Ceriodaphnia dubia is a small (0.2 to 1.5 mm, juvenile to adult) freshwater zooplankton which feeds by filtering microscopic algi and organic material from the water. These zooplankton are common and important members of freshwater ecosystems throughout the world, and form the basis of food chains for fish and birds. Under optimal environmental conditions, the typical C. dubia hatches, matures and produces three or four broods of offspring during the first seven days of its 30-day life-span. Growth, reproduction and mortality of C. dubia are all affected by water pollution from sewage and toxic waste, and toxicity tests of C. dubia survival and reproduction are accepted standards for assessing the toxicity of effluents and chemicals.

A typical toxicity test is initiated with <12-hour-old juveniles and measures the survival and reproductive output of 10 individuals in each of four concentrations of toxicant, plus a control, for a seven-day period.

A reproductive toxicity test was conducted on nitrofen, a herbicide used for the control of broad-leaved and grass weeds in cereals and rice. Although relatively non-toxic to mammals, nitrofen is a significant teratogen and mutagen, and is acutely toxic to C. dubia. In fact, because of the risk of human exposure and its persistence in water, nitrofen has been withdrawn from commercial use in the US.

The test was conducted on 50 females, 10 at each nitrofen concentration 0, 80, 160, 235 and 310 µg/L. The total number of offspring born to each female (in three broods) was observed, and the data are set out in the file dubia.dat which is available from my web site.

Your task is to investigate the relationship between the toxicity, as measured by reproduction, and concentration of nitrofen.

(a) Plot the total number of offspring for each female by concentration of nitrofen. Describe the relationship.

What is the natural model for these data?

(b) Tabulate the means and variances for the total number of offspring by nitrofen concentration group, and comment.

(c) A generalized linear model has been proposed for these data, which is of the form

\[ E(Y_{ij}) = \mu_i = \exp \left( \sum_{m=0}^{K} \beta_mC_i^m \right), \]
where $Y_{ij}$ is the total number of offspring produced by individual $j$ at concentration $C_i$, the $\beta$'s are unknown coefficients, and $K$ is some positive integer.

Use Akaike's Information Criterion to select the best fitting model. Use residual plots to assess the goodness of fit of your fitted model, and add the fitted curve to your plot in (a).

(d) Summarize and interpret your results, justifying your choice of model.

2. Feigl and Zelen (Biometrics, 1965) analysed data on the survival time of patients with leukaemia. The data are survival time in weeks from diagnosis and the effects of two covariates: $x_1$, white blood cell count (WBC) at diagnosis, and $x_2$, an indicator for AG positive or AG negative, also at diagnosis. AG positive patients are identified by the presence of Auer rods and/or significant granularity of the leukaemic cells at diagnosis, but these factors are absent in the AG negative group.

You will fit regression models to the (possibly) transformed response, survival time, and compare models using cross-validation.

(a) Determine whether transformation(s) of the data would be appropriate.

(b) Two models which have been proposed for these data are firstly, a regression model with common slope for log(WBC), and secondly, a regression model which allows separate slopes and intercepts. Fit both these models to the data, and summarize and interpret your results.

(In your answer, ensure you include the model equations and a summary of the R commands used.)

(c) Comparing the crossvalidation estimates of prediction error, which model is the better predictor of survival time?

The data are in the file leuk.dat which is available from my web site.

This assignment is due by: 4pm Friday May 28 (Week 11).

Patty Solomon
10 May 2004