

Generalized Linear Mixed Models

Jonathan Harrington

GLMM: Generalized linear mixed models

Das gleiche wie logistische Regression, aber zusätzlich mit Random Factor(s).

Wird eingesetzt, wenn wir die Wirkung einer oder mehrerer Variablen ausklammern wollen (siehe Mixed-Models). Oft in der Perzeption wenn wir mehrere Hörer haben (Hörer als Random Factor).

Die Funktion: `lmer()` mit `family=binomial`

GLMM: Ein Fixed Faktor mit 2 Stufen

```
alt = read.table(url("http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/alter.txt"))
```

	Hoerer	Alter	Response	Wort	Correct	Phon	Sprecher
2	EAU	j	fortis	lei	TRUE	t	HSE

Reihe 2: Hörer EAU (jung) beurteilte leiten (Wort = lei und Phon = t) richtig (Correct = TRUE) als *leiten*.

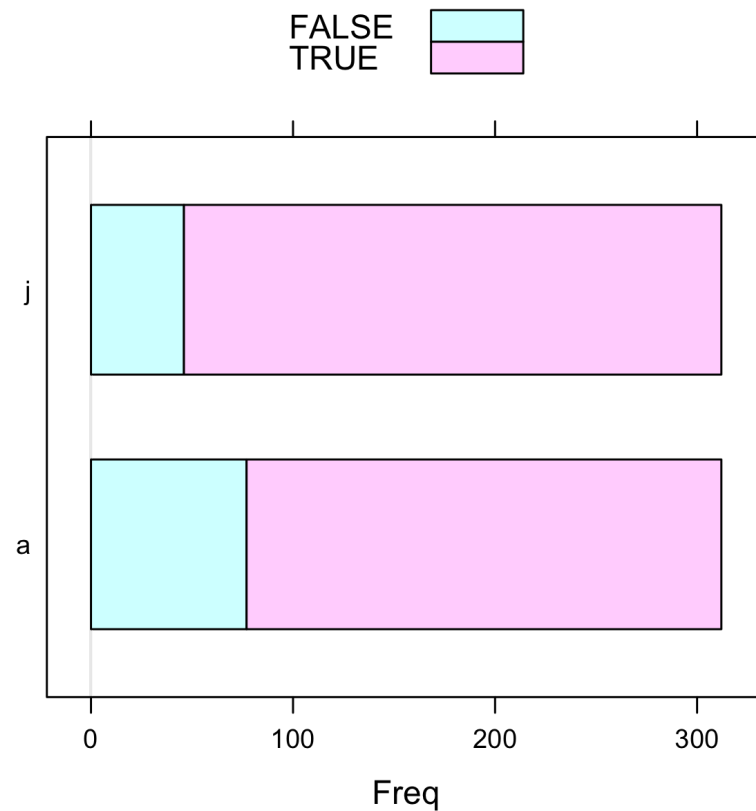
Inwiefern wird die Entscheidung (TRUE/FALSE) vom Alter beeinflusst?

Fixed Factor: Alter

Random factor(s): Hoerer, Sprecher

GLMM: Ein Fixed Faktor mit 2 Stufen

```
tab = with(alt, table(Alter, Correct))  
barchart(tab, auto.key=T)
```



GLMM-Berechnung

`a = lmer(Correct ~ Alter + (1|Hoerer) + (1|Sprecher), family=binomial, data = alt)`

eingeschätzter
log-odd = Neigung + Intercept + Hörer-
spezifischer Intercept + Sprecher-
spezifischer Intercept

$$L = b * \text{Alter} + k + k_H + k_S$$

`fixef(a)`

`ranef(a)`

(Intercept) Alterj
1.4288235 0.7488927

\$Hoerer
(Intercept)
EAU
0.13556251

\$Sprecher
(Intercept)
HSE
-0.9330030

`contrasts(alt$Alter)` j
 a 0
 j 1

$$L = 0.7488927 * 1 + 1.4288235 + 0.13556251 - 0.9330030$$

Die eingeschätzte Proportion von T wenn Hörer EAU Sprecher HSE hörte

$$p = \exp(L) / (1 + \exp(L)) \quad \text{fitted(a)[2]}$$

[1] 0.7990353

GLMM Prüftest: Ein Fixed Faktor mit 2 Stufen

`summary(a)`

```
Alterj          0.7489      0.2745      2.728    0.00637 **
```

das gleiche

`summary(glht(a, linfct=mcp(Alter="Tukey")))`

```
j - a == 0      0.7489      0.2745      2.728    0.00637 **
```

Alter hatte einen signifikanten Einfluss auf die True/False Verteilung ($z = 2.7$, $p < 0.01$)

Ein Fixed-Faktor mit 3 Stufen

Inwiefern hat der Vokal-Kontext einen Einfluss auf die Präaspiration – d.h. ob präaspiert wurde oder nicht? (Daten von Mary Stevens, 120 Sprecher).

