# Gemischte Lineare Modelle

## Linear Mixed Effect Models

Fritz Günther

SFB833, Projekt Z2

March 20, 2015

# Übersicht

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

# Literatur

**Tutorial:**

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:**

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., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

# Lineare Modelle

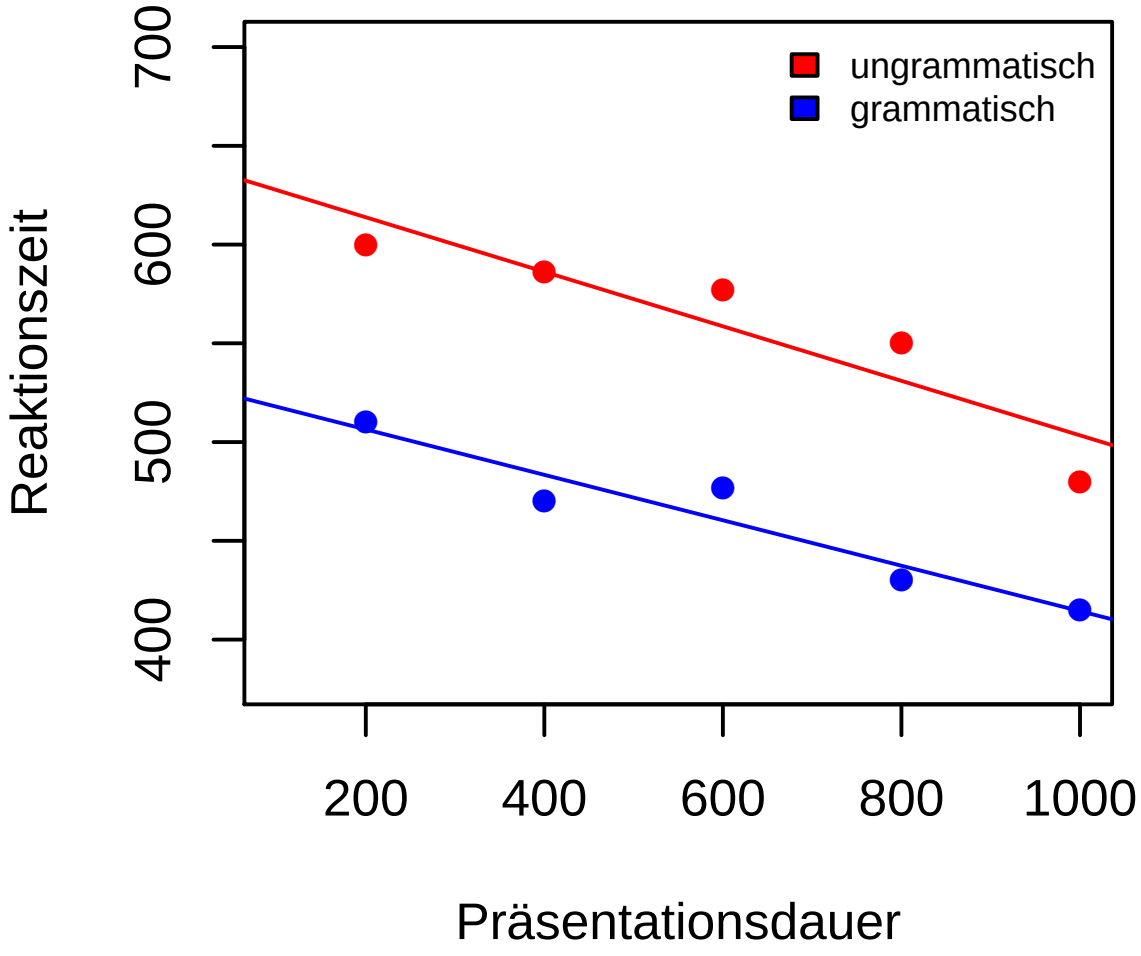
$$Y = a + b + \epsilon$$

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

*Beispiel:*

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

# Beispieldaten

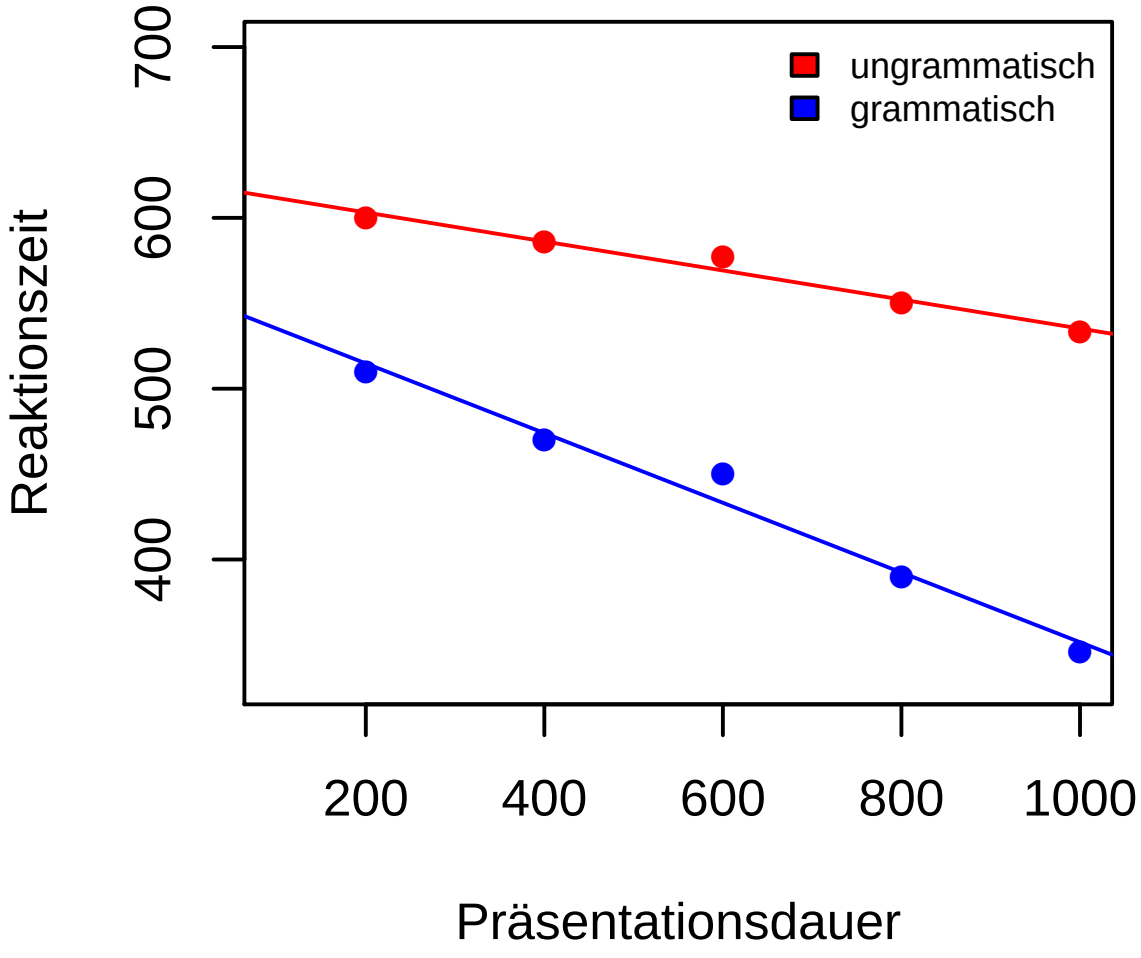


# Lineare Modelle: Interaktionen

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

$Y$: Abhängige Variable ("Kriterium")
$a$: Unabhängige Variable 1 ("Prädiktor 1")
$b$: Unabhängige Variable 2 ("Prädiktor 2")
$ab$: Interaktionseffekt beider Prädiktoren
$\epsilon$: 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

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

*Beispiele:* Versuchspersonen (VPs), Items

$\Longrightarrow$ Generalisierung erwünscht

# Grundproblem

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

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

Wieso nicht gleich?

