STAT3015/4030/7030 Generalised Linear Modelling

Tutorial 10

1. **Emulating Jane Austen's Writing Style** (Ramsey and Schafer, 2013). When she died in 1817, the English novelist Jane Austen had not yet finished the novel *Sanditon*, but she did leave notes on how she intended to conclude the book. The novel was completed by a ghost writer, who attempted to emulate (that is, copy or imitate) Austen's style. In 1978, a researcher reported counts of some words found in chapters of books written by Austen and in chapters written by the ghost writer. The data is in the file `austen.txt` .

Word	**Book**			
	Sense and Sensibility	Emma	Sanditon I	Sanditon II
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4

   Was Jane Austen consistent in the three books in her relative uses of these words? Did the ghost writer do a good job in terms of matching the relative rates of occurrence of these six words? In particular, did the ghost writer match the relative rates that Austen used the words in the first part of *Sanditon*?

2. **El Nino and Hurricanes** (Ramsey and Schafer, 2013). The data set `elnino.txt` contains data on the number of tropical storms and hurricanes each year from 1950 to 1997. Data is also recorded on whether the year was a cold, warm or neutral El Nino year; a constructed numerical variable `temperature` that takes on the values -1, 0, and 1, according to whether the El Nino temperature is cold, neutral or warm; and a variable indicating whether West Africa was wet or dry that year. It is thought that wet years in West Africa often bring more hurricanes, and that the warm phase of El Nino suppresses hurricanes while a cold phase encourages them. Use a poisson log-linear regression to describe the distribution of (a) number of storms and (b) number of hurricanes as a function of `temperature` and West African wetness.

3. Schriener, Gregoire and Lawrie (1962) conducted an experiment to examine the effect of supposedly inert gases on fungal growth. The molecular weight (MW) of each inert gas and the fungal growth rate (in millimeters per hour) for 10 samples are given below:

Gas	He	Ne	N_2	N_2	Ar	Ar	Kr	Kr	Xe	Xe
MW	4.0	20.2	28.2	28.2	39.9	39.9	83.8	83.8	131.3	131.3
mm/hr	3.51	3.14	3.03	2.83	2.71	2.76	2.27	2.17	1.88	1.85

Schriener et al. report a relationship of: $mm/hr = 3.87 - 0.1774\sqrt{MW}$.

(a) Plot each of three transformations of the response variable $(y, \log y, 1/y)$ versus each of five transformations of the predictor variable $(x, \sqrt{x}, \log x, x^{1/3}, x^{2/3})$ and include the correlation cofficient and normal linear regression line on each of the fifteen plots. Which of the relationships seem the most linear? Is the Schriener et al. model one of these?

(b) Assess the suitability of the normal linear model which regresses the inverse of the rate $(1/y)$ on the untransformed molecular weight (x) using appropriate diagnostics.

(c) The structure of the model in part (b) brings to mind a GLM using which distribution (HINT: think about canonical links)? Why would this model make sense for this data? Fit this GLM to the data and assess its suitability using appropriate diagnostics. Do you think this model is preferable to the normal model of part (b)?

(d) The element Radon is an inert gas with a molecular weight of 222. Predict the fungal growth rate in the presence of Radon gas using the normal linear regression from part (b) as well as the GLM from part (c) and the Schriener et al. normal linear model. Also, find 95% confidence intervals for this fungal growth rate using each of the three models.

References

F. L. Ramsey and D. W. Schafer. The statistical sleuth: a course in methods of data analysis. Brooks/Cole, 2013.