Generalised Linear Models – 2nd Homework Assignement

Airborne Bacteria

1. Consider again the data on the bacteria counts in the air in and around Graz. The predictor should be based on humidity and temperature (both as linear and quadratic effects) as also on the factor site. Now try to model the counts directly by using a glm based on a normal distribution but a log-link function. How does the model fit compare with the one when using a standard linear model (identity link)?
2. Will a gamma model with log-link even give better results? Try to graphically compare the prediction regions under the normal and under the gamma model (both based on using the log-link) for site 6 and humidity 60% depending on temperature (shown as horizontal axis).

Titanic Data

1. One of the best known data sets with categorical content is based on notes about the 2201 passengers on board of the Titanic. This data set is available on our webpage and contains information about the class (`Class`) with levels `First`, `Second`, `Third`, and `Crew`, the passengers age (`Age`) with the two levels `Adult` and `Child`, the gender (`Gender`) with `Female` or `Male`, as also about the survival status (`Survived`) with level `Yes` or `No`. Find a suitable model for the survival status and interpret, how survival depends from other attributes.
2. Calculate all relevant *odds*, and the *odds ratios* that are relevant for you (probably for someone with `Age = Adult` and the respective gender level for `Gender`) as a passenger.
3. Determine the odds ratio of your favorite class that allows for a comparison of age (to answer the question: would it have been better, if you would have been a child instead of an adult?)
4. Which type of passengers has generally (under the model you've found) the best and which the worst chance to survive.