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

# Hands On: Einlesen der Beispieldaten

- Auf folgende Seite gehen:
  http://www.lingexp.uni-tuebingen.de/z2/LMEM/
- Speichern von LMEMdat.txt in einen auffindbaren Ordner
- R starten
- Verzeichnis setzen:
  ```
  setwd("AuffindbarerOrdner")
  ```
- Daten einlesen:
  ```
  dat <- read.table("LMEMdat.txt")
  ```

# Hands On: Berechnen eines LMEM

- Das Paket `lme4` installieren:
  `install.packages("lme4")`
- Das Paket `lme4` laden:
  `library(lme4)`
- 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)
```

# Hands On: Ergebnisse anschauen

- Inspektion der Modellparameter:

```
> summary(model)
Linear mixed model fit by REML ['lmerMod']
Formula: RT ~ Gramm + PresT + Gramm:PresT + (1 | VP) + (1 | Item)
   Data: dat

REML criterion at convergence: 56514.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9801 -0.6892 -0.0057  0.6770  3.7385

Random effects:
 Groups   Name        Variance  Std.Dev.
 VP       (Intercept) 2.296e-12 1.515e-06
 Item     (Intercept) 0.000e+00 0.000e+00
 Residual             3.992e+02 1.998e+01
Number of obs: 6400, groups:  VP, 32; Item, 20

Fixed effects:
                    Estimate Std. Error t value
(Intercept)       608.751083   0.828320   734.9
Grammungramm      -59.627539   1.171422   -50.9
PresT              -0.048002   0.001249   -38.4
Grammungramm:PresT -0.051296   0.001766   -29.0

Correlation of Fixed Effects:
            (Intr) Grmmng PresT
Grammungrmm -0.707
PresT       -0.905  0.640
Grmmngrm:PT  0.640 -0.905 -0.707
```

# Hands On: Ergebnisse anschauen

- Konfidenzintervalle:

```
> confint(model,method="Wald")
                              2.5 %       97.5 %
(Intercept)          607.12760525 610.37456082
Grammungramm         -61.92348295 -57.33159435
PresT                 -0.05044907  -0.04555410
Grammungramm:PresT    -0.05475740  -0.04783486
```

# Hypothesentests und Signifikanzen

Konfidenzintervalle können wie folgt interpretiert werden:
Enthält das Intervall für einen Parameter nicht 0, so hat der entsprechende Prädiktor einen Einfluss auf das Kriterium

Auch die t-Werte bieten eine (grobe!) Faustregel:
Einfluss ist vorhanden bei $t > 2$

# Hypothesentests und Signifikanzen

Aber wie kommt man an Hypothesentests auf Signifikanz?

Antwort: **Likelihood-Ratio-Tests**

Hierfür benötigen wir noch drei Konzepte:

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

# Hierarchisch Geschachtelte Modelle

Zwei Modelle sind *hierarchisch geschachtelte Modelle* genau dann, wenn ein Modell ein Spezialfall des anderen Modells ist

bzw.

Zwei Modelle sind *hierarchisch geschachtelte Modelle* genau dann, wenn ein Modell alle Parameter des anderen Modells enthält und noch mehr

Beispiel:

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

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

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

# 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)

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"*)

**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)
```

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

Benötigt man Parameter für Grammatikalität **und** Präsentationsdauer im Modell?:

```
m3 <- lmer(RT ~ Gramm + PresT +
                (1 |VP) + (1 |Item), dat, REML = F)
anova(m3,m1)
anova(m3,m2)
```

# Hands On: Ergebnisse anschauen

- Die Ergebnisse:

```
> anova(m3,m1)
Data: dat
Models:
m1: RT ~ Gramm + (1 | VP) + (1 | Item)
m3: RT ~ Gramm + PresT + (1 | VP) + (1 | Item)
   Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
m1  5 61604 61638 -30797    61594
m3  6 57296 57336 -28642    57284  4310      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
>
> anova(m3,m2)
Data: dat
Models:
m2: RT ~ PresT + (1 | VP) + (1 | Item)
m3: RT ~ Gramm + PresT + (1 | VP) + (1 | Item)
   Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
m2  5 68233 68267 -34112    68223
m3  6 57296 57336 -28642    57284 10939      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

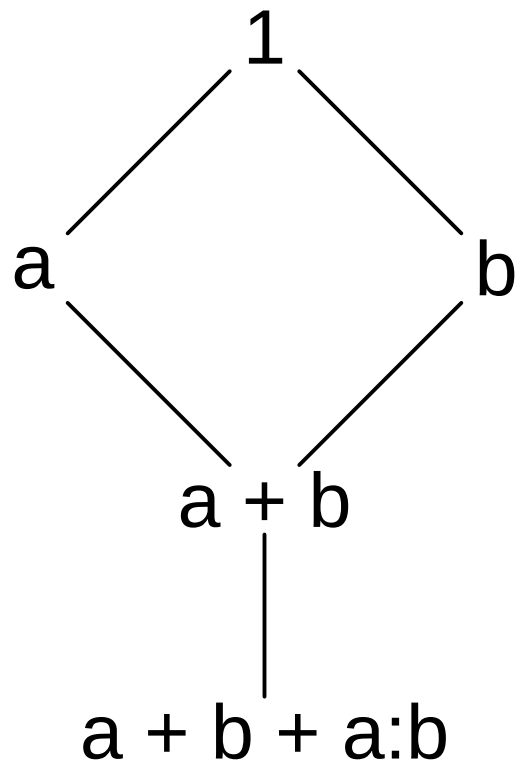
# Hands On: Modellvergleiche

Test auf Interaktion:

