MTH5131 Actuarial Statistics

Coursework 6

This coursework is not to be turned in. You may ask questions about the coursework in tutorial or by email.

Exercise 1. Explain the difference between the two types of covariate: a variable and a factor.

Exercise 2. Explain why the link function $g(\mu) = \ln \mu$ is appropriate for the Poisson distribution by considering the range of values that it results in μ_i taking.

Exercise 3. An insurer wishes to use a generalised linear model to analyse the claim numbers on its motor portfolio. It has collected the following data on claim numbers y_i, $i = 1, 2, \ldots, 35$ from three different classes of claim numbers

Class I	1	2	0	2	1	0	0	2	2	1										
Class II	1	0	1	1																
Class III	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0

For these data values,

$$\sum_{i=1}^{10} y_i = 11, \quad \sum_{i=11}^{15} y_i = 3, \quad \sum_{i=16}^{35} y_i = 4.$$

The company wishes to use a Poisson model to analyse these data.

1. Show that the Poisson distribution is a member of the exponential family of distributions.
2. The insurer decides to use a model (Model A) for which

$$\ln \mu_i = \begin{cases} \alpha & i = 1, 2, \ldots, 10; \\ \beta & i = 11, \ldots, 15; \\ \gamma & i = 16, \ldots, 35. \end{cases}$$

   where μ_i is the mean of the relevant Poisson distribution. Derive the likelihood function for this model, and hence find the maximum likelihood estimates for α, β, and γ.

3. The insurer now analyses the simpler model $\ln \mu_i = \alpha$, for all policies.
   (a) Calculate the maximum likelihood estimate for μ_i under this model (Model B).
   (b) Show that the scaled deviance for Model A is 24.93, and calculate the scaled deviance for Model B. It can be assumed that $f(y) = y \ln y$ is equal to zero when $y = 0$.
   (c) Compare Model A directly with Model B, by calculating an appropriate test statistic.

Exercise 4. A small insurer wishes to model its claim costs for motor insurance using a simple generalised linear model based on the three factors:

$$YO_i = \begin{cases} i = 1 & \text{for 'young' drivers;} \\ i = 0 & \text{for 'old' drivers;} \end{cases}$$

$$FS_j = \begin{cases} j = 1 & \text{for 'fast' cars;} \\ j = 0 & \text{for 'slow' cars;} \end{cases}$$

$$TC_k = \begin{cases} k = 1 & \text{for 'town' areas;} \\ j = 0 & \text{for 'country' areas;} \end{cases}$$

The data is the 8 aggregate claim costs for each combination of factors.
The insurer is considering three possible models for the linear predictor:

$$\text{Model1}: \quad YO + FS + TC$$

$$\text{Model2}: \quad YO + FS + YO.FS + TC$$

$$\text{Model1}: \quad YO * FS * TC$$

1. Write each of these models in parameterised form, stating how many non-zero parameter values are present in each model.

2. Explain why Model 1 might not be appropriate and why the insurer may wish to avoid using Model 3.

3. The student fitting the models has said 'We are assuming a normal error structure and we are using the canonical link function.' Explain what this means.

4. The table below shows the student's calculated values of the scaled deviance for these three models and the constant model. The degrees of freedom is defined to be the amount of data minus the number of parameters fitted.

Model	Scaled Deviance	Degrees of Freedom
1	50	7
$YO + FS + TC$	10	
$YO + FS + YO.FS + TC$	5	
$YO * FS * TC$	0	

Complete the table by filling in the missing entries in the degrees of freedom column and carry out the calculations necessary to determine which model would be the most appropriate.