STA 312f22 Assignment Five[1]

Please bring hard copy of your complete R input and output from Question 8 to the quiz. The non-computer questions are practice for the quiz on Friday Oct. 28th, and are not to be handed in.

---

1. This question and the next one could have been on Assignment 4. That is, they are based on Contingency Tabes Part One, not Two. Let $X_1, \ldots, X_{n_1} \stackrel{iid}{\sim}$ Bernoulli(π_1), and independently, $Y_1, \ldots, Y_{n_2} \stackrel{iid}{\sim}$ Bernoulli(π_2). That is, we have independent random samples from two Bernoulli distributions, and we want to test whether the two probabilities π_1 and π_2 are different.

   (a) Derive the likeihood ratio statistic for testing $H_0 : \pi_1 = \pi_2$. Your final answer is a formula. You may use the fact that the unrestricted MLE of π_1 is p_1 and the unrestricted MLE of π_2 is p_2.

   (b) Of 140 study participants receiving a placebo, 22.1% got sick, while only 12.2% of the 139 participants who took Vitamin C became ill. Calculate your test statistic and the p-value, and state your conclusion. The percentages are not quite exact (they are rounded), so it may help to start by recovering the observed frequencies.

   (c) How could you have done this problem with a lot less work? Hint: Your G^2 statistic (the actual number) is in the lecture slides.

2. Please use the notation shown in the following table. Assume it's a cross-sectional study.

x	$a-x$	a
$b-x$	$1-a-b+x$	$1-a$
b	$1-b$	1

   (a) How many free (unknown) parameters are there under the null hypothesis of independence? In the notation of the table, what are they?

   (b) Write the likelihood function, restricted by the null hypothesis of independence. Use the symbols $n_{11}, n_{12}, n_{21}, n_{22}$ for the cell frequencies.

   (c) Find the MLE of the free parameters of the restricted model. Show your work.

   (d) Write down the entire restricted MLE (four quantities) in a 2×2 table.

   (e) Based on your work, give a formula for $\widehat{\mu}_{ij}$. Simplify.

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

3. Denote the probability of an event by π, and the odds of the event by d.

   (a) Give an expression for π as a function of d. Show your work.

   (b) Show that d is a strictly increasing function of π. So, the greater the probability, the greater the odds. For this item, you could differentiate either the odds or the log odds.

4. Using the notation in this table and assuming a cross-sectional study,

	$Y = 1$	$Y = 2$	Total
$X = 1$	π_{11}	π_{12}	$\pi_{11} + \pi_{12}$
$X = 2$	π_{21}	π_{22}	$\pi_{21} + \pi_{22}$
Total	$\pi_{11} + \pi_{21}$	$\pi_{12} + \pi_{22}$	

   (a) What are the odds of $Y = 1$ given $X = 1$? Just write the answer down.

   (b) What are the odds of $Y = 1$ given $X = 2$? Just write the answer down.

   (c) What are the odds of $Y = 1$ given $X = 1$ *divided by* the odds of $Y = 1$ given $X = 2$? Simplify.

   (d) Show that $P(Y = 1|X = 1) = P(Y = 1|X = 2)$ if and only if $\theta = 1$.

   (e) Show that $P(X = 1|Y = 1) = P(X = 1|Y = 2)$ if and only if $\theta = 1$.

   (f) The two-by-two table is a matrix. Show that the odds ratio equals one if and only if the determinant of the matrix is zero.

5. Make a table like the table in Question 4, except for a prospective design. Show that the odds ratio equals the cross-product ratio.

6. Make a table like the table in Question 4, except for a *retrospective* design. This time, the odds ratio of interest is the odds of $X = 1$ given $Y = 1$ divided by the odds of $X = 1$ given $Y = 2$. Show that the odds ratio equals the cross-product ratio.

7. 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.

8. Data from the 1912 sinking of the Titanic are available in a built-in R dataset. See `help(Titanic)` for details. For this question, you might want to look at the Contingency Tables with R lecture. Also, I found the following R functions to be helpful: `as.data.frame`, `subset`, `xtabs`, and of course `chisq.test`.

   (a) What kind of design is this? Introspective?

   (b) Make a 2×2 table of Sex by Survived, just for adult passengers. Display the observed frequencies.

      i. Use `prop.table` to get a table of the relevant proportions. Round to 3 decimal places.

      ii. Calculate the sample odds ratio (not the restricted MLE produced by `fisher.test`). In words, what does this number represent?

      iii. Carry out a Pearson chi-squared test. Display the test statistic, degrees of freedom and p-value.

      iv. In plain, non-statistical language, what do you conclude?

   (c) Now make a $3 \times 2 \times 2$ table of Class by Sex by Survived, again just for adult passengers.

      i. Use `prop.table` to get a table of proportions. What you want is a table showing the proportions of males and females who survived, for each class separately.

      ii. Carry out a Pearson chi-squared test separately for each Class. In plain, non-statistical language, what do you conclude? So you can check, I get $X^2 = 59.159$ for third class.

      iii. Calculate the sample odds ratio for each sub-table. So you can check, I get 72.45614 for first class. Do these odds ratios look equal? Later, we will be able to test it.

(d) You have done a lot of the coding already, so it's easy to look at just the children.

   i. Start with a $3 \times 2 \times 2$ table of Class by Sex by Survived, just for children. Comment.
   ii. How about estimated odds ratios?
   iii. Do the only analysis you really can do. What proportion of female 3d class children survived? Males? Is the difference statistically significant at $\alpha = 0/05$ using the Pearson test?
   iv. Describe your findings for the children (all 3 passenger classes) in plain, non-statistical language.

Please bring hard copy of your complete R input and output from Question 8 to the quiz.
