

THE CHINESE UNIVERSITY OF HONG KONG
Department of Statistics

STAT4006: Categorical Data Analysis
Problem Sheet 2

The deadline for this Problem Sheet is 5.30pm on Monday 26th October, i.e. before class starts. Please submit your solutions via the link provided on the course Blackboard page - if you must submit your solutions in hard copy, please contact me at jawright@sta.cuhk.edu.hk in advance. **No late submissions will be accepted. A late submission will receive a mark of zero.** Students may discuss set problems with others, but their final submissions must be their own work.

Please answer the following problems. All these questions should be answered using a pen, paper, calculator (good practice for your midterm and final). That said, you may use any software you like to find percentiles (i.e. for finding p -values). Show your working.

1. Table 1 is from a report on the relationship between aspirin use and myocardial infarction (heart attacks) by the Physicians' Health Study Research Group at Harvard Medical School. Find the p -value for testing that

Group	Myocardial Infarction	
	Yes	No
Placebo	158	10410
Aspirin	71	10321

Table 1: Heart Attack Data

the incidence of heart attacks is independent of aspirin intake using the chi-square test. Interpret your results.

2. An analysis of campus accident data was made to determine the distribution of numbers of fatal accidents for automobiles of two sizes. The data for 16 accidents are given in Table 2. Do the data indicate that the frequency of fatal accidents is independent of the size of automobiles? Choose a test, justify your choice,

	Size of auto		
	Small	Large	Total
Fatal	1	7	8
not Fatal	5	3	8
Total	6	10	16

Table 2: Campus Accident Data

perform it and interpret the results.

3. Table 3 is based on records of accidents in 1988 compiled by the Department of Highway Safety and Motor Vehicles in Florida.

Safety Equipment in Use	Injury	
	Fatal	Non-fatal
None	1598	162526
Seat belt	502	421360

Table 3: Highway Safety Data

- (a) Find and interpret the difference of proportions, relative risk, and odds ratio. Why are the relative risk and odds ratio approximately equal?
- (b) Construct 95% confidence intervals for the difference of proportions; the relative risk and the odds ratio. Interpret.

Highest Degree	Religious Beliefs		
	Fundamentalist	Moderate	Liberal
Less than high school	178	183	108
High school or junior college	570	648	442
Bachelor or graduate	138	252	252

Table 4: Religious Belief Data

4. Table 4, from a recent General Social Survey, cross-classifies the degree of fundamentalism of subjects' religious beliefs by their highest degree of education.
 - (a) Find and interpret a measure of association between subjects' religious beliefs and their education level.
 - (b) Test the independence of the religious beliefs and the degree using two-sided test with $\alpha = 0.05$. (*Hint: use scores for Highest degree 1, 2, 3, and scores for religious belief 1, 2, 3.*)
 - (c) What is the advantage for the test method we adopted in part (b), compared with the X^2 and G^2 tests?
5. Give a "real world" example of three variables X , Y , and Z , for which you expect X and Y to be marginally associated but conditionally independent, controlling for Z .
6. The following table refers to applicants to graduate school at the University of California, Berkeley for the fall 1973 session. Admissions decisions are presented by gender of applicant, for the three largest graduate departments. Denote the three variables by A = whether admitted, G = gender, and D = department.

Department	Whether Admitted			
	Male		Female	
	Yes	No	Yes	No
1	478	302	80	23
2	365	199	16	7
3	117	203	204	385
4	133	276	127	250
5	50	138	94	299
6	22	351	24	317

Table 5: Berkeley Data

- (a) Find the sample AG conditional odds ratios, and compare them with the sample AG marginal odds ratio. Why are they so different?
- (b) Conduct the Cochran-Mantel-Haenszel test. Specify the hypothesis and interpret. Comment on the applicability of this data to this data.
- (c) Calculate and interpret an estimate of the common AG conditional odds ratio.

THE END