

4 Mehrkategoriale Regressionsmodelle

Lösung zu Aufgabe 3

```
# Daten einlesen:
load("alligator.rda")

str(alligator)

## 'data.frame': 63 obs. of 3 variables:
## $ Nahrung : Factor w/ 3 levels "A","F","I": 3 2 2 2 3 2 3 2 3 3 ...
## $ Laenge : num 1.3 1.32 1.32 1.4 1.42 1.42 1.47 1.47 1.5 1.52 ...
## $ Geschlecht: Factor w/ 2 levels "M","W": 1 1 1 1 1 1 1 1 1 1 ...

attach(alligator)

# Überblick: Bivariater Zusammenhang von Geschlecht und Nahrung
table(Geschlecht,Nahrung)

##           Nahrung
## Geschlecht A  F  I
##           M  7 22 10
##           W  3 11 10
```

(a) *# Package für Multinomiales Modell (und mehr)*

```
library(nnet)

# Multinomialles Modell mit Haupteffekten.
Multinom1 <- multinom(Nahrung ~ Geschlecht + Laenge)

## # weights: 12 (6 variable)
## initial value 69.212574
## iter 10 value 52.779702
## final value 52.779015
## converged

summary(Multinom1)

## Call:
## multinom(formula = Nahrung ~ Geschlecht + Laenge)
##
## Coefficients:
## (Intercept) GeschlechtW Laenge
## F 0.9428283 0.1583906 0.08718304
## I 5.4546440 1.3657067 -2.84340795
##
## Std. Errors:
## (Intercept) GeschlechtW Laenge
## F 1.197997 0.7830207 0.4847348
## I 1.849104 0.9245794 1.0180082
##
## Residual Deviance: 105.558
## AIC: 117.558
```

Die erste Kategorie von Nahrung (Andere) wird als Referenz verwendet. Die Chance auf Fisch als Nahrung gegenüber Andere verändert sich also kaum (vgl. Standardfehler) in Abhängigkeit von Geschlecht und Größe. Dagegen fressen weibliche Alligatoren offenbar lieber Invertebrata als männliche, auch nimmt die Chance auf Invertebrata als Nahrung mit der Größe des Alligators ab.

```
# Quantifizierung:

# Bsp.1: geschätzte (P(I | weibl) / P(A | weibl)) / (P(I | männl) / P(A | männl))
exp(summary(Multinom1)$coefficients[2,2])

## [1] 3.918491

# Bsp.2: geschätzte (P(F | weibl) / P(I | weibl)) / (P(F | männl) / P(I | männl))
exp(summary(Multinom1)$coefficients[1,2] - summary(Multinom1)$coefficients[2,2])

## [1] 0.2989987

# Bsp.3: geschätzte (P(F | +1m) / P(I | +1m)) / (P(F | .) / P(I | .))
exp(summary(Multinom1)$coefficients[1,3] - summary(Multinom1)$coefficients[2,3])

## [1] 18.7387
```

(b) *# Modell mit Wechselwirkungen*

```
Multinom2 <- multinom(Nahrung ~ Geschlecht * Laenge)

## # weights: 15 (8 variable)
## initial value 69.212574
## iter 10 value 51.705243
## iter 20 value 51.302832
## iter 30 value 51.296149
## final value 51.295865
## converged

# Vergleich über LQ-Test
anova(Multinom1, Multinom2)

##           Model Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1 Geschlecht + Laenge      120   105.5580      NA      NA      NA      NA
## 2 Geschlecht * Laenge      118   102.5917 1 vs 2     2  2.966301 0.2269216

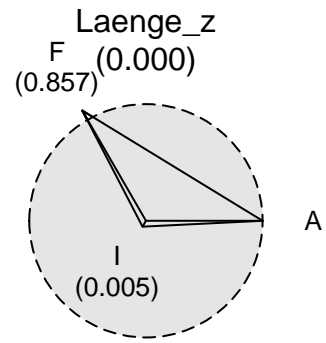
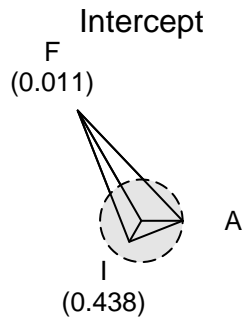
# Der p-Wert zeigt an, dass Wechselwirkungen zu keiner signifikanten
# Modellverbesserung führen
```

(c) *# lade EffectStars Package*

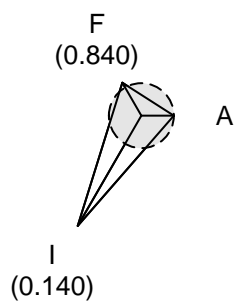
```
library(EffectStars)

# EffectStars funktionieren viel besser, wenn man alle metrischen Kovariablen
# zentriert.
Laenge_z <- Laenge - mean(Laenge)

star.nominal(Nahrung ~ Laenge_z + Geschlecht, data = alligator, symmetric = F)
```



Geschlecht
(1: W, 0.177)



```
## $odds
##   (Intercept)  Laenge_z  GeschlechtW
## A  1.0000000  1.0000000    1.000000
## F  3.0846078  1.09107432   1.171661
## I  0.5872551  0.05822557   3.918719
##
## $coefficients
##   (Intercept)  Laenge_z  GeschlechtW
## A  0.0000000  0.0000000    0.000000
## F  1.1264245  0.08716283   0.1584225
## I -0.5322959 -2.84343069   1.3657648
##
## $se
##   (Intercept)  Laenge_z  GeschlechtW
## A  0.0000000  0.0000000    0.000000
## F  0.4446342  0.4847371    0.7830328
## I  0.6861570  1.0180108    0.9245924
##
## $pvalues
##   (Intercept)  Laenge_z  GeschlechtW
## A           NaN           NaN           NaN
## F  0.01129704  0.857298099   0.8396673
## I  0.43788834  0.005220162   0.1396345
##
## $p_rel
##           Laenge_z  GeschlechtW
## [1,] 0.0001204243  0.1771527
##
## $xlim
## [1] 10.18867 39.18719
##
## $ylim
## [1] 9.75761 39.10881
```

Der Kreis entspricht einem Effekt β_{jr} von 0, d.h. einer Strahllänge von $exp(0) = 1$. Sternspitzen innerhalb dieses Einheits- oder Nulleffekt-Kreises bedeuten negativen Einfluss einer Kovariable auf die jeweilige Responsekategorie, Sternspitzen ausserhalb des Kreises einen positiven Effekt. Beachte: Die Referenzkategorie ist hier Andere, dementsprechend liegen die Effekte für Andere immer exakt auf dem Nullkreis. Die Zahlen unter den jeweiligen Effekten sind die p-Werte für die einzelnen Koeffizienten einer Variable für eine Responsekategorie, die Zahlen unter dem Variablennamen geben den p -Wert dieser Variable als ganzes an. (Beachte: hier gehen alle Kovariablen grundsätzlich mit mehreren Koeffizienten ein!)