

Mixed models

Jonathan Harrington

Die R-Befehle: mixed.txt

```
library(lme4)
```

```
lex = read.table(paste(pfad, "lex.txt", sep="/"))
```

Mixed models

Baayen, R.H. (2009) *Analyzing Linguistic Data: A practical introduction to Statistics*. Kapitel 7 <http://www.ualberta.ca/~baayen/publications/baayenCUPstats.pdf>

Artikel in einem Special Issue im *Journal of Memory and Language*, Vol. 59. insbesondere: Baayen, Davidson & Bates (2008); Quene & van den Bergh (2008); Jaeger (2008).

2 Präsentationen hier vorhanden

<http://hlplab.wordpress.com/2009/05/03/multilevel-model-tutorial/>

Levy & Jaeger (2009) A Brief and Friendly Introduction to Mixed-Effects Models in Psycholinguistics

Frank & Jaeger (April, 2009) Post hoc comparisons Additional Issues: Random effects diagnostics, multiple comparisons

Erste Veröffentlichung: Pinheiro & Bates (2000).

<http://www.amazon.com/Mixed-Effects-Models-S-S-Plus/dp/0387989579>

Mixed model (MM)

Ein MM ist eine Art von Regression in dem ein Response (abhängige Variable) aus einer Kombinationen von gewichteten Faktoren eingeschätzt wird.

Lineares Modell, Minimierung vom Abstand zwischen tatsächlichen und eingeschätzten Werten.

(Das Verfahren um dies zu tun, ist nicht least-squares wie in Regression sondern maximum-likelihood)

Mixed Model und Faktoren

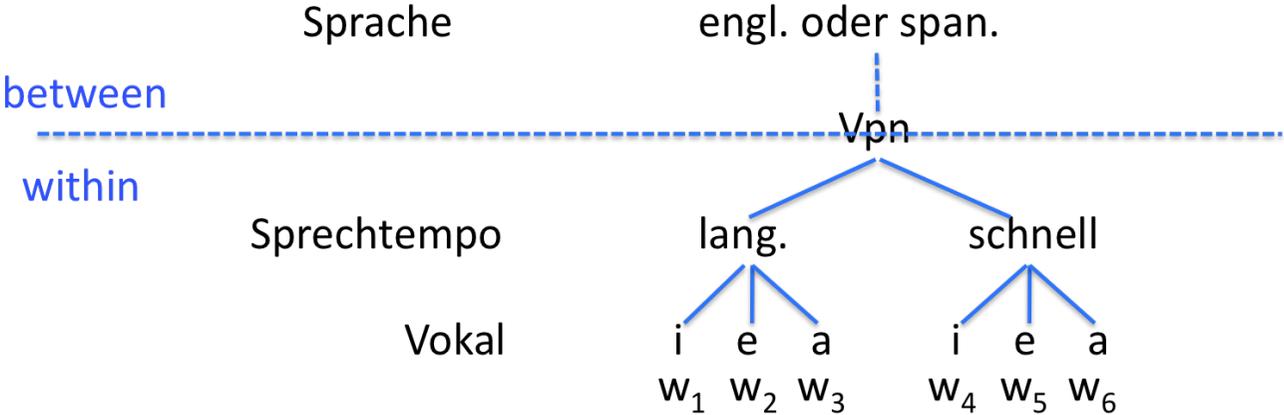
Im MM-Verfahren wird prinzipiell zwischen 2 verschiedene Sorten von Faktoren differenziert

- Fixed: Faktoren, die geprüft werden sollen
- Random: Faktoren, die ausgeklammert werden sollen

Vergleich: MM und RM-ANOVA

Die Kieferposition wurde in 3 Vokalen /i, e, a/ und jeweils zu 2 Sprechtempi (langsam, schnell) gemessen. Die Messungen sind von 8 mit Muttersprache spanisch, 8 mit Muttersprache englisch aufgenommen worden.

RM-Anova



MM

Fixed: Sprache, Sprechtempo, Vokal

soll geprüft werden

Random: Sprecher

soll ausgeklammert werden

Mixed model (MM) und RM-ANOVA

F1 und die Dauer wurden in 3 Vokalen /i, e, a/ und jeweils in betonten und unbetonten Silben gemessen. Die Messungen sind von 10 Vpn aufgenommen worden. Inwiefern wird F1 von der Dauer, Vokal und Betonung beeinflusst?

