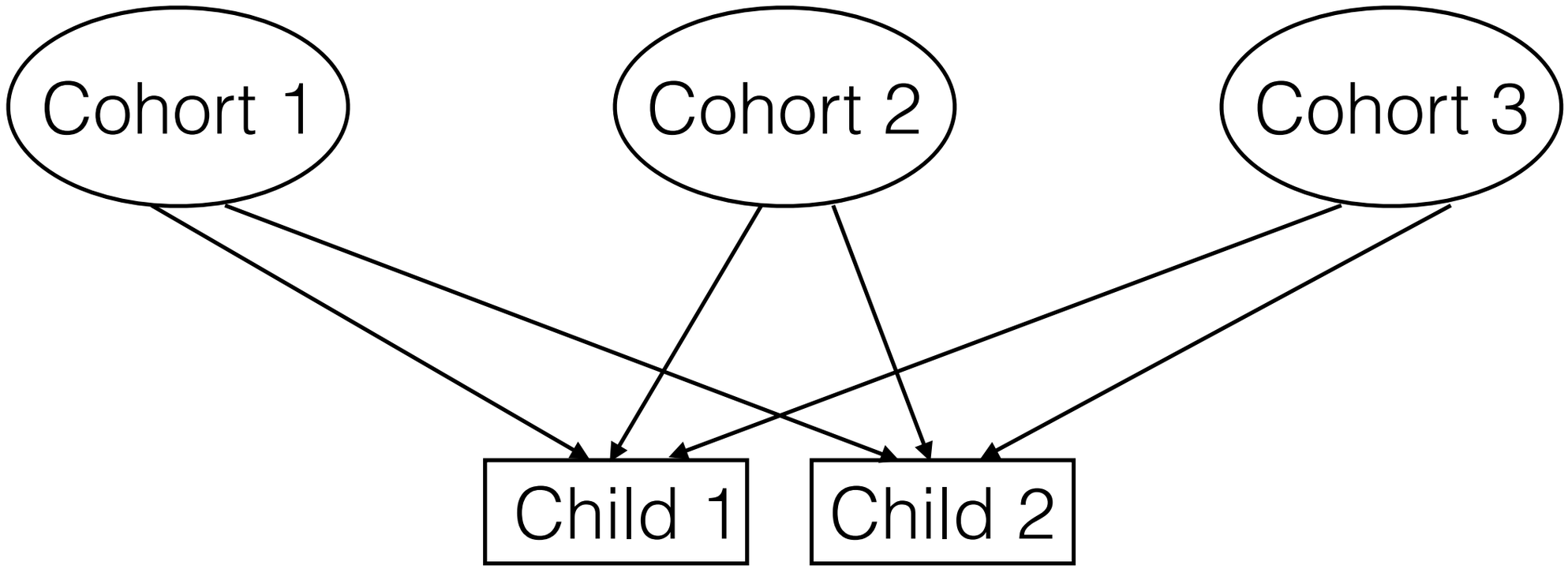
Nested



Child has an identifier that refers to a different child in each Cohort. Each child appears only in one Cohort. Child is nested within Cohort so random effects structure would be

```
(1+Factor| Cohort/Child)
```