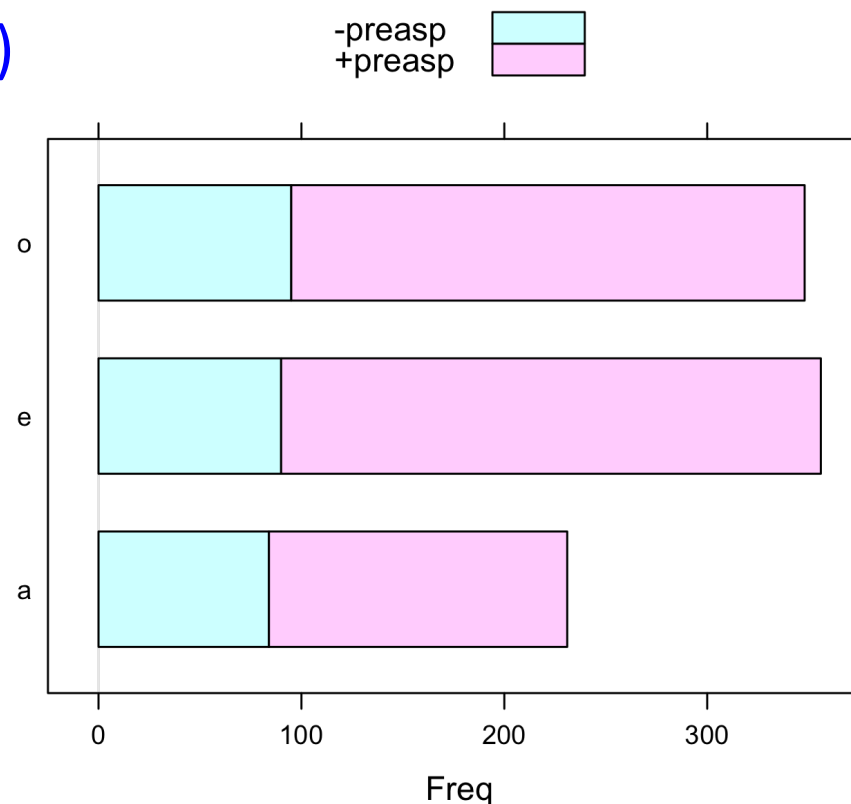
```
mes = read.table(url("http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/mes.txt"))
```

```
str(mes)
```

```
tab = with(mes, table(vtype, Pre))
```

```
barchart(tab, auto.key=T)
```

```
o = lmer(Pre ~ vtype + (1|spk),  
family=binomial, data=mes)
```



Ein Faktor mit mehreren Stufen: Prüfstatistik

summary(o)

liefert /a/ vs /e/ und /a/ vs. /o/

/e/ vs /o/ fehlt

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.6801	0.1787	-3.807	0.000141	***
vtypee	-0.6560	0.1979	-3.314	0.000919	***
vtypeo	-0.5014	0.1961	-2.556	0.010574	*

weil /a/ die **Basis-Stufe** ist

levels(mes\$vtype)

[1] "a" "e" "o"

neue Basis-Stufe

neuv = relevel(mes\$vtype, "o")

o2 = update(o, ~ . -vtype + neuv)

summary(o2)

neuva 0.5013 0.1961 2.556 0.0106 *

neuve -0.1547 0.1848 -0.837 0.4025

/e/ vs /o/

Das Gleiche:

o2 = lmer(Pre ~ neuv + (1|spk),
family=binomial, data=mes)

Ein Faktor mit mehreren Stufen: Prüfstatistik

Alternativ

```
summary(glht(o, linfct=mcp(vtype = "Tukey")))
```

```
Estimate Std. Error z value Pr(>|z|)
e - a == 0   -0.6560     0.1979   -3.314  0.00259 **
o - a == 0   -0.5014     0.1961   -2.556  0.02850 *
o - e == 0    0.1546     0.1848    0.837  0.68002
(Adjusted p values reported -- single-step method)
```

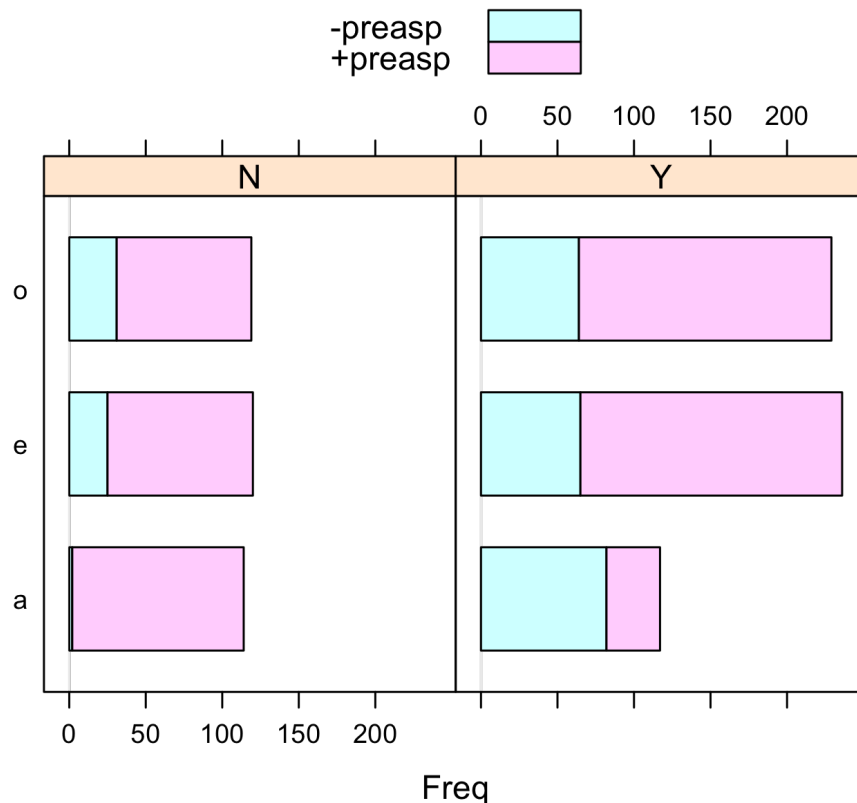
Die Wahrscheinlichkeiten sind höher, da für die Anzahl der Tests (Tukey-Verfahren) korrigiert wird.

Zwei Fixed Faktoren

Inwiefern wird die Preäspiration vom Vokal und von Pretonic (ob die nächste Silbe betont war oder nicht) beeinflusst?

```
tab = with(mes, table(vtype, ptonic, Pre))
```

```
barchart(tab, auto.key=T)
```



Vokal sig?
Pretonic sig?
Interaktion?

