

## Musterloesung Aufgabe 5.2S: GLMM

---

Lese-Empfehlung Kapitel 4.3.1 von Christopher Molnar

Interessierte hier oder hier

Download R-Skript

Download PDF

---

### kommentierter Lösungsweg

```
df <- nova # kopiert originaler Datensatz

# Generiert eine Dummyvariable: Fleisch 1, kein Fleisch 0
df %<>%
  # entfernt Personen die sich ein Buffet Teller gekauft
  filter(label_content != "Hot and Cold") %>%
  # ihr könnt keine Angabe vernachlässigen, sind (nur) 54 Personen
  filter(age_group != "keine Angaben") %>%
  mutate(label_content = str_replace_all(.$label_content,
                                          c("Fisch|Geflügel"),"Fleisch")) %>%
  mutate(meat = if_else(.$label_content == "Fleisch", 1, 0)) %>%
  # setzt andere Reihenfolge für die Hochschulzugehörigkeit, nur für die Interpretation
  # nützlich: neu Referenzkategorie Studierende (vorher Mitarbeitende)
  mutate(member = factor(.$member, levels = c("Studierende", "Mitarbeitende")))

# wie viele NA's hat es dirn (uns interessiert v.a. die responsevariable: meat)
sum(is.na(df$meat)) #Amelia::missmap(df_)
```

```
## [1] 0
```

```
# sieht euch die Verteilung zwischen Fleisch und kein Fleisch an,
# beide kategorien kommen nicht gleich häufig vor, aber nicht super tragisch
prop.table(table(df$meat)) # gibt die prozente an
```

```
##
##      0      1
## 0.4782123 0.5217877
```

```
table(df$meat) # gibt die absoluten werte an
```

```
##
##      0      1
## 8560 9340
```

```

# definiert das logistische Modell mit ccrs als random intercept und
# wendet es auf den Datensatz an

library(lme4)
#dauert ein paar Sekunden
mod0 <- glmer(meat ~ gender + member + age_group + (1|ccrs),
              data = df, binomial("logit"))

# lasst euch das Modell anzeigen: sieht so aus, als ob v.a. Geschlecht eine
# Rolle spielt
# wahrnehmung kann vernachlässigt werden (aufgrund der unicode resp.
# umlaute in den variablen)
summary(mod0)

```

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: meat ~ gender + member + age_group + (1 | ccrs)
## Data: df
##
##      AIC      BIC   logLik deviance df.resid
## 22611.1 22665.6 -11298.5 22597.1    17893
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.9879 -0.8253  0.4831  0.7618  3.2980
##
## Random effects:
## Groups Name          Variance Std.Dev.
## ccrs (Intercept) 1.047    1.023
## Number of obs: 17900, groups: ccrs, 1427
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.61178    0.07146  -8.561 < 2e-16 ***
## genderM         0.85769    0.07384  11.615 < 2e-16 ***
## memberMitarbeitende 0.14252    0.11044   1.290 0.19690
## age_group26 bis 34-jährig -0.23126    0.08914  -2.594 0.00948 **
## age_group35 bis 49-jährig -0.07394    0.13687  -0.540 0.58905
## age_group50 bis 64-jährig -0.01069    0.17538  -0.061 0.95138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Warning in abbreviate(rn, minlength = 11): abbreviate mit nicht-ASCII Zeichen
## genutzt

##
## Correlation of Fixed Effects:

## Warning in abbreviate(rn, minlength = 6): abbreviate mit nicht-ASCII Zeichen
## genutzt

```

```
##          (Intr) gendrM mbrMt a_26b3 a_35b4
## genderM      -0.658
## mbrMtrbtnd -0.076  0.029
## ag_26b34-Ã¼ -0.411 -0.025 -0.353
## ag_35b49-Ã¼ -0.281  0.041 -0.663  0.467
## ag_50b64-Ã¼ -0.208  0.020 -0.598  0.393  0.517
```

*## erste Interpretation: Geschlecht (Mann) und Alter (junge Personen) scheinen den Fleischkonsum positiv zu beeinflussen, d.h. könnte man vernachlässigen. Ich lasse aus inhaltlichen Gründen aber im Modell drin*

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

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.
```

```
## Warning: The null model is correct only if all variables used by the original
## model remain unchanged.
```

```
##          R2m      R2c
## theoretical 0.03760018 0.2699647
## delta      0.03240612 0.2326719
```

