Homework on Categorical Data

1

Discrimination? Wabash Tech has two professional schools, business and law. Here are two-way tables of applicants to both schools, categorized by gender and admission decision. (Although these data are made up, similar situations occur in reality.)[13]

	Business			Law	
	Admit	Deny		Admit	Deny
Male	480	120	Male	10	90
Female	180	20	Female	100	200

(a) Make a two-way table of gender by admission decision for the two professional schools together by summing entries in these tables.

(b) From the two-way table, calculate the percent of male applicants who are admitted and the percent of female applicants who are admitted. Wabash admits a higher percent of male applicants.

(c) Now compute separately the percents of male and female applicants admitted by the business school and by the law school. Each school admits a higher percent of female applicants.

(d) This is Simpson's paradox: both schools admit a higher percent of the women who apply, but overall Wabash admits a lower percent of female applicants than of male applicants. Explain carefully, as if speaking to a skeptical reporter, how it can happen that Wabash appears to favor males when each school individually favors females.

2

Majors for men and women in business. A study of the career plans of young women and men sent questionnaires to all 722 members of the senior class in the College of Business Administration at the University of Illinois. One question asked which major within the business program the student had chosen. Here are the data from the students who responded:[6]

	Female	Male
Accounting	68	56
Administration	91	40
Economics	5	6
Finance	61	59

This is an example of a single sample classified according to two categorical variables (gender and major).

(a) Test the null hypothesis that there is no relationship between the gender of students and their choice of major. Give a P-value and state your conclusion.

(b) Verify that the expected cell counts satisfy the requirement for use of chi-square.

(c) Describe the differences between the distributions of majors for women and men with percents, with a graph, and in words.

(d) Which two cells have the largest terms of the chi-square statistic? How do the observed and expected counts differ in these cells? (This should strengthen your conclusions in (b).)

(e) What percent of the students did not respond to the questionnaire? The nonresponse weakens conclusions drawn from these data.

3

Course grades. Most students in a large statistics course are taught by teaching assistants (TAs). One section is taught by the course supervisor, a senior professor. The distribution of grades for the hundreds of students taught by TAs this semester was

Grade	A	B	C	D/F
Probability	0.32	0.41	0.20	0.07

The grades assigned by the professor to students in his section were

Grade	A	B	C	D/F
Count	22	38	20	11

(These data are real. We won't say when and where, but the professor was not the author of this book.)

(a) What percents of each grade did students in the professor's section earn? In what ways does this distribution of grades differ from the TA distribution?

(b) Because the TA distribution is based on hundreds of students, we are willing to regard it as a fixed probability distribution. If the professor's grading follows this distribution, what are the expected counts of each grade in his section?

(c) Does the chi-square test for goodness of fit give good evidence that the professor follows a different grade distribution? (