Categorical Data Analysis

(STAT343)

Assignment 2

Due March 28, 2019

* **NOTE: Do calculations by hand, SAS, and/or R to answer the following questions.**

1. The following table shows the results of a study comparing radiation therapy with surgery in treating cancer of the larynx. The response indicates the cancer was controlled for at least two years following treatment. Let θ be the odds ratio in the 2×2 table. Significance level 0.05 is used for testing.

	Controlled	Not controlled
Surgery	21	2
Radiation therapy	8	9

   (a) Calculate and interpret the difference of proportions (D), relative risk (RR), and odds ratio (OR), including their confidence intervals.

   (b) Calculate and interpret the Pearson and likelihood-ratio (LR) chi-squared statistics for testing $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$.

   (c) Calculate and interpret the p-value for Fisher's exact test with $H_0 : \theta = 1$ and $H_1 : \theta \neq 1$.

2. A study aspirations of high school students measured aspirations with the scale (some high school, high school graduate, some college, college graduate). The data are shown in the following table. The table also contains the estimated expected frequencies under independence and the standardized Pearson residuals in the parentheses.

Study Aspirations	Family Income low	middle	high	Total
some high school	9	11	9	29
	(?, ?)	(11.47, -0.19)	(9.45, -0.19)	
high school graduate	44	52	41	137
	(38.14, 1.58)	(54.20, -0.54)	(44.66, -0.95)	
some college	13	23	12	48
	(13.36, -0.13)	(18.99, 1.30)	(15.65, -1.24)	
college graduate	10	22	27	59
	(16.43, -2.11)	(23.34, -0.40)	(?, ?)	
Total	76	108	89	273

(a) Test independence of educational aspirations and family income using the Pearson χ^2 statistic (X^2), and the likelihood ratio χ^2 statistic (G^2).

(b) Complete the above table. Do the residuals suggest any association pattern?

(c) Calculate and interpret the odds ratio (OR) from the cells at the four corners.

3. Consider the previous data for investigating the association between educational aspirations and family income. Assign the scores to study aspiration and family income in the following manner: 1=some high school, 2=high school graduate, 3=some college, 4=college graduate and 1=low, 2=middle, 3=high.

   (a) Calculate the sample correlation between educational aspirations and family income, and test whether there is an increasing or descreasing trend.

   (b) Let 10=some high school, 20=high school graduate, 30=some college, 40=college graduate and -1=low, 0=middle, 1=high. Do the results change?

   (c) Let 1=some high school, 3=high school graduate, 5=some college, 10=college graduate and 1=low, 5=middle, 10=high. Do the results change?

4. The following table refers to applicants to graduate school at the University of California at Berkeley, for fall 1973. It presents admissions decisions by gender of applicant for the size largest graduate departments. Denote the three variables by A = whether admitted, G = gender, and D = department.

	Whether Admitted			
	Male		Female	
Department	Yes	No	Yes	No
1	512	313	89	19
2	353	207	17	8
3	120	205	202	391
4	138	279	131	244
5	53	138	94	299
6	22	351	24	317
Total	1198	1493	557	1278

(a) Compute and interpret the sample AG conditional odds ratios.

(b) Compute and interpret the sample AG marginal odds ratios.

(c) Explain why they give such different indications of the AG association.