Auswertung - Linear Mixed Effect Models

Fritz Günther
fritz.guenther@uni-tuebingen.de

Universität Tübingen

June 30, 2021

Übersicht

- Ausreißerkorrektur
- Lineare Modelle allgemein
- Gemischte Lineare Modelle
- Hypothesentests/ Modellvergleiche
- Berichten der Ergebnisse

Verfahren zur Ausreißerkorrektur

Das wichtigste zuerst

Das Verfahren zur Ausreißerkorrektur sollte **vor** der Auswertung festgelegt werden

Ausschluss von Versuchspersonen

Datensätze ganzer Versuchspersonen sollten nur nach eindeutig vorher festgelegten Regeln ausgeschlossen werden

Fehlertrials

Für gewöhnlich sollten Trials mit fehlerhafter Antwortreaktion ausgeschlossen werden

Begründung: Man beobachtet nicht den Prozess, der interessiert

Ausschließen aller Trials für die gilt:

$$\text{RT} < y \text{ ms oder RT} > z \text{ ms}$$

Vorteile?
Nachteile?

Ausschließen der Trials mit den x % höchsten und/ oder niedrigsten RTs

oder

Ausschließen aller Trials, deren RT mehr als eine (zwei, drei) Standardabweichungen über/ unter dem Mittelwert liegt

Vorteile?
Nachteile?

Problem: Angepasster Cut-Off wie eben besprochen kann z.B. besonders langsame Vps systematisch ausschließen

Lösungsansatz: Vor Anwendung des angepassten Cut-Offs zunächst die Daten innerhalb jeder VP Z-transformieren

Problem: Dies kann immer noch dazu führen, dass besonders langsame Bedingungen (z.B. inkongruente) oder besonders langsame Items systematisch ausgeschlossen werden

Lösungsansatz (siehe Kaup, Lüdtke & Zwaan, 2006):

- Zunächst Werte innerhalb jeder VP Z-transformieren
- Dann die erhaltenen Werte innerhalb jedes Items innerhalb jeder Bedingung Z-transformieren (z.B. in R durch geschachtelte for-Schleifen)
- Dann angepassten Cut-Off anwenden

Ausschluss von Trials

Harte und angepasste Cut-Offs können auch kombiniert werden

Linear Mixed Effect Models

Literatur

Tutorials:

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science, 4,* 1–19.

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]

Literatur (Einführung):

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.

Lineare Modelle

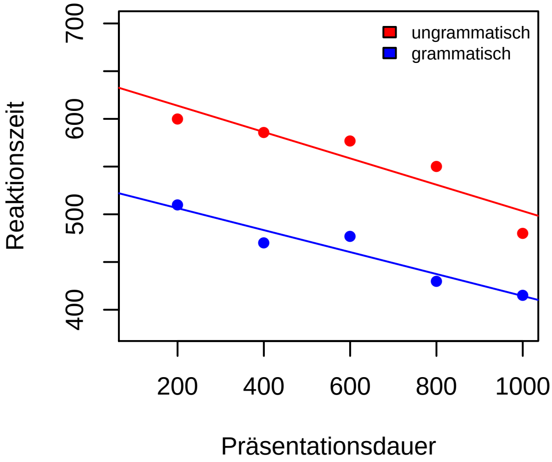
$$Y = a + b + \epsilon$$

Y: Abhängige Variable ("Kriterium")
a: Unabhängige Variable 1 ("Prädiktor 1")
b: Unabhängige Variable 2 ("Prädiktor 2")
ϵ: Zufälliger Fehler

Beispiel:

$$\text{Reaktionszeit} = \text{Präsentationsdauer} + \text{Grammatikalität} + \epsilon$$

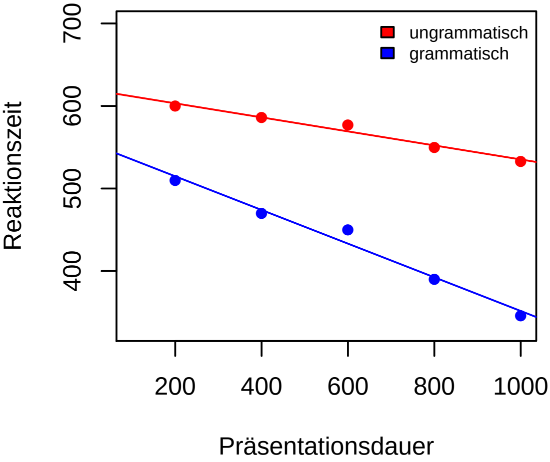
Beispieldaten



$$Y = a + b + ab + \epsilon$$

Y: Abhängige Variable ("Kriterium")
a: Unabhängige Variable 1 ("Prädiktor 1")
b: Unabhängige Variable 2 ("Prädiktor 2")
ab: Interaktionseffekt beider Prädiktoren
ϵ: Zufälliger Fehler

