STAT3015/4030/7030 Generalised Linear Modelling
Tutorial Week 12

1. A weekly lottery is conducted in which numbered and colored balls are chosen by a physical randomising device to determine the winners. The randomising machine contains 54 balls (consisting of six different colored sets of balls numbered 1 through 9) which are mixed by a draught of air in a closed, transparent container, and six balls are allowed to escape, one at a time. The randomisation of the machine is to be examined, and the number of times that each ball appears in the weekly winning six is tabulated over a one year period (52 weeks). The total number of times that each ball is in the winning six is tabulated in the data file `Lot.txt` on Wattle.

   (a) Clearly, if the machine is truly producing random balls, then the number and color of the winning balls should be independent of each other. Test this fact by directly calculating expected ball counts and constructing the Pearson chi-squared statistic.

   (b) Test the independence again, this time using a Poisson GLM approach. In other words, fit a Poisson model with link structure:

   $$\ln(\mathbb{E}\{Y_{ij}\}) = \beta_0 + \beta_1 r_2 + \cdots + \beta_5 r_6 + \beta_6 c_2 + \cdots + \beta_{13} c_9 + \beta_{14} r_2 c_2 + \cdots + \beta_{53} r_6 c_9,$$

   where r_i is the indicator for the i^{th} row and c_j is the indicator for the j^{th} column, and test whether $\beta_{14} = \cdots = \beta_{53} = 0$.

   (c) If the winning balls are truly random, there should also be no row or column effects (i.e., the marginal distributions should be uniform, so that the chance of being in any row or column is the same for all the rows and columns, meaning that $\beta_1 = \cdots = \beta_5 = 0$ and $\beta_6 = \cdots = \beta_{13} = 0$). Use your Poisson GLM from part b to test these hypotheses. Refit the model without the interaction terms and examine the estimated row and column effects. Do there appear to be any rows or columns which are out of line with the hypothesis of true randomness?

   (d) Suppose that we are now told that the orange balls were only added after 24 weeks of the lottery had been run. Thus, clearly there will be a row effect for this color ball. However, if the data is truly random, then the other rows should have no effect. Fit a Poisson GLM to test whether the other rows are uniform (i.e., test whether the model $\ln(\mathbb{E}\{Y_{ij}\}) = \beta_0 + \beta_5 r_6$ is adequate for this data). Moreover, a bit of algebra shows that if the orange balls were added after week twenty-four and the balls were truly random than it should be the case that $\beta_5 = -0.7073$. Test whether the observed data is consistent with this value.

2. The data file `HWCon.txt` is located on Wattle contains counts concerning the performance on homework assignments of schoolchildren learning from five different teaching styles and philosophies. The different teaching regimes were labeled A through E and the quality of the submitted homeworks was rated as either high, moderate or low.

   (a) Test whether the two categorical variables (i.e., teaching regime and homework quality) are independent using a Pearson chi-squared test and treating both variables as nominal.

   (b) Clearly, the quality of the submitted homework is an ordinal variable. Using this variable as the response, fit the proportional hazards model (i.e., the weighted complementary log-log binomial GLM to the appropriate transformed proportions) and test whether there is a significant column effect (i.e., that there is a difference in the quality of submitted homeworks for the different teaching regimes). Also, examine the suitability of the proportional hazards model for these data using the residual deviance.

   (c) Plot the fitted cumulative probabilities for each teaching regime. Which regimes are the best and which the worst (at least in terms of the assessed quality of the submitted homework)?