```
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



# 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`

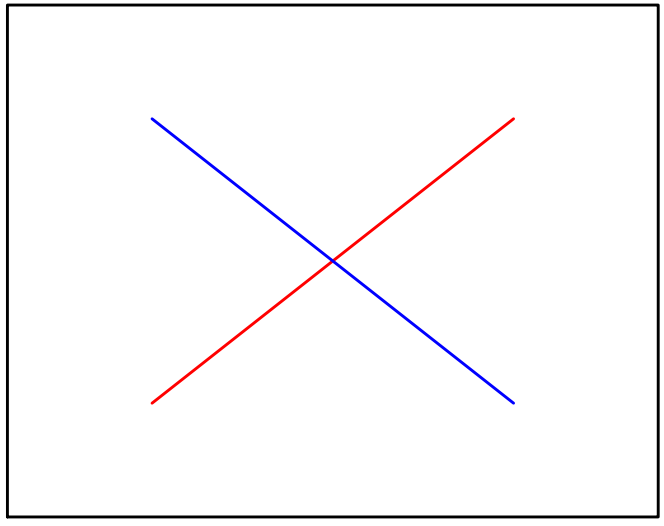
# Reihenfolge der Tests

- Soll man das Modell `m3` (`Gramm + PresT`) gegen `m1` (`Gramm`) testen oder gegen `m2` (`PresT`)?
- Sinnvoll: Gegen das informativere Modell (geringeres AIC bzw. BIC)

  AIC und BIC verrechnen Modellpassung (Likelihood) mit Modelkomplexität (Anzahl Parameter)

# Reihenfolge der Tests

Was ist mit folgendem Fall?



$\Longrightarrow$ Keine Haupteffekte, aber Interaktion vermutet

# Reihenfolge der Tests

Wenn Tests auf beide Haupteffekte nicht signifikant sind, kann dennoch eine Interaktion vorhanden sein

Folgender Modellvergleich testet eine solche "reine" Interaktion:

(1) RT $\sim$ + (1 |VP) + (1 |Item)
vs
(2) RT $\sim$ a + b + a:b + (1 |VP) + (1 |Item)

# Kovariaten

Gemischte Modelle erlauben einfaches Einfügen von Kovariaten in das Modell

Beispiel:

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

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

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

# Berichten der Ergebnisse - Ein Beispiel

"We used the *lme4* package (Bates, Maechler & Bolker, 2014) for R (R Core Team, 2014) to perform a linear mixed effects analysis for the influence of grammaticality stimulus duration on reaction times. As fixed effects, we entered grammaticality and stimulus duration into the model. As random effects, we entered random intercepts for subjects as well as items.
We tested for the significance of our fixed effects by performing likelihood ratio tests of the full model with the effect in question against the model without the effect in question."

# Berichten der Ergebnisse - Ein Beispiel

"The analysis yielded a significant effect of grammaticality ($\chi^2(1) = 7655.5, p < .001$) as well as an additional effect of stimulus duration ($\chi^2(1) = 4310, p < .001$).
Furthermore, we found a significant interaction between both variables ($\chi^2(1) = 793.02, p < .001$). The model parameters of the final model (containing both main effects and interaction effect) and their confidence intervals are shown in Table 1."

