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 9
Generalized Linear Models

Discussion of the tutorial exercises on January 9 and 12, 2022

Preparation Download the data set `insurance.dat` from Moodle and read its description.

Problem 1 (*) The data set `insurance.dat` consists of car insurance claims in Sweden from 1977, and the response variable is `Claims`. The data set contains four covariates with information on the policy holders of each group: `Kilometres`, `Zone`, `Bonus` and `Make` along with the number of policyholders `Insured`. The variable `Insured` corresponds to the exposure time in the Poisson model.

a) Characterize the scale (quantitative, ordinal, nominal) of all four covariates. Transform all nominal and ordinal variables into factor variables using `as.factor`.

b) Perform an exploratory data analysis to investigate the main effects of the four covariates using the function `cat_plot`. Remember to correctly specify the number of policyholders in the function. Can some of the covariate levels be aggregated here? If yes, do so.

c) Consider only the main effects of the four covariates. Select the Poisson model for the data for which the covariates are ordered by significance with respect to the AIC. Do that by performing a forward selection with the R function `step()`. Make sure that you also specify the number of policyholders in the offset. In the remainder, this model is referred to as `model.main`.

d) Which effects in `model.main` are significant at $\alpha = 0.05$?

e) Perform partial deviance tests for the nested models of `model.main` and explain the outcome with $\alpha = 0.05$.

f) We consider a group of 100 policyholders that travel 15000 to 20000 kilometers annually, live in Gotland, are in bonus category 4, and drive car model 1. What is the expected number of claims for this group of policyholders in `model.main`?

Problem 2 (Additional) Consider the model `small.model.agg` from Sheet 8 and the model `small.model` which uses the same covariates but with the data that is not aggregated.

a) Analyze the `small.model.agg`'s deviance using a rule of thumb discussed in class and perform a residual deviance test at $\alpha = 0.05$. Are there any indications of over-dispersion?

b) Analyze the `small.model`'s deviance using a rule of thumb discussed in class and perform a residual deviance test at $\alpha = 0.05$. Are there any indications of over-dispersion?

Problem 3 (Additional) Consider the following distribution for count variables Y_i.

$$Y_i|Z_i = z_i \quad \sim \quad Pois(z_i) \qquad \text{independent for } i = 1, \ldots, n$$

with

$$Z_i \quad \sim \quad Gamma(\mu_i, \nu).$$

Show that Y_i follows a negative binomial distribution with

$$\begin{aligned} E(Y_i) &= \mu_i \\ Var(Y_i) &= \mu_i + \frac{\mu_i^2}{\nu}. \end{aligned}$$