Zwei Fixed Faktoren

1. Interaktion prüfen

```
o = lmer(Pre ~ vtype * ptonic + (1|spk), family=binomial, data=mes)
```

```
ohne = lmer(Pre ~ vtype + ptonic + (1|spk), family=binomial, data=mes)
```

```
anova(o, ohne)
```

```
114.92      2    < 2.2e-16 ***
```

2. Wenn die Interaktion signifikant ist, dann Faktoren kombinieren

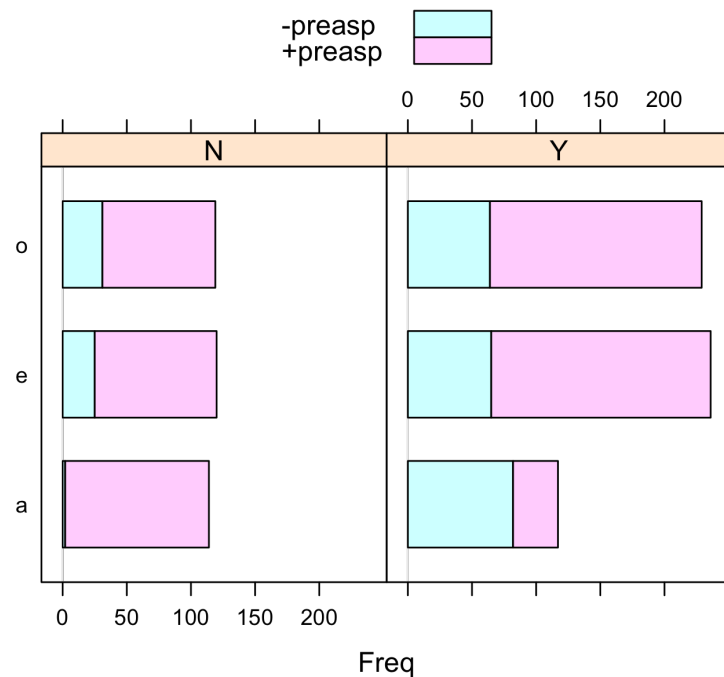
```
plabs = factor(with(mes, paste(mes$vtype, mes$ptonic)))
```

3. Modell

```
beide = lmer(Pre ~ plabs + (1|spk), family=binomial, data=mes)
```

```
summary(glht(beide, linfct=mcp(plabs = "Tukey")))
```

e N - a N == 0	3.25578	0.88213	3.691	0.00253	**
o N - a N == 0	3.73341	0.87842	4.250	< 0.001	***
o N - e N == 0	0.47764	0.35485	1.346	0.73657	
e Y - a Y == 0	-2.59039	0.29623	-8.745	< 0.001	***
o Y - a Y == 0	-2.54048	0.29698	-8.555	< 0.001	***
o Y - e Y == 0	0.04991	0.23892	0.209	0.99994	
o Y - o N == 0	0.14881	0.29416	0.506	0.99534	
a Y - a N == 0	6.42271	0.88253	7.278	< 0.001	***
e Y - e N == 0	0.57654	0.31151	1.851	0.40326	



Die Proportion [+preasp] zu [-preasp] unterschied sich signifikant in /e/ vs. /a/ (p < 0.01) und in /o/ vs /a/ (p < 0.001) aber nicht in /o/ vs /e/. Der Einfluss von Pretonic auf die Proportion war signifikant nur in /a/-Vokalen (p < 0.001) aber nicht in /e/ noch /o/ Vokalen.

Zwei Fixed Faktoren

Alternativ mit summary()

summary(o)

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)		
(Intercept)	-5.1793	0.8577	-6.039	1.55e-09	***	
vtypee	3.2558	0.8821	3.691	0.000224	***	eN vs aN
vtypeo	3.7334	0.8784	4.250	2.14e-05	***	oN vs aN
ptonicY	6.4227	0.8825	7.278	3.40e-13	***	aY vs aN
vtypee:ptonicY	-5.8462	0.9325	-6.269	3.63e-10	***	
vtypeo:ptonicY	-6.2739	0.9301	-6.746	1.52e-11	***	

lab = relevel(mes\$vtype, "o")

o2 = lmer(Pre ~ lab * ptonic + (1|spk), family=binomial, data=mes)

summary(o2)

Fixed effects:

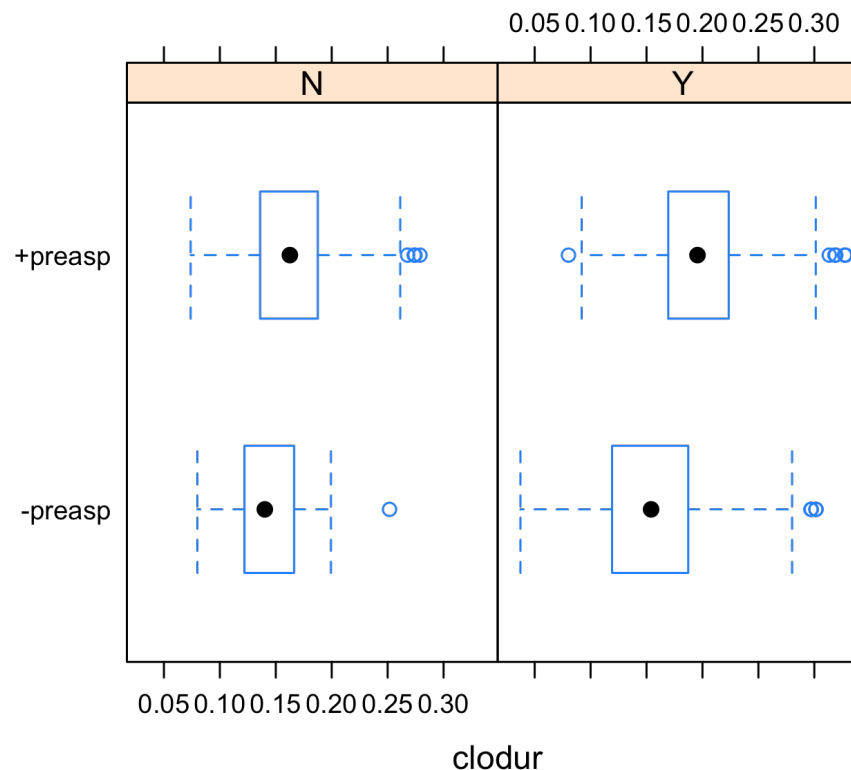
	Estimate	Std. Error	z value	Pr(> z)		
(Intercept)	-1.4459	0.2823	-5.123	3.01e-07	***	
laba	-3.7334	0.8784	-4.250	2.14e-05	***	oN vs aN
labe	-0.4776	0.3548	-1.346	0.178		oN vs eN
ptonicY	0.1489	0.2942	0.506	0.613		oN vs oY
laba:ptonicY	6.2738	0.9301	6.745	1.53e-11	***	
labe:ptonicY	0.4275	0.4278	0.999	0.318		

Zwei Fixed Faktoren – einer numerisch

Inwiefern haben Verschlussdauer (clodur) und Pretonic einen Einfluss auf die Verteilung, ob präaspiert wurde oder nicht?

Da ein Faktor numerisch ist, benötigen wir einen boxplot am besten mit bwplot():

```
bwplot(Pre ~ clodur | ptonic, data = mes)
```



Zwei Fixed Faktoren – einer numerisch

1. Prüfen ob Interaktionen vorliegen

```
o = lmer(Pre ~ clodur * ptonic + (1|spk), family=binomial, data=mes)
```

```
ohne = lmer(Pre ~ clodur + ptonic + (1|spk), family=binomial, data=mes)
```

```
anova(o, ohne)
```

$\chi^2_1 = 0.0068$ $p = 0.9343$

Nur wenn keine Interaktionen vorliegen, dann 2. und 3.

2. Pretonic sig?

Closure dur sig?

```
o2 = update(ohne, ~ . - ptonic)
```

```
o3 = update(ohne, ~ . - clodur)
```

```
anova(ohne, o2)
```

```
anova(ohne, o3)
```

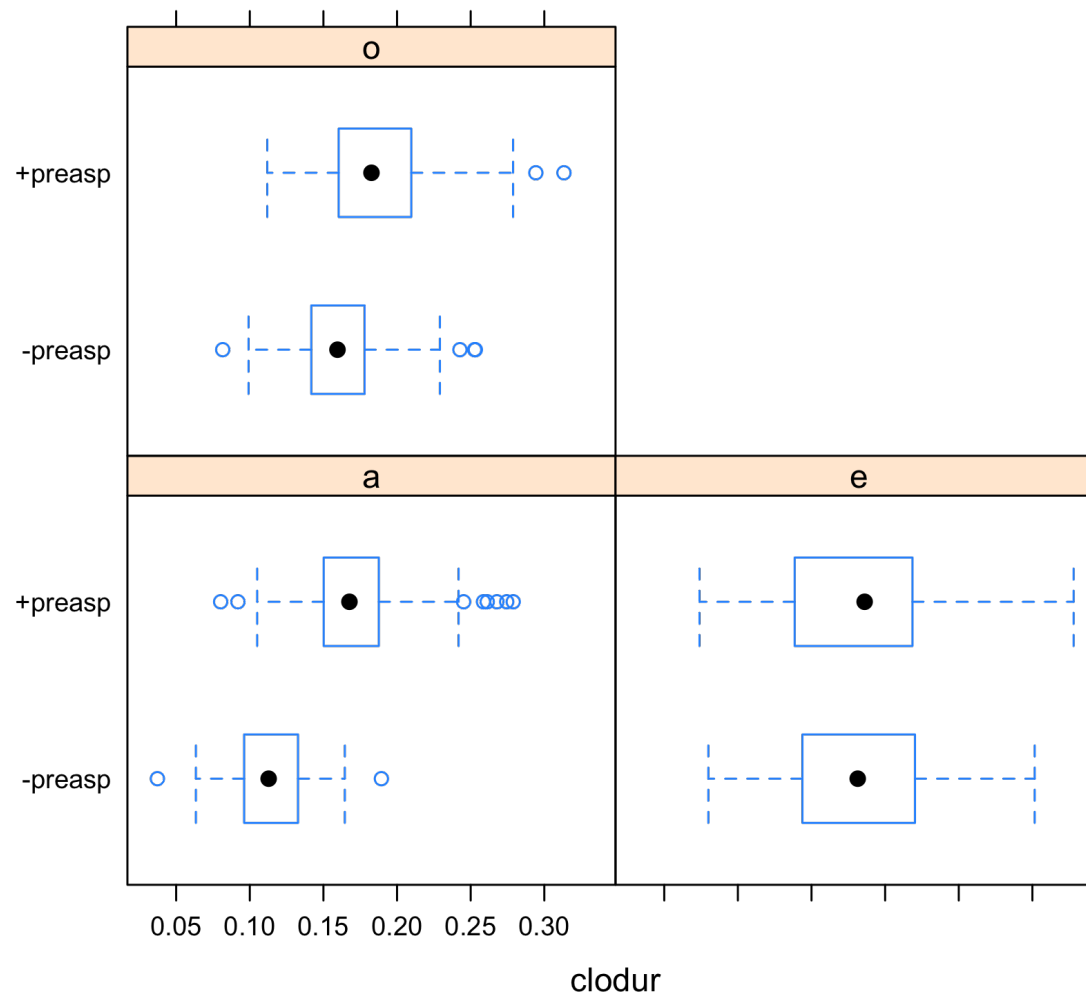
114.45 1 < 2.2e-16 **

158.14 1 < 2.2e-16 **

Zwei Fixed Faktoren – einer numerisch (mit Interaktion)

