STA 312f10 Assignment 3

Please start by reading Chapter 2. You may skip Section 2.7 and 2.8. They are brief discussions of things we are *not* going to cover.

Do this assignment in preparation for the quiz on Friday, Oct. 1st. The non-computer parts are practice for the quiz, and are not to be handed in. But please bring your R printouts from Questions 12 and 13. One or both may be handed in (or maybe not).

1. Do Problem 2.4 using a calculator. The problem is looking for the Pearson chi-square test of independence. It is a good idea to check your calculation with R's `chisq.test` function.

2. In the Z-test for the data of Table 2-1 (the vitamin C study), how is the null hypothesis being used in the denominator of the test statistic?

3. If the random variable Z has a standard normal distribution, what is the distribution of the random variable Z^2? You don't have to show any work; just write down the answer. This is something you should know. What is the connection to Expression (2.5)?

4. Look at the bottom section of Table 2-2 on page 11. Where does the number 33.9 come from? Show your work (about one line).

5. For the data of Table 2-3 on p. 13,

   (a) State the null hypothesis in words. It is something about lizards.

   (b) State the null hypothesis in symbols. It is something about p_{ij} for $i = 1, 2$ and $j = 1, 2$.

   (c) Use R's `chisq.test` function to reproduce the value of X^2 from the text. Or you could do it by hand if you insist.

   (d) What is the critical value of the test statistic at $\alpha = 0.05$? It's in an Appendix.

   (e) What do you conclude at the $\alpha = 0.05$ significance level, in words. Begin your answer with "These results are consistent with . . . ."

6. What is the most appropriate sampling model for the following data sets? Your answer should be either "multinomial," "Poisson" or "product multinomial." Be able to explain and justify your choice. In some cases there could be a legitimate difference of opinion.

   (a) Table 1-2 on page 6.

   (b) Table 2-1 on page 8.

   (c) Race of Prisoner and Race of Victim example from lecture.

   (d) Table 2-2 on page 11.

   (e) Table 2-6 on page 20.

(f) Data of Problem 2.1.

(g) Data of Problem 2.3.

(h) Data of Problem 2.5.

7. Consider a log-linear model for a one-dimensional table – a single multinomial variable with k categories, $\mathbf{x} \sim M(N, \mathbf{p})$. The model is based on

$$\mu = \frac{1}{k}\sum_{j=1}^{k} \log m_j \textbf{ and } \log m_j = \mu + \mu_{(j)}, \text{ where } \sum_{j=1}^{k} \mu_{(j)} = 0.$$

   (a) Write p_j in terms of μ and $\mu_{(j)}$. Show a little work.

   (b) Prove that $\mu = \log\left(\frac{N}{\sum_{i=1}^{k} e^{\mu_{(i)}}}\right)$; show your work. Hint: start with the fact that the probabilities add up to one.

   (c) Prove that $p_j = \frac{e^{\mu_{(j)}}}{\sum_{i=1}^{k} e^{\mu_{(i)}}}$. Show your work.

   (d) Derive the log likelihood $\ell = \sum_{j=1}^{k} \mu_{(j)} x_j - N \log \sum_{i=1}^{k} e^{\mu_{(i)}}$. Show your work.

8. Consider a log-linear model for the following two-dimensional table with $N = 250$.

$p_{11} = 0.1$	$p_{12} = 0.2$
$p_{21} = 0.3$	$p_{22} = 0.4$

Calculate the following. The answers are all numbers. If a question like this appears on the quiz, the lecture slide with the formulas you need will be provided.

   (a) μ

   (b) $\mu_{2(1)}$

   (c) $\mu_{1(2)}$

   (d) α

   (e) The odds of an observation being being in column one given it is in row one.

   (f) The odds of an observation being being in column one given it is in row two.

   (g) The odds ratio.

   (h) $\mu_{12(11)}$

   (i) $\mu_{12(12)}$

   (j) $\mu_{12(22)}$

Want to check your answers? Do it by hand first (as on the quiz), then try this.

```
p = rbind(c(.1,.2),
          c(.3,.4))
N = 250
m = p*N
```

```
logm = log(m)
mu = mean(logm); mu
help(apply)
eff1 = apply(logm,1,mean) - mu; eff1
eff2 = apply(logm,2,mean) - mu; eff2
alpha = m[1,1]*m[2,2]/(m[1,2]*m[2,1]); alpha
int11 = log(alpha)/4; int11
```

9. For the data of Table 2-1 (the vitamin C study), the (estimated) odds of a cold are _____ times as great for the skiers who did not take ascorbic acid supplements. The answer is a number.

10. Prove that independence in a 2×2 table implies $\alpha = 1$. Show your work.

11. You have just shown that independence in a 2×2 table implies $\alpha = 1$. Does $\alpha = 1$ imply independence? Answer Yes or No. Justify your answer with a calculation starting $(p_{11} + p_{12})(p_{11} + p_{21}) - p_{11} = \ldots$

    Because of well-known probability rules like A and B are independent if and only if A and B^c are independent and so on, you only need to consider one cell in the table.

12. Do problem 2.1 from the text. Base your answer on the likelihood ratio test; calculate G^2 and the p-value with R, and bring the printout to the quiz.

13. Do problem 2.5 from the text using R. You can stick to Pearson chi-square tests. Don't bother with part (d). Instead of (d), use Excel, Open Office (www.openoffice.org), or some other software of your choice to plot the log observed frequencies. You could even use R if you want, but it's not recommended. Bring the printouts to the quiz.