Beispieldaten mit Interaktion



Feste vs zufällige Effekte

Feste Effekte: Erhobene Faktorstufen sind *Vollerhebung* der interessierenden Faktorstufen

Beispiele: Präsentationsdauer, Präsentation eines grammatischen vs ungrammatischen Satzes
$\Longrightarrow$ Keine Generalisierung nötig

Feste vs zufällige Effekte

Feste Effekte: Erhobene Faktorstufen sind *Vollerhebung* der interessierenden Faktorstufen

Beispiele: Präsentationsdauer, Präsentation eines grammatischen vs ungrammatischen Satzes
$\Longrightarrow$ Keine Generalisierung nötig

Zufällige Effekte: Erhobene Faktorstufen sind *Teilstichprobe* der interessierenden Faktorstufen

Beispiele: Versuchspersonen (VPs), Items
$\Longrightarrow$ Generalisierung erwünscht

Grundproblem

Welches sind die Annahmen linearer Modelle?

Grundproblem

Welches sind die Annahmen linearer Modelle?

- Modellresiduen $\sim N(0, \sigma^2)$

  Annahme zur Fehlerverteilung

  - Normalverteilt
  - Unsystematisch
  - Homoskedastizität

Grundproblem

Welches sind die Annahmen linearer Modelle?

- Modellresiduen $\sim N(0, \sigma^2)$

  Annahme zur Fehlerverteilung

  - Normalverteilt
  - Unsystematisch
  - Homoskedastizität

- **Unabhängigkeit der Datenpunkte!**

Grundproblem

Messwiederholungen liefern Datenpunkte, die *nicht* unabhängig sind; entsprechende Parameter müssen ins Modell aufgenommen werden

Grundproblem

Messwiederholungen liefern Datenpunkte, die *nicht* unabhängig sind; entsprechende Parameter müssen ins Modell aufgenommen werden

Verschiedene VPs sind i.A. unterschiedlich schnell
$\Longrightarrow$ Zufällige Effekte für VPs im Modell (F_1 ANOVA)

$$Y = a + b + ab + (1|subject) + \epsilon$$

Grundproblem

Messwiederholungen liefern Datenpunkte, die *nicht* unabhängig sind; entsprechende Parameter müssen ins Modell aufgenommen werden

Verschiedene VPs sind i.A. unterschiedlich schnell
$\Longrightarrow$ Zufällige Effekte für VPs im Modell (F_1 ANOVA)

$$Y = a + b + ab + (1|subject) + \epsilon$$

Verschiedene Items werden i.A. unterschiedlich schnell bearbeitet
$\Longrightarrow$ Zufällige Effekte für Items im Modell (F_2 ANOVA)

$$Y = a + b + ab + (1|item) + \epsilon$$

Grundproblem

Gemischte Lineare Modelle

$$Y = a + b + ab + (1|subject) + (1|item) + \epsilon$$

Gemischte Lineare Modelle

$$Y = a + b + ab + (1|subject) + (1|item) + \epsilon$$

Wieso nicht gleich?

$$Y = a + b + ab + (1|subject) + (1|item) + \epsilon$$

Wieso nicht gleich?

- Schätzung der Modelle war lange sehr aufwendig

Gemischte Lineare Modelle

$$Y = a + b + ab + (1|subject) + (1|item) + \epsilon$$

Wieso nicht gleich?

- Schätzung der Modelle war lange sehr aufwendig
- Implementiert "an sich" keine Signifikanztests

Berechnen eines LMEM

- Die Pakete `lme4` und `lmerTest` installieren:
  ```
  install.packages("lme4")
  install.packages("lmerTest")
  ```

Berechnen eines LMEM

- Die Pakete `lme4` und `lmerTest` installieren:
  ```
  install.packages("lme4")
  install.packages("lmerTest")
  ```
- Die Pakete laden:
  ```
  library(lme4)
  library(lmerTest)
  ```

Berechnen eines LMEM

- Wichtig: Alle kategorialen Variablen als Faktoren definieren
```
dat$PresT <- as.factor(dat$PresT)
dat$Gramm <- as.factor(dat$Gramm)
```

