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

Additional Exercise Sheet
Generalized Linear Models

Install the dataset `Atemwegserkrankungen von Schulkindern (Respiratory diseases of children)` from https://data.ub.uni-muenchen.de/13/. There is an explanation of all the variables in german but we give an english translation for the variables we are considering in the following.

Do an analysis on this data set using a binomial response function. The response variable is `lubro` and the covariates we are considering are:

- `zone`: environmental damage at place of residence (1 = high, 2 = less, 3 = ozone),
- `aller`: allergic respiratory diseases like asthma (y/n),
- `kehle`: laryngitis (y/n),
- `schnu`: often having a cold (y/n),
- `huste`: often coughing (y/n),
- `sex`: 1 = male, 2 = female,
- `gewi`: weight.

Transform the categorical covariates into factor variables. Also group the data in order to get a binomial response instead of a binary response.

Your analysis should include the following steps:

- EDA (nonlinearity, main effects and interaction effects).
- Fitting of models.
- Comparison of models.
- Residual analysis.
- Interpretation.
- Checking for overdispersion.