

## Model Answers for Part B, 2010

26. Parts (a), (b) and (c) required defining the concepts as they were given on the lecture slides. Refer to the slides for the definitions.

(d) We use the model with the smallest CV so that the indicated model is the one containing all the variables except raindays. Note that the description of the variable raindays was omitted from the question. [ 2 marks for selecting the correct model, 2 marks for the reason]

27. Parts (a) and (b) are taken directly from the lecture slides.

(c) The test statistic is the difference in the deviances, so its value is  $30.4 - 21.7 = 8.7$ . The p-value for the test is calculated using the chi-square distribution with  $30 - 23 = 7$  degrees of freedom i.e. the p-value is  $1 - \text{pchisq}(8.7, 7)$ . [6 marks: 2 for the difference in deviances, 2 for the difference in df, 2 for the p-value.]

(d) The first model is a regression model with all two and three factor interactions. It has a deviance of 64.96 in 76 df. The model without the variable Revival has a deviance of 77.86 on 84 df. The test that the second model is adequate (i.e. that revival has no effect) has test statistic  $77.86 - 64.96 = 12.90$  on  $84 - 76 = 8$  df, with a corresponding p-value of  $1 - \text{pchisq}(12.90, 8) = 1 - 0.884663 = 0.115337$ . Thus there is not much evidence that revivals have an effect on the number of Tony awards won. [4 marks: one each for the test statistic, the degrees of freedom, the p-value and the conclusion]

28. Parts (a) and (b) are taken directly from the lecture slides.

(c) There doesn't seem to be much evidence of a 3-factor interaction so the homogeneous association model would appear to fit well. This is also the best model of those considered in terms of AIC. In this model, the association between pairs of factors conditional on the third is the same for all levels of the conditioning factor. The association graph is like a triangle, with edges AB, AC and BC. [4 marks: 2 for picking the homogeneous association model, 1 for the interpretation in terms of equal odds ratios, 1 for the graph]

(d) the point estimate for the log-odd ratio between gender and being still born for age group 37-41 is  $\exp(0.09929 + 0.17077) = 1.310$ .

The other groups are

33-36:  $\exp(0.09929 - 0.09682) = 1.002$ .

29-32:  $\exp(0.09929 - 0.03526) = 1.066$ .

25-28:  $\exp(0.09929 - 0.05433) = 1.046$ .

$\leq 24$ :  $\exp(0.09929) = 1.104$ .

These are quite similar (which is why the homogeneous association model is reasonable here.) The odds ratios are all  $> 1$  which suggests that the odds of a stillbirth for a female child are a bit greater than those for a male child.

[4 marks, 2 for calculating the OR's, 2 for the interpretation]

(e) A better approach would be to base the estimate of the (common) odds ratio by exponentiating the interaction between gender and still born obtained from the regression summary of the homogeneous association model.

R output:

```
summary(glm(counts~agegroup*stillbirth*gender -  
agegroup:stillbirth:gender, family=poisson))  
OR = exp(*****)
```

where \*\*\*\*\* is the stillbirth\*gender interaction.

*[4 marks: 2 for suggesting using the homogeneous association model, 2 for the R code]*