Generalized linear model

Multinomial – Ordinal data

1. In page 153 of Dobson and Barnett (2008), An introduction to generalized linear models, data about the importance of air-conditioning are presented. The dependent variable is preference of having airconditioning in a car having three categories. Sex and the age of the respondent are also recorded.

A. Consider that data comes from multinomial distributions the categories of which have no order. Fit a logistic model using the last category as reference category (very important), that includes the main effect of factor sex and the linear effect of age.

B. Interpret the coefficients of sex and age effects.

C. Check the statistical significance of sex and age and decide if it is necessary to remove one of them.

D. Now assume the data as ordinal. Fit a logit model for the cumulative probabilities of importance of having airconditioning in a car, assuming proportionality of odds and using the main effect of sex only. What is the interpretation of the parameter of sex's effect?

E. Add the linear effect of age in the model. What is the interpretation of the parameters of this effect? Does the value of the parameter that measures the effect of sex change at all, compared to the value you got from (D)?

F. Assuming a model for the non-parallelism, check the assumption of non-proportionality. What do you conclude?

2. Consider the data of exercise 8.3 (page 164, Dobson and Barnett (2008)). Here are presented in Table 8.6:

Table 8.6 *Tumor responses to two different treatments: numbers of patients in each category.*

Treatment	Sex	Progressive disease	No change	Partial remission	Complete remission
Sequential	Male	28	45	29	26
	Female	4	12	5	2
Alternating	Male	41	44	20	20
	Female	12	7	3	1

A. Fit a proportional odds model for examining the effect of the treatment and the effect of the sex on the response of disease's progress. What is the interpretation of model's parameters?
B. Use residuals for checking goodness-of-fit of the model.
C. Use a more complex model to compare it with the model from (A), as an alternative way to test goodness-of-fit.
D. Check the proportional odds assumption of the model at (A).

3. Consider the data that refers to the degree of pneumonia developed by miners as a function of their exposure over t years. The severity of the disease is measured using three levels. Data can be found in Chapter 5 of McCullagh and Nelder, 1989, Generalized Linear Models, Chapman and Hall.

A. Calculate the empirical logits and make a plot of them against time. Which model could you propose for relating logits of cumulative probabilities with the time of exposure in the mine?

B. A researcher proposed to transform the time of exposure with the logarithm. Check visually if this suggestion is plausible.

C. Fit a model that assumes the effect of the logarithm of exposure time in the mine. What is the interpretation of the coefficient of this effect?

4. The following data represents the relation between the duration of stay in hospital and the frequency of visits from relatives in 132 chronic schizophrenic patients. The duration of stay is categorized as 2-10 years, 10-20 years and over 20 years.

	Duration of stay		
Frequency of visits	(2-10)	(10-20)	20+
Normal	43	16	3
No normal	6	11	10
Never	9	18	16

Analyze the data with two ways.

The first way considers duration of stay as the dependent ordinal variable. Fit a proportional odds model for checking the effect of the frequency of visits. What are your conclusions?

The second way assumes the dependent variable as a multinomial variable. Fit a multinomial regression model, where frequency of visits is the independent variable. What are your conclusions?