STA 312f12 Assignment Three[1]

Please bring all R printouts to the quiz.The non-computer questions are practice for the quiz on Friday Oct. 5th, 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. Here is a repeat of the question from the last assignment. This time, you are asked to do it with technology from the multinomial.

   For years, brand awareness for Big Red chewing gum has been stuck at about 6%, meaning that about 6% of consumers who chew gum say they remember hearing about Big Red gum. The company started an advertising campaign to increase brand awareness, and after it had been running for a few weeks, they interviewed a random sample of 200 gum chewers, and found that twenty had heard of Big Red.

   (a) State the model and the null hypothesis, or look at your answer from last week. Is your model a special case of the multinomial?

   (b) Give the expected frequencies under H_0. Your answer is a set of numbers. How many numbers?

   (c) What is $\widehat{\mu}_1$ What is $\widehat{\mu}_2$

   (d) Carry out the Pearson X^2 and large-sample likelihood test G^2. Use the formula sheet and a calculator. In each case, do you reject H_0 at $\alpha = 0.05$?

   (e) State your conclusion from each test in plain, non-statistical language.

2. Recall the beer study, in which 120 beer drinkers each tasted 6 beers and indicated which one they liked best. The numbers preferring each beer are as follows.

	Preferred Beer					
	1	2	3	4	5	6
Frequency	30	24	22	28	9	7

   Last week, you found that preference for the 6 beers was not the same. But it seems that the first 4 beers are lagers and the last two are ales. No one would expect preference for lagers and ales to be the same.[2] So let's test whether preference for the 4 lagers is different, and at the same time, whether preference for the 2 ales is different.

   (a) State the null hypothesis in symbols. It is a statement about the p_js.

   (b) What are the degrees of freedom of the test? The answer is a number.

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

[2]Actually, I am making all this up with only a vague idea of what these terms mean.

(c) Differentiating the log likelihood function, obtain the maximum likelihood estimator of the parameter $\boldsymbol{\pi}$, under the null hypothesis. Show all your work. The answer is a symbolic expression, a vector of length 6. If it is not absolutely the most natural thing you can imagine, then either it's wrong or you have not simplified enough.

(d) Give the maximum likelihood estimate of $\boldsymbol{\pi}$ under the null hypothesis for this particular data set. The answer is a set of 6 numbers. Note the difference between an estimator and an estimate.

(e) What are the expected frequencies under H_0? Your answer is a set of 6 numbers. Are these estimated expected values, or exact expected values?

(f) Using the formula sheet, calculate the likelihood ratio test statistic G^2. Show your work. Your answer is a number. This is something you should be able to do with a calculator if necessary on the quiz. You would be silly not to check it with R.

(g) Calculate the Pearson X^2 statistic for these data. The answer is a number.

(h) What is the critical value for this test at $\alpha = 0.05$? The answer is a number.

(i) Do you reject the null hypothesis at $\alpha = 0.05$ based on the likelihood ratio test? Answer Yes or No.

(j) Do you reject the null hypothesis at $\alpha = 0.05$ based on the Pearson Chisquare test? Answer Yes or No.

(k) Based on this analysis, are you able to conclude that there is any difference in preference among the 4 lagers or between the 2 ales? Answer Yes or No. (In applied situations, beware of concluding that there is no effect, or absolutely no difference at all in the population.)

Just so you can check your answer to this question, my p-value for the likelihood ratio test is 0.7735209.

3. A sample of 150 students each try to solve two difficult logic problems. The problems are complicated, but it's multiple choice, so the answers are either right or wrong. For each student, the data file indicates whether he or she got Question 1 right, and whether he or she got Question 2 right. It is expected that students who get one question correct will also tend to get the other correct; this is not the issue. The issue is whether the one question is more difficult than the other.

   At first glance, this seems like just a problem of comparing two proportions, but there is a twist. Each student answers *both* questions. In a multinomial model, there are N *independent* observations, so each case (person or whatever) must contribute only one frequency to the table.

We will set up the problem as a multinomial with four categories, as follows.

	Question 1 Correct	Question 1 Incorrect
Question 2 Correct	p_1	p_2
Question 2 Incorrect	p_3	p_4

The observed frequencies are:

	Question 1 Correct	Question 1 Incorrect
Question 2 Correct	66	41
Question 2 Incorrect	30	13

(a) State the null hypothesis in symbols. It is a statement about the π_js. Simplify.

(b) What are the degrees of freedom for this test?

(c) Differentiating the log likelihood function, obtain the maximum likelihood estimator of the parameter $\boldsymbol{\pi}$, under the null hypothesis. Show all your work. The answer is a symbolic expression, a vector of length 4.

(d) Give the maximum likelihood estimate of $\boldsymbol{\pi}$ under the null hypothesis for this particular data set. The answer is a set of 4 numbers.

(e) What are the expected frequencies under H_0? Your answer is a set of 4 numbers. Are these estimated expected values, or exact expected values?

(f) Using the formula sheet, calculate the likelihood ratio test statistic G^2. Show your work. Your answer is a number. This is something you should be able to do with a calculator if necessary on the quiz.

(g) Calculate the Pearson X^2 statistic for these data. The answer is a number. This is something you should be able to do with a calculator if necessary on the quiz.

(h) What is the critical value for this test at $\alpha = 0.05$? The answer is a number.

(i) Do you reject the null hypothesis at $\alpha = 0.05$ based on the likelihood ratio test? Answer Yes or No.

(j) Do you reject the null hypothesis at $\alpha = 0.05$ based on the Pearson Chisquare test? Answer Yes or No.

(k) What percent of students answered Question 1 correctly? The answer is a number.

(l) What percent of students answered Question 2 correctly? The answer is a number.

(m) Does this study provide solid evidence that the two questions differ in their difficulty? Answer Yes or No.

4. This is a follow-up on the Quebec polling question from last week. A polling firm plans to ask a random sample of registered voters in Quebec whether Quebec should separate from Canada and become an independent nation: Yes or No. Suppose we intend to

test whether the true percent favouring independence is different from 50%, and in fact the true percent is 53%. What is the minimum sample size required to reject the null hypothesis at $\alpha = 0.05$ with probability at least 0.80, using the 2-sided Z_2 test from your formula sheet? Do the calculation with R. Please use the non-central chi-squared distribution this time. *You may be asked to hand in your printout, so please print this R session on a separate sheet of paper.* It is perfectly okay to use my R code from lecture.

5. This is about the Big Red brand awareness study if Question 1, *before* the data were collected.

   The advertising agency has a problem. With the budget they have been given to purchase media (air time and so on), they are confident they can move brand awareness a little – perhaps to 8%. In the old days, they could tell the client they had increased awareness by 33% and start to celebrate, but now the client has fallen under the influence of a U of T graduate who insists that a null hypothesis be rejected at the $\alpha = 0.05$ level with a non-directional test before they admit that anything actually worked. So the advertising agency has to decide how many people they need to survey when they measure brand awareness, in order to have a good chance of rejecting the null hypothesis. It's important, because if the client thinks the advertising didn't work, they might get a new advertising agency.

   (a) Suppose they want to be 90% sure of rejecting H_0 if they increase brand awareness to 8%. What sample size do they need if they use a Pearson X^2 test? Please obtain the answer using R and bring your printout to the quiz.

   (b) What sample size do they need if they use a large-sample likelihood ratio G^2 test? Please obtain the answer using R and bring your printout to the quiz.

   You may be asked to hand in your printout, so please print this R session on a separate sheet of paper.
