Department of Statistics

COURSE STATS 330

Assignment 5, 2008

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on **Monday October 20.**

In 1912, the White Star liner Titanic sank in the North Atlantic after striking an iceberg. The data set titanic.txt on the web site contains details of 633 of the passengers on board.

This data set includes the following variables:

`pclass`	A factor giving the class of the passenger: one of 1st, 2nd, 3rd.
`age`	The age of the passenger in years.
`sex`	Passenger's gender: female or male
`age.group`	Passengers age group, one of 0-9 , 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79
`survived`	Passenger's survival (1=survived, 0=did not survive)

Your task in the first part of this assignment is to assess the effect of `class, age and sex` on the survival. Specifically,

1. Ignoring the variable age, group the data into groups corresponding to each age-group/sex/class combination. Compute the logits for each combination as we did for the plum tree data. Make a data frame containing the logits, and the categorical variables. You should have one line in the data frame for each combination of the factor levels. Show the R code used.[8 marks]

   Hints: the function **`tapply`** will be useful to calculate the r and n values, and hence the logits. This will result in a 3-dimensional array which you can turn into a vector with the function **`as.vector`**. The function **`expand.grid`** is useful for generating all possible combinations of factor levels. See Tutorial 9 for an example.

2. Draw suitable plots and use them to assess graphically how the factors `class, age and sex` are related to the logits. Report your conclusions.[8 marks]

3. Fit a suitable model to the data, using the ages rather than the age groups (i.e. using the original data frame). Are there significant interactions between the explanatory variables? [8 marks]

4. Use a confidence interval and the model fitted in 3. to estimate the probability that a female 1st class passenger aged 50 will survive. [8 marks]

5. What is the relationship between the three factors age-group, sex and class? Are any independent of the others? Any pair conditionally independent given the third? Fit a contingency table model for this part. (see Lecture 28) [8 marks]

Question for STATS 763 only.

For a general I x J contingency table, show mathematically that when factors A and B are independent, then the interaction in the corresponding Poisson regression model is zero.