Mixed-effects regression models

Martijn Wieling

Center for Language and Cognition Groningen, University of Groningen

Seminar on Statistics and Methodology, Groningen - April 24, 2012

Overview

- Part I
  - Introduction
  - Recap: multiple regression
  - Mixed-effects regression analysis: explanation
  - Case-study: Dutch dialect data (with Harald Baayen and John Nerbonne)
  - Conclusion
- Part II
  - Extended case-study: Analyzing eye tracking data (with Hanneke Loerts)[1]

[1] All slides will be made available after this lecture.

Introduction

- Consider the following situation (taken from Clark, 1973):
  - Mr. A and Mrs. B study reading latencies of verbs and nouns
  - Each randomly selects 20 words and tests 50 participants
  - Mr. A finds (using a sign test) verbs to have faster responses
  - Mrs. B finds nouns to have faster responses
- How is this possible?

Introduction

- Consider the following situation (taken from Clark, 1973):
  - Mr. A and Mrs. B study reading latencies of verbs and nouns
  - Each randomly selects 20 words and tests 50 participants
  - Mr. A finds (using a sign test) verbs to have faster responses
  - Mrs. B finds nouns to have faster responses
- How is this possible?

The language-as-fixed-effect fallacy

- The problem is that Mr. A and Mrs. B disregard the variability in the words (which is huge)
  - Mr. A included a difficult noun, but Mrs. B included a difficult verb
  - Their set of words does not constitute the complete population of nouns and verbs, therefore their results are limited to their words
- This is known as the language-as-fixed-effect fallacy (LAFEF)
  - Fixed-effect factors have repeatable and a small number of levels
  - Word is a random-effect factor (a non-repeatable random sample from a larger population)

Why linguists are not always good statisticians

- LAFEF occurs frequently in linguistic research until the 1970's
  - Many reported significant results are wrong (the method is anti-conservative)!

- Clark (1973) combined a by-subject (F_1) analysis and by-item (F_2) analysis in a measure called *min F'*
  - Results are significant and generalizable across subjects and items when *min F'* is significant
  - Unfortunately many researchers (>50%!) incorrectly interpreted this study and may report wrong results (Raaijmakers et al., 1999)
  - E.g., they only use F_1 and F_2 and not *min F'* or they use F_2 while unneccesary (e.g., counterbalanced design)

Our problems solved...

- Apparently, analyzing this type of data is difficult...
- Fortunately, using mixed-effects regression models solves all our problems!
  - The method is easier than using the approach of Clark (1973)
  - Results can be generalized across subjects and items
  - Mixed-effects models are robust to missing data (Baayen, 2008, p. 266)
  - We can easily test if it is necessary to treat item as a random effect
- But first some words about regression...

Our problems solved...

- Apparently, analyzing this type of data is difficult...
- Fortunately, using mixed-effects regression models solves all our problems!
  - The method is easier than using the approach of Clark (1973)
  - Results can be generalized across subjects and items
  - Mixed-effects models are robust to missing data (Baayen, 2008, p. 266)
  - We can easily test if it is necessary to treat item as a random effect
- But first some words about regression...

Regression vs. ANOVA

- Most people either use ANOVA or regression
  - ANOVA: categorical predictor variables
  - Regression: continuous predictor variables
- Both can be used for the same thing!
  - ANCOVA: continuous and categorical predictors
  - Regression: categorical (dummy coding) and continuous predictors
- Why I use regression as opposed to ANOVA
  - No temptation to dichotomize continuous predictors
  - Intuitive interpretation (your mileage may vary)
  - Your design does not have to be completely balanced
  - Mixed-effects analysis is relatively easy to do and does not require a balanced design (which is generally necessary for repeated-measures ANOVA)
- This talk will focus on regression

Recap: multiple regression

- Multiple regression: predict one numerical variable on the basis of other independent variables (numerical or categorical)
  - (*Logistic* regression is used to predict a categorical dependent)
- We can write a regression formula as $y = I + ax_1 + bx_2 + ...$
- E.g., predict the reaction time of a participant on the basis of word frequency, word length and speaker age:
  $RT = 200 - 5WF + 3WL + 10SA$

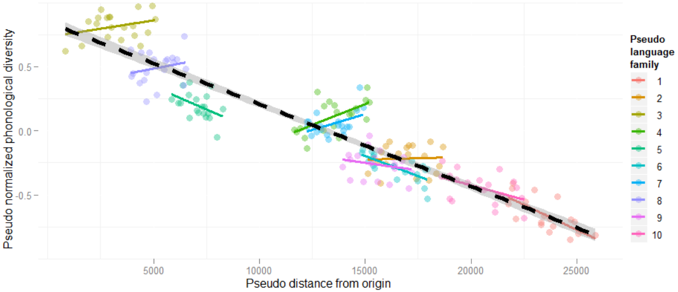
Mixed-effects regression modeling: introduction

- Mixed-effects regression modeling distinguishes fixed-effects and random-effects factors
- Fixed-effects factors:
  - Repeatable levels
  - Small number of levels (e.g., Gender, Word Category)
  - Same treatment as in multiple regression (treatment coding)
- Random-effects factors:
  - Levels are a non-repeatable random sample from a larger population
  - Often large number of levels (e.g., Subject, Item)

What are random-effects factors?

- Random-effect factors are factors which are likely to introduce systematic variation
  - Some participants have a slow response (RT), while others are fast
    = Random Intercept for Subject
  - Some words are easy to recognize, others hard
    = Random Intercept for Item
  - The effect of word frequency on RT might be higher for one participant than another: non-native speakers might benefit more from frequent words than native speakers
    = Random Slope for Word Frequency per Subject
  - The effect of speaker age on RT might be different for one word than another: modern words might be recognized easier by younger speakers
    = Random Slope for Speaker Age per Item
- Note that it is essential to test for random slopes!

Random slopes are necessary!



		Estimate	Std. Error	t value	Pr(>|t|)
Linear regression	DistOrigin	-6.418e-05	1.808e-06	-35.49	<2e-16 ***
+ Random intercepts	DistOrigin	-2.224e-05	6.863e-06	-3.240	<0.001 ***
+ Random slopes	DistOrigin	-1.478e-05	1.519e-05	-0.973	n.s.