Inwiefern haben der Verschlussdauer (clodur) und Vokal (vtype) einen Einfluss auf die Verteilung, ob präaspiert wurde oder nicht?

```
bwplot(Pre ~ clodur | vtype, data = mes)
```



Zwei Fixed Faktoren – einer numerisch (mit Interaktion)

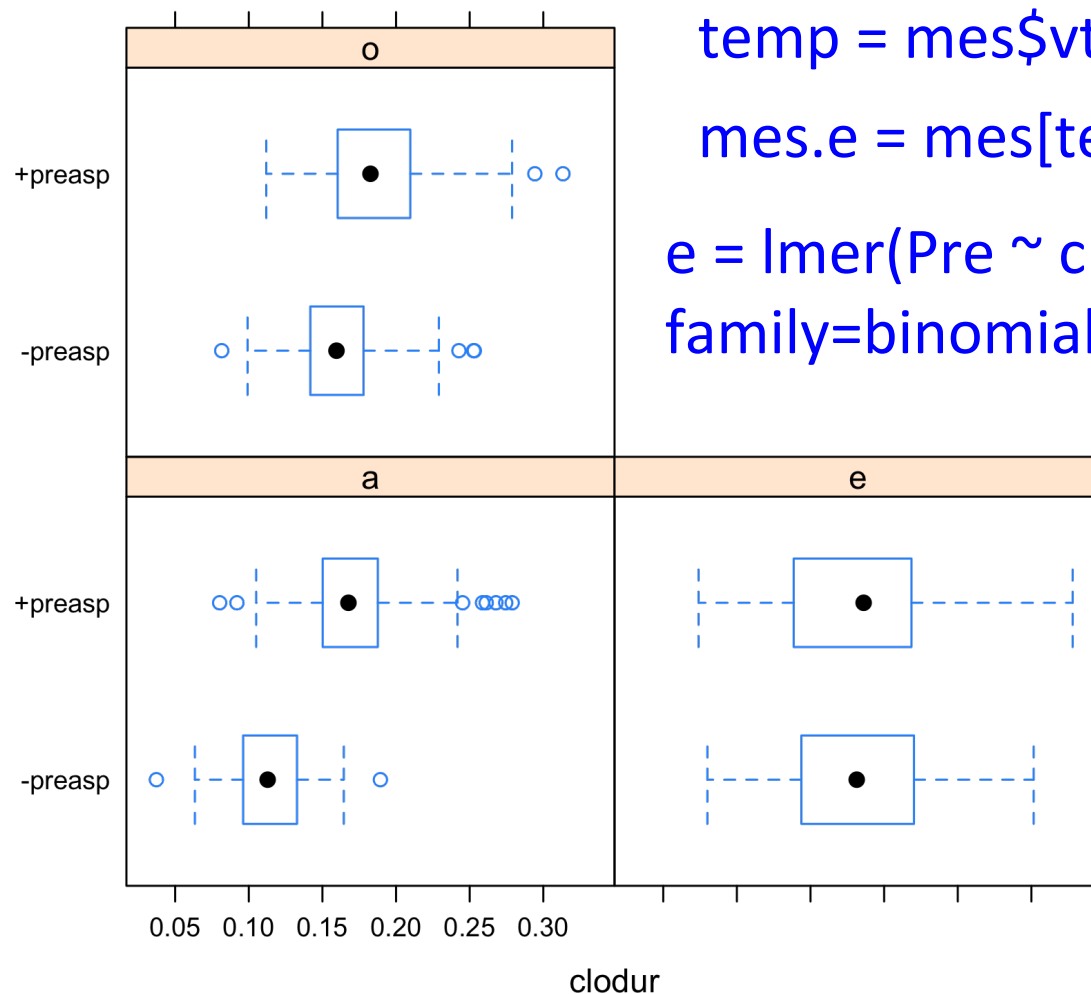
Prüfen ob Interaktionen vorliegen

```
o = lmer(Pre ~ clodur * vtype + (1|spk), family=binomial, data = mes)
o2 = lmer(Pre ~ clodur + vtype + (1|spk), family=binomial, data = mes)
anova(o, o2)
```

Interaktion hoch signifikant. 136.18 2 < 2.2e-16 ***

Zwei Fixed Faktoren – einer numerisch (mit Interaktion)

2. Hat clodur einen Einfluss auf die $[\pm\text{preasp}]$ -Verteilung getrennt in den Vokalen? z.B. unterscheiden sich $[\pm\text{preasp}]$ in /e/ (unten rechts)?



```
temp = mes$vtype == "e"
```

```
mes.e = mes[temp,]
```

```
e = lmer(Pre ~ clodur + (1|spk),  
family=binomial, data = mes.e)
```

```
summary(e)
```

Verschlussdauer hatte
keine Einfluss auf die
Verteilung von
 $[\pm\text{preasp}]$ in /e/ ($z =$
0.001 $p = 0.9992$)