- Das Modell schätzen:
```
model <- lmer(RT ~ Gramm + PresT + Gramm:PresT +
                   (1 |VP) + (1 |Item), dat)
```

Berechnen eines LMEM

- Wichtig: Alle kategorialen Variablen als Faktoren definieren
```
dat$PresT <- as.factor(dat$PresT)
dat$Gramm <- as.factor(dat$Gramm)
```
- Das Modell schätzen:
```
model <- lmer(RT ~ Gramm + PresT + Gramm:PresT +
                   (1 |VP) + (1 |Item), dat)
```
  oder
```
model <- lmer(RT ~ Gramm*PresT +
                   (1 |VP) + (1 |Item), dat)
```

(Anmerkung: Bei RTs bietet sich oft log-Transformation an (Baayen & Milin, 2010))

Ergebnisse anschauen

```
summary(model)
```

```
Fixed effects:
                        Estimate Std. Error        df t value Pr(>|t|)    
(Intercept)             599.4039     0.7899 6390.0000 758.818  < 2e-16 ***
Grammungramm            -70.7890     1.1171 6390.0000 -63.368  < 2e-16 ***
PresT400                -10.0899     1.1171 6390.0000  -9.032  < 2e-16 ***
PresT600                -19.3762     1.1171 6390.0000 -17.345  < 2e-16 ***
PresT800                -29.5120     1.1171 6390.0000 -26.418  < 2e-16 ***
PresT1000               -38.2906     1.1171 6390.0000 -34.276  < 2e-16 ***
Grammungramm:PresT400    -8.1590     1.5798 6390.0000  -5.164 2.48e-07 ***
Grammungramm:PresT600   -20.0520     1.5798 6390.0000 -12.692  < 2e-16 ***
Grammungramm:PresT800   -28.9892     1.5798 6390.0000 -18.349  < 2e-16 ***
Grammungramm:PresT1000  -40.8810     1.5798 6390.0000 -25.877  < 2e-16 ***
```

Likelihood-Ratio-Tests

Likelihood-Ratio-Tests

Für diese Tests benötigen wir drei Konzepte:

- Geschachtelte Modelle (Nested Models)
- Modellpassung/ Likelihood
- Modellvergleiche

Hierarchisch geschaltete Modelle genau dann, wenn

Hierarchisch Geschachtelte Modelle

Hierarchisch geschaltete Modelle genau dann, wenn

- ein Modell ein Spezialfall des anderen Modells ist

Hierarchisch Geschachtelte Modelle

Hierarchisch geschaltete Modelle genau dann, wenn

- ein Modell ein Spezialfall des anderen Modells ist

  bzw.

- ein Modell alle Parameter des anderen Modells enthält und noch mehr
  content...

Hierarchisch Geschachtelte Modelle

Hierarchisch geschaltete Modelle genau dann, wenn

- ein Modell ein Spezialfall des anderen Modells ist

  bzw.

- ein Modell alle Parameter des anderen Modells enthält und noch mehr
  content...

Beispiel:
(1) $Y = a + b \qquad + (1|subject) + (1|item) + \epsilon$
(2) $Y = a + b + \mathbf{ab} + (1|subject) + (1|item) + \epsilon$

Hierarchisch Geschachtelte Modelle

Hierarchisch geschaltete Modelle genau dann, wenn

- ein Modell ein Spezialfall des anderen Modells ist

  bzw.

- ein Modell alle Parameter des anderen Modells enthält und noch mehr
  content...

Beispiel:
(1) $Y = a + b \qquad + (1|\mathit{subject}) + (1|\mathit{item}) + \epsilon$
(2) $Y = a + b + \mathbf{ab} + (1|\mathit{subject}) + (1|\mathit{item}) + \epsilon$

Hier ist (1) das *einfachere* Modell und (2) das *komplexere*, da (1) weniger Parameter enthält

Likelihood ist definiert als

$$L(Parameter) = P(Daten|Parameter)$$

Modellpassung/ Likelihood

Likelihood ist definiert als

$$L(Parameter) = P(Daten|Parameter)$$

Es wird genau jenes Parameterset als Modellparameter geschätzt, das das Auftreten der Daten am wahrscheinlichsten macht (Maximum-Likelihood-Schätzung)

Modellpassung/ Likelihood

Likelihood ist definiert als

$$L(Parameter) = P(Daten|Parameter)$$

Es wird genau jenes Parameterset als Modellparameter geschätzt, das das Auftreten der Daten am wahrscheinlichsten macht (Maximum-Likelihood-Schätzung)