This example is explained at Florian Jaeger's blog: http://hlplab.wordpress.com

Specific models for every observation

- Mixed-effects regression analysis allow us to use random intercepts and slopes to make the regression formula as precise as possible for every individual observation in our random effects
  - A single parameter (standard deviation) models this variation for every random slope or intercept
  - The actual random intercepts and slopes are derived from this value
  - Likelihood-ratio tests assess whether the inclusion of random intercepts and slopes is warranted
- Note that multiple observations for each level of a random effect are necessary for mixed-effects analysis to be useful (e.g., participants respond to multiple items)

Specific models for every observation

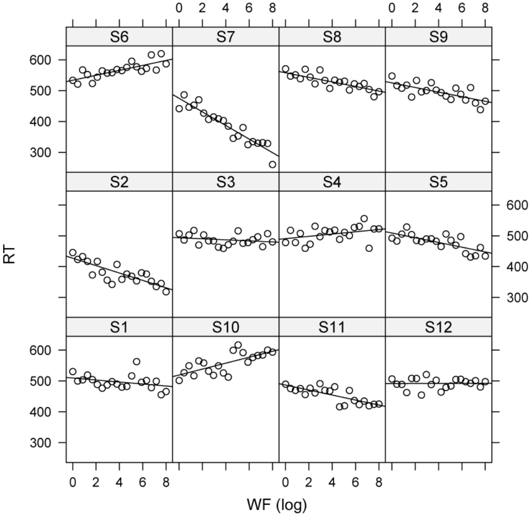
- $RT = 200 - 5WF + 3WL + 10SA$ (general model)
  - The intercepts and slopes may vary (according to the estimated standard variation for each parameter) and this influences the word- and subject-specific values

- $RT = 400 - 5WF + 3WL - 2SA$ (word: scythe)
- $RT = 300 - 5WF + 3WL + 15SA$ (word: twitter)
- $RT = 300 - 7WF + 3WL + 10SA$ (subject: non-native)
- $RT = 150 - 5WF + 3WL + 10SA$ (subject: fast)

- And it is easy to use!

```
> lmer( RT ~ WF + WL + SA + (1+SA|Wrd) + (1+WF|Subj) )
```

Specific models for every subject



Case study: Dutch dialects w.r.t. standard Dutch

- The goal of this study is to investigate which factors predict the dialect distances of 562 words in 424 locations from standard Dutch
  - This study was a collaboration with John Nerbonne and Harald Baayen and is published as Wieling, Nerbonne and Baayen (2011, PLoS ONE).
- We use a mixed-effects regression model for this purpose
  - Random-effects factors: Location, Word and Transcriber
- Several location-, speaker- and word-related factors are investigated
  - E.g., number of inhabitants, average age of inhabitants, speaker age, speaker gender, word frequency and word category

Geographic distribution of locations

Determining dialect distances

- We use phonetic transcriptions of 562 words in 424 locations in NL
- These are compared to standard Dutch transcriptions using the Levenshtein algorithm (Levenshtein, 1965)
  - The Levenshtein algorithm measures the minimum number of insertions, deletions and substitutions to transform one string into another

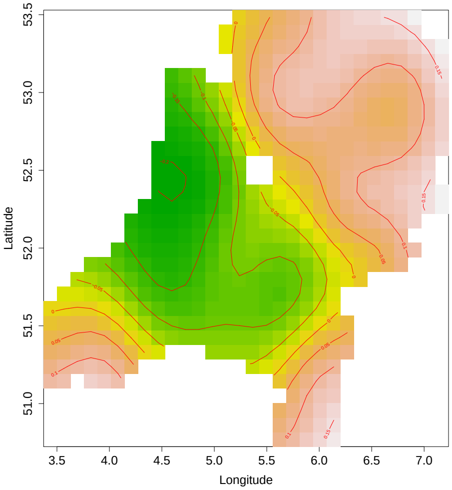
b		ɪ	n	d	ə	n
b	ɛ	i	n	d	ə	
	1	1				1

- The distance between the dialectal and standard Dutch pronunciation is based on the total cost of the operations (above: 3)
- We actually use linguistically validated sensitive sound distances: e.g., [a]:[ɑ] versus [a]:[i] (Wieling et al., 2012, JPHON)

The influence of geography

- An important determinant for dialect variation is geographic location (people in nearby locations have more contact than in distant locations)
- We include geography by predicting dialect distances with a Generalized Additive Model (GAM) which models the non-linear interaction between longitude and latitude
  - The fitted values of this GAM are included as a predictor in our model
  - (The details of this procedure are outside the scope of this lecture)

Fitted GAM for dialect distance from standard Dutch



Final model: fixed-effects

	Estimate	Std. Error	t-value
Intercept	-0.0153	0.0105	-1.4561
GAM distance (geography)	0.9684	0.0274	35.3239
Population size (log)	-0.0069	0.0026	-2.6386
Population average age	0.0045	0.0025	1.8049
Population average income (log)	-0.0005	0.0026	-0.1988
Noun instead of Verb/Adjective	0.0409	0.0122	3.3437
Word frequency (log)	0.0198	0.0060	3.2838
Vowel-consonant ratio (log)	0.0625	0.0059	10.5415

*t-values indicate significance if $|t| > 2$ (two-tailed) or $|t| > 1.65$ (one-tailed)

Final model: random effects

Factors	Rnd. effects	Std. Dev.	Cor.	
Word	Intercept	0.1394		
	Pop. size (log)	0.0186		
	Pop. avg. age	0.0086	-0.856	
	Pop. avg. income (log)	0.0161	0.867	-0.749
Location	Intercept	0.0613		
	Word freq. (log)	0.0161	-0.084	
	Noun instead of Verb/Adjective	0.0528	-0.595	0.550
Transcriber	Intercept	0.0260		
Residual		0.2233		

*The inclusion of all random slopes and intercepts was warranted by likelihood-ratio tests

*A richer random effect structure is likely possible, but not computationally feasible (now: 24 CPU hours!)