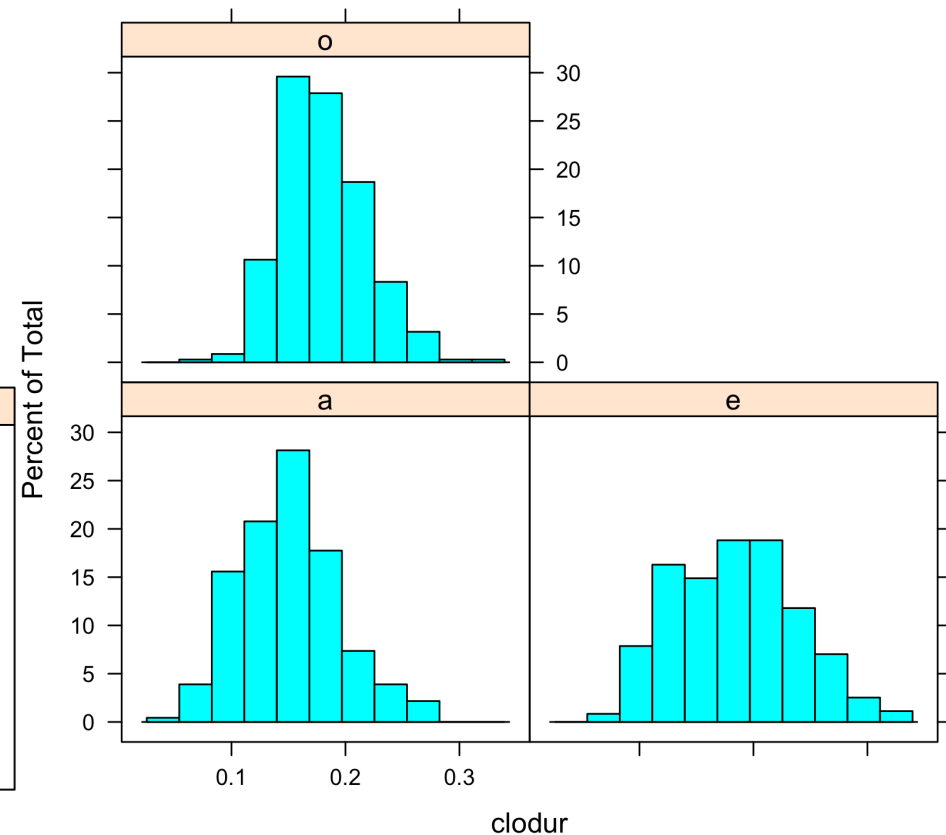
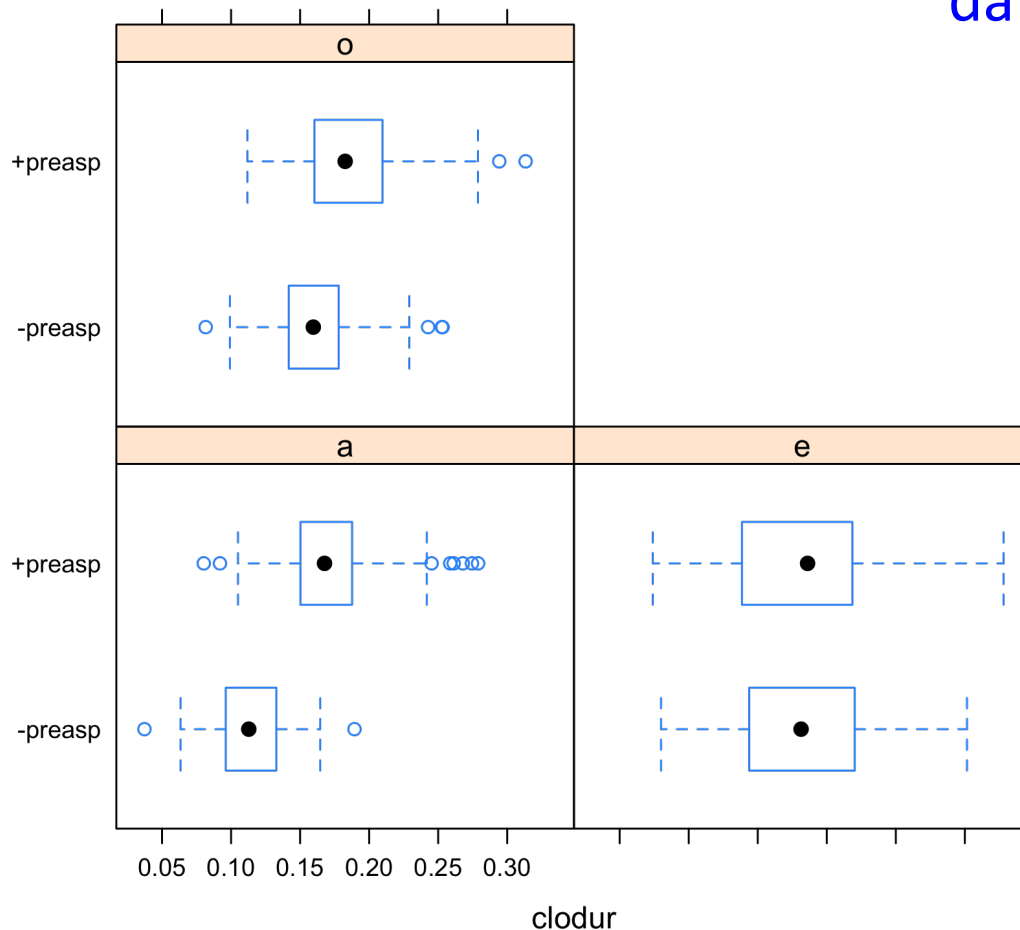
Zwei Fixed Faktoren – einer numerisch (mit Interaktion)

3. Unterscheiden sich /e, a, o/ in der $\pm$ [preasp]-Verteilung?

Sind die $\pm$ [preasp]-Verhältnisse hier anders?

oder hier?

```
histogram(Pre ~ clodur | vtype,  
data = mes)
```



3. Unterscheiden sich /e, a, o/ in der $\pm$ [preasp]-Verteilung?

Diese Antwort ist hier:

`summary(o)`

vtypee	12.564	1.537	8.174	2.98e-16	***	/a/ vs /e/
vtypeo	5.531	1.709	3.237	0.001207	**	/a/ vs. /o/

`lab = relevel(mes$vtype, "o")`

`o3 = lmer(Pre ~ clodur * lab + (1|spk), family=binomial, data = mes)`

`summary(o3)`

laba	-5.531	1.709	-3.237	0.001207	**	/o/ vs /a/
labe	7.033	1.098	6.407	1.48e-10	***	/o/ vs /e/

Vokal hatte einen signifikanten Einfluss auf die $\pm$ [preasp] Verteilung (/a/-/e/: $z = 8.2$, $p < 0.001$; /a/-/o/: $z = 3.2$, $p < 0.01$; /o/-/e/: $z = 6.4$, $p < 0.001$).

Umkipppunkte: Zwei Kontinua vergleichen

```
yuki = read.table(url("http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/yuki.txt"))
```

Die Grundfrequenz wurde zeitlich zwischen 2 Endpunkten in einem 11-stufigen Kontinuum manipuliert. In Stimulus 1 war ein f₀-Gipfel früh im Vokal und in Stimulus 11 war der Gipfel spät (und die Stimuli 2-10 waren stufenweise intermediär). Zu jedem Stimulus mussten 39 Versuchspersonen in einem forced-choice Experiment beurteilen, ob ein Stimulus wie eine Aussage oder Erstaunen wirkte. Es gab zwei verschiedene Kontinua: in einem war der mittlere Grundfrequenz H (hoch) und in dem andere L (tief). Inwiefern unterscheiden sich die H- und L-Kontinua in dem Stimulus-Umkipppunkt?

Vorgang

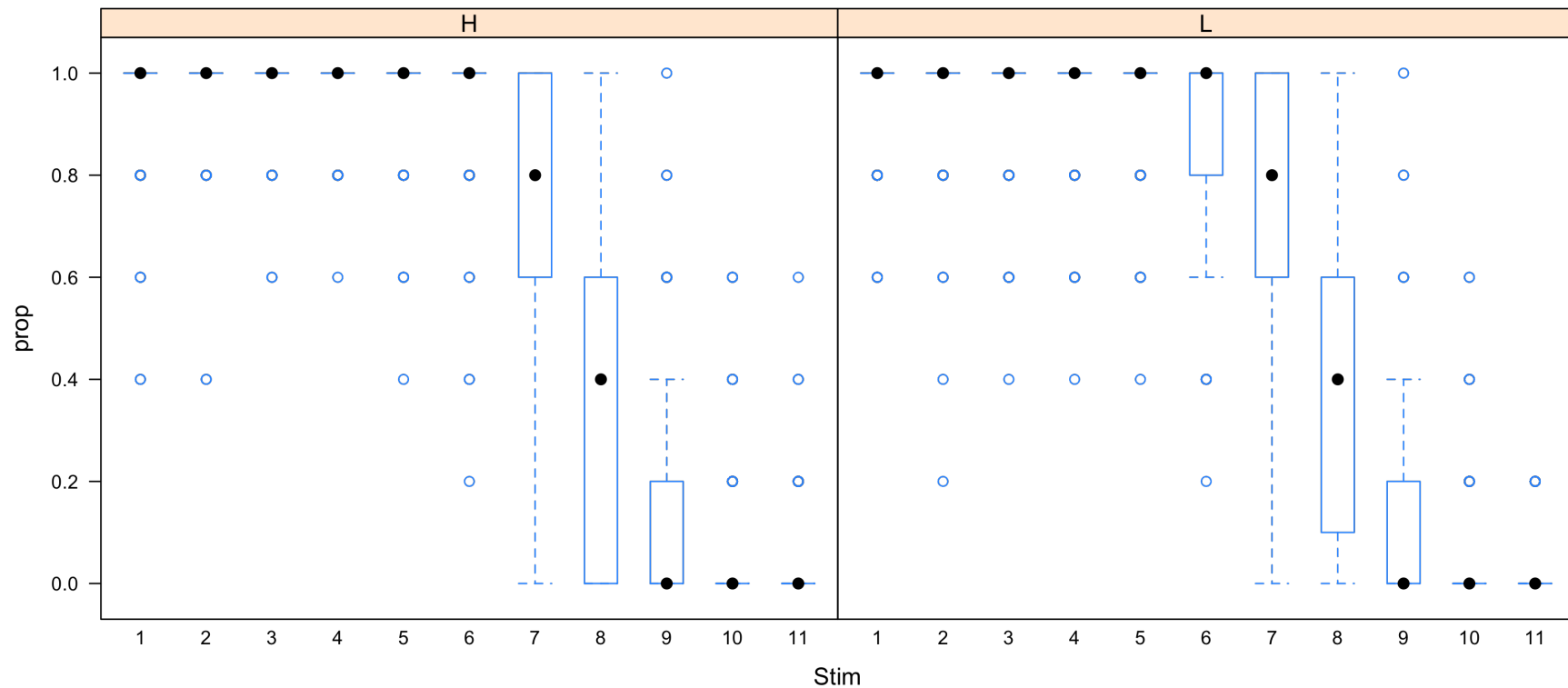
1. Abbildung (rauhe Daten)
2. Statistik: Stimulus x Kontinuum
3. Umkipppunkt und Neigung
4. Psychometrische Kurven

Abbildung (Rauhe Daten)

```
prop = with(yuki, Aus/(Aus+Ers))
```

```
yuki = cbind(yuki, prop)
```

```
bwplot(prop ~ Stim | Typ, data = yuki, horizontal=F)
```



Interaktion zwischen Stimuli und Typ

Wenn Typ (H vs L) einen Einfluss auf die Stimulus-Verteilung hat (siehe vorige Abbildung), dann müsste die Interaktion Stim * Typ signifikant sein

```
o = lmer(cbind(Aus, Ers) ~ Stim * Typ + (1|Vpn), family=binomial,  
data = yuki)  
o2 = lmer(cbind(Aus, Ers) ~ Stim + Typ + (1|Vpn), family=binomial,  
data = yuki)  
anova(o, o2)
```

Der Verlauf der proportionalen Verteilung Aussage:Erstaunen über Stimulus 1-11 wird vom Kontinuum signifikant ($\chi^2[1] = 8.0$, $p < 0.01$) beeinflusst.