RM-Anova

MM

Fixed: Dauer, Vokal, Sprechtempo

Random: Sprecher

Vergleich: MM und RM-ANOVA

F1 von /a:/ wurde in 100 verschiedenen Wörtern gemessen (*Bart, Pfad, mager, maßgebend, erstarrt...*). Die Wörter wurden von 10 Vpn produziert sowohl phrasenmedial als auch phrasenfinal. Inwiefern wird F1 von Phrasenposition beeinflusst (N.B. F1 variiert sehr stark wegen Kontext, also von Wort zu Wort).

(a) Wir wollen die Sprechervariation ausklammern (interessiert uns nicht).

(b) Wir wollen aber auch die Wortvariation ausklammern (dass F1 unterschiedliche Werte hat in *Bart vs. mager* usw. interessiert uns nicht).

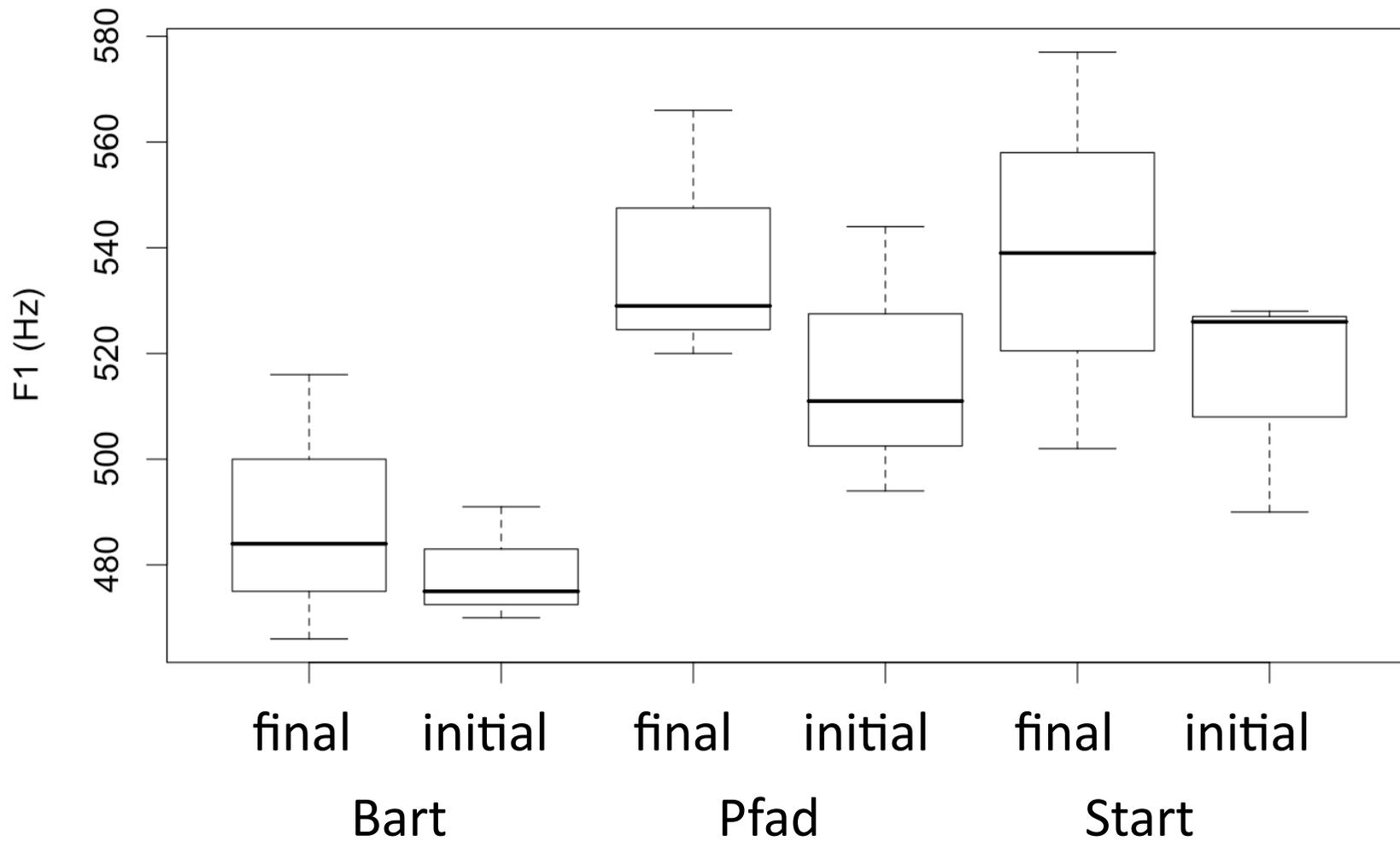
(a) und (b) gleichzeitig ausklammern in einem RM-Anova geht nicht.