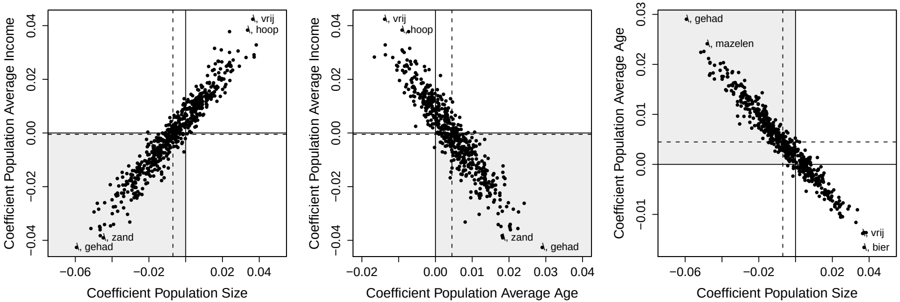
Final model: interesting correlational structure

Factors	Rnd. effects	Std. Dev.	Cor.	
Word	Intercept	0.1394		
	Pop. size (log)	0.0186		
	Pop. avg. age	0.0086	-0.856	
	Pop. avg. income (log)	0.0161	0.867	-0.749
Location	Intercept	0.0613		
	Word freq. (log)	0.0161	-0.084	
	Noun instead of Verb/Adjective	0.0528	-0.595	0.550
Transcriber	Intercept	0.0260		
Residual		0.2233		

*The inclusion of all random slopes and intercepts was warranted by likelihood-ratio tests

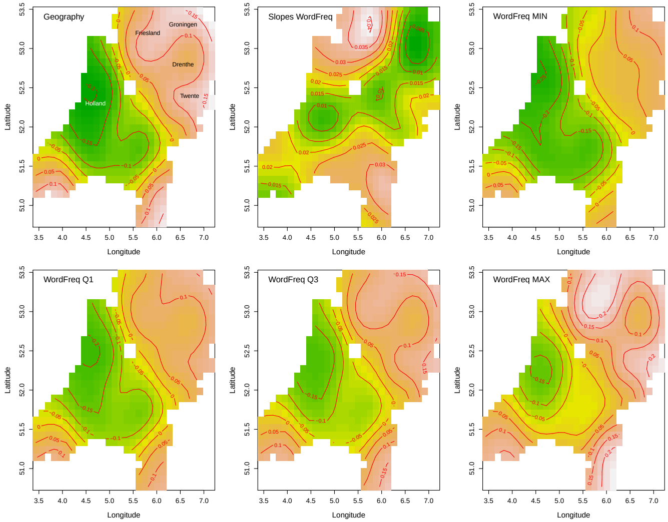
*A richer random effect structure is likely possible, but not computationally feasible (now: 24 CPU hours!)

Correlation structure of by-word random slopes

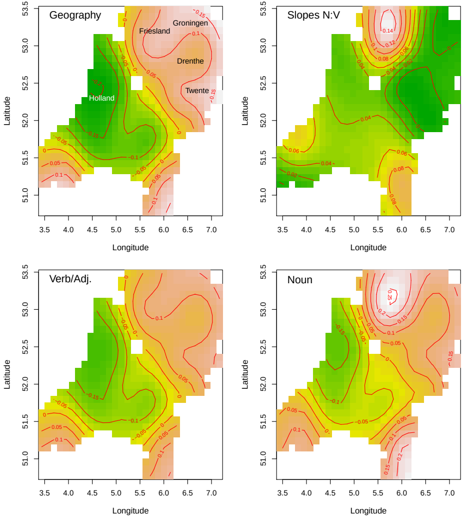


- $LD = -0.0069 PS - 0.0005 PI + 0.0045 PA + \ldots$ (general model)
- $LD = -0.0600 PS - 0.0420 PI + 0.0290 PA + \ldots$ (*gehad*: extreme pattern)
- $LD = \quad 0.0380 PS + 0.0420 PI - 0.0110 PA + \ldots$ (*vrij*: inverted pattern)

By-location random slopes for word frequency



By-location random slopes for Noun-Verb contrast



Case study conclusions

- Our model explained about 45% of the variation in the data with respect to the distance from standard Dutch
- We identified a number of location- and word-related variables playing an important role in predicting the dialect distance from standard Dutch
  - Geography (i.e. social contact between locations)
  - Location-related factors: population size and average age
  - Word-related factors: word category, word frequency and vowel-cons. ratio
- Using a mixed-effects regression approach ensures our results are generalizable and enabled us to quantify and study the variation of individual words and speakers

Time for a short break!

Case study: gender processing in Dutch

- The goal of this study is to investigate if Dutch people use grammatical gender to anticipate upcoming words
  - This study was conducted together with Hanneke Loerts (who did most of the work) and is currently under review (Loerts, Wieling and Schmid, submitted)
- What is grammatical gender?
  - Gender is a property of a noun
  - Nouns are divided into classes: masculine, feminine, neuter, ...
  - E.g., *hond* ('dog') = common, *paard* ('horse') = neuter
- The gender of a noun can be determined from the forms of other elements syntactically related to it (Matthews, 1997: 36)

Gender in Dutch

Gender	Definite article	Adjective in definite NPs	Adjective in Indefinite NPs
Common	De hond	De mooie hond	Een mooie hond
English equivalent	*The$_{com}$ dog$_{com}$*	*The$_{com}$ beautiful dog$_{com}$*	*A beautiful$_{com}$ dog$_{com}$*
Neuter	Het huis	Het mooie huis	Een mooi huis
English equivalent	*The$_{neu}$ house$_{neu}$*	*The$_{neu}$ beautiful house$_{neu}$*	*A beautiful$_{neu}$ house$_{neu}$*

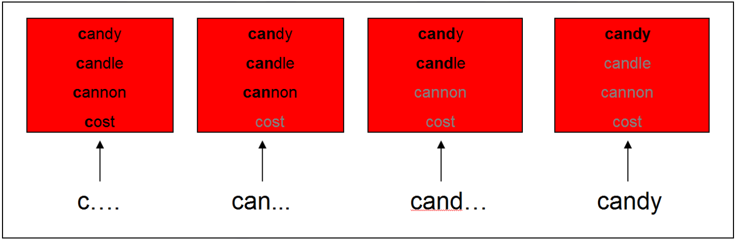
- Gender in Dutch: 70% common, 30% neuter
  - When a noun is diminutive it is always neuter
- Gender is unpredictable from the root noun and hard to learn
  - Children overgeneralize until the age of 6 (Van der Velde, 2004)

Why use eye tracking?