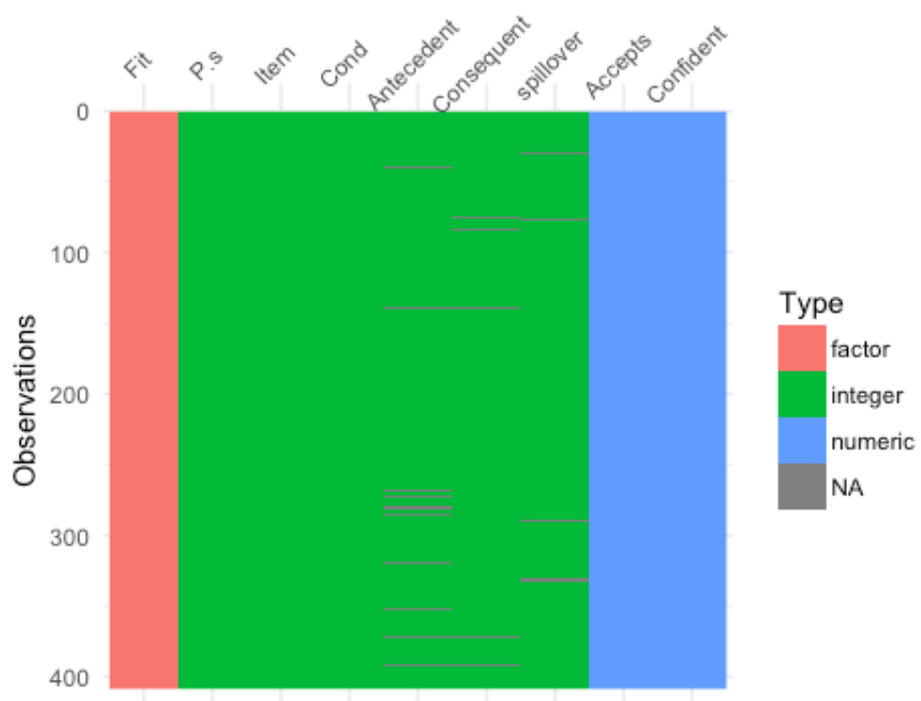
Crossed

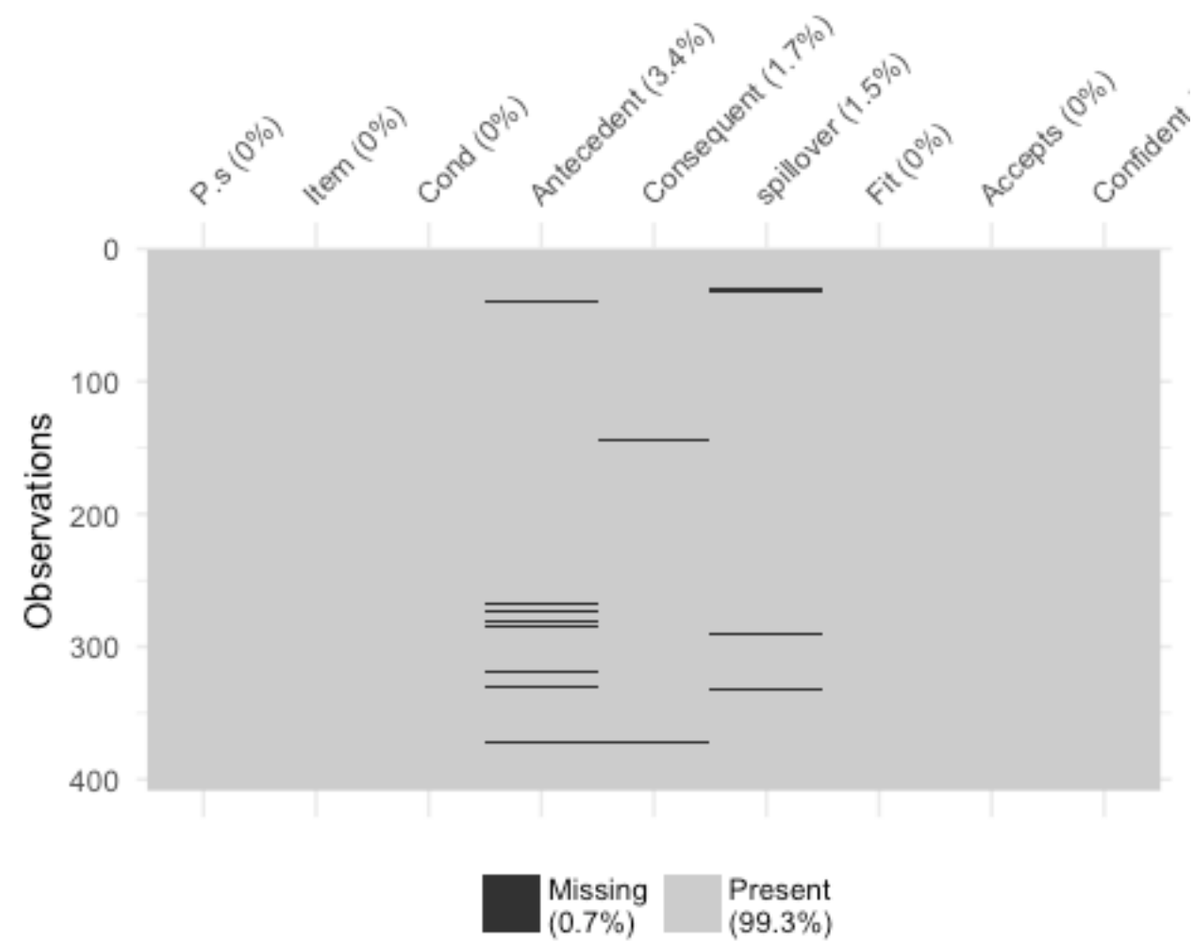


Each child appears in each Cohort. The levels are crossed so random effects structure would be `(1+Factor| Cohort) + (1+Factor| Child)`

Visualising our whole dataset using “visdat"

```
install.packages ("visdat")
library (visdat)

#read in data file
RPs_plus_ratings <- read_csv("~/Desktop/Air Work/R analyses/Igor study/RPs_plus_ratings.csv")

#make Fit a factor
RPs_plus_ratings$Fit <- factor (RPs_plus_ratings$Fit)

#create an index so we can remove item 9
index <- RPs_plus_ratings$Item !="9"

#visualise the data
vis_dat(RPs_plus_ratings[index,])

#visualise missing data
vis_miss(RPs_plus_ratings[index,])
```



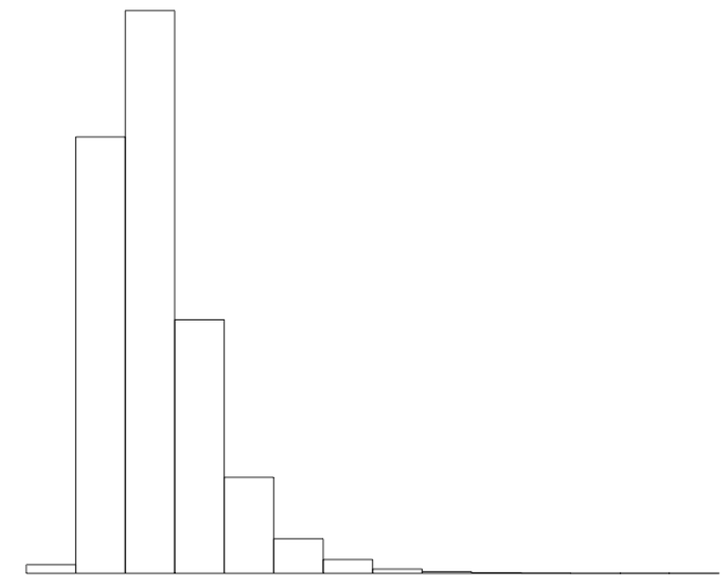


What about normality?

