Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Statistik
Prof. Claudia Czado, Ph.D. (cczado@ma.tum.de)
Ariane Hanebeck (ariane.hanebeck@tum.de)

WS 2022/2023

Exercise Sheet 11
Generalized Linear Models

Discussion of the tutorial exercises on January 23 and 26, 2022

Problem 1 (*) Consider the data about the perching position of two species of lizards in the table below. The following data were observed of 409 lizards: species `species`, perch diameter `PD` and perch height `PH`. The entries in the cells correspond to the number of lizards in the respective categories.

		`PH` $>$ 4.75 feet	`PH` $\leq$ 4.75 feet
`species` = Anoli	`PD` $\leq$ 4 inches	32	86
	`PD` $>$ 4 inches	11	35
`species` = Distichus	`PD` $\leq$ 4 inches	61	73
	`PD` $>$ 4 inches	41	70

We are interested in the joint relationship of all variables.

a) We define a data set (`dat`) with the four columns `species` (with values Anoli or Disticus), `PD` (perch diameter with values small or high), `PH` (perch height with values high and low) and `FREQ` (frequency of lizards that have the respective combination of `species`, `PD` and `PH`) using the following `R` command:

```
dat <- cbind(expand.grid(Species = c("Anoli", "Disticus"),
PD = c("small","large"), PH = c("high", "low")),
FREQ = c(32, 61, 11, 41, 86, 73, 35, 70))
```

b) Investigate the pairwise dependencies of the three variables: Compute the odds-ratio for each of the three possible pairs. Then infer conclusions about dependence of the two variables of each pair from its odds-ratio. Apply Theorem 4.10 to test if the dependence is significant.

In the following, we use the auxiliary Poisson regression to estimate the parameters of the log linear model to describe the data of the contingency table. We use FREQ as the response variable and use the variables species, PD and PH as the covariates.

c) Perform an exploratory data analysis to investigate the main effects using the R-function `cat_plot` and interpret them.

d) Perform an exploratory data analysis to investigate the interaction effects using the R-function `cat_plot` and interpret them.

e) Consider only the main effects of the covariates. Write down its log-linear model using lecture notations, fit it, and discuss the goodness of the fit performing a residual deviance test at level $\alpha = 0.05$.

f) Consider all pairwise interaction terms of the covariates. Write down its log-linear model using lecture notations, fit it, and discuss the goodness of the fit performing a residual deviance test at level $\alpha = 0.05$.

g) **(Additional)** Write down all possible log-linear models using lecture notations. Find the best log-linear model based on a deviance table. If necessary, make use of partial deviance tests at level $\alpha = 0.05$.

Problem 2 (Additional) Which of the following log-linear models is a hierarchical log-linear model?

a) $l_{ijk} = \lambda + \lambda_i^1 + \lambda_k^3 + \lambda_{ik}^{13}$

b) $l_{ijk} = \lambda + \lambda_i^1 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23}$

c) $l_{ijkl} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123} + \lambda_{ijl}^{124} + \lambda_{jkl}^{234}$

d) $l_{ijkl} = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_l^4 + \lambda_{ij}^{12} + \lambda_{il}^{14} + \lambda_{jl}^{24} + \lambda_{ijl}^{124}$