MM

fixed: Phrasenposition

random: Sprecher, Wort

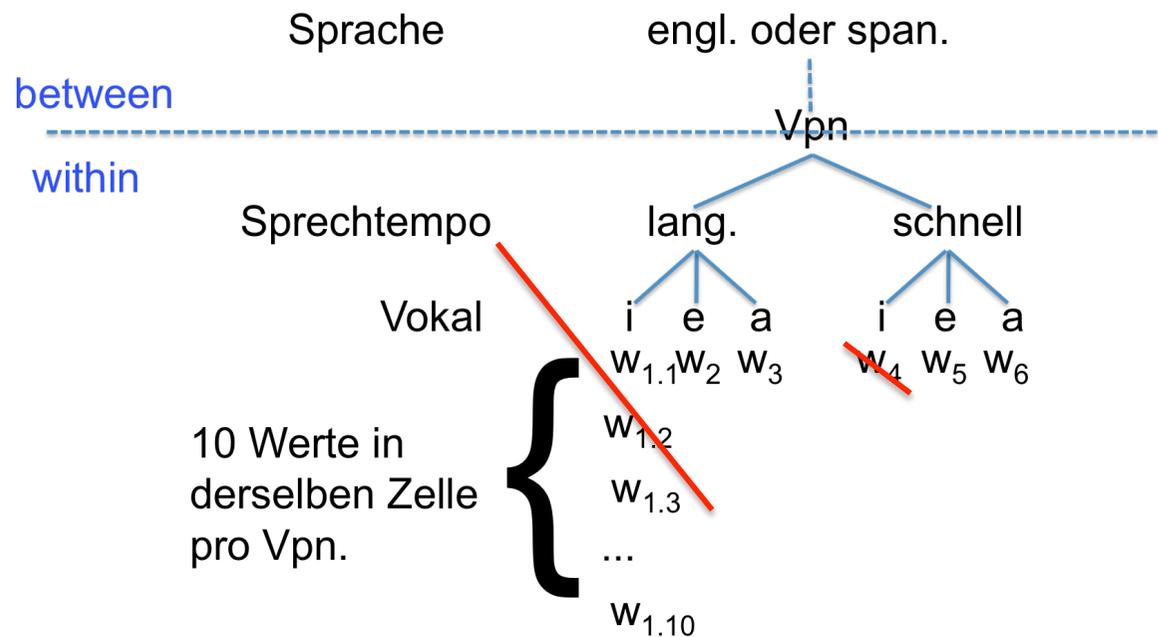
(b) Wir wollen auch die Wortvariation ausklammern (dass F1 unterschiedliche Werte hat in *Bart* vs. *mager* usw. interessiert uns nicht).



Vergleich: MM und RM-ANOVA

In einem MM:

- müssen die Zellen nicht vollständig sein
- es muss nicht über Wiederholungen gemittelt werden



- MM basiert auf maximum likelihood – dass die Werte normalverteilt sind, wird nicht vorausgesetzt

Nachteile von einem Mixed-Model

library(lme4) und mixed modelling (MM) überhaupt sind noch in der Entwicklungsphase. Daher bugs, häufige code Änderungen und einige Teile des Verfahrens sind in R noch nicht ganz vollständig.

Mit MM können zwar Werte aus der t- und F-Verteilung berechnet werden, aber diese lassen sich nur schwierig und vielleicht sogar ungenau in Wahrscheinlichkeiten umsetzen – weil die Freiheitsgrade nicht eindeutig berechnet werden können.

Wann soll der MM verwendet werden?

