STA 312f10 Assignment 6

Do this assignment in preparation for the quiz on Friday, Oct. 22nd. Please bring your R printout to the quiz; part or all of it may be handed in. Please do *not* write anything on your printout before the quiz, except possibly your name and student number.[1]

This assignment uses the Titanic data, which are included with R as 4-dimensional array called "Titanic." Type `help(Titanic)` at the R prompt for more information. As agreeed in class, we will *limit the analysis to passengers for now.*

Clearly, there are three explanatory variables and one response variable. Please adopt the *conditional* approach described in Lecture (Log-linear Part 5) and in Chapter 6 of the text. This will simplify the task of model building quite a bit.

1. Explore the data and find an acceptable conditional model. "Acceptable" means that the p-value for G^2 is above 0.05, and the model is as simple as possible. Simple means having fewer terms, subject of course to the constraint that it's a conditional model. Your printout for this part of the assignment should consist of 2 parts.

   - The R code documenting your exploration. Please just give the code, and *not the output* for this part. You do not need to be completely systematic in your exploration; I was not, and I like my model. The only way you can lose marks for the exploration is if you have a model that seems to come from nowhere. In this case I urge Christine to give you a zero on the quiz, and to look hard for evidence of copying someone else's work.
   - Type the name of your model to display it, and also calculate G^2, the degrees of freedom and the p-value.

2. Now, using your chosen model as the alternative hypothesis, test the null hypothesis that the highest order interaction among the explanatory variables equals zero. Give change in G^2, the degrees of freedom and the p-value. Hint: $df = 2$. Would you end up with different models using the conditional approach and the unconditional approach?

3. For every 3-factor interaction involving survival in your model, carry out a test of independence in each sub-table; calculate percentages, G^2, df and the p-value. Be able to describe the results in plain, non-statistical language, but *don't write any of this on your printout in advance!*

4. Finally, look at the 2-dimensional marginal table of Class by Survival. Calculate Pearson's X^2, the p-value and the appropriate percentages. Now follow up this highly significant result by testing all pairwise differences between survival percentages. Use the Pearson X^2 test on each 2×2 table (that was a hint). Be able to state your conclusions in plain, non-statistical language, but *don't write them on your printout in advance.*

   I used the `summary` function for this question, but the Titanic data are in an array, not technically a table. So I used code like `summary(as.table(classwar))`.

[1] Here are the usual suggestions about the computer work. It would be smart to compose your commands in a text file, and drag them to R a bit at a time, debugging as you go. If I were you I would put the question numbers (but *not* the answers to the questions, please!) in comment statements. Save the text file. This way if you discover a mistake or omission, it will be easy to fix.