- Eye tracking reveals incremental processing of the listener during the time course of the speech signal
- As people tend to look at what they hear (Cooper, 1974), lexical competition can be tested

Testing lexical competition using eye tracking

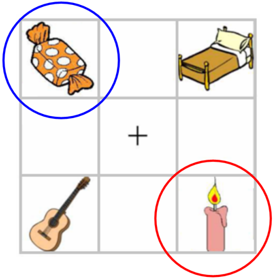
- Cohort Model (Marslen-Wilson & Welsh, 1978): Competition between words is based on word-initial activation



- This can be tested using the visual world paradigm: following eye movements while participants receive auditory input to click on one of several objects on a screen

Support for the Cohort Model

- Subjects hear: “Pick up the candy” (Tanenhaus et al., 1995)
- Fixations towards target (Candy) *and* competitor (Candle): support for the Cohort Model

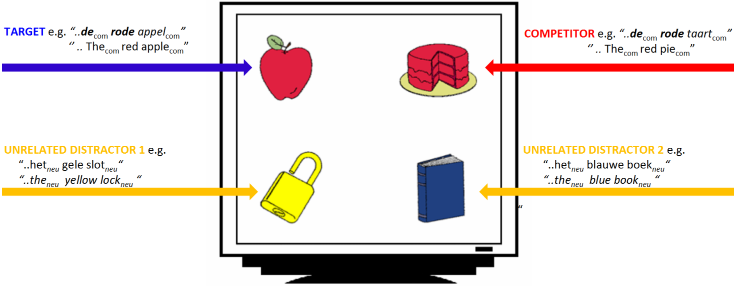


Lexical competition based on syntactic gender

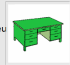
- Other models of lexical processing state that lexical competition occurs based on all acoustic input (e.g., TRACE, Shortlist, NAM)
- Does gender information restrict the possible set of lexical candidates?
  - I.e. if you hear *de*, will you focus more on an image of a dog (*de hond*) than on an image of a horse (*het paard*)?
  - Previous studies (e.g., Dahan et al., 2000 for French) have indicated gender information restricts the possible set of lexical candidates
- In the following, we will investigate if this also holds for Dutch with its difficult gender system using the visual world paradigm
  - We analyze the data using mixed-effects regression in `R`

Experimental design

- 28 Dutch participants heard sentences like:
  - *Klik op de rode appel* ('click on the red apple')
  - *Klik op het plaatje met een blauw boek* ('click on the image of a blue book')
- They were shown 4 nouns varying in color and gender
  - Eye movements were tracked with a Tobii eye-tracker (E-Prime extensions)

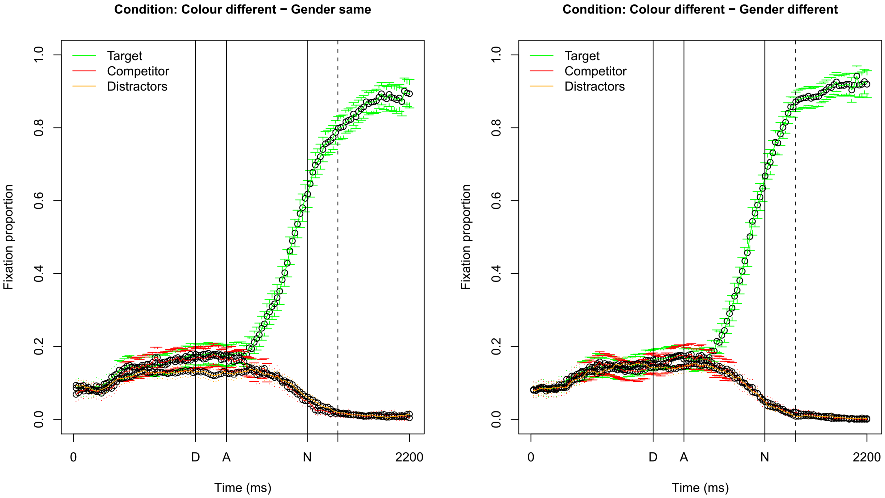


Experimental design: conditions

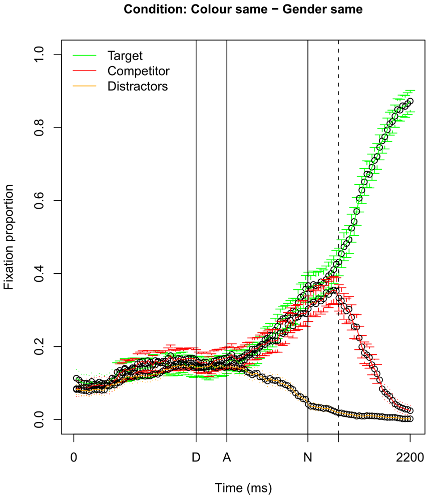
Target	Competitor		Gender Competitor	Colour Competitor
De$_{com}$ rode appel$_{com}$ *The$_{com}$ red apple$_{com}$*	Het$_{neu}$ groene bureau$_{neu}$ *The$_{neu}$ green desk$_{neu}$*		*Different*	*Different*
	De$_{com}$ gele zon$_{com}$ *The$_{com}$ yellow sun$_{com}$*		*Same*	*Different*
	Het$_{neu}$ rode hart$_{neu}$ *The$_{neu}$ red heart$_{neu}$*		*Different*	*Same*
	De$_{com}$ rode taart$_{com}$ *The$_{com}$ red cake$_{com}$*		*Same*	*Same*

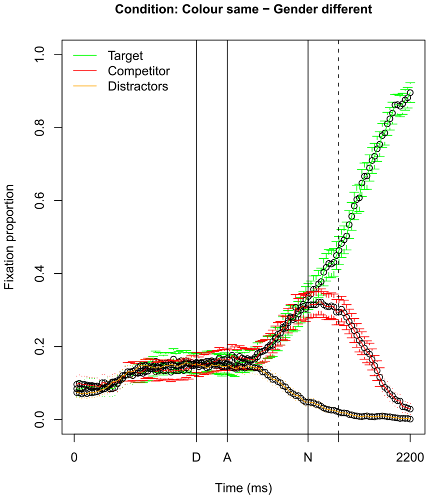
- Subjects were shown 96 different screens
  - 48 screens for indefinite sentences (*klik op het plaatje met een rode appel*)
  - 48 screens for definite sentences (*klik op de rode appel*)

Visualizing fixation proportions: different color



Visualizing fixation proportions: same color





Which dependent variable?

