STA 312f10 Assignment 5

Do this assignment in preparation for the quiz on Friday, Oct. 15th. 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.

Here are a few 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.

This assignment uses the Kidney Stone data from the first set of lecture slides — the Introduction. For all tests, be able to give the value of chisquare, the degrees of freedom and the p-value.

1. Make a nicely labelled 3-dimensional table in which rows are treatments, columns are effective vs. not, and layers are size of stone. Adding over the third dimension and transposing the result (use the `t()` function), you will get a marginal table you can check against one of the slides.

2. Try the model of complete independence. Does it fit? Base your conclusion on G^2.

3. In the 2-dimensional marginal table of Treatment by Effectiveness, which treatment *seems* to be more effective? Make a table that shows the percent of cases where A was effective, and so on.

4. Do a common Pearson chisquare test of independence on your 2-dimensional marginal table. State your conclusions (if any) in words. Use simple, non-technical language. Avoid use of the word "independent."

5. Investigate the association between Treatment and Effectiveness *controlling* for Size of the patient's kidney stones. Follow these steps.

   (a) Display the two sub-tables. For each one, use the `prop.table` function to calculate the percent of patients for whom Treatment A was effective, and the percent of patients for whom Treatment B was effective. Carry out a likelihood ratio chisquare test of independence for each sub-table. What, if anything, do you conclude.

   (b) State the log-linear model for this question, using bracket notation.

   (c) In μ notation, what terms are present in the saturated model but absent from this one?

(d) Fit the model and carry out the likelihood ratio test for model fit. Here is one way to think about it. The null hypothesis that this model holds is the same as a null hypothesis in which one of the 2-variable associations is missing, but because the models are hierarchical, the three-factor effect must be missing as well. Those are the two terms in Question 5c. So it's a way of testing that association, allowing for all other possible effects in the model. What do you conclude? Is there good evidence of a relationship between Treatment and Effectiveness controlling for Stone size?

(e) Check your work by adding the two G^2 statistics from Question 5a. They should add up the the G^2 statistic from Question 5d.

(f) Does this model of conditional independence fit adequately? Answer Yes or No.

6. Now try the model with all possible two-variable associations. Is it an improvement on the model of Question 5? Base your conclusion on change in G^2.

7. At this point I like the model of Question 5. It has two associations, and we need to see what they mean. So please look at a couple of two-dimensional marginal tables in the usual way, calculating Pearson chisquare tests of independence as well as percentages. Describe each conclusion in words.

8. Now try dropping each term, one at a time. Does the model get significantly worse each time? Base your conclusions on change in G^2.

9. Finally we have a model, and it looks like we know what the associations mean. But rather than completely trusting 2-dimensional marginal tables, please display and examine the estimated μ values for all 2-factor associations in the model. Are they positive and negative in the right places?

10. Suppose that Treatment A is more expensive than Treatment B. Do you have any suggestions that would make the government (say, OHIP) very happy with you and want to hire you again next time as their statistical consultant? Do you have any reservations about this advice?