

# Lineare Multilevel Regression in R - Teil 1

Benjamin Schlegel

12. Dezember 2016

In diesem Artikel wird erklärt, wie man ein lineares Mehrebenenmodell in R rechnen kann. Dazu wird der Datensatz "eidgenössische Abstimmungen auf Gemeindeebene" verwendet, welcher unter [Data](#) heruntergeladen werden kann.

Zuerst wird der Datensatz eingelesen und in ein `data.table` Objekt umgewandelt. Anschliessend werden die Anzahl Abstimmungen jedes Abstimmungstermins eruiert [Ein Erklärung wie man in R mit `data.table` aggregieren kann, findet man im Blog [Aggregieren mit data.table in R](#)]. Nach dem aggregieren wird die neu gewonnene Information wieder in den ursprünglichen Datensatz überführt.

```
data = read.csv("gemeinden_abstimmungen.csv", stringsAsFactors = F)
library(data.table)
data = data.table(data)

datum.vorlagen = data[,.(anzahl.vorlagen=round(length(vorlage_nr)/2291)),by=Datum]

data = merge(data, datum.vorlagen, by="Datum")
```

Um pro Abstimmung nur eine Abstimmung zu haben und die Fallzahl zu reduzieren fürs Beispiel, wird für die aktuellsten 60 Abstimmungen im Datensatz jeweils das Thema genommen, wo der Autor davon ausgeht, dass es am meisten mobilisiert hat. [In einer wissenschaftlichen Arbeit müsste die Auswahl begründet werden. Um die Methode zu demonstrieren soll diese Vorgehensweise genügen.]

```
abstimmungen = data[vorlage_nr %in% c(6010,5970,5940,5910,5880,5860,5840,
5800,5750,5720,5710,5680,5660,5650,5600,5550,5540,5521,5510,
5500,5470,5430,5420,5400,5350,5320,5300,5290,5280,5270,5250,
5220,5200,5190,5170,5160,5130,5080,5060,5010,4930,4910,4891,
4880,4850,4830,4770,4740,4700,4670,4640,4630,4550,4530,4520,
4480,4440,4400,4370,4340), ]
```

Zu guter letzt wird noch die Variable X aus dem Datensatz entfernt und die Variablen Stimmbeteiligung und ja.proz in numerische Variablen konvertiert. Nun sind die Daten bereit.

```
abstimmungen[,c("X", "ja.proz", "Stimmbeteiligung") :=
.(NULL, as.numeric(ja.proz), as.numeric(Stimmbeteiligung))]
```

Es gibt verschiedene Bibliotheken in R, um Multilevel Modelle zu rechnen. Eine davon ist [lme4](#), welche hier demonstriert wird. Mit der Funktion `lmer` kann ein lineares Multilevel-Modell geschätzt werden. Das Ziel im Beispiel ist es, denn Einfluss des Abstimmungsthemas auf die Stimmbeteiligung herauszufinden, wobei die Daten auf Gemeindeebene vorliegen und als Level zwei die Kantonsebene verwendet wird. Im ersten Schritt kann das Anova-Modell gerechnet werden, um die Intraclass-Korrelation zu berechnen. Dabei wird die zweite Ebene in der Klammer hinter dem | angegeben. Alles vor dem | in der Klammer ist die, was über die zweite Ebene variiert, alles vor der Klammer die, was nicht über der zweiten Ebene konstant ist. Beim Anova-Modell wird jeweils bei beiden Ebenen nur der Achsenabschnitt

(1) verwendet.

```
library(lme4)
model.anova = lmer(Stimmbeteiligung ~ 1 + (1 | Kanton), data=abstimmungen, REML=F)
summary(model.anova)
```

```
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: Stimmbeteiligung ~ 1 + (1 | Kanton)
Data: abstimmungen
```

AIC	BIC	logLik	deviance	df.resid
1050382.2	1050411.7	-525188.1	1050376.2	136879

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.5758	-0.6589	-0.0016	0.6702	4.9918

Random effects:

Groups	Name	Variance	Std.Dev.
Kanton	(Intercept)	31.24	5.589
	Residual	125.77	11.215

Number of obs: 136882, groups: Kanton, 26

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	46.682	1.098	42.51

Um die Intraclass-Korrelation (ICC) zu berechnen, wird die Varianz des Achsenabschnitts auf zweiter Ebene durch die Summe dieser Varianz mit der Varianz der Residuen dividiert. Die Varianz des Random-Achsenabschnitts ist 31.24, die Varianz der Residuen 125.77.

