Name: *Caleb McWhorter — Solutions*
MAT 222
Fall 2019
Homework 7

"Don't do anything I would do, and definitely don't do anything I wouldn't do. There's a little grey area there, and that's where you operate."

–Tony Stark, Spider-Man Homecoming

Problem 1: What are the conditions on the expected counts for a Chi-Squared test are required?

The average expected cell count should be 5 or greater and each expected count should be 1 or greater.

Problem 2: Suppose you are performing a Chi-squared analysis test between factors 'hours studied' and 'exam grade'. If you reject the the null hypothesis, does mean that the data suggests that the amount of hours studied predicts your exam grade or influences your exam grade?

No! Rejecting the null hypothesis merely means that you are stating that the data is more consistent with there being a dependent relationship between the number of hours studied for the exam and the exam grade. This does not mean that one of them 'causes' the other or that there is a predictive relationship between them. This is like the fact that correlation does not implies causation. There could be hidden variables or other complicated connections between the two. Do not jump on the 'simple' explanation for Chi-squared tests! Chi-squared tests alone are not enough to establish causality, i.e. a casual relationship between the variables!

Problem 3: Benford's Law is an observation that the frequency distribution of leading digits in many 'real-life' datasets follows a certain distribution, given below:

Digit	1	2	3	4	5	6	7	8	9
Proportion	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

This is even used as a starting point for detecting fraud. Suppose you have a financial data set with leading digit counts given below:

Digit	1	2	3	4	5	6	7	8	9
Count	41	22	13	12	9	8	8	8	7

Use a chi-square analysis to determine if the given dataset is consistent with Benford's Law. Be sure to state your degrees of freedom, test statistic, p-value, and to interpret your conclusion.

There were a total of 128 numbers examined. Then we find for our dataset has an expected value table of For instance, the entry for 6 was calculated via $0.067 \cdot 128 = 8.576$. *Then we have a contribution*

Digit	1	2	3	4	5	6	7	8	9
Expected	38.53	22.53	16.00	12.42	10.11	8.58	7.42	6.53	5.89

to χ^2*-table of which gives test statistic*

Digit	1	2	3	4	5	6	7	8	9
Contribution	0.159	0.012	0.563	0.014	0.122	0.039	0.045	0.332	0.210

$$X^2 = 0.159 + 0.012 + 0.563 + 0.014 + 0.122 + 0.039 + 0.045 + 0.332 + 0.210 = 1.496$$

We have degrees of freedom $9 - 1 = 8$ *so that this gives* p*-value* $p > 0.25$. *Therefore, we fail to reject the null hypothesis. The dataset is consistent with Benford's Law.*

Problem 4: A graduate program is trying to evaluate the quality of success of students accepted to their Masters program to make better future admissions decisions. To determine the quality of success of the students, they use whether students achieve a 'High Pass' or 'Low Pass' on their Master's Exams. The data is broken down by the type of undergraduate university the student attended.

	High Pass	Low Pass
Public	15	17
Private	19	14
International C	17	16

Use an appropriate test to determine if there is a relationship between the type of university a student attended and whether the achieved a 'High Pass' or 'Low Pass' on their Master's Exams. Be sure to state your null and alternative hypotheses in the context, give your degrees of freedom, test statistic, and p-value, and state your conclusion in the context of the problem.

We test the hypotheses

$$\begin{cases} H_0 : \text{there is no association between university type and pass type} \\ H_a : \text{there is some association between university type and pass type} \end{cases}$$

The table with totals is

	High Pass	Low Pass	Total
Public	15	17	32
Private	19	14	33
International	17	16	33
Total	51	47	98

We have expected count table:

	High Pass	Low Pass
Public	16.65	15.35
Private	17.17	15.83
International	17.17	15.83

For example, the middle left entry was calculated by $\frac{51 \cdot 33}{98} = 17.17$. *Then we have a contribution to* χ^2*-table as follows:*

	High Pass	Low Pass
Public	0.164	0.178
Private	0.194	0.211
International	0.002	0.002

For example, the middle left entry was calculate by $\frac{(19 - 17.17)^2}{17.17} = 0.195$. *Then the test statistic is*

$$X^2 = 0.164 + 0.194 + 0.002 + 0.178 + 0.211 + 0.002 = 0.751$$

We have degrees of freedom $(3-1)(2-1) = 2$. *Then we have* p*-value* $p > 0.25$. *As* $p > \alpha = 0.10$, *we fail to reject the null hypothesis. The data is consistent with the fact that there being no association between the undergraduate university the student attended and whether or not they received a 'High Pass' on their Master's Exams. Observe the expected values were all at least 5 so that the assumptions of the chi-squared test were satisfied.*