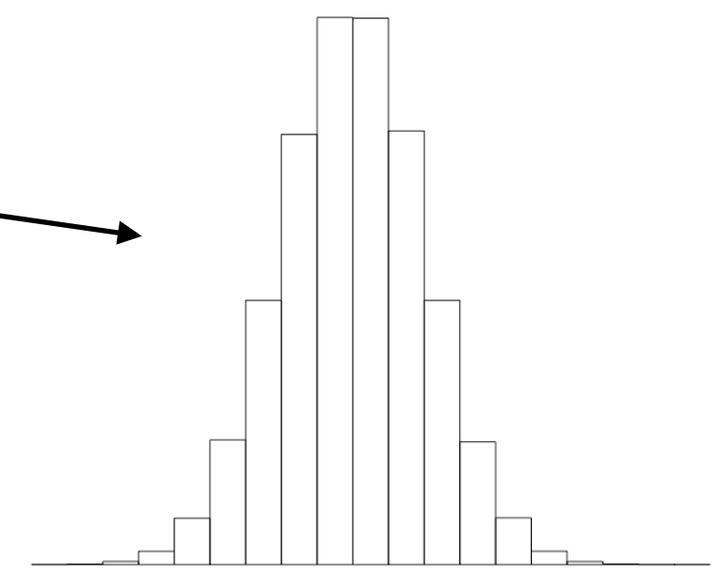
- In LMMs (as with the GLM) we need to worry about the normality of the residuals…
- You can check normality in a number of ways.
- Graphically, you can use the *qqnorm* function (which produces a Q-Q plot), and *hist* (which produces a histogram) applied to the model residuals.
- Statistically, you could use the *shapiro.test* function applied to a distribution of data. Be aware that for large datasets, even small deviations from normality will result in a significant Shapiro test. So best not to use this…

Log transform

Typically, RT data are non-normal and more often the DV looks like this.

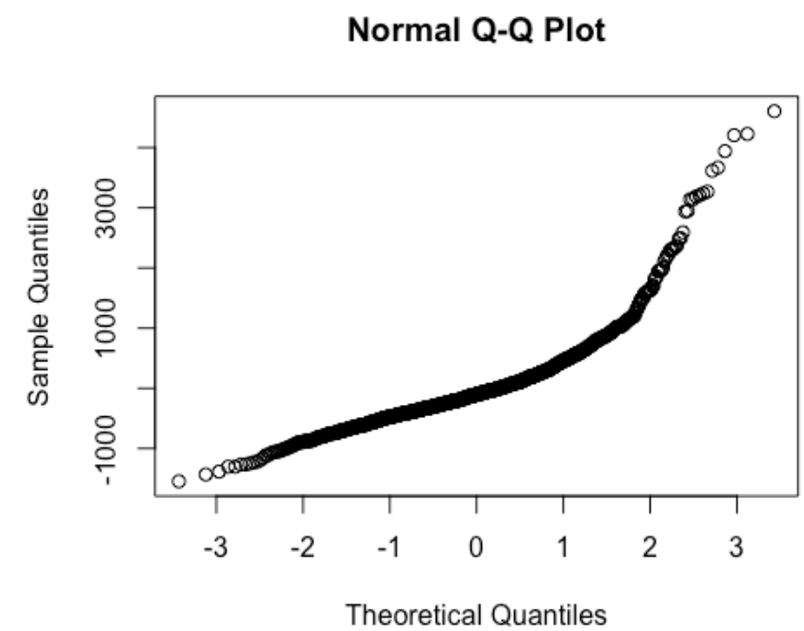


We can log transform our DV to approximate something that looks a bit more like the normal distribution (could also look at inverse RT). But there are risks around transforming data (Lo & Andrews, 2015)



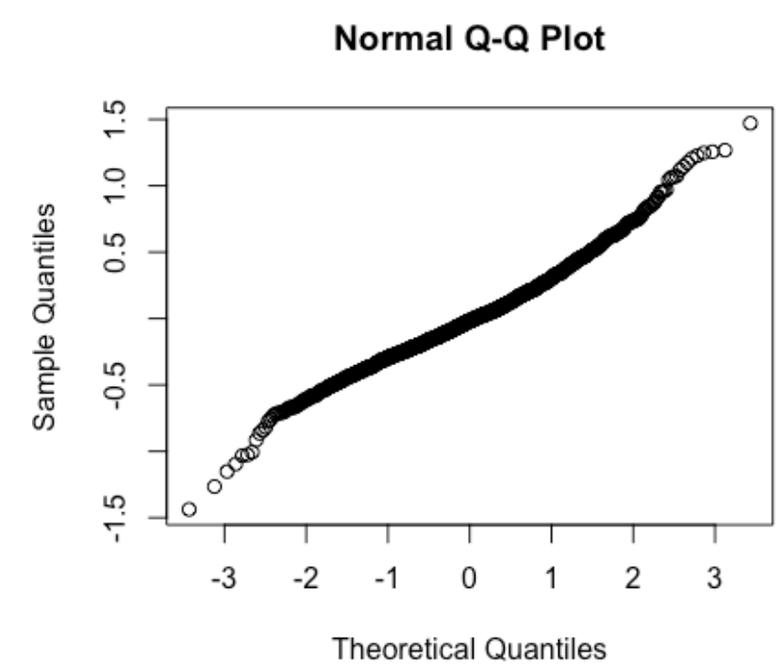
- Normality test on the model residuals from the untransformed data:

```
> qqnorm(residuals(model.full))
> shapiro.test(residuals(model.full))

        Shapiro-Wilk normality test

data:  residuals(model.full)
W = 0.84583, p-value < 2.2e-16
```



- Normality test on the model residuals from the log transformed data:

```
> model.full <- lmer (log(RT) ~ Sentence*Context + (1+Sentenc
e*Context|Subject) + (1+Sentence*Context |Item), data=DV, REM
L=TRUE)
> qqnorm(residuals(model.full))
> shapiro.test(residuals(model.full))

        Shapiro-Wilk normality test

data:  residuals(model.full)
W = 0.98321, p-value = 4.626e-13
```



- The original analysis on the untransformed data:

```
Fixed effects:
                   Estimate Std. Error      df t value Pr(>|t|)
(Intercept)         1568.75      76.24   50.07  20.577   <2e-16 ***
Context1             -36.20      86.01   29.77  -0.421   0.6768
Sentence1            -69.01      39.87   25.93  -1.731   0.0954 .
Context1:Sentence1  -168.73      80.36   25.51  -2.100   0.0458 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- The new analysis on the log transformed data:

```
Fixed effects:
                    Estimate Std. Error       df t value
(Intercept)          7.23975    0.04967 49.13000 145.761
Sentence1            0.01392    0.05278 29.03000   0.264
Context1             0.04316    0.02258 28.62000   1.911
Sentence1:Context1 -0.09333    0.04618 25.55000  -2.021
```

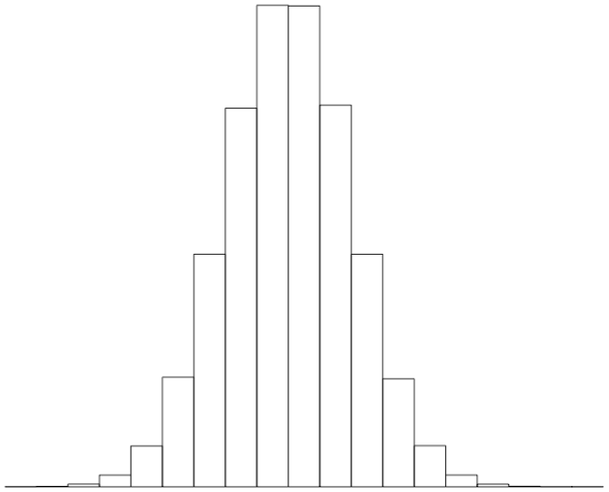
t-value of the interaction smaller than in analysis over untransformed data. With similar dfs, *p* will be bigger.

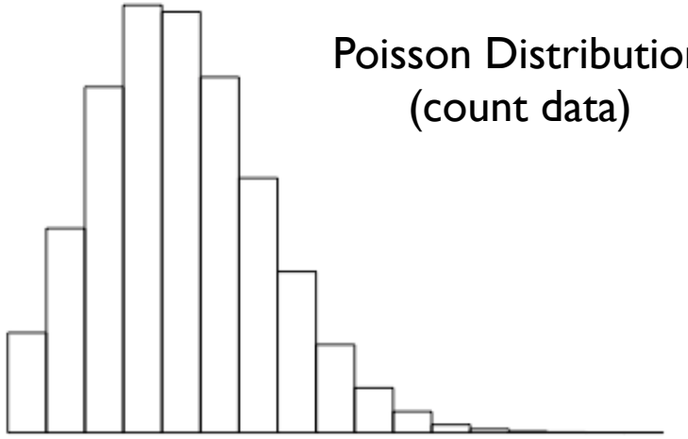
Other distributions under the GLMM via the function `glmer` are available…

Usage

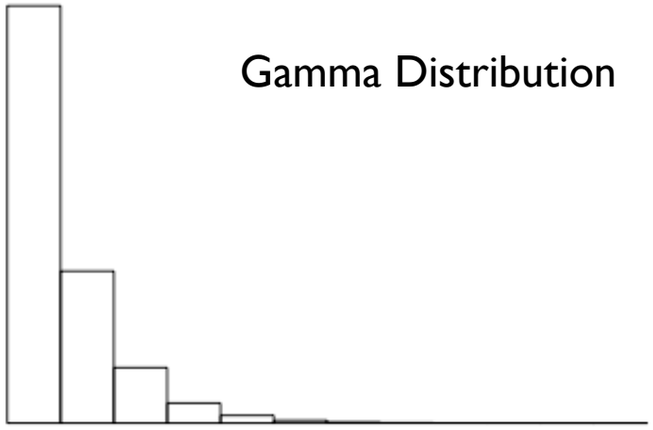
```
family(object, ...)

binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
quasi(link = "identity", variance = "constant")
quasibinomial(link = "logit")
quasipoisson(link = "log")
```

Normal (Gaussian) Distribution





Poisson Distribution
(count data)



Gamma Distribution

- Standard linear model assumes a normal distribution of residuals. In the *generalised* linear mixed model, we can assume a distribution in our model that doesn't involve a normal distribution. We have already looked at the binomial.
- Gamma distribution is another possibility (see Kliegl et al. 2010, Lo & Andrews, 2015, for discussion).

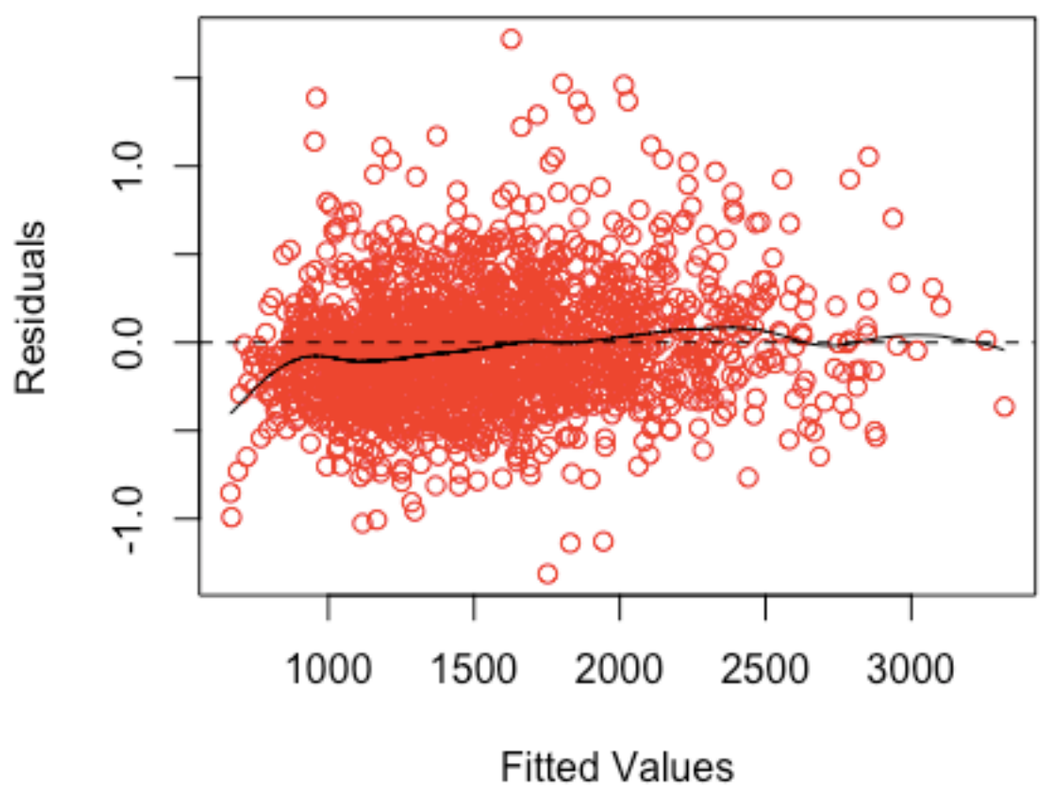
```
model1 <- glmer (RT ~ Sentence*Context + (1+Sentence*Context|Subject) + (1+Sentence*Context|Item), data=DV, family=Gamma)
summary (model1)
```

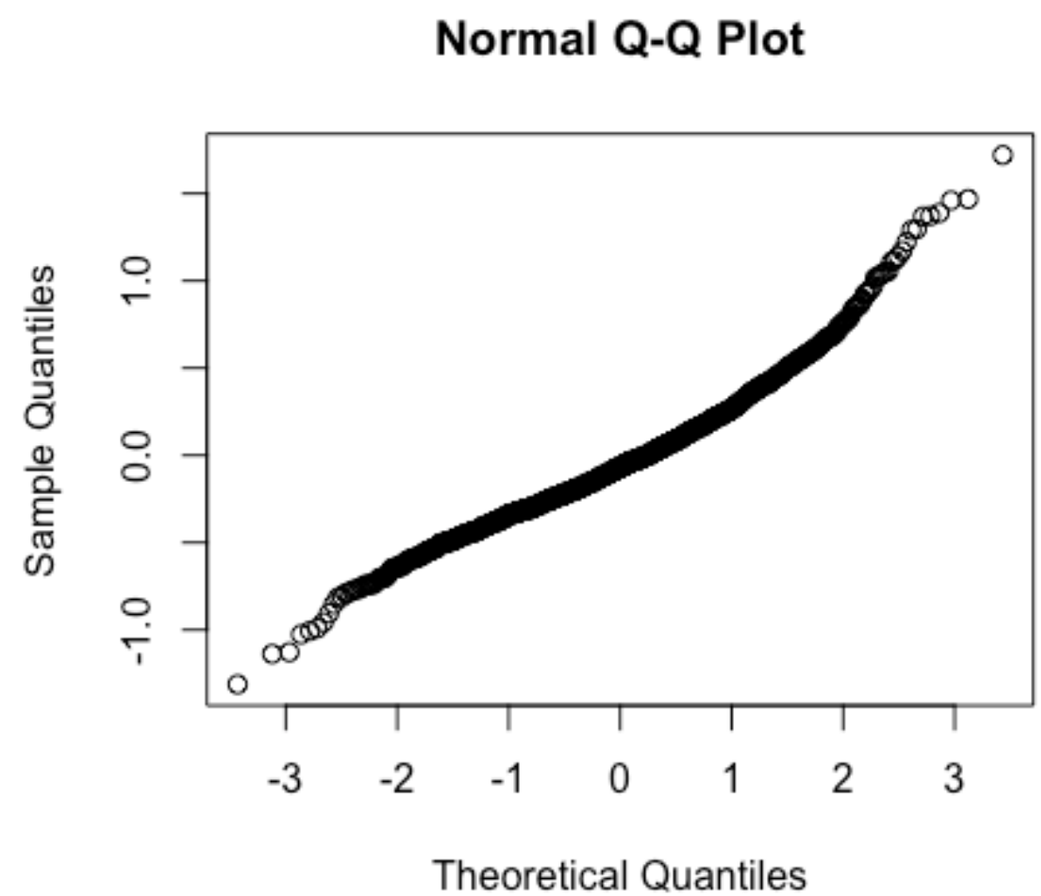
```
Fixed effects:
                   Estimate Std. Error t value
(Intercept)         7.28232    0.06731  108.20
Sentence1           0.02284    0.07679    0.30
Context1            0.04276    0.01701    2.51
Sentence1:Context1 -0.10806    0.03403   -3.18
```

t-value of the interaction larger than in previous analysis.

- Normality test on the model residuals from the data assuming sampling from the Gamma distribution:

```
> qqnorm (residuals (model1))
```





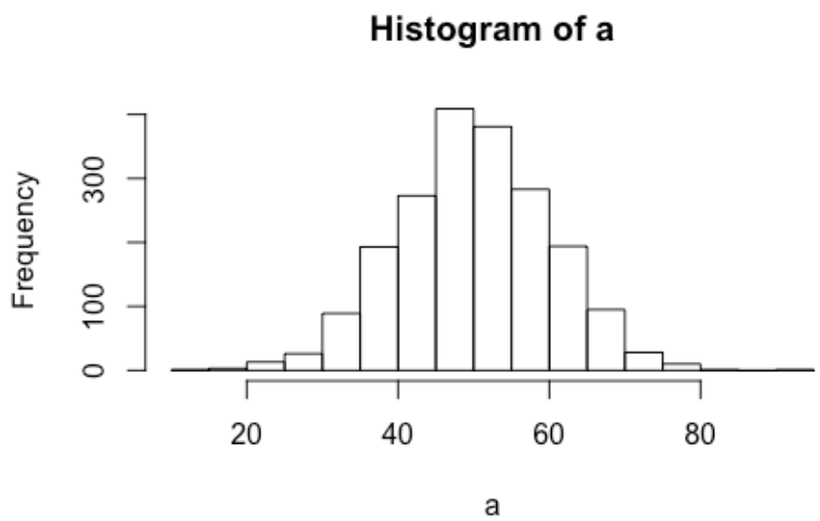
So what to do?

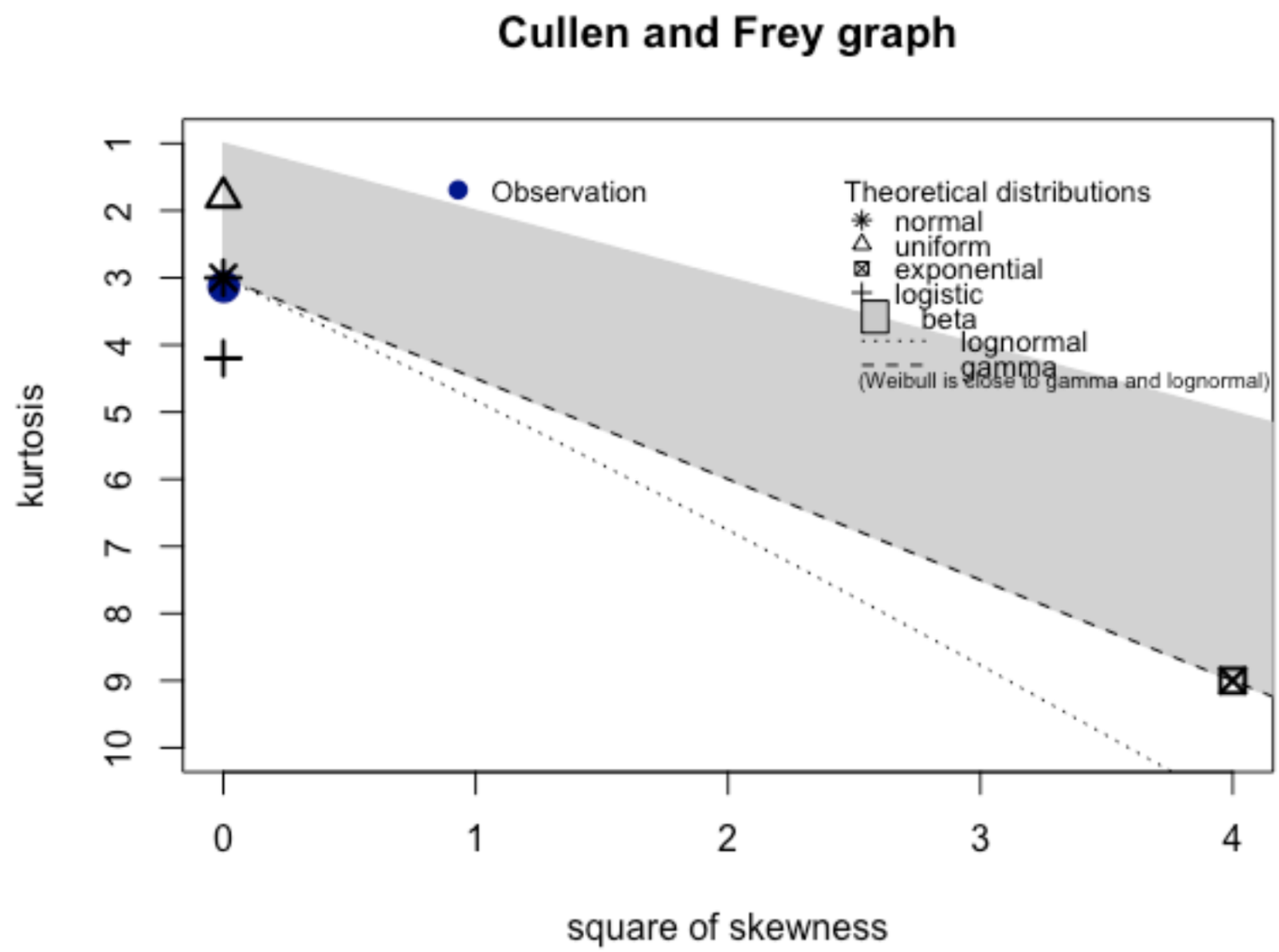
- In this example, all three analyses told basically the same story - there is an effect in our interaction term. They differ in terms of the value of the t-statistic associated with testing this.
- It's an issue - but if each possible way of analysing the data (incl. log transform and GLMM under the Gamma distribution) produces the same story, probably don't need to worry too much.
- Key is to be transparent in the write-up (did you transform the data? If so, how? What distribution do you assume your data come from?). **Most importantly, publicly archive your data and analysis code so it can be examined by others.**

Determining the likely distribution of our data

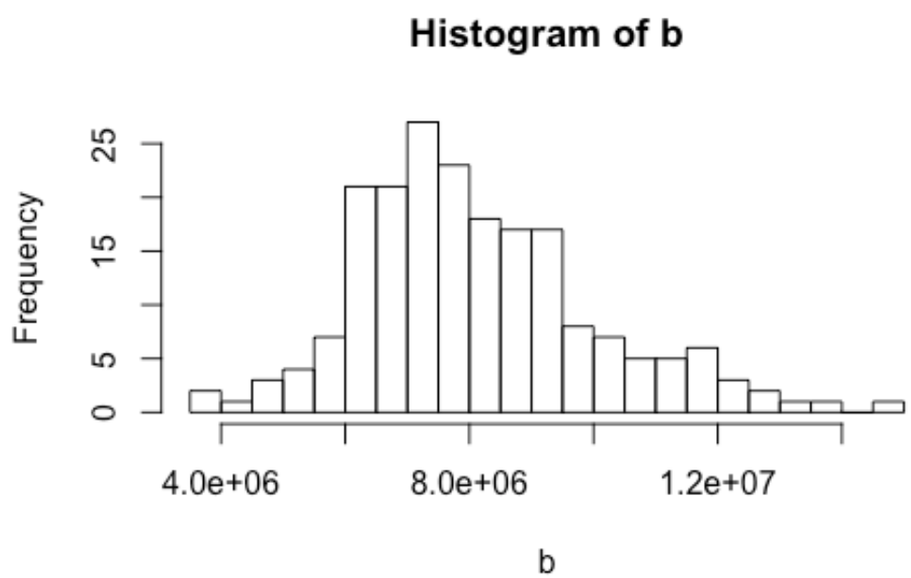
- We can use the function `descdist` from the package `fitdistrplus` to plot any set of data on a Cullen and Frey graph - this will help us determine what known distribution of data our data match.
- First, I'm going to create some data drawn from the normal distribution and plot that sample…