1. Wenn ein RM-Anova nicht/kaum einsetzbar ist:

- Ein oder mehrere Faktoren sind kontinuierlich
- mehr als ein Random Faktor vorhanden sind.
- unvollständige Daten ('fehlende Zellen')
- man will nicht über Wiederholungen mitteln
- der Response ist kategorial (0 vs 1; ja vs nein) – also eine Form logistischer Regression

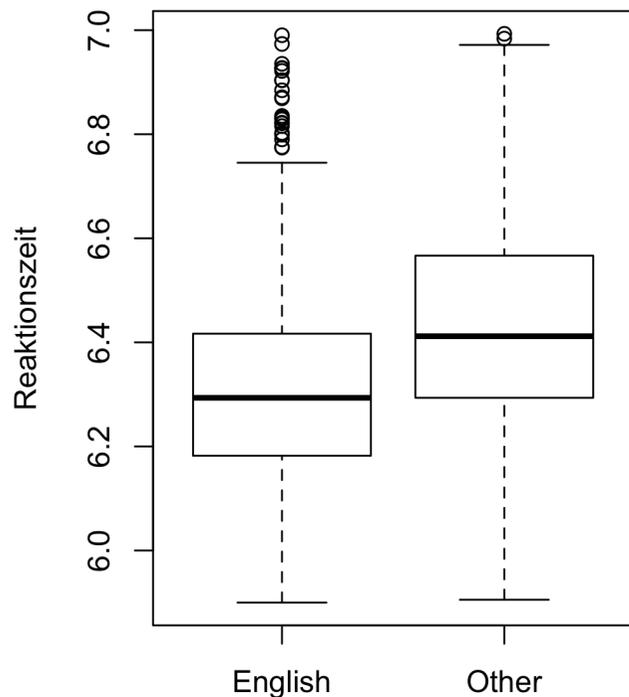
2. Ein MM ist nur robust bei einer großen Anzahl von Stichproben – mindestens insgesamt 200-300, je mehr umso besser.

Abbildungen

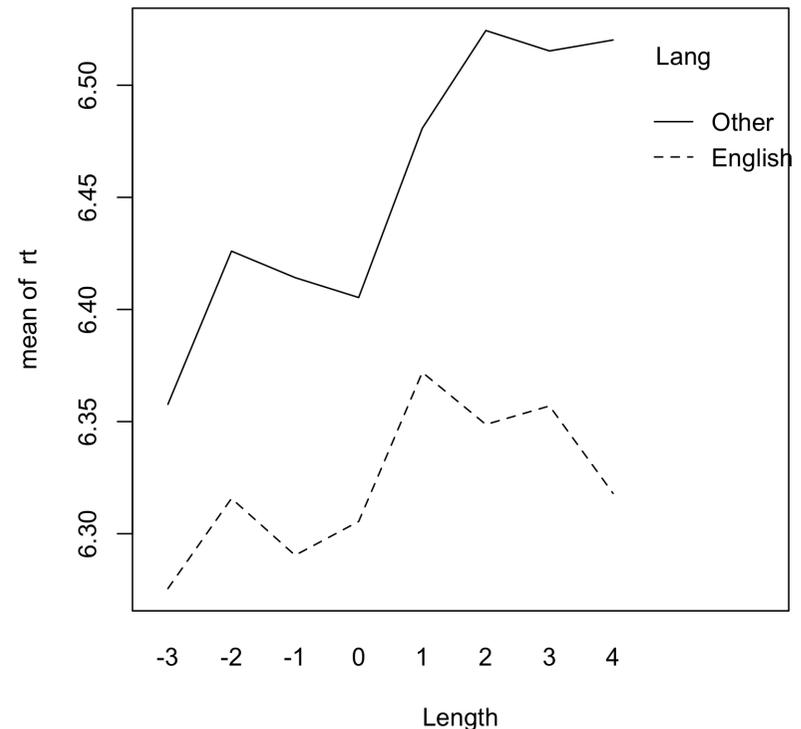
boxplot() (oder bwplot()) wenn der Fixed-Factor kategorial ist,
sonst durch interaction.plot() einschätzen

```
with(lex, boxplot(rt ~ Lang, ylab = "Reaktionszeit"))
```

```
with(lex, interaction.plot(Lang, Length, rt))
```



Lang signifikant?



Length sig? (= liegt ein Trend vor?)

Lang * Length Interaktion? (bedeutet...?)

