STA 312f12 Assignment Five[1]

Please bring your R printouts from the last question to the quiz. The non-computer questions are practice for the quiz on Friday Oct. 19th, and are not to be handed in.

On the quiz, you will be given a copy of the formula sheet, so please use the formula sheet while doing your homework, and let me know if there are problems. There is a link to the formula sheet from the course home page, in case the one in this document does not work.

1. In a study of attempts to increase the success rate for entering first year students, 100 entering students are randomly assigned to be paired up with an advanced student mentor who is supposed to give advice and show a good example. Another 100 students are randomly assigned to a control condition in which they are shown a movie about how to succeed in university. The outcomes of interest are (1) In good academic standing at the end of first year, (2) On academic probation at the end of first year, and (3) Gone.

   (a) What is the explanatory variable? What is the response variable?

   (b) Is the study cross-sectional, prospective or retrospective?

   (c) How do you know it's an experimental study rather than a purely observational study?

   (d) Write a product-multinomial likelihood for this problem. Are the parameters probabilities, or are they conditional probabilities? How many free parameters are there in the unrestricted model?

   (e) The null hypothesis is that the explanatory variable and the response variables are unrelated (which is quite possible). Write the null hypothesis in symbols.

   (f) Write another product-multinomial model that applies under the null hypothesis. How many free parameters are there in the restricted model?

   (g) So what are the degrees of freedom of the test?

   (h) Letting n_{ij} denote the observed cell frequencies, write down the unrestricted maximum likelihood estimators of your model parameters. You don't have to show any work.

   (i) Derive the restricted maximum likelihood estimators. Show your work, or if you realize you have done this problem several times already, then just be ready to show your work on the quiz if asked.

   (j) Make a table with six cells. In each cell, write the estimated expected frequencies under the null hypothesis. Do they look familiar? Comment.

[1]Copyright information is at the end of the last page.

2. Look at Table 2.3 on p. 27 of the text. Why does your work for Question 1 make you comfortable doing a fast likelihood ratio chi-squared test? Do it with a calculator.

   (a) What is the critical value of the test statistic at $\alpha = 0.05$?

   (b) What is your computed value of G^2? The answer is a number. Show your work.

   (c) Do you reject H_0? Answer Yes or No.

   (d) In plain, non-statistical language, what do you conclude?

   Notice how different this analysis is from the one in the book. Are the conclusions also different?

3. Look at Problem 2.24 in the text. Instead of exactly doing the problem,

   (a) Show the equivalence of the formula to (2.8).

   (b) The idea behind the question in the text is correct, but the statement is not quite right. Show you understand by correcting the question. Hint: There is an exception to what the author says.

4. Do Problems 2.25, 2.37 and 2.39.

5. This question is about Fisher's exact test, In the 2×2 table below, the integers a, b and n are fixed.

	Response=1	Response=2	
Explanatory=1	x	$a - x$	a
Explanatory=2	$b - x$	$n - a - b + x$	$1 - a$
	b	$1 - b$	n

   (a) How many ways are there for a of the cases to have the explanatory variable equal to one and b of the cases to have the response variable equal to one? Express your answer in terms of binomial coefficients.

   (b) How many ways are there to observe the four cell frequencies shown in the table? Express your answer as a multinomial coefficient.

   (c) Note that if x is specified, the other 3 cell frequencies are determined. If all the ways of sorting the observations subject to the constraints are equally likely (that's what H_0 says), what is the probability of $n_{11} = x$. Express your answer in terms of binomial coefficients.

   (d) Show that the cross-product ratio θ is an increasing function of x. This means that tail probabilities for any observed cross-product ratio can be obtained by summing over x in your last answer.

   (e) Noting that each of the four cells in the table must be non-negative, what is the range of possible values for the random variable n_{11}? Show some work.

6. Do 2.30 and 2.31(a) using R. Bring your printout to the quiz.

7. In your own words, explain the difference between conditional independence and marginal independence. Marginal independence is defined in the text, not the lecture notes. But the term is self-explanatory.

8. Do Problem 2.35

9. Take a look at the Death Penalty Data. Try the "Data Sets" link on the course home page if the one in this document does not work. The (non-random) sample here consists of prisoners in U.S. Jails (in Florida, I believe) who were convicted of first degree murder. The prisoners are either Black or White, the victims were either Black or White, and the prisoner either got the death penalty or not. If you are interested in more detail, the reference is *American Sociological Review* 1981, **46**, 918-927. Of course you do not need to go to the library to do this assignment.

   To read the data, I suggest `read.table` with the `header=T` option. Once you start digging into the data, you will see a zero cell frequency that causes trouble for the chi-squared tests because of low expected frequencies. So let's stick with 2-sided Fisher exact tests.

   (a) What is the response variable? What are the explanatory variable?

   (b) Is the study cross-sectional, prospective or retrospective?

   (c) Is the study experimental, or purely observational? Why?

   (d) Look at all the 2-way marginal tables, and do tests. In each case, what do you conclude? Try `help(prop.table)` for a very handy way (better than in lecture) of calculating things like the proportion of Black prisoners compared to whites who got the death penalty.

   (e) Look at the relationship of Prisoner's race to Death penalty *controlling for* Victim's race. Check the null hypothesis of conditional independence; you don't need to do a pooled test. In plain language, what do you conclude?

   (f) Look at the relationship of Victim's race to Death penalty controlling for Prisoner's race. Check the null hypothesis of conditional independence; you don't need to do a pooled test. In plain language, what do you conclude?

   Naturally, the `table` function will generate 3-way tables if you ask nicely. See `help(table)`, or just experiment with it. Bring your complete R printout to the quiz.