Beispiel: $p_{Niete} = 0.9$ bei 18 Nieten und 2 Gewinnlosen

Modellpassung/ Likelihood

Likelihood ist definiert als

$$L(Parameter) = P(Daten|Parameter)$$

Es wird genau jenes Parameterset als Modellparameter geschätzt, das das Auftreten der Daten am wahrscheinlichsten macht (Maximum-Likelihood-Schätzung)

Beispiel: $p_{Niete} = 0.9$ bei 18 Nieten und 2 Gewinnlosen

Generell: Je höher die Likelihood, desto besser beschreibt ein Modell die Daten

Modellvergleiche

Geschachtelte Modelle können anhand ihrer Likelihood miteinander verglichen werden (Likelihood-Ratio-Test)

Modellvergleiche

Geschachtelte Modelle können anhand ihrer Likelihood miteinander verglichen werden (Likelihood-Ratio-Test)

Die Likelihood des komplexeren Modells ist **immer** größer (oder gleich) der des einfacheren

Modellvergleiche

Geschachtelte Modelle können anhand ihrer Likelihood miteinander verglichen werden (Likelihood-Ratio-Test)

Die Likelihood des komplexeren Modells ist **immer** größer (oder gleich) der des einfacheren

Aber: Ist sie signifikant größer?

Modellvergleiche

Trade-Off
(vgl. Occam's Razor: *"Entia non sunt multiplicanda praeter necessitatem"*)

Trade-Off
(vgl. Occam's Razor: "*Entia non sunt multiplicanda praeter necessitatem*")

Nutzen: Passung des Modells (Likelihood)

Modellvergleiche

Trade-Off
(vgl. Occam's Razor: *"Entia non sunt multiplicanda praeter necessitatem"*)

Nutzen: Passung des Modells (Likelihood)
Kosten: Zusätzliche Parameter im Modell

Modellvergleiche

Trade-Off
(vgl. Occam's Razor: *"Entia non sunt multiplicanda praeter necessitatem"*)

Nutzen: Passung des Modells (Likelihood)
Kosten: Zusätzliche Parameter im Modell

Der Nutzen muss die Kosten rechtfertigen!
(Der Likelihood-Ratio-Test implementiert dieses Prinzip)

Hands On: Modellvergleiche

Start: *Nullmodel*

```
m0 <- lmer(RT ~ (1 |VP) + (1 |Item), dat, REML = F)
```

Hands On: Modellvergleiche

Start: *Nullmodel*

```
m0 <- lmer(RT ~ (1 |VP) + (1 |Item), dat, REML = F)
```

Test auf Signifikanz für Grammatikalität:

```
m1 <- lmer(RT ~ Gramm +
                (1 |VP) + (1 |Item), dat, REML = F)
anova(m0,m1)
```

Hands On: Modellvergleiche

Start: *Nullmodel*

```
m0 <- lmer(RT ~ (1 |VP) + (1 |Item), dat, REML = F)
```

Test auf Signifikanz für Grammatikalität:

```
m1 <- lmer(RT ~ Gramm +
                (1 |VP) + (1 |Item), dat, REML = F)
anova(m0,m1)
```

Test auf Signifikanz für Präsentationsdauer:

```
m2 <- lmer(RT ~ PresT +
                (1 |VP) + (1 |Item), dat, REML = F)
anova(m0,m2)
```

Hands On: Ergebnisse anschauen

- Die Ergebnisse:

```
> anova(m1,m0)
Data: dat
Models:
m0: RT ~ 1 + (1 | VP) + (1 | Item)
m1: RT ~ Gramm + (1 | VP) + (1 | Item)
   Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
m0  4 69257 69284 -34625    69249
m1  5 61604 61638 -30797    61594 7655.5      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
>
> anova(m2,m0)
Data: dat
Models:
m0: RT ~ 1 + (1 | VP) + (1 | Item)
m2: RT ~ PresT + (1 | VP) + (1 | Item)
   Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
m0  4 69257 69284 -34625    69249
m2  5 68233 68267 -34112    68223 1026.3      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hands On: Modellvergleiche

Test auf Interaktion:

```
m3 <- lmer(RT ~ Gramm + PresT +
               (1 |VP) + (1 |Item), dat, REML = F)
m4 <- lmer(RT ~ Gramm + PresT + Gramm:PresT
               (1 |VP) + (1 |Item), dat, REML = F)
anova(m4,m3)
```

```
> anova(m4,m3)
Data: dat
Models:
m3: RT ~ Gramm + PresT + (1 | VP) + (1 | Item)
m4: RT ~ Gramm + PresT + Gramm:PresT + (1 | VP) + (1 | Item)
   Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
m3  6 57296 57336 -28642    57284
m4  7 56505 56552 -28245    56491 793.02      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Reihenfolge der Tests

- Der Interaktionsparameter `Gramm:PresT` ist nur dann sinnvoll, wenn das Modell schon die Parameter `Gramm` und `PresT` enthält!