MM Anwendung in R

keine Interaktion

$\text{Imer}(\text{response} \sim \text{fixed1} * \text{fixed2} * \dots + (1 | \text{Random1}) + (1 | \text{Random2}) + \dots$

$\text{lex.Imer} = \text{Imer}(\text{rt} \sim \text{Lang} * \text{Length} + (1 | \text{Subj}) + (1 | \text{Word}), \text{data} = \text{lex})$

$\text{fitted}(\text{lex.Imer})$

eingeschätzter RT

$$\hat{\text{rt}} = b_1 \text{Lang} + b_2 \text{Length} + b_3 (\text{Lang} \times \text{Length}) + k + k_{\text{Subj}} + k_{\text{Word}}$$

by-subject
intercept
adjustment

by-word
intercept
adjustment

Hier werden 3 Neigungen und 3 Intercepts (also 6 Parameter) berechnet, sodass der Abstand zwischen rt und $\hat{\text{rt}}$ minimiert wird.

Beitrag der Fixed Factors

Haben Lang und/oder Length einen Einfluss auf die Reaktionszeiten?

`anova(lex.lmer)`

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value
Lang	1	0.13993	0.13993	6.2512
Length	1	0.47000	0.47000	20.9972
Lang:Length	1	0.36477	0.36477	16.2961

Je größer der F-Wert, umso wichtiger ist der Faktor. Es werden aber keine Wahrscheinlichkeiten berechnet, da die Freiheitsgrade in einem MM-Modell im Nenner nicht festgestellt werden können.

MM und Wahrscheinlichkeiten

Reubold, Harrington & Kleber, 2010, *Speech Communication*

"One of the difficulties with mixed models is in determining the number of degrees of freedom in the denominator. An anti-conservative estimate can be obtained from $df = n - k - 1$ where n is the number of observations and k the number of degrees of freedom (Baayen, 2008). Instead of using this anti-conservative estimate, we set df to be equal to the more conservative value of 60 and chose an alpha level of 0.01. Part of the motivation for choosing a somewhat arbitrary value of 60 is that for $df > 60$ there is a fairly small change to the F-value for which significance is obtained. For example, the F values at $\alpha = 0.01$ are $F[1, 60] = 8.49$ and $F[1, 600] = 7.94$, i.e., an F-value change of 0.55 for a change in df from 60 to 600."

also: so lange man genügend Stichproben hat, sagen wir dass alles über ca. $F = 8.49$ "signifikant" ist.

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value
Lang	1	0.13993	0.13993	6.2512
Length	1	0.47000	0.47000	20.9972
Lang:Length	1	0.36477	0.36477	16.2961

Length hatte einen signifikanten Einfluss auf die Reaktionszeiten ($F = 21.0$, $p < 0.01$) und es gab eine signifikante Length x Language Interaktion ($F = 16.3$, $p < 0.01$). Der Einfluss von Language auf die Reaktionszeiten war nicht signifikant.

Interaktionen prüfen

Es gibt keine robuste Methodik, um post-hoc Tests nach einem MM anzuwenden. Daher einfach den MM noch einmal getrennt für die Faktoren durchführen.

d.h. hat Length einen Einfluss auf die Reaktionszeiten (a) für englisch; (b) für andere Sprachen?

```
temp = with(lex, Lang) == "English"
```

```
engl = lex[temp,]          other = lex[!temp,]
```

```
engl.lmer = lmer(rt ~ Length + (1 | Subj) + (1 | Word), data = engl)
```

```
other.lmer = lmer(rt ~ Length + (1 | Subj) + (1 | Word), data = other)
```

```
anova(engl.lmer)
```

	Df	Sum Sq	Mean Sq	F value
Length	1	0.188	0.188	8.1661

NS

```
anova(other.lmer)
```

	Df	Sum Sq	Mean Sq	F value
Length	1	0.62451	0.62451	29.523

Sig, $p < 0.01$

Der Data-Frame asp enthält Werte der Aspirationsdauer von silbeninitialem /t/ und /k/ aus gelesenen Sätzen in dem Kielcorpus. Diese Dauern sind für 55 Versuchspersonen und 287 Wörter erhoben worden. (Die Versuchspersonen produzierten nicht alle dieselben Wörter). Inwiefern wird die Aspirationsdauer von der Artikulationsstelle (/k/, /t/) oder von der Silbenbetonung ("betont", "unbetont") beeinflusst?