- Difficulty 1: choosing the dependent variable
  - Fixation difference between Target and Competitor
  - Fixation proportion on Target - requires transformation to empirical logit, to ensure the dependent variable is unbounded: $log(\frac{(y+0.5)}{(N-y+0.5)})$
  - ...
- Difficulty 2: selecting a time span
  - Note that about 200 ms. is needed to plan and launch an eye movement
  - Taking every individual sampling point into account is possible but computationally expensive
- In this lecture we use:
  - The difference in fixation between Target and Competitor
  - Averaged over the time span starting 200 ms. after the onset of the determiner and ending 200 ms. after the onset of the noun (about 800 ms.)
  - This ensures that gender information has been heard and processed, both for the definite and indefinite sentences

Independent variables

- Variable of interest
  - Competitor gender vs. target gender
- Variables which could be important
  - Competitor color vs. target color
  - Gender of target (common or neuter)
  - Definiteness of target
- Participant-related variables
  - Gender (male/female), age, education level
  - Trial number
- Design control variables
  - Competitor position vs. target position (up-down or down-up)
  - Color of target
  - ... (anything else you are not interested in, but potentially problematic)

Some remarks about data preparation

- Check if variables correlate highly
  - If so: exclude variable, or transform variable
  - See Chapter 6.2.2 of Baayen (2008)
- Check if numerical variables are normally distributed
  - If not: try to make them normal (e.g., logarithmic transformation)
  - See Chapter 2.2 of Baayen (2008)
  - Note that your dependent variable does not need to be normally distributed (the residuals of your model do!)
- Center your numerical predictors when doing mixed-effects regression
  - If the predictor is not centered, a different random slope (i.e. coefficient) will directly result in a different intercept and this will result in uninformative correlations of this random slope and the random intercept
  - See Chapter 7.1 and Figure 7.5 of Baayen (2008)

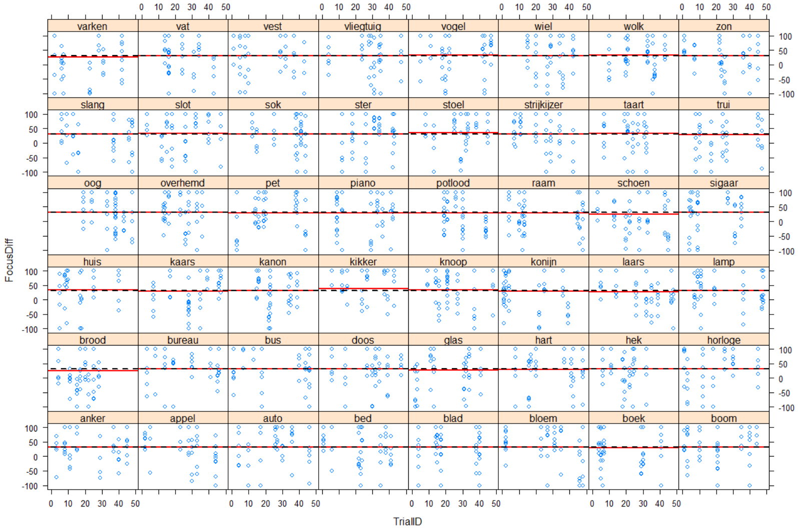
Our data

```
> head(eye)

  Subject  Item TargetDefinite TargetNeuter TargetColor TargetBrown TargetPlace
1    S300 appel              1            0         red           0           1
2    S300 appel              0            0         red           0           2
3    S300   vat              1            1       brown           1           4
4    S300   vat              0            1       brown           1           1
5    S300  boek              1            1        blue           0           4
6    S300  boek              0            1        blue           0           1
  TargetTopRight CompColor CompPlace TupCdown CupTdown TrialID Age IsMale
1              0       red         2        0        0      44  52      0
2              1     brown         4        1        0       2  52      0
3              0    yellow         2        0        1      14  52      0
4              0     brown         3        1        0      43  52      0
5              0      blue         3        0        0       5  52      0
6              0    yellow         3        1        0      30  52      0
  Edulevel SameColor SameGender TargetPerc  CompPerc  FocusDiff
1        1         1          1   40.90909  6.818182  34.090909
2        1         0          0   63.63636  0.000000  63.636364
3        1         0          0   47.72727 43.181818   4.545455
4        1         1          0   27.90698  9.302326  18.604651
5        1         1          0   11.11111 25.000000 -13.888889
6        1         0          1   23.80952 50.000000 -26.190476
```

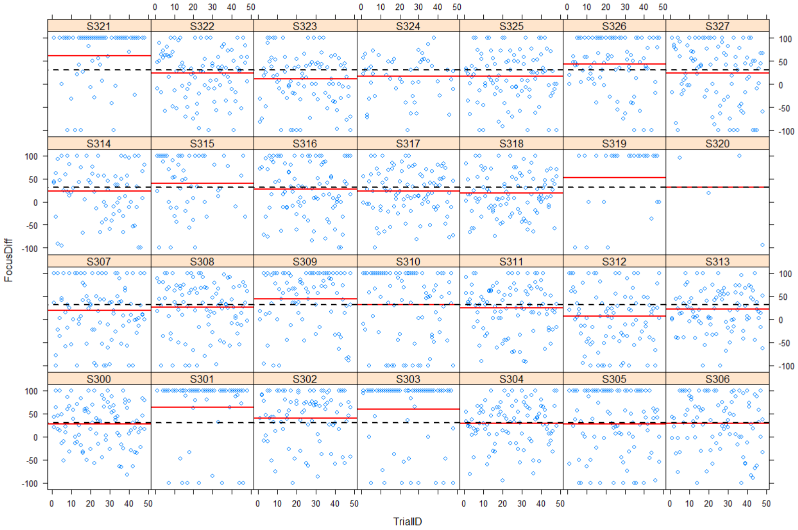
Our very first mixed-effects regression model

```
# Comments are preceded by #
# Commands are preceded by >
# Results are preceded by nothing

# A model having only random intercepts for Subject and Item
> model = lmer( FocusDiff ~ (1|Subject) + (1|Item) , data=eye )

# Show the results of the model
> print( model, corr=F )

[...]
Random effects:
 Groups   Name        Variance Std.Dev.
 Item     (Intercept)   22.968  4.7925
 Subject  (Intercept)  257.111 16.0347
 Residual             3275.691 57.2336
Number of obs: 2280, groups: Item, 48; Subject, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)   30.867      3.377    9.14
```