# Fortführung

- Messwiederholungen - Random Effect Structures
- Modellvergleiche bei mehr als 2 Prädiktoren
- Binäre Kriteriumsvariablen

# Multivariate Designs

**Grundproblem:**

Wie findet man bei einem Experiment mit $n$ Prädiktoren heraus, welche Haupt- und Interaktionseffekte signifikant sind?

Beispiel: $n = 4$ Drei Prädiktoren $a, b, c$ und Kovariate $v$

# Multivariate Designs

Beste Lösung: *Hypothesengeleitetes* Vorgehen

Angenommen, man hat folgende Hypothesen:

1. Haupteffekt a
2. Interaktion b:c
3. Dreifachinteraktion a:b:c

Dabei sollen mögliche Einflüsse von $v$ kontrolliert werden

# Multivariate Designs

Diese Hypothesen können wie folgt überprüft werden:

1. $RT \sim v + (1 \mid VP) + (1 \mid Item)$

   vs.

   $RT \sim v + a + (1 \mid VP) + (1 \mid Item)$

2. $RT \sim v + \textit{(a)} + b + c + (1 \mid VP) + (1 \mid Item)$

   vs.

   $RT \sim v + \textit{(a)} + b + c + b{:}c + (1 \mid VP) + (1 \mid Item)$

   In 2. sollte a vorkommen, wenn sich in 1. ein signifikanter Effekt für a ergeben hat

# Multivariate Designs

3. $RT \sim v + a + b + c + a{:}b + b{:}c + a{:}c + (1\,|VP) + (1\,|Item)$

vs.

$RT \sim v + a + b + c + a{:}b + b{:}c + a{:}c + a{:}b{:}c + (1\,|VP) + (1\,|Item)$

Kann auch geschrieben werden als:

$RT \sim v + a{*}b + b{*}c + a{*}c + (1\,|VP) + (1\,|Item)$

vs.

$RT \sim v + a{*}b{*}c + (1\,|VP) + (1\,|Item)$

# Multivariate Designs

Hypothesengeleitete Verfahren bezeichnet man als *konfirmatorisch*. Im Falle unklarer Hypothesen sind auch *explorative* Verfahren möglich.

Im Folgenden wird die *forward selection* als exploratives Verfahren besprochen. Dabei wird, ausgehend vom einfachsten möglichen Modell, schrittweise überprüft, durch welche zusätzlichen Parameter das Modell am meisten verbessert wird.

Ein zusätzlicher Parameter wird also genau dann ins Modell aufgenommen, wenn er

- Das Modell signifikant verbessert
- Ein informativeres Modell liefert als alle anderen möglichen zusätzlichen Parameter

# Forward Selection: Beispiel

$p$-Werte aus LR-Tests, Informationskriterium: AIC

Nullmodell (*Schritt 0*): $\text{RT} \sim \text{v} + (1\,|\text{VP}) + (1\,|\text{Item})$

*Schritt 1*:

1. $\text{RT} \sim \text{v} + \text{a} + (1\,|\text{VP}) + (1\,|\text{Item})$
   $p = .04, AIC = 1000$

2. $\text{RT} \sim \text{v} + \text{b} + (1\,|\text{VP}) + (1\,|\text{Item})$
   $p = .02, AIC = 990$

3. $\text{RT} \sim \text{v} + \text{c} + (1\,|\text{VP}) + (1\,|\text{Item})$
   $p = .10, AIC = 1100$

$\Longrightarrow$ Wähle 2.

# Forward Selection: Beispiel

*Schritt 1*: $\mathrm{RT} \sim \mathrm{v} + \mathrm{b} + (1\,|\mathrm{VP}) + (1\,|\mathrm{Item})$

*Schritt 2*:

1. $\mathrm{RT} \sim \mathrm{v} + \mathrm{b} + \mathrm{a} + (1\,|\mathrm{VP}) + (1\,|\mathrm{Item})$
   $p = .03, AIC = 980$

2. $\mathrm{RT} \sim \mathrm{v} + \mathrm{b} + \mathrm{c} + (1\,|\mathrm{VP}) + (1\,|\mathrm{Item})$
   $p = .24, AIC = 1020$

$\Longrightarrow$ Wähle 1.