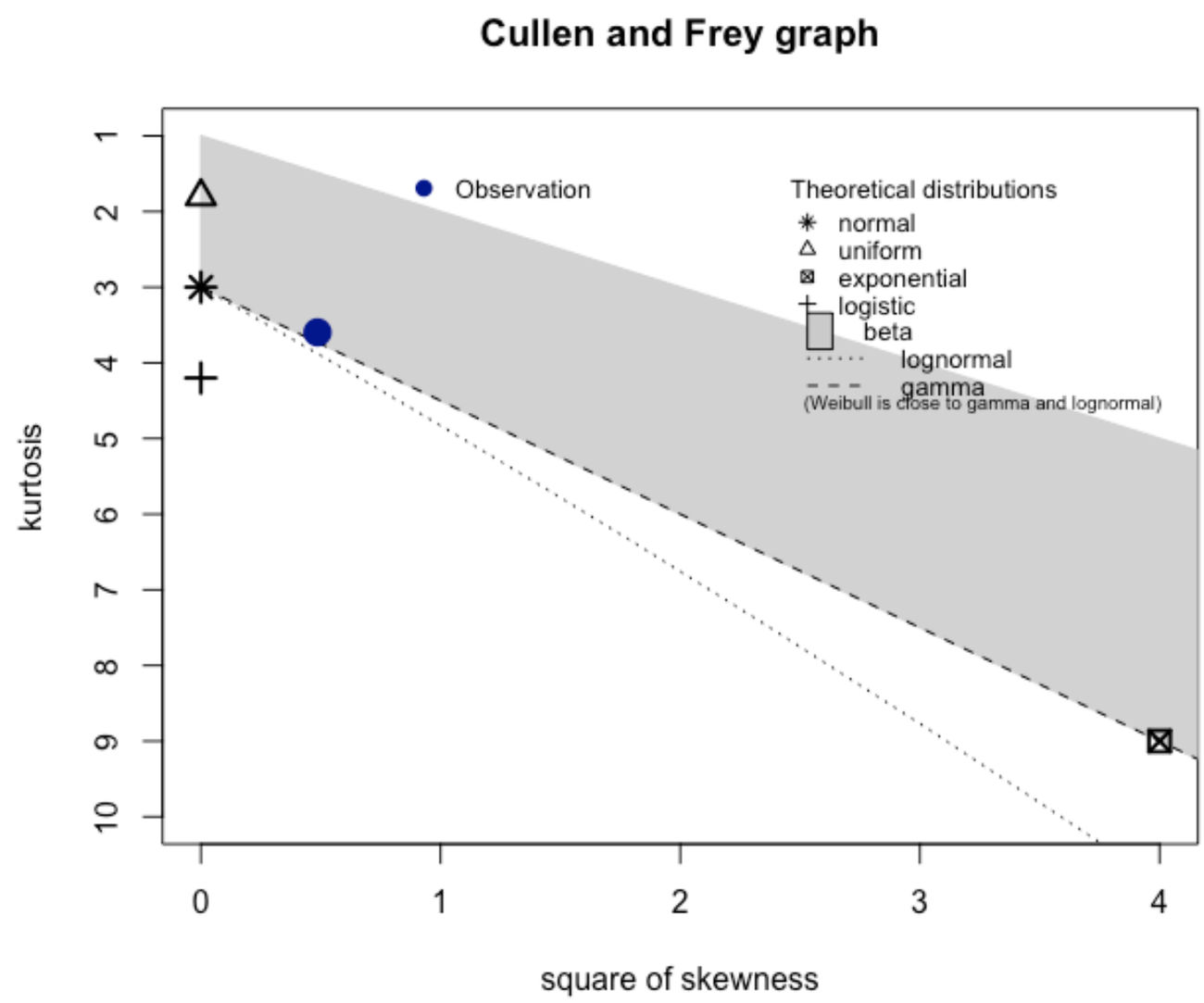
```
> library(fitdistrplus)
> a <- rnorm (2000, mean=50, sd=10)
> descdist(a)
```



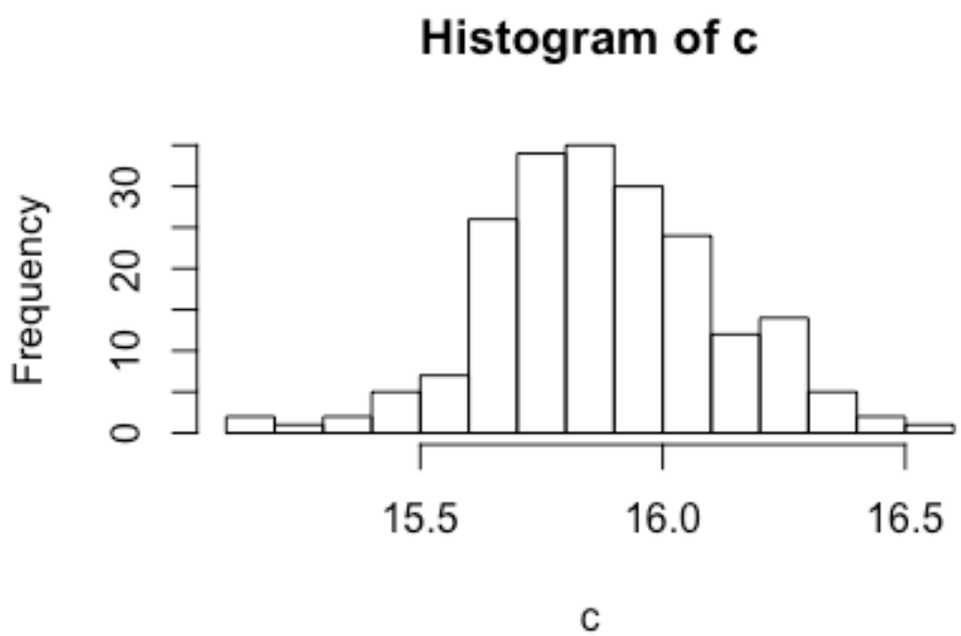


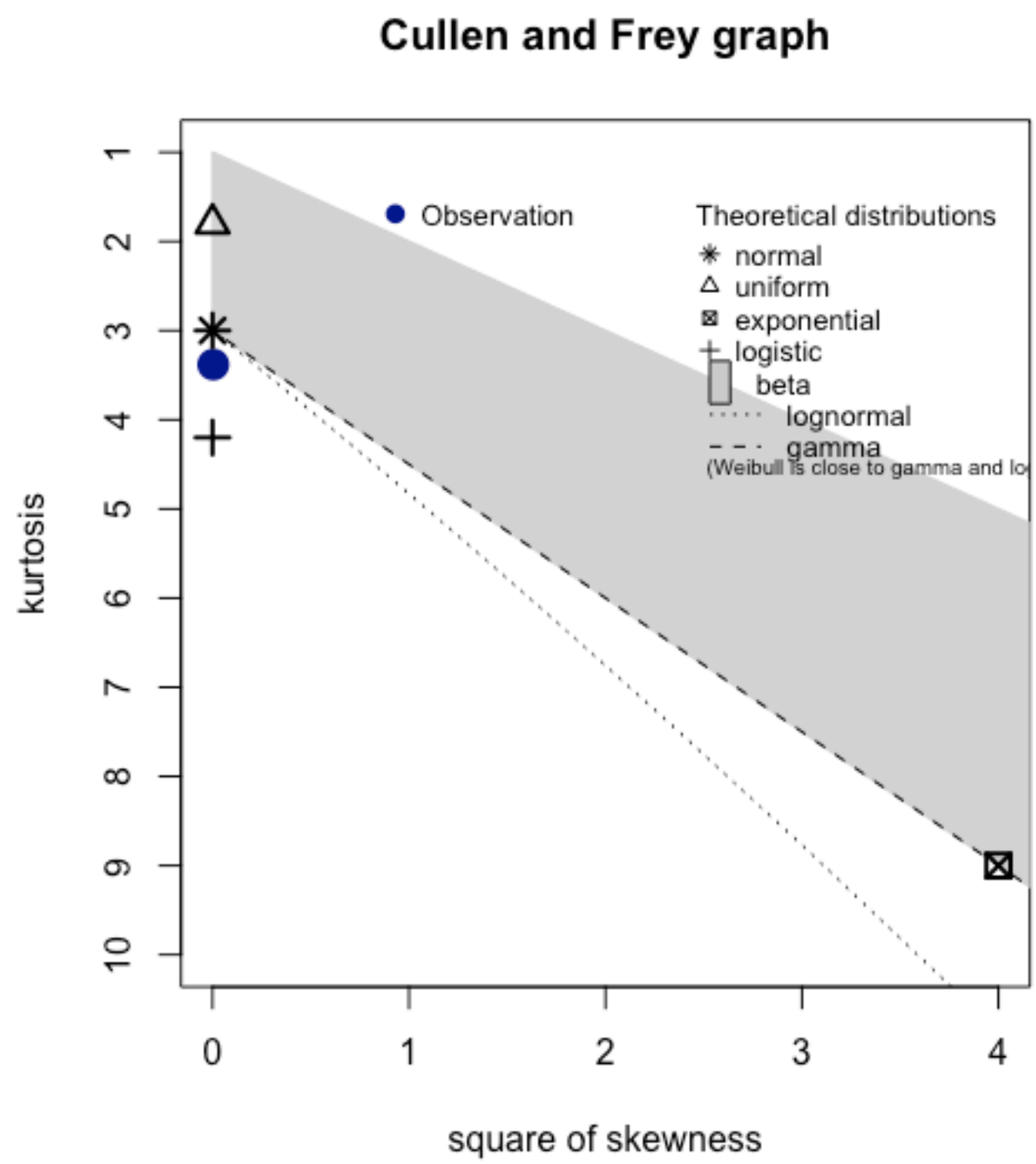
- Now some positively skewed data:





- We can log transform the data and then view on a Cullen and Frey graph…





Understanding the distribution our data is likely sampled from will help us avoid analysis pitfalls.

General R Tips

- Restart R whenever you start a new analysis - and create a new Project for each analysis - you don't want old variable names clogging up your workspace.
- Make sure you remember to install the library packages you need.
- If you get really stuck, look at some of the R advice forums.
- Chances are you are making a mistake related to syntax, capitalisation, or trying to run a package you haven't installed/updated.

Some examples where I've used LMMs in my own work - code is on my GitHub page.

Stewart, A.J., Wood, J.S., Le-luan, E., Yao, B., & Haigh, M. (in press). "It's hard to write a good article." The online comprehension of excuses as indirect replies. *Quarterly Journal of Experimental Psychology.*

Stewart, A.J., Le-luan, E., Yao, B., Wood, J., & Haigh, M., (in press). Comprehension of indirect requests is influenced by their degree of imposition. *Discourse Processes.*

McGarrigle, R.A., Dawes, P., Stewart, A.J., Kuchinsky, S.E., & Munro, K.J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology, 161,* 95-112.

McGarrigle, R.A., Dawes, P., Stewart, A.J., Kuchinsky, S.E., & Munro, K.J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology, 54,* 193-203.

Wood, J., Haigh, M., & Stewart, A.J. (2016). "This isn't a promise, it's a threat": eye movements reveal pragmatic scope differences in conditional inducements. *Experimental Psychology, 63,* 89-97.

Further Reading

Baayen, R.H., Davidson, D.J., Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Barr, D.J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Kliegl, R., Masson, M. E. J., and Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18,* 655–681.

Lo, S., and Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Frontiers in Psychology, 6:1171.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. New York, NY: CRC Press.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]