Umkipppunkt und Neigung

Hier ist das Ziel festzustellen, ob sich die H- und L-Kontinua im Umkipppunkt und in der Neigung unterscheiden.

Dafür müssen pro Vpn. einen Intercept und Neigung getrennt pro Kontinuum berechnet werden

1|Vpn Nur Intercept

1+Stim|Vpn Intercept und Neigung

H- und L-Kontinua trennen

```
temp = yuki$Typ == "L"
```

```
yukiL = yuki[temp,]; yukiH = yuki[!temp,]
```

Umkipppunkt und Neigung

Modell

```
H = lmer(cbind(Aus, Ers) ~ Stim + (1 + Stim | Vpn), family=binomial, data = yukiH)
```

Koeffiziente

```
w.H = coef(H)[[1]]
```

Intercept

```
k.H = w.H[,1]
```

Neigung

```
m.H = w.H[,2]
```

Umkipppunkt

```
u.H = -k.H/m.H
```

Versuchspersonen

```
vpn.H = rownames(w.H)
```

Das gleiche für L

```
L = lmer(cbind(Aus, Ers) ~ Stim + (1 + Stim | Vpn),  
family=binomial, data = yukiL)
```

```
w.L = coef(L)[[1]]
```

```
k.L = w.L[,1]
```

```
m.L = w.L[,2]
```

```
u.L = -k.L/m.L
```

```
vpn.L = rownames(w.L)
```

Umkipppunkt und Neigung

Prüfen ob die Reihenfolge stimmt

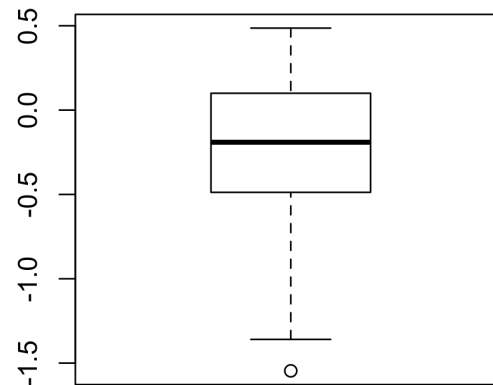
```
all(vpn.L == vpn.H)
```

```
par(mfrow=c(1,2))
```

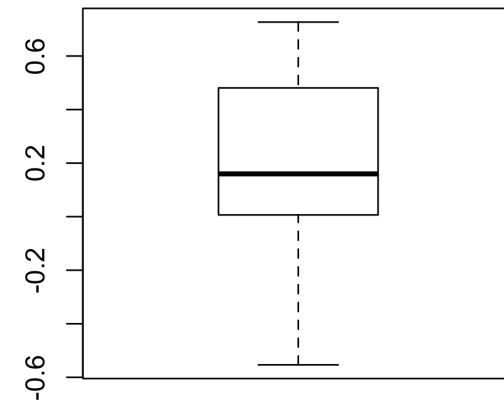
```
boxplot(u.L - u.H)
```

```
boxplot(m.L - m.H)
```

Umkipppunkt (Unterschied)



Neigung (Unterschied)



```
t.test(u.L, u.H, paired=T)
```

```
t.test(m.L, m.H, paired=T)
```

Kontinuum hatte einen signifikanten Einfluss auf den Umkipppunkt ($t_{38} = 3.2$, $p < 0.01$) und auf die Neigung ($t_{38} = 3.6$, $p < 0.01$)

Psychometrische Kurven

Mittelwert Intercept

$K.L = \text{mean}(k.L)$

Mittelwert Neigung

$M.L = \text{mean}(m.L)$

Mittelwert Umkipppunkt

$U.L = \text{mean}(-K.L/M.L)$

Mittelwert Proportion

$\text{Prop.L} = \text{with}(\text{yukiL}, \text{tapply}(\text{prop}, \text{Stim}, \text{mean}))$

Das Gleiche für H

$K.H = \text{mean}(k.H)$

$M.H = \text{mean}(m.H)$

$U.H = \text{mean}(-K.H/M.H)$

$\text{Prop.H} = \text{with}(\text{yukiH}, \text{tapply}(\text{prop}, \text{Stim}, \text{mean}))$

Psychometrische Kurven

Für L

```
ylim = c(0,1); xlim = c(1,11)
```

```
# Kurve
```

```
curve(exp(M.L * x + K.L) / (1 + exp(M.L * x + K.L)), xlim=xlim,  
ylim=ylim, ylab="Aussage (Proportion)", xlab = "Stimulus Nummer")
```

```
# Mittelwert rauhe Proportionen
```

```
points(1:11, Prop.L)
```

```
abline(v=U.L, lty=2)
```

Für H

```
par(new=T)
```

```
curve(exp(M.H * x + K.H) / (1 + exp(M.H * x + K.H)), add=T, col=2)
```

```
points(1:11, Prop.H, col=2)
```

```
abline(v=U.H, lty=2, col=2)
```

Psychometrische Kurven

