

# MSc. Research Methods - Statistikteil Lösungen 2018

Gian-Andrea Egeler

November 2018

## Übung 5.2S: GLMM

Führt mit dem Datensatz novanimal.csv eine logistische Regression durch, wobei ihr die einzelnen Käufer (single campus\_card holder) als weitere Variable mitberücksichtigt. Kann der Fleischkonsum durch das Geschlecht, die Hochschulzugehörigkeit und das Alter erklärt werden?

Ähnliches Vorgehen wie bei der Übung 4.2S:

- Generiert eine neue Variable "Fleisch" (0 = kein Fleisch, 1 = Fleisch)
- Entfernt fehlende Werte aus der Variable "Fleisch"
- Lasst für die Analyse den Menü-Inhalt «Buffet» weg
- Definiert ein Modell mit einem random intercept und wendet es auf den Datensatz an
- Interpretiert die Ergebnisse

## Musterlösung Übung 5.2S: GLMM

```
# Generiert eine Dummyvariable: Fleisch 1, kein Fleisch 0
df <- nova # kopiert originaler Datensatz
df$meat <- ifelse(nova$label_content == "Fleisch", 1, 0)
df_ <- df[df$label_content != "Buffet", ] # entfernt Personen die sich ein Buffet Teller gekauft haben

# Löscht alle Missings bei der Variable "Fleisch"
df_ <- df_[!is.na(df_$meat), ]

# setzt andere Reihenfolge für die Hochschulzugehörigkeit
df_$member <- factor(df_$member, levels = c("Studierende", "Mitarbeitende"))

# sieht euch die Verteilung zwischen Fleisch und kein Fleisch an
table(df_$meat)

##
##      0      1
## 387 564

# definiert das logistische Modell mit card_num als random intercept und wendet es auf den Datensatz an
mod0 <- glmer(meat ~ gender + member + age + (1|card_num), data = df_, binomial("logit"))
summary(mod0)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: binomial ( logit )
## Formula: meat ~ gender + member + age + (1 | card_num)
##   Data: df_
##
##      AIC      BIC    logLik deviance df.resid
## 1246.0   1270.3   -618.0   1236.0      946
##
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -1.6942 -0.8774  0.5235  0.6446  1.3379
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   card_num (Intercept) 0.9459   0.9726
## Number of obs: 951, groups:  card_num, 563
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.21059    0.34861   0.604   0.546
## genderM           0.93317    0.19349   4.823 1.41e-06 ***
## memberMitarbeitende 0.31142    0.25672   1.213   0.225
## age              -0.01664    0.01219  -1.365   0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) gendrM mmbrMt
## genderM      -0.346
## mmbrMtrbtnd  0.443  0.116
## age          -0.871 -0.027 -0.708
```

```
# Pseudo R2
library(MuMin)
r.squaredGLMM(mod0)
```

```
##              R2m      R2c
## theoretical 0.04639524 0.2593393
## delta      0.03906337 0.2183557
```

```
# das marginale R2 gibt uns die erklärte Varianz der fixen Effekte
# das conditionale R2 gibt uns die erklärte Varianz für das ganze Modell (mit fixen und variablen Effekten)
# für weitere Informationen: https://rdrr.io/cran/MuMin/man/r.squaredGLMM.html
```

```
# zusätzliche Informationen, welche für die Interpretation gut sein kann
# berechnet den Standardfehler
se <- sqrt(diag(vcov(mod0)))
```

```
# zeigt eine Tabelle der Schätzer mit 95% Konfidenzintervall => falls 0 enthalten dann ist der Unterschied signifikant
tab1 <- cbind(Est = fixef(mod0), LL = fixef(mod0) - 1.96 * se, UL = fixef(mod0) + 1.96 * se)
```

```
# erzeugt die Odds Ratios
exp(tab1)
```

```
##              Est      LL      UL
## (Intercept)  1.2344121 0.6233350 2.444549
## genderM      2.5425651 1.7400936 3.715109
## memberMitarbeitende 1.3653596 0.8255046 2.258263
## age          0.9835007 0.9602809 1.007282
```

## Methoden

Die Responsevariable “Fleischkonsum” ist eine binäre Variable. Demnach wird eine multiple logistische Regression mit den Prädiktoren “Alter”, “Geschlecht” und “Hochschulzugehörigkeit” gerechnet. Da in den Daten gewisse Individuen mehrmals vorkommen, wird das Individuum (Variable `card_num`) als variabler Effekt in das Modell aufgenommen.

## Ergebnisse

Table 1: Modellschätzer und das dazugehörige 95% Konfidenzintervall

	Est	LL	UL
Intercept	0.21	-0.47	0.89
Geschlecht (Mann)	0.93	0.55	1.31
Hochschulzugehörigkeit (Mitarbeitende)	0.31	-0.19	0.81
Alter	-0.02	-0.04	0.01

Lediglich das Geschlecht nimmt einen signifikanten Einfluss auf den Fleischkonsum (siehe Table 1). Die Prädiktoren Alter und Hochschulzugehörigkeit nehmen keinen signifikanten Einfluss auf den Fleischkonsum. Die Chance, dass Männer Fleisch konsumieren ist 2.5mal höher als bei Frauen (siehe Table 2). Das marginale  $\text{pseudo-}R^2$  zeigt uns, dass die fixen Effekte kaum zur Varianzaufklärung beitragen.

Table 2: Odds Ratios und das dazugehörige 95% Konfidenzintervall

	Est	LL	UL
Intercept	1.23	0.62	2.44
Geschlecht (Mann)	2.54	1.74	3.72
Hochschulzugehörigkeit (Mitarbeitende)	1.37	0.83	2.26
Alter	0.98	0.96	1.01