Applied Statistics With R
Exercises: Generalized Linear Models

John Fox

WU-Wien May/June 2006

Do the Poisson-regression problem and, if you wish, one of the two contingency-table problems.

1 Poisson Regression

1.1 Long's Data on the Research Productivity of Doctoral Students in Biochemistry

Long (1990) reports data on the productivity of doctoral students in biochemistry during the last three years of their PhD programmes. The data are in the file `Long.txt` on the course web site. The response variable is the number of articles published during this period (`art`). Explanatory variables include the gender of the student (`fem` = 1 for women, 0 for men); the student's marital status (`mar` = 1 if married, 0 otherwise); the student's number of children five years old or younger (`kid5`); a rating of the prestige of the student's PhD department (phd); and the number of articles published by the student's mentor during the three-year period (ment).

(a) Examine the distribution of the response variable. Does least-squares linear regression appear a promising strategy for these data?

(b) Perform a Poisson regression of number of articles published on the explanatory variables. What conclusions would you draw from this analysis?

(c) Consider the possibility of over-dispersion, either by fitting an over-dispersed Poisson model. Is there evidence for over-dispersion? How, if at all, do the results change when over-dispersion is taken into account?

2 Loglinear Models for Contingency Tables

2.1 Harris, Luginbuhl, and Fishbein's Data; the Berkeley Admissions Data

Analyze either (or both) of these contingency tables (given in the homework exercises on logit and probit models) employing loglinear models, but keeping in mind that response in the first case and admission in the second are the response variables. Show that likelihood-ratio tests for these loglinear models are the same as for the logit models previously fit to the data. Remember always to fit the highest-order term for relationships among the explanatory variables.