STAT3015/4030/7030 Generalised Linear Modelling

Tutorial 2

1. The file `productivity.csv` contains data regarding improvements in productivity for 27 business firms. Each firm was classifed according to whether their average expenditure for research and development in the past three years was high, moderate or low. In addition, the productivity improvement (measured on a scale of 0 to 100) was recorded for each firm.

   (a) Plot the productivity improvement scores versus the level of R&D expenditure. Do you think that heteroscedasticity will be a problem?

   (b) Fit a one-way analysis of variance model and test whether there is a difference between the expected productivity increases for each of the three factor levels. Construct a normal q-q plot to investigate the suitability of the normal error assumption.

   (c) Suppose we wish to test whether the difference in productivity improvement between firms with low and moderate levels of research and development funding is the same as the difference in productivity improvement for firms with moderate and high R&D funding (note that this is a sort of analog to linearity in the ANOVA setting). Write down the appropriate null hypothesis using parameters defined for your model. Test the null hypothesis at the $\alpha = 0.01$ level.

2. There are 8 distinct blood types among human beings, classified according to whether certain proteins are present or not. The blood types are: O-, O+, A-, A+, B-, B+, AB-, AB+. The appearance of an A in the name of the blood type indicates that the individual's blood contains the A-antigen, while the appearance of a B in the name indicates that the blood contains the B-antigen, and a + in the name indicates the existence of the so-called Rhesus (or Rh) factor in the blood. Thus, a person with AB- blood has both antigens in their blood stream, but no Rh factor, while a person with O+ blood has neither the A-antigen or the B-antigen, but does have the Rh factor. Suppose that a random sample of people is gathered and each person's blood type as well as the value for some quantitative biological trait are measured. The data are:

Blood Type	Responses			
O-	9	11		
O+	20	19	23	19
A-	12	10		
A+	17	18	21	20
B-	16			
B+	24	28	25	
AB-	15			
AB+	25			

(a) Fit a one-way ANOVA model to this data and test whether there is any difference in the response variable for individuals of different blood types.

(b) Construct a linear combination of the μ_i's to see whether the ability to produce the A- antigen has any effect on the response variable. Repeat this process for the B-antigen and the Rh factor.

(c) Suppose that we now believe that the ability to make the A-antigen is not important in explaining variation in the response. Create a new factor based on the original one which has only 4 levels and ignores the ability to make the A-antigen. Refit an ANOVA model using this new factor and test whether there is any overall significance in explaining the response variation. Also, construct constrasts for this new model to test whether the ability to make the B-antigen and the presence of the Rh factor have the same explanatory power.