Project 3: Beetle Mortality Data – Generalized Linear Models

Due on October 26, 2015

The table below shows the numbers of beetles dead (y_i out of n_i) after five hours' exposure to gaseous carbon disulphide (CS_2) at various concentrations (x_i with unit $\log_{10} CS_2 mgl^{-1}$). The data was first reported by Bliss (Annals of Applied Biology, 1935), and later on used by Dobson & Barnett in their book *An Introduction to Generalized Linear Models*.

Concentration x_i	Number of beetles n_i	Number of killed y_i
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

1. Model the rate of beetle mortality with predictor x_i and fit your generalized linear model. Check the residual plots to see whether your model assumptions seem reasonable. Report your findings.

2. Try at least two other models/link functions. Use cross-validation to compare these models with your previous one in Part 1.

3. What is your recommended model(s) for this dataset? Does your conclusion reply on the particular partition of training vs. testing in Part 2?

Notes:

[1] Students are required to work in groups on course projects and submit their reports in pdf or doc format. The group size can be 1, 2 or 3.

[2] Each group is required to submit one hard copy of the report in the class of the due day. A list of names of the group members should be put on the cover page of the report.

[3] Each report should include a nontechnical part and a technical part. The nontechnical part should be fitted into one page. The technical part should include the statement of the problem, model assumptions, formulas used, results, conclusion, suggestion and discussion if necessary.

[4] Any statistical software and other tools can be used. If R program is used, the attachment of R code is recommended.