By-item random intercepts



By-subject random intercepts



Is a by-item analysis necessary?

```
# comparing two models
> model1 = lmer(FocusDiff ~ (1|Subject), data=eye)
> model2 = lmer(FocusDiff ~ (1|Subject) + (1|Item), data=eye)
> anova( model1 , model2 )

Data: eye
Models:
model1: FocusDiff ~ (1 | Subject)
model2: FocusDiff ~ (1 | Subject) + (1 | Item)
       Df   AIC   BIC logLik  Chisq Chi Df Pr(>Chisq)
model1  3 25001 25018 -12497
model2  4 25000 25023 -12496 2.0772      1     0.1495
```

- `anova` always compares the simplest model (above) to the more complex model (below)
- The *p*-value > 0.05 indicates that there is no support for the by-item random slopes
  - This indicates that the different conditions were very well controlled in the research design

Adding a fixed-effect factor

```
# model with fixed effects, but no random-effect factor for Item
> model3 = lmer(FocusDiff ~ SameColor + (1|Subject), data=eye)
> print(model3, corr=F)

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept)  211.22  14.534
 Residual             2778.65  52.713
Number of obs: 2280, groups: Subject, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)   53.067      3.189   16.64
SameColor    -45.858      2.217  -20.69
```

- `SameColor` is highly important as $|t| > 2$
  - negative estimate: more difficult to distinguish target from competitor
- We need to test if the effect of `SameColor` varies per subject
  - If there is much between-subject variation, this will influence the significance of the variable in the fixed effects

Testing for a random slope

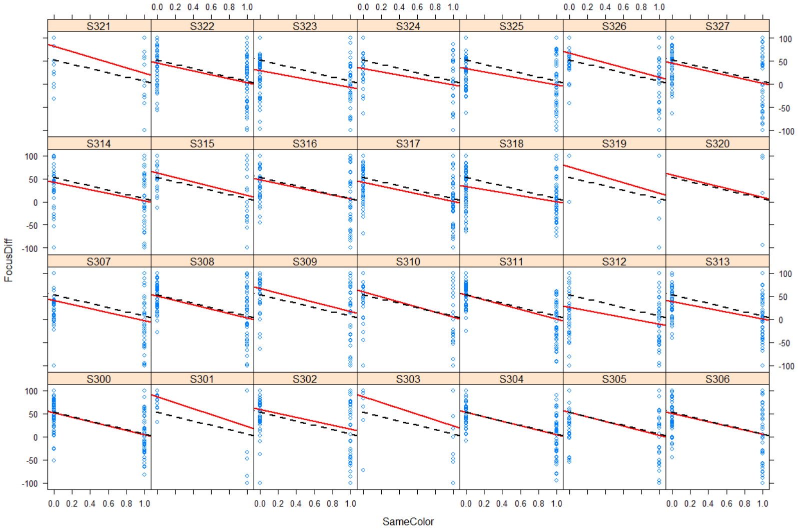
```
# a model with an uncorrelated random slope for SameColor per Subject
> model4 = lmer(FocusDiff ~ SameColor + (1|Subject) + (0+SameColor|Subject),
                data=eye)
> anova(model3,model4)

model3: FocusDiff ~ SameColor + (1 | Subject)
model4: FocusDiff ~ SameColor + (1 | Subject) + (0 + SameColor | Subject)
       Df   AIC   BIC logLik Chisq Chi Df Pr(>Chisq)
model3  4 24610 24633 -12301
model4  5 24611 24640 -12301 0.738      1     0.3903


# model4 is no improvement, what about a model with a random slope for
# SameColor per Subject correlated with the random intercept
> model5 = lmer(FocusDiff ~ SameColor + (1+SameColor|Subject), data=eye)
> anova(model3,model5)

model3: FocusDiff ~ SameColor + (1 | Subject)
model5: FocusDiff ~ SameColor + (1 + SameColor | Subject)
       Df   AIC   BIC logLik  Chisq Chi Df Pr(>Chisq)
model3  4 24610 24633 -12301
model5  6 24603 24637 -12295 11.111      2   0.003866 **
```

Investigating the model structure

```
> print(model5, corr=F)

Linear mixed model fit by REML
Formula: FocusDiff ~ SameColor + (1 + SameColor | Subject)
   Data: eye
   AIC   BIC logLik deviance REMLdev
 24595 24629 -12292    24591   24583
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept)  335.73  18.323
          SameColor    115.33  10.739   -0.855
 Residual             2754.06  52.479
Number of obs: 2280, groups: Subject, 28

Fixed effects:
            Estimate Std. Error t value
(Intercept)   53.021      3.849   13.78
SameColor    -46.511      3.035  -15.32
```

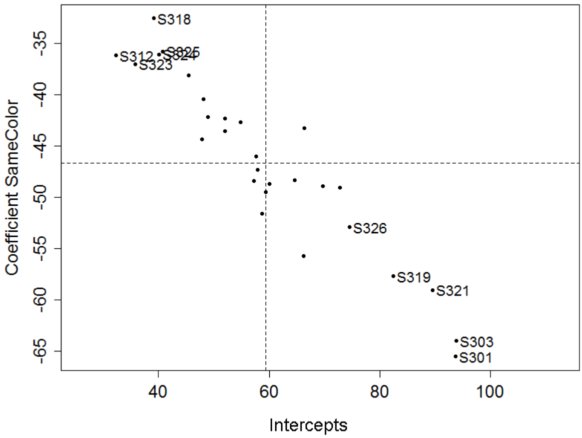
- Note `SameColor` is still highly significant as the $|t| > 2$ (absolute value)