# Forward Selection: Beispiel

*Schritt 2*: $\text{RT} \sim \text{v} + \text{b} + \text{a} + (1\,|\text{VP}) + (1\,|\text{Item})$

*Schritt 3*:

1. $\text{RT} \sim \text{v} + \text{b} + \text{a} + \text{c} + (1\,|\text{VP}) + (1\,|\text{Item})$
   $p = .41, AIC = 1030$

$\Longrightarrow$ Bleibe bei Modell aus Schritt 2.

# Forward Selection: Beispiel

*Schritt 3*: RT $\sim$ v + b + a + (1 |VP) + (1 |Item)

*Schritt 4*:

1. RT $\sim$ v + a*b + (1 |VP) + (1 |Item)
   $p = .002, AIC = 920$

$\Longrightarrow$ Wähle 1.

# Forward Selection: Beispiel

*Schritt 4*: RT $\sim$ v + a*b + (1 |VP) + (1 |Item)

*Schritt 5*:

1. RT $\sim$ v + a*b + b*c (1 |VP) + (1 |Item)
   $p = .012, AIC = 860$

2. RT $\sim$ v + a*b + a*c (1 |VP) + (1 |Item)
   $p = .06, AIC = 900$

$\Longrightarrow$ Wähle 1.

Dadurch ist der Parameter c doch Teil des Modells, denn

$$\text{b*c} = \text{b} + \text{c} + \text{b:c}$$

# Forward Selection: Beispiel

*Schritt 5*: RT $\sim$ v + a*b + b*c + (1 |VP) + (1 |Item)

*Schritt 6*:

1. RT $\sim$ v + a*b + b*c + a*c + (1 |VP) + (1 |Item)
   $p = .10, AIC = 850$

$\Longrightarrow$ Bleibe bei Modell aus Schritt 5.

# Forward Selection: Beispiel

*Schritt 6*: RT $\sim$ v + a*b + b*c + (1 |VP) + (1 |Item)

*Schritt 7*:

1. RT $\sim$ v + a*b*c + (1 |VP) + (1 |Item)
   $p = .55, AIC = 900$

$\Longrightarrow$ Bleibe bei Modell aus Schritt 6.
$\Longrightarrow$ Wähle dieses Modell als endgültiges Modell.

*Hinweis*: Hier wurde angenommen, dass die Kovariate $v$ nicht mit $a, b, c$ interagiert.

# Zusatzliteratur

Schöne Beschreibung von Modellvergleichen und Modellselektion (sowie v.a. kategorialen Kriteriumsvariablen):

Wickens, T.D. (1989). *Multiway Contingency Tables Analysis for the Social Sciences.* New York, NY: Erlbaum.