*# das marginale R<sup>2</sup> (r2m) gibt uns die erklärte Varianz der fixen Effekte: hier 4% (das ist sehr wenig)  
 # das conditionale R<sup>2</sup> (r2c) gibt uns die erklärte Varianz für das ganze Modell  
 # (mit fixen und variablen Effekten): hier 27% (ganz ok, aber auch nicht super mega)  
 # für weitere Informationen: <https://rdrr.io/cran/MuMIn/man/r.squaredGLMM.html>*

*# zusätzliche Informationen, welche für die Interpretation gut sein kann  
 # berechnet den Standardfehler (mehr infos: <https://www.youtube.com/watch?v=r-txC-dpI-E> oder hier: <http://stats.stackexchange.com/questions/26650/how-do-i-reference-a-regression-models-standard-error>)*  
 se <- sqrt(diag(vcov(mod0)))

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

```
# erzeugt die Odds Ratios
tab2 <- exp(tab1)
```

---

## Methoden

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

---

## Ergebnisse

Das Geschlecht und das Alter nehmen einen signifikanten Einfluss auf den Fleischkonsum (siehe Table 1): Männer kaufen signifikant häufiger ein fleischhaltiges Gericht als Frauen; junge Personen (15 bis 25-jährig) kaufen signifikant häufiger ein fleischhaltiges Gericht in der Mensa. Es sieht so aus, als ob die Hochschulzugehörigkeit auf den ersten Blick keinen Einfluss nimmt. Aber man müsste auch die Interaktion zwischen Geschlecht und Hochschulzugehörigkeit berücksichtigen, um ein abschliessendes Bild zu bekommen. Das kleine marginale Pseudo- $R^2$  zeigt auf, dass es nicht das "beste" Modell ist. Insbesondere die tiefe Varianzaufklärung für die randomisierte Variable (r2c; ccrs) scheint mit (nur) 4% sehr gering. Das sind Hinweise dafür, dass das Modell ggf. noch weitere Ebenen haben könnte (z.B. Standort Mensa).

Table 1: Modellschätzer (Coefficients) mit dazugehörigem 95% Konfidenzintervall

	Coefficients	Lower Limit (LL)	Uppewr Limit (UL)
Intercept	-0.61	-0.75	-0.47
Männer	0.86	0.71	1.00
Mitarbeitende	0.14	-0.07	0.36
26 bis 34-jährig	-0.23	-0.41	-0.06
35 bis 49-jährig	-0.07	-0.34	0.19
50 bis 64-jährig	-0.01	-0.35	0.33

Die Chance, dass Männer ein fleischhaltiges Gericht kaufen ist 2.36mal (+136%) höher als bei Frauen (siehe Table 2). Die Chance, dass 26 bis 34-jährige Personen ein fleischhaltiges Gericht kaufen ist kleiner (-21%) als bei den 15 bis 25-jährigen Personen.

Table 2: Odds Ratio (OR) mit dazugehörigem 95% Konfidenzintervall

	OR	Lower Limit (LL)	Uppewr Limit (UL)
Intercept	0.54	0.47	0.62
Männer	2.36	2.04	2.72
Mitarbeitende	1.15	0.93	1.43
26 bis 34-jährig	0.79	0.67	0.95
35 bis 49-jährig	0.93	0.71	1.21
50 bis 64-jährig	0.99	0.70	1.40