By-subject random slopes



Correlation of random intercepts and slopes

$r = -0.855$

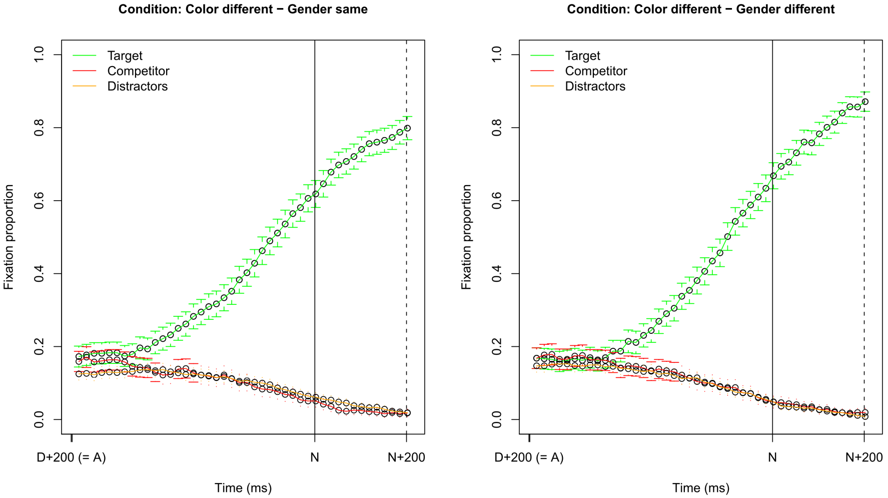


Investigating the gender effect

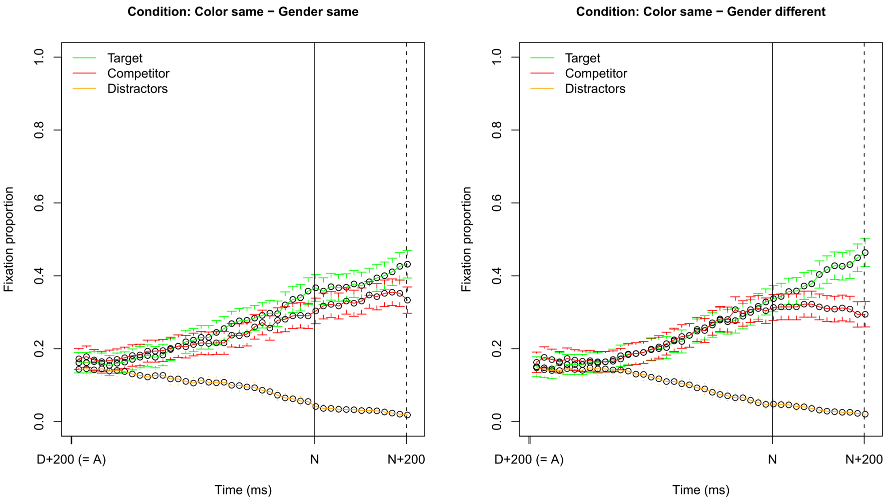
```
> model6 = lmer(FocusDiff ~ SameColor + SameGender + (1+SameColor|Subject),
                data=eye)
> print(model6, corr=F)

Fixed effects:
            Estimate Std. Error t value
(Intercept)  53.1401     3.9952  13.301
SameColor   -46.5076     3.0348 -15.325
SameGender   -0.2454     2.2008  -0.112
```

- It seems there is no gender effect...
- Perhaps we can take a look at the fixation proportions again (now within our time span)

Visualizing fixation proportions: different color



Visualizing fixation proportions: same color



There might be an interaction?

```
> model7 = lmer(FocusDiff ~ SameColor + SameGender * TargetDefinite +
                           (1+SameColor|Subject), data=eye)
> print(model7, corr=F)

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                53.0880     4.3022  12.340
SameColor                 -46.5098     3.0344 -15.327
SameGender                 -0.5754     3.1305  -0.184
TargetDefinite              0.1003     3.1082   0.032
SameGender:TargetDefinite   0.6581     4.4011   0.150
```

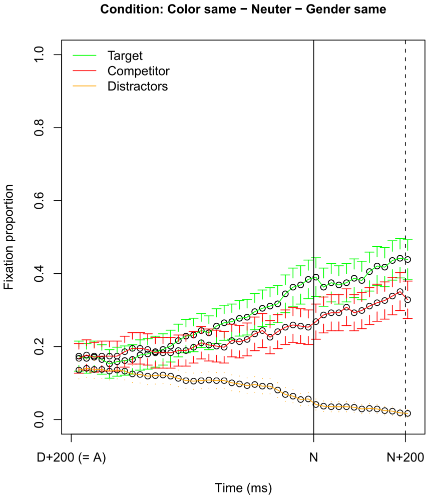
- An interaction is specified by a $*$ in the model specification
- No interaction present between definiteness of the target and same gender of competitor and target (*t*-values lower than |2|)...

