

4 Mehrkategoriale Regressionsmodelle (II)

Aufgabe 1

Der Datensatz `alligator` enthält Informationen über 63 Alligatoren (gefangen im Lake George, Florida) und deren bevorzugte Nahrung. Letztere ist klassifiziert nach: *Fisch*, wirbellose Tiere (*Invertebrata*) und *Andere*.

- Fitten Sie ein multinomiales Logit-Modell mit Haupteffekten, das den Einfluss von Geschlecht und Größe auf die Wahl der Nahrung beschreibt, und interpretieren Sie das Ergebnis. *Hinweis:* Verwenden Sie die Funktion `multinom` im R-Package `nnet`.
- Überprüfen Sie mit einem geeigneten Test, ob auch Wechselwirkungen zwischen den Prädiktoren im Modell berücksichtigt werden sollten.
- Visualisieren Sie die Parameter des Modells mit Hilfe von sogenannten “Effect Stars”. Verwenden Sie dazu das R-Package `EffectStars`.

Aufgabe 2

Gegeben seien k latente Variablen (sog. Random Utilities) der Form

$$U_j = u_j + \epsilon_j, \quad j = 1, \dots, k.$$

Dabei seien $\epsilon_1, \dots, \epsilon_k$ iid Zufallsvariablen, die der Verteilung F folgen. Nach dem Prinzip des maximalen zufälligen Nutzens ist die beobachtbare Response-Variable Y gegeben durch

$$Y = r \Leftrightarrow U_r = \max_j U_j.$$

Man zeige, dass

$$P(Y = r) = \frac{\exp(u_r)}{\sum_{j=1}^k \exp(u_j)}$$

gilt, falls die ϵ_j einer Gumbel-Verteilung folgen, d.h. falls $F(x) = \exp(-\exp(-x))$.

(*Tipp: Stellen Sie zunächst $P(Y = r)$ als Funktion der u_j und der Zufallsvariablen ϵ_j dar und integrieren Sie dann über ϵ .)*

Aufgabe 3

Das Proportional-Hazards-Modell zur Modellierung einer ordinalen Response-Variable Y mit k Ausprägungen ist gegeben durch

$$P(Y \leq r | \mathbf{x}) = 1 - \exp(-\exp(\theta_r + \mathbf{x}'\boldsymbol{\gamma})), \quad r = 1, \dots, k - 1.$$

Man zeige, dass dieses Modell alternativ durch

$$P(Y = r | Y \geq r, \mathbf{x}) = 1 - \exp(-\exp(\tilde{\theta}_r + \mathbf{x}'\boldsymbol{\gamma})), \quad r = 1, \dots, k - 1,$$

mit $\tilde{\theta}_r = \log(\exp(\theta_r) - \exp(\theta_{r-1}))$ dargestellt werden kann.