# Messwiederholungen

# Messwiederholungen

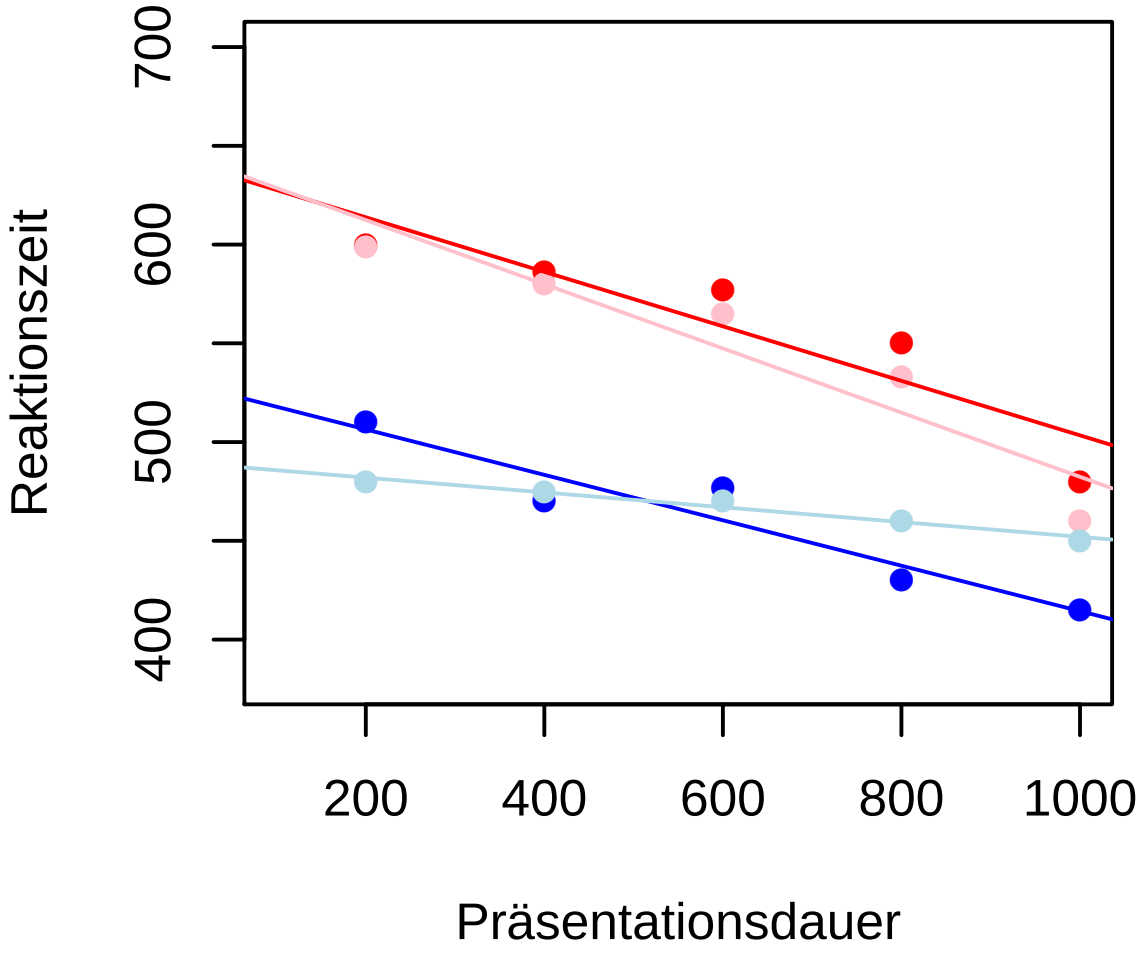
Bisher haben wir nur *Random Intercepts* betrachtet:

Für jede VP bzw. jedes Item wird ein bestimmter konstanter Einfluss auf die RTs angenommen
(zB kann VP3 generell 50ms 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



# 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:

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

# Messwiederholungen: Between- und Within-Designs

**Within: VPs und Items**

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

**Within: VPs, Between: Items**

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

**Within: Items, Between: VPs**

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

**Between: VPs und Items**

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

# Messwiederholungen: Between- und Within-Designs

Die Random Effect Structure spiegelt also direkt das Experimentaldesign wieder!

*Beispiel:* Das Material besteht aus semantisch sinnlosen und sinnvollen Sätzen, wobei keine Minimalpaare nötig 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?

# Messwiederholungen: Between- und Within-Designs

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

# Messwiederholungen: Hypothesentests

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

Was, wenn das Modell nicht konvergiert?

Vereinfachung der Random Effect Structure auf ein noch zu rechtfertigendes Format (durch inhaltliche Punkte oder durch forward- oder backward selection).

# Messwiederholungen: Hypothesentests

Welche Random Effect Structure für Hypothesentests??

Angenommen, es interessiert der feste Effekt a:b in

$$RT \sim a + b + a{:}b + (a + b + a{:}b \,|VP) + (1 \,|Item)$$

Testet man folgenden Vergleich:

$$RT \sim a + b + a{:}b + (a + b + a{:}b \,|VP) + (1 \,|Item)$$
vs
$$RT \sim a + b + \quad\quad + (a + b + a{:}b \,|VP) + (1 \,|Item)$$

oder

$$RT \sim a + b + a{:}b + (a + b \,|VP) + (1 \,|Item)$$
vs
$$RT \sim a + b + \quad\quad + (a + b \,|VP) + (1 \,|Item)$$

# Messwiederholungen: Hypothesentests

Welche Random Effect Structure für Hypothesentests??

Diese Frage scheint noch nicht geklärt

Persönliche Präferenz: Option 2, und anschließend testen, ob Modell mit random slope (a*b |VP) das Modell noch zusätzlich verbessert

Auf jeden Fall: Genau berichten, was getestet wurde!

# 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
- 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

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