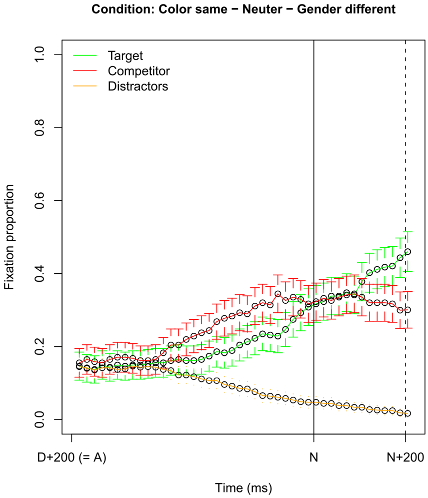
Another interaction?

```
> model8 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                           (1+SameColor|Subject), data=eye)
> print(model8, corr=F)

Fixed effects:
                        Estimate Std. Error t value
(Intercept)               59.373      4.290  13.839
SameColor                -46.692      3.033 -15.394
SameGender                -7.622      3.080  -2.475
TargetNeuter             -12.464      3.096  -4.027
SameGender:TargetNeuter   14.974      4.386   3.414
```

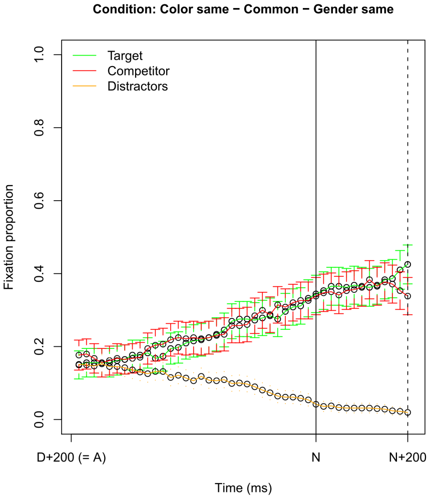
- There is clear support for an interaction (all $|t| > 2$)
- Can we see this in the fixation proportion graphs?

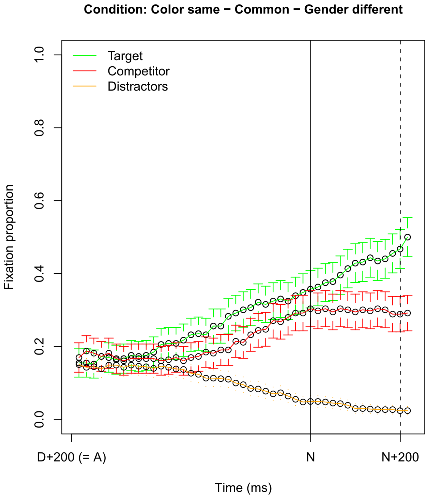
Visualizing fixation proportions: target neuter





Visualizing fixation proportions: target common





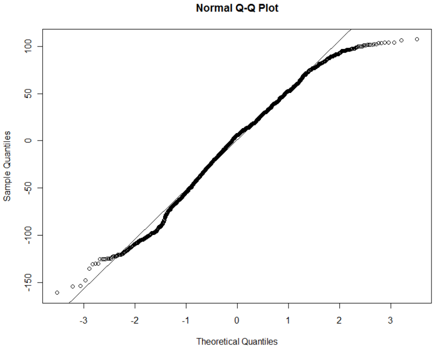
Testing if the interaction yields an improved model

```
# To compare models differing in fixed effects, we specify REML=F.
# We compare to the best model we had before, and include TargetNeuter as
# it is also significant by itself.

> model8a = lmer(FocusDiff ~ SameColor + TargetNeuter +
                             (1+SameColor|Subject), data=eye, REML=F)
> model8b = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                             (1+SameColor|Subject), data=eye, REML=F)
> anova(model8a,model8b)

        Df   AIC   BIC logLik  Chisq Chi Df Pr(>Chisq)
model8a  7 24600 24640 -12293
model8b  9 24592 24644 -12287 11.656      2   0.002944 **
```

- The interaction improves the model significantly
  - Unfortunately, we do not have an explanation for the strange neuter pattern
- Note that we still need to test the variables for inclusion as random slopes (we do this in the lab session)

How well does the model fit?

```
# "explained variance" of the model (r-squared)
> cor( eye$FocusDiff , fitted( model8 ) )^2
[1] 0.2347549

> qqnorm( resid( model8 ) )
> qqline( resid( model8 ) )
```



Adding a factor and a continuous variable

```
# set a reference level for the factor
> eye$TargetColor = relevel( eye$TargetColor , "brown" )
> model9 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                            TargetColor + Age +
                            (1+SameColor|Subject), data=eye)
> print(model9, corr=F)

Fixed effects:
                        Estimate Std. Error t value
(Intercept)              81.7602    17.7363   4.610
SameColor               -46.8085     3.0221 -15.489
SameGender               -7.5730     3.0641  -2.472
TargetNeuter            -12.4801     3.0794  -4.053
TargetColorblue          11.6108     3.5878   3.236
TargetColorgreen         15.3901     3.5768   4.303
TargetColorred           16.5083     3.5798   4.612
TargetColoryellow        16.0423     3.5931   4.465
Age                      -0.7301     0.3591  -2.033
SameGender:TargetNeuter  14.8669     4.3637   3.407
```

Converting the factor to a contrast

```
> eye$TargetBrown = (eye$TargetColor == "brown")*1
> model10 = lmer(FocusDiff ~ SameColor + SameGender * TargetNeuter +
                             TargetBrown + Age +
                             (1+SameColor|Subject), data=eye)
> print(model10, corr=F)
Fixed effects:
                        Estimate Std. Error t value
(Intercept)              96.6001    17.5691   5.498
SameColor               -46.8328     3.0300 -15.456
SameGender               -7.5911     3.0635  -2.478
TargetNeuter            -12.5225     3.0789  -4.067
TargetBrown             -14.8923     2.9256  -5.090
Age                      -0.7284     0.3588  -2.030
SameGender:TargetNeuter  14.8965     4.3626   3.415

# model9b and model10b: REML=F
> anova(model9b, model10b)
         Df   AIC   BIC logLik  Chisq Chi Df Pr(>Chisq)
model10b 11 24566 24629 -12272
model9b  14 24570 24650 -12271 2.6164      3     0.4546
```

Many more things to do...

- We need to see if the significant fixed effects remain significant when adding these variables as random slopes per subject
- There are other variables we should test (e.g., education level)
- There are other interactions we can test
- We will experiment with these issues in the lab session (Friday, 9-11)!
  - We use a subset of the data (only color competitors)
  - Simple `R`-functions are used to generate all plots

What you should remember...

- Mixed-effects regression models offer an easy-to-use approach to obtain generalizable results even when your design is not completely balanced
- Mixed-effects regression models allow a fine-grained inspection of the variability of the random effects, which may provide additional insight in your data
- Mixed-effects regression models are easy in `R`
  - Lab session: this Friday, 9:00 - 11:00

Thank you for your attention!