**Section 6.2 practice problems**

The answers given here are sometimes only partial answers. Please see the answer keys for projects and tests for examples of full answers.

Note: Many of the practice problems are based on exercises given in Alan Agresti’s “Introduction to Categorical Data Analysis” book.

1. Below is a contingency table that has many 0 counts (this is often referred to as being “sparse”).

> c.table <- array(data = c(0, 1, 0,

 7, 1, 8,

 0, 1, 0,

 0, 1, 0,

 0, 1, 0,

 0, 1, 0,

 0, 1, 0,

 1, 0, 0,

 1, 0, 0), dim=c(3,9))

> c.table

 [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]

[1,] 0 7 0 0 0 0 0 1 1

[2,] 1 1 1 1 1 1 1 0 0

[3,] 0 8 0 0 0 0 0 0 0

Complete the following using this data:

* 1. Perform a Pearson chi-square test for independence using a  distribution approximation.

> chisq.test(x = c.table, correct = FALSE)

 Pearson's Chi-squared test

data: c.table

X-squared = 22.286, df = 16, p-value = 0.1342

Warning message:

In chisq.test(x = c.table, correct = FALSE) :

 Chi-squared approximation may be incorrect

* 1. Determine how well the  distribution approximation performs.

Note that because there are no row and column names in c.table, R will use letters.

A  approximation does a poor job in estimating the exact distribution for X2.



* 1. Perform Fisher’s exact test.

> fisher.test(x = c.table)

 Fisher's Exact Test for Count Data

data: c.table

p-value = 0.001505

alternative hypothesis: two.sided

* 1. Perform a permutation test for independence.

The p-value is very low.

* 1. Find estimates of the probabilities for the exact PMF for X2.

> temp <- table(X.sq.star.save)/B

> data.frame(X.sq = round(as.numeric(names(temp)),4), rel.freq =

 round(as.numeric(temp),4))

 X.sq rel.freq

1 15.6637 0.055

2 15.7976 0.113

3 15.8929 0.040

4 15.9018 0.132

5 16.1429 0.043

6 16.1875 0.012

7 16.3333 0.015

8 16.4018 0.029

9 16.4643 0.107

10 16.7143 0.066

11 16.7351 0.059

12 16.7500 0.002

13 16.8304 0.008

14 16.9018 0.108

15 17.1875 0.007

16 17.3304 0.016

17 17.7143 0.010

18 17.8304 0.038

19 17.8929 0.041

20 18.0208 0.001

21 18.0476 0.029

22 18.2351 0.013

23 18.4762 0.005

24 18.6637 0.023

25 18.7500 0.002

26 19.2262 0.001

27 19.6875 0.004

28 19.9018 0.004

29 20.0000 0.004

30 20.0833 0.007

31 20.1429 0.004

32 20.4643 0.001

33 21.1875 0.001

* 1. Compare the results for the three hypothesis test methods.

The  distribution approximation would result in a different conclusion than the other testing methods!

* 1. Using the Monte Carlo simulation methods of Section 3.2, try to perform a test for independence. What types of problems do you encounter?

Because the row and column totals are not fixed, we have a large number of simulated contingency tables without the same size as what was observed!