STA 312f12 Assignment Two[1]

For Question 3d and Question 11, you are asked to do some calculations with R. Please bring the printouts to the quiz. The rest of the questions are practice for the quiz on Friday Sept. 28th, and are not to be handed in.

1. Customers arrive at a Tim Hortons according to a Poisson process with rate $\lambda = 30$ per hour. What is the probability that exactly 8 customers arrive during a 10-minute period?

2. 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 marketing department is planning an advertising campaign to increase brand awareness, in the hope that increased brand awareness will lead to increased sales. Once the campaign was running a few weeks, they interviewed a random sample of 200 gum chewers, and found that twenty had heard of Big Red.

   (a) State a reasonable model for these data.

   (b) What is the parameter space $\mathcal{B}$?

   (c) Without any derivation, estimate the brand awareness for Big Red, in percent. Your answer is a number between zero and one hundred.

   (d) Give an approximate 95% confidence interval for the brand awareness in percent. Your answer is a set of two numbers.

   (e) What is the null hypothesis corresponding to the *main question*, in symbols?

   (f) What is the critical value (or values) of the test statistic at $\alpha = 0.05$ for a 2-sided test? The answer is a number or a pair of numbers.

   (g) Calculate the Wald test statistic Z_2. What is the value of the test statistic? Your answer is a number. Show some work.

      i. Do you reject H_0 at $\alpha = 0.05$? Answer Yes or No.

      ii. Using R, calculate the p-value. Make sure it's on the printout you bring to the quiz.

      iii. Do the data provide convincing evidence against the null hypothesis?

      iv. In plain, non-statistical language, what do you conclude? Your answer is a statement about brand awareness.

   (h) Some clever person suggests that for this problem, the test based on Z_1 will always be bigger than Z_2 as long as $p > \pi_0$. Try it and see. What is the value of Z_1? Your answer is a number. Show some work.

      i. Do you reject H_0 at $\alpha = 0.05$? Answer Yes or No.

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

ii. Using R, calculate the p-value. Make sure it's on the printout you bring to the quiz.

iii. Do the data provide convincing evidence against the null hypothesis?

iv. In plain, non-statistical language, what do you conclude? Your answer is a statement about brand awareness.

(i) But is Z_1 *always* bigger than Z_2 when $p > \pi_0$, as claimed? To answer this question, it is easier to base the test on Z^2 rather than Z. The null hypothesis is rejected when $|Z| > 1.96$, which occurs if and only if $Z^2 > 1.96^2 = \chi^2_0.05(1)$ – so it's the same test. Now find the value of p for which the denominator of Z_2^2 is greatest. Make a rough sketch of the function. Now you can answer the question: Is Z_1 *always* bigger than Z_2 when $p > \pi_0$?

3. 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. They would like to be able to say that their results are expected to be accurate within three percentage points, nineteen times out of twenty.

   (a) Suppose the population percent favouring independence is 25%. What sample size is required to achieve the desired margin of error?

   (b) Suppose the population percent favouring independence is 40%. What sample size is required to achieve the desired margin of error?

   (c) What sample size would be required if you were unwilling to make any assumptions about the true percentage favouring independence?

   (d) 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 a 2-sided Z_2 test? Do the calculation with R. *You may be asked to hand in your printout, so please print this R session on a separate sheet of paper.*

4. Ten friends have a party right after graduating from university. At the time, none of them has ever been married. The party includes a visit by a fortune teller, who says "Five years from now, 3 of you will still be unmarried, 3 of you will be married for the first time, 2 will be divorced, one will be married for the second time, and one will be widowed."

   How many ways are there for this to happen? The answer is a number. Show your work.

5. Students entering U of T have to choose a division: Humanities, Social Sciences, or Sciences.

(a) Of the 25 students from a particular high school, how many ways are there for 8 to choose the Humanities, 14 to choose the Social Sciences and 3 to choose the Sciences? The answer is a number. Show your work.

(b) Of the 3 students from another high school, how many ways are there for 1 to choose the Humanities, 1 to choose the Social Sciences and 1 to choose the Sciences? The answer is a number. Show your work.

6. Please do problem 1.6 from the text.

7. A fair die is tossed 8 times. What is the probability of observing the numbers 3 and 4 twice each, and the others once each? The answer is a number.

8. A box contains 5 red, 3 white and two blue marbles. A sample of six marbles is drawn with replacement. Find the probability that

   (a) 3 are red, 2 are white and one is blue

   (b) 2 are red, 3 are white and 1 is blue

   (c) 2 of each colour appears.

   All the answers are numbers.

9. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from a $M\left(1, (\pi_1, \ldots, \pi_c)\right)$ distribution. Show why the likelihood function is written $\ell(\boldsymbol{\pi}) = \pi_1^{n_1}\pi_2^{n_2}\cdots\pi_c^{n_c}$.

10. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from a $M\left(1, (\pi_1, \pi_2, \pi_3)\right)$ distribution. Find the maximum likelihood estimator of (π_1, π_2, π_3). Show *all* your work.

11. Under carefully controlled conditions, 120 beer drinkers each tasted 6 beers and indicated which one they liked best. Here are the numbers preferring each beer.

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

The main question is whether preference for the 6 beers is different in the population from which this sample was taken. Use R whenever possible. *You may be asked to hand in your printout of the R parts, so please print this R session on a separate sheet of paper.*

(a) State a reasonable model for these data.

(b) What is the parameter space $\mathcal{B}$?

(c) State the null hypothesis in symbols. It is a statement about the π_js. Please be specific. The research question allows you to give a specific numerical value for each π_j under H_0.

(d) What is the restricted parameter space $\mathcal{B}_0$?

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

(f) 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?

(g) Calculate the likelihood ratio test statistic G^2. Show some work. This is something you should be able to do with a calculator if necessary on the quiz. Your answer is a number.

(h) Now calculate G^2 again using R.

(i) Obtain the critical value at $\alpha = 0.05$? with R.

(j) Calculate the p-value using R. Print out all the R output and bring it to the quiz.

(k) Do you reject the null hypothesis at $\alpha = 0.05$? Answer Yes or No.

(l) It is tempting to ask you to state your conclusion in words. But all you can conclude without further testing is that preference for all the beers is not equal. It *looks* like preference for beers 1 through 4 is greater than preference for 5 and 6, and this is what you would tell your management or client in a job situation.

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

(n) Now calculate X^2 again using R.

(o) Do you reject the null hypothesis at $\alpha = 0.05$? Answer Yes or No. Just so you can check your answer to this question, my p-value for X^2 is 0.0002479085.