Reihenfolge der Tests

- Der Interaktionsparameter `Gramm:PresT` ist nur dann sinnvoll, wenn das Modell schon die Parameter `Gramm` und `PresT` enthält!

- Eine Interaktion höherer Ordnung benötigt immer alle "niedrigeren" Parameter

Reihenfolge der Tests

- Der Interaktionsparameter `Gramm:PresT` ist nur dann sinnvoll, wenn das Modell schon die Parameter `Gramm` und `PresT` enthält!
- Eine Interaktion höherer Ordnung benötigt immer alle "niedrigeren" Parameter
- Beispiel: Dreifachinteraktion `a:b:c` benötigt notwendig auch folgende Parameter im Modell: `a, b, c, a:b, b:c, a:c`

Tests auf Faktoren – Gesamtübersicht

```
> anova(model)
Type III Analysis of Variance Table with Satterthwaite's method
              Sum Sq  Mean Sq NumDF DenDF  F value    Pr(>F)    
Gramm       13076965 13076965     1  6390 32746.36 < 2.2e-16 ***
PresT        2777471   694368     4  6390  1738.78 < 2.2e-16 ***
Gramm:PresT   338016    84504     4  6390   211.61 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hier passiert nicht exakt das gleiche wie bei einzelnen Modellvergleichen mit `anova()`

Kovariaten

Gemischte Modelle erlauben einfaches Einfügen von Kovariaten in das Modell

Kovariaten

Gemischte Modelle erlauben einfaches Einfügen von Kovariaten in das Modell

Beispiel:

$$\begin{aligned}\text{RT} \sim \text{Gramm} + \text{PresT} + \text{Gramm:PresT} \\ + \text{Satzlänge} + \text{Muttersprache} \\ + (1\ |\text{VP}) + (1\ |\text{Item})\end{aligned}$$

Kovariaten

Gemischte Modelle erlauben einfaches Einfügen von Kovariaten in das Modell

Beispiel:

$$\begin{aligned}\mathrm{RT} \sim \mathrm{Gramm} + \mathrm{PresT} + \mathrm{Gramm{:}PresT} \\ + \mathrm{Satzlänge} + \mathrm{Muttersprache} \\ + (1\ |\mathrm{VP}) + (1\ |\mathrm{Item})\end{aligned}$$

Hypothesentests sind auch mit Kovariaten mglich. Für Test auf Interaktion vergleiche das obige Modell mit

$$\begin{aligned}\mathrm{RT} \sim \mathrm{Gramm} + \mathrm{PresT} + \\ + \mathrm{Satzlänge} + \mathrm{Muttersprache} \\ + (1\ |\mathrm{VP}) + (1\ |\mathrm{Item})\end{aligned}$$

Günther, Nguyen, Chen, Dudschig, Kaup & Glenberg, 2020:
LR-Test

Using these data, we estimated Linear Mixed Effect Models (LMEMs Baayen, Davidson, & Bates, 2008) using the R packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017). Our model to predict log-transformed release times (Baayen & Milin, 2010) included a fixed effect for congruency, random intercepts for participants and items, and by-participant random slopes for congruency.[4]

This model including the congruency fixed effect significantly

outperformed a model without this effect in a likelihood-ratio test ($\chi^2(1) = 10.41$, $p = .001$, $b = 0.03$, $t = 3.23$), indicating a significant congruency effect.[5] Release times in the congruent condition ($M = 1604$ ms) were faster than in the incongruent condition ($M = 1648$ ms), amounting to a congruency effect of 44 ms (see Fig. 3).[6] Applying the

Berichten der Ergebnisse - Beispiel 2

Günther, Smolka, & Marelli, 2019: ANOVA-Tabelle

Table 1 – Model parameters for the effects in the three-way interaction model, along with F-tests, estimated on the trimmed data set after outlier removal.

Parameter	β	F	D.df	p
Intercept	.60			
Vector type	.23	7.64	23.73	.011
Semantic transparency	−.12	61.47	27.16	<.001
Language	.19	38.12	38.32	<.001
Vector type: ST	−.33	23.64	123.14	<.001
Vector type: language	−.12	.13	23.73	.072
ST: language	−.01	4.74	27.16	.038
Vector type: ST: language	.27	10.79	123.14	.001

Note: Reference levels for our dummy variables were vector type: compositional, semantic transparency: transparent, and language: English, so the intercept is the model prediction for the condition compositional/transparent/English. Model predictions for the other conditions are obtained by adding the respective β-values to this intercept. Denominator degrees of freedom for the F-tests are estimated via Satterthwaite approximation with the *lmerTest* package (Kuznetsova et al., 2017), all numerator degrees of freedom are $N.df = 1$.

Berichten der Ergebnisse - Beispiel 2

Günther & Marelli, 2020: Modellparameter

Table 1 – Model parameters for the effects in the three-way interaction model, along with F-tests, estimated on the trimmed data set after outlier removal.

