STAT3015/4030/7030 Generalised Linear Modelling

Tutorial 8

1. For the purposes of setting hull construction standards, the rate of damage by waves to certain types of cargo ships needs to be determined. Ships of three different types were examined and the cumulative data is given below, as well as in the file `Wave.txt` on Wattle:

Ship Type	Year of Construction	Period of Operation	Aggregate Months Service	Number of Damage Incidents
A	1960-64	1960-74	127	0
A	1960-64	1975-79	63	0
A	1965-69	1960-74	1095	3
A	1965-69	1975-79	1095	4
A	1975-79	1975-79	2244	11
B	1960-64	1960-74	44882	39
B	1960-64	1975-79	17176	29
B	1965-69	1960-74	28609	58
B	1965-69	1975-79	20370	53
B	1975-79	1975-79	7117	18
C	1960-64	1960-74	1179	1
C	1960-64	1975-79	552	1
C	1965-69	1960-74	781	0
C	1965-69	1975-79	676	1
C	1975-79	1975-79	274	1

Clearly, a good approximation to the distribution of the number of damage incidents would be the Poisson distribution. Without any other information, it seems reasonable to start by employing the canonical link function, which is the logarithm in this case. Of course, the number of damage incidents is not necessarily the best reponse variable in this case, since clearly the total amount of time in service is important. So, modelling the rates of damage incidents appears to be a more pertinent approach. Our model for the expected response then becomes:

$$\log\left(\frac{\texttt{dmge}}{\texttt{mnths}}\right) = \beta_0 + \beta_1\texttt{typb} + \beta_2\texttt{typc} + \beta_3\texttt{cons65} + \beta_4\texttt{cons75} + \beta_5\texttt{opr75},$$

where typb and typc are indicators of the second two ship types, respectively, and cons65, cons75 and opr75 are indicators of the obvious categories for year of construction and period of service.

(a) Use R to fit this model, recalling that the rates actually have a Poisson$(\lambda T)/T$ distribution, which means that we must take account of the different number months of observation for each data point by employing the weights option.

(b) Examine the potential need for an interaction term between type of ship and period of operation in the model by plotting the logarithms of the observed averages within each ship type and period of service category combination (i.e., ignore interaction with year of construction by averaging the values for the three different levels of year of construction) against ship type and connecting the points associated with similar period of service category (note that this is the analog of the so-called "cell-means" plot for two-way ANOVA). What should this plot look like if there is no interaction between the two variables? Do you think an interaction term is necessary?

(c) Fit the model with an interaction term between ship type and period of operation and examine whether the effect appears statistically significant. Do the results bear out your visual assessment in part (b)?

(d) The initial model form can clearly be rewritten as:

$$\log(\texttt{dmge}) = \beta_0 + \beta_1\texttt{typb} + \beta_2\texttt{typc} + \beta_3\texttt{cons65} + \beta_4\texttt{cons75} + \beta_5\texttt{opr75} + \log(\texttt{mnths}).$$

So, we could fit a Poisson generalised linear model to dmge using the predictors typb, typc, cons65, cons75, opr75 and log(mnths). What would we expect the coefficient for the predictor log(mnths) to be? Fit the appropriate model and test whether this value is compatible with the actual observed data.

2. An experiment to determine the effects of temperature and storage time on the loss of ascorbicacid (vitamin C) in snap-beans was performed and the observed concentrations are shown below:

Temp ($^\circ F$)	Weeks of Storage			
	2	4	6	8
0	45	47	46	46
10	45	43	41	37
20	34	28	21	16

(a) Suppose that ascorbic acid concentration decays exponentially, and that the expected concentration after t weeks for the beans at temperature T is $\mu_T = \exp(\alpha + \beta_T t)$, where the initial concentration of acid, e^α, is assumed to be the same for each temperature group and the decay rate, β_T, is dependent on the temperature group. Fit a gamma generalised linear model with logarithmic link. Examine the dispersion parameter estimate. Does the data look consistent with the idea of an exponential distribution? [HINT: This model is a bit unusual in that it contains an interaction between time and temperature, but no main effect of temperature, and we can re-write the model as:

$$\log \mu = \alpha + \beta_0 t + \beta_{10} t z_1 + \beta_{20} t z_2,$$

where z_1 and z_2 are indicators for the temperature categories 10 $^\circ F$ and 20 $^\circ F$, repectively.]

(b) Estimate the time taken for the concentration to be halved at each temperature. [HINT: Recall that the initial concentration is assumed to be e^α, and we want to find the time when the predicted concentration is $0.5e^\alpha$.]

(c) Suppose that we do not assume that the initial concentrations were the same for each temperature. Create additional indicators and fit this model. Do you think the assumption of equal initial concentrations is reasonable?