$$ICC = \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + \sigma_{\epsilon}^2}$$

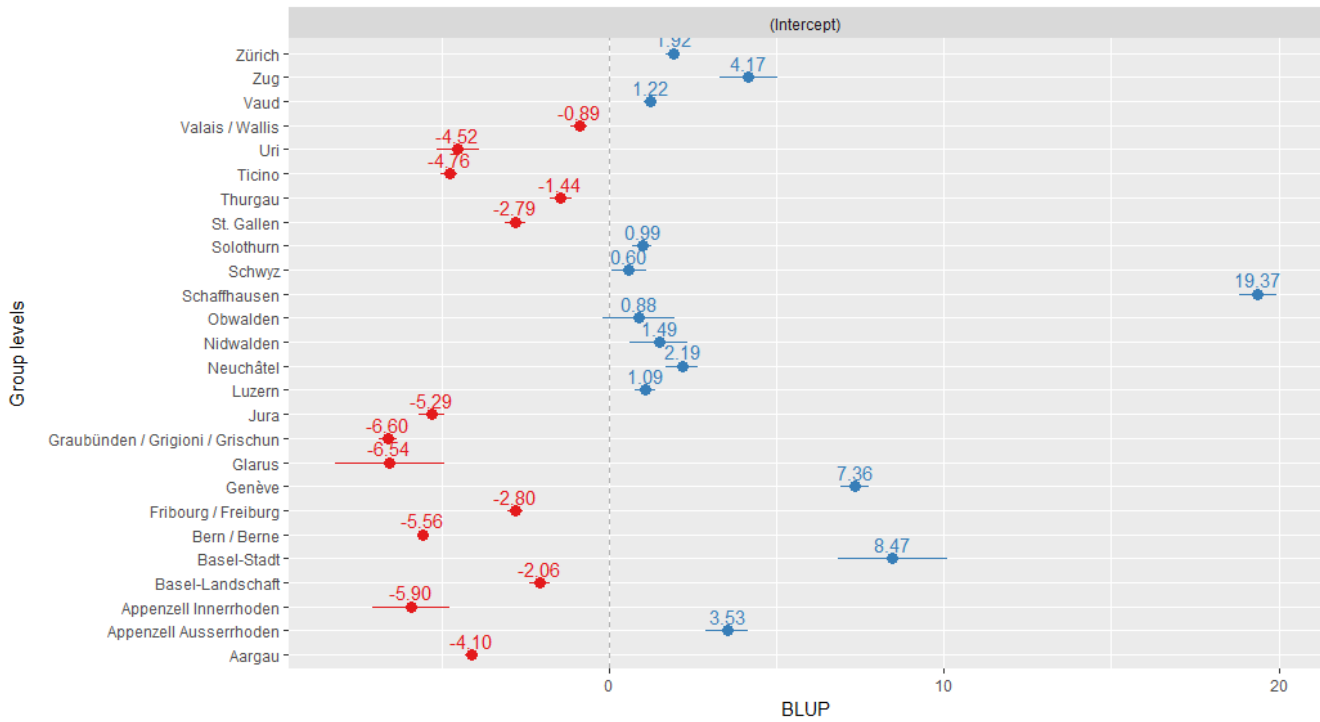
```
31.24 / (31.24 + 125.77)
```

Die Intraclass-Korrelation ist 0.199. Der Wert liegt immer zwischen 0 und 1, wobei 0 bedeutet, dass die Fälle innerhalb einer Gruppe (hier Kanton) nichts miteinander zu tun haben, 1 bedeutet, dass sie perfekte Kopien voneinander sind. Ein Werte von 0.2 bedeutet, dass sie zwar Gemeinsamkeiten haben, trotzdem aber recht verschieden sind.

Im nächsten Schritt werden die kantonalen Unterschiede dargestellt. Dazu kann die Funktion [sjp.lmer](#) aus der Bibliothek

sjPlot verwendet werden.

```
library(sjPlot)
sjp.lmer(model.anova)
```



Es ist klar erkennbar, dass Schaffhausen oben aussticht und mehr eine höhere Partizipation aufweist. Die Stimmpflicht hat also zumindest bei diesem Punkt ihr Wirkung entfaltet. Unterdurchschnittlich nehmen hingegen die Glarner und Appenzell Innerrödler teil, was an den Landsgemeinden liegen könnte.

Der 2. Teil kann [hier](#) gelesen werden.