Generalisierte Lineare Modelle – Übungen: Blatt 3

1. Zeige, dass bei einem logistischen Regressionsmodell mit beliebig vielen Faktoren im linearen Prädiktor η_i (sowohl als Haupt- als auch als Interaktionseffekte) für $y_i \sim \text{Bernoulli}(\mu_i)$ $(i = 1, \ldots, n)$ mit $\text{logit}(\mu_i) = \eta_i$ immer deterministisch $X^2 = n$ resultiert.
(Hinweis: betrachte zuerst eine einfach Zufallsstichprobe und zeige, dass zumindest dafür diese Eigenschaft hält. Verallgemeinere die Modellannahme auf einen Prädiktor mit einem k-stufigen Effekt und weise dafür nach, dass wiederum $X^2 = n$ gilt. Da jede Kombination von Haupteffekten und deren Interaktionen alternativ auch durch 1 mehrstufigen Faktor darstellbar ist, folgt damit die Behauptung.)

2. (Agresti Problem 7.3) Mit den folgenden Daten wird die politische Einstellung einer Person mit Geschlecht und Rasse verglichen.

		Politische Einstellung		
Geschlecht	Rasse	demokratisch	republikanisch	unabhängig
männlich	weiß	132	176	127
	schwarz	42	6	12
weiblich	weiß	172	129	130
	schwarz	56	4	15

   (a) Finde ein Logit-Modell welches auf Abweichungen von einer Referenzzelle basiert und gut passt. Verwende dazu `multinom` aus der `library(nnet)`.
   (b) Schätze dasselbe Modell nur unter Verwendung von `glm` und vergleiche die Parameterschätzungen.
   (c) Interpretiere die geschätzten Effekte auf die Chancen (odds), dass die politische Gesinnung *demokratisch* anstatt *republikanisch* ist.

3. (Agresti Problem 8.5) Aufzeichnungen über Autounfälle in Florida in 1988 ergaben:

Verwendung der Sicherheitsgurte	aus dem Auto geschleudert	Unfall endet nicht tödlich	tödlich
angeschnallt	ja	1105	14
	nein	411111	483
nicht angeschnallt	ja	4624	497
	nein	157342	1008

   (a) Finde ein Loglineares Modell, das die Daten gut beschreibt. Argumentiere die Suche.
   (b) Vergleiche die Beobachtungen mit den Prognosen unter den folgenden 5 Modellen: $B + E + F$, $B : E + F$, $B : E + B : F$, $B : E + B : F + E : F$, $B : E : F$.
   Erstelle dazu eine Tabelle, deren ersten drei Spalten die *yes, no* Stufen der drei Faktoren B (belt/Gurte), E (ejected/rausgeschleudert), und F (fatal/tödlicher Ausgang) beinhalten, gefolgt von den entsprechenden Prognosen unter diesen Modellen. Was fällt hierbei auf?
   (c) Schätze unter Verwendung aller obigen Modelle die sogenannten *konditionalen odds ratios*, bei denen jeweils ein Faktor auf einer Stufe (z.B. auf der ersten Stufe) fest gehalten ist, d.h.

$$\frac{\mu_{11(1)}\mu_{22(1)}}{\mu_{12(1)}\mu_{21(1)}}, \quad \frac{\mu_{1(1)1}\mu_{2(1)2}}{\mu_{1(1)2}\mu_{2(1)1}}, \quad \frac{\mu_{(1)11}\mu_{(1)22}}{\mu_{(1)12}\mu_{(1)21}}.$$

Wie sehen die entsprechenden Schätzungen aus, wenn der festgehaltene Faktor auf seine zweite Stufe fixiert wird. Interpretiere alle erzielten Resultate im Vergleich zum dafür verwendeten Modell.

(d) Betrachte nun als Responsevariable die Information, ob eine Person getötet wurde oder nicht. Bezeichne Y diese 2-stufige Variable und X und Z zwei erklärende Faktoren. Für ein loglineares Modell mit allen zweifachen Interaktionen ($X : Y$, $X : Z$, $Y : Z$) gilt dann für den Erwartungswert von Y

$$\begin{aligned}
\log \frac{P(Y=1|X=i,Z=k)}{P(Y=2|X=i,Z=k)} &= \log \frac{\mu_{i1k}}{\mu_{i2k}} = \log \mu_{i1k} - \log \mu_{i2k} \\
&= \Big(\mu + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ}\Big) \\
&\quad -\Big(\mu + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ}\Big) \\
&= (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ})\,.
\end{aligned}$$

Der erste Term ist konstant, der zweite hängt nur von der Stufe i von X ab, und der dritte von der Stufe k von Z. Somit resultiert für das obige Logit-Modell

$$\text{logit } P(Y=1|X=i,Z=k) = \alpha + \beta_i^X + \beta_k^Z\,.$$

Passe unter Verwendung dieser Erkenntnis ein zum *unter (a) gefundenen optimalen Modell* äquivalentes Logit Modell an. Interpretiere damit die Effekte.