Parameter	β	F	D.df	p
Intercept	.60			
Vector type	.23	7.64	23.73	.011
Semantic transparency	−.12	61.47	27.16	<.001
Language	.19	38.12	38.32	<.001
Vector type: ST	−.33	23.64	123.14	<.001
Vector type: language	−.12	.13	23.73	.072
ST: language	−.01	4.74	27.16	.038
Vector type: ST: language	.27	10.79	123.14	.001

Note: Reference levels for our dummy variables were vector type: compositional, semantic transparency: transparent, and language: English, so the intercept is the model prediction for the condition compositional/transparent/English. Model predictions for the other conditions are obtained by adding the respective β-values to this intercept. Denominator degrees of freedom for the F-tests are estimated via Satterthwaite approximation with the *lmerTest* package (Kuznetsova et al., 2017), all numerator degrees of freedom are $N.df = 1$.

Günther & Marelli, 2020: Modellparameter

Table 5. Parameters for the final model predicting log-transformed rejection times for novel compounds in both tasks (Experiment 1: timed sensibility judgements, Experiment 2: lexical decision judgements), after model criticism. The timed sensibility task serves as the reference level for the factor Experiment.

Parameter	*b*	*t*	*p*
Intercept	6.45	199.27	<.001
Modifier frequency	0.004	3.19	.001
As-modifier frequency	0.01	3.93	<.001
As-head frequency	0.01	3.95	<.001
Modifier family size	0.01	2.69	.007
Length	0.02	14.53	<.001
Experiment: as-head frequency	0.003	3.67	<.001
Experiment: length	–0.01	–3.97	<.001
Modifier composition	0.06	4.14	<.001

Fortführung

Messwiederholungen - Random Effect Structures

Messwiederholungen

Bisher haben wir nur *Random Intercepts* betrachtet:

Messwiederholungen

Bisher haben wir nur *Random Intercepts* betrachtet:

Für jede VP bzw. jedes Item wird ein bestimmter konstanter Einfluss auf die RTs angenommen
(zB kann VP3 generell 50 ms langsamer sein als der Durchschnitt)

Messwiederholungen

Bisher haben wir nur *Random Intercepts* betrachtet:

Für jede VP bzw. jedes Item wird ein bestimmter konstanter Einfluss auf die RTs angenommen
(zB kann VP3 generell 50 ms langsamer sein als der Durchschnitt)

In den Beispieldaten ist aber jede VP (und jedes Item) in jeder Experimentalbedingung (vollständige Messwiederholung)

Messwiederholungen

Bisher haben wir nur *Random Intercepts* betrachtet:

Für jede VP bzw. jedes Item wird ein bestimmter konstanter Einfluss auf die RTs angenommen
(zB kann VP3 generell 50 ms langsamer sein als der Durchschnitt)

In den Beispieldaten ist aber jede VP (und jedes Item) in jeder Experimentalbedingung (vollständige Messwiederholung)

$\Longrightarrow$ Was, wenn die Bedingungen für verschiedene VPs unterschiedlich starken Einfluss haben?

Messwiederholungen

Bisher haben wir nur *Random Intercepts* betrachtet:

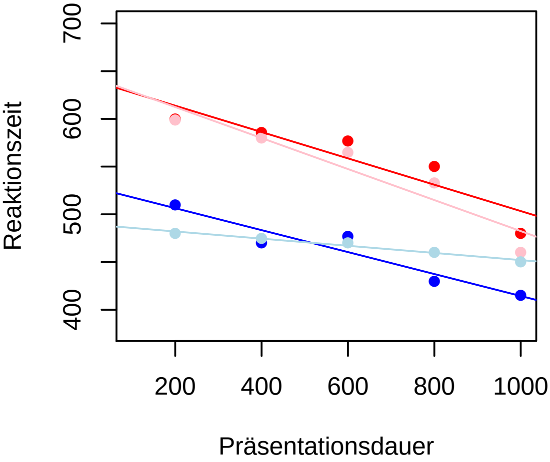
Für jede VP bzw. jedes Item wird ein bestimmter konstanter Einfluss auf die RTs angenommen
(zB kann VP3 generell 50 ms langsamer sein als der Durchschnitt)

In den Beispieldaten ist aber jede VP (und jedes Item) in jeder Experimentalbedingung (vollständige Messwiederholung)

$\Longrightarrow$ Was, wenn die Bedingungen für verschiedene VPs unterschiedlich starken Einfluss haben?

Die Unabhängigkeit der Datenpunkte ist in einem solchen Fall also noch weiter eingeschränkt

Messwiederholungen



Da bei Messwiederholungen dieser Fall nicht ausgeschlossen werden kann, sollten Random Slopes ins Modell mit aufgenommen werden (Barr et al. 2013)

Messwiederholungen

Da bei Messwiederholungen dieser Fall nicht ausgeschlossen werden kann, sollten Random Slopes ins Modell mit aufgenommen werden (Barr et al. 2013)

Dies entspricht individuellen Steigungen für jede VP bzw jedes Item, auf dem eine Messwiederholung stattfindet

Messwiederholungen

Da bei Messwiederholungen dieser Fall nicht ausgeschlossen werden kann, sollten Random Slopes ins Modell mit aufgenommen werden (Barr et al. 2013)

Dies entspricht individuellen Steigungen für jede VP bzw jedes Item, auf dem eine Messwiederholung stattfindet

Das vollständige Modell für die Beispieldaten sieht also folgendermaßen aus:

$$\text{RT} \sim \text{Gramm*PresT} + (\text{Gramm*PresT} \,|\text{VP}) + (\text{Gramm*PresT} \,|\text{Item})$$

Within: VPs und Items

$$\text{RT} \sim \text{Gramm*PresT} \\ + (\text{Gramm*PresT } |\text{VP}) + (\text{Gramm*PresT } |\text{Item})$$

Messwiederholungen: Between- und Within-Designs

Within: VPs und Items

$$\mathrm{RT} \sim \mathrm{Gramm*PresT} \\ \quad + (\mathrm{Gramm*PresT}\ |\mathrm{VP}) + (\mathrm{Gramm*PresT}\ |\mathrm{Item})$$

Within: VPs, Between: Items

$$\mathrm{RT} \sim \mathrm{Gramm*PresT} \\ \quad + (\mathrm{Gramm*PresT}\ |\mathrm{VP}) + (1\ |\mathrm{Item})$$

Within: VPs und Items

$$\mathrm{RT} \sim \mathrm{Gramm*PresT} + (\mathrm{Gramm*PresT}\ |\mathrm{VP}) + (\mathrm{Gramm*PresT}\ |\mathrm{Item})$$

Within: VPs, Between: Items

$$\mathrm{RT} \sim \mathrm{Gramm*PresT} + (\mathrm{Gramm*PresT}\ |\mathrm{VP}) + (1\ |\mathrm{Item})$$

Within: Items, Between: VPs

$$\mathrm{RT} \sim \mathrm{Gramm*PresT} + (1\ |\mathrm{VP}) + (\mathrm{Gramm*Pres}\ |\mathrm{Item})$$

Within: VPs und Items

RT ~ Gramm*PresT
+ (Gramm*PresT |VP) + (Gramm*PresT |Item)

Within: VPs, Between: Items

RT ~ Gramm*PresT
+ (Gramm*PresT |VP) + (1 |Item)

Within: Items, Between: VPs

RT ~ Gramm*PresT
+ (1 |VP) + (Gramm*Pres |Item)

Between: VPs und Items

RT ~ Gramm*PresT
+ (1 |VP) + (1 |Item)

Die Random Effect Structure spiegelt also direkt das Experimentaldesign wieder!

Die Random Effect Structure spiegelt also direkt das Experimentaldesign wieder!

Beispiel: Das Material besteht aus semantisch sinnlosen und sinnvollen Sätzen, wobei keine Minimalpaare möglich sind. Man kann also nicht annehmen, dass es einen Satz als sinnlose und sinnvolle Variante gibt. Jede Person sieht jeden Satz des Materials. Dabei sind die Hälfte der VPs L1-Sprecher, die andere Hälfte L2-Sprecher.

Was für Random Slopes sollten daher ins Modell aufgenommen werden?

Antwort:

RT $\sim$ Sinn*Sprache + (Sinn |VP) + (Sprache |Item)

Jedes Item wird von L1- und L2- Sprechern bearbeitet, liefert also hier Werte für beide Bedingungen von Sprache
Jede VP bearbeitet sinnlose und sinnvolle Sätze, liefert also hier Werte für beide Bedingungen von Sinn

Konvergenz

Ein konvergierendes Modell ist in jedem Fall wichtiger als eine vollständige Random Effect Structure!

Konvergenz

Ein konvergierendes Modell ist in jedem Fall wichtiger als eine vollständige Random Effect Structure!

Was, wenn das Modell nicht konvergiert?

Konvergenz

Ein konvergierendes Modell ist in jedem Fall wichtiger als eine vollständige Random Effect Structure!

Was, wenn das Modell nicht konvergiert?

$\Longrightarrow$ Vereinfachung der Random Effect Structure auf ein von den Daten unterstütztes Format

Konvergenz

Ein konvergierendes Modell ist in jedem Fall wichtiger als eine vollständige Random Effect Structure!

Was, wenn das Modell nicht konvergiert?

$\Longrightarrow$ Vereinfachung der Random Effect Structure auf ein von den Daten unterstütztes Format

Eine mögliche Strategie: Schrittweise Vereinfachung zuerst der Item Random Effect Structure und dann der Subject Random Effect Structure, unter Einhaltung der Interaktionshierarchie

Konvergenz

Ein konvergierendes Modell ist in jedem Fall wichtiger als eine vollständige Random Effect Structure!

Was, wenn das Modell nicht konvergiert?

$\Longrightarrow$ Vereinfachung der Random Effect Structure auf ein von den Daten unterstütztes Format

Eine mögliche Strategie: Schrittweise Vereinfachung zuerst der Item Random Effect Structure und dann der Subject Random Effect Structure, unter Einhaltung der Interaktionshierarchie

(Empfohlen von Jaeger, 2009; in einem Vergleich verschiedener bestehender Strategien von Kimball et al., 2016 als sinnvoll befunden)

Beispiel: Hypothesentest auf Interaktion zweier Faktoren

Beispiel: Hypothesentest auf Interaktion zweier Faktoren
Beide Faktoren haben within-subjects Messwiederholung

Beispiel: Hypothesentest auf Interaktion zweier Faktoren
Beide Faktoren haben within-subjects Messwiederholung
Nur Faktor A hat within-items Messwiederholung

Beispiel: Hypothesentest auf Interaktion zweier Faktoren
Beide Faktoren haben within-subjects Messwiederholung
Nur Faktor A hat within-items Messwiederholung

```
m0 <- lmer(RT ~ a + b +
                (a*b |VP) + (a |Item), dat, REML = F)
m1 <- lmer(RT ~ a + b + a:b
                (a*b |VP) + (a |Item), dat, REML = F)
anova(m0,m1,test="Chisq")
```

Fortführung

Binäre Kriteriumsvariablen

Binäre Kriteriumsvariablen

Beispiel: Beurteilung von VPs über Korrektheit von Sätzen
Drei Prädiktoren a, b, c

Daten:

VP	Item	a	b	c	Answer
1	1	1	13	1	1
1	2	1	14	2	0
1	3	1	8	1	1
1	4	1	7	2	1

Binäre Kriteriumsvariablen

Theoretischer Hintergrund:

- Lineare Modelle setzen *stetige* Kriterien voraus, nicht kategoriale

Theoretischer Hintergrund:

- Lineare Modelle setzen *stetige* Kriterien voraus, nicht kategoriale
- Über die *Wahrscheinlichkeit* des Beobachtens einer Kategorie lassen sich jedoch stetige Variablen erzeugen (z.B. sogenannte *Logits*)

Binäre Kriteriumsvariablen

Theoretischer Hintergrund:

- Lineare Modelle setzen *stetige* Kriterien voraus, nicht kategoriale
- Über die *Wahrscheinlichkeit* des Beobachtens einer Kategorie lassen sich jedoch stetige Variablen erzeugen (z.B. sogenannte *Logits*)
- Diese können durch lineare Modelle vorhergesagt werden

Binäre Kriteriumsvariablen

Beispiel für Umsetzung in R mit `glmer`:

```
model <- glmer(Answer ~ a*b*c + (1 |VP) + (1 |Item),
                          dat, family = "binomial")
```

Ausführliche Anleitung unter:
http://www.ats.ucla.edu/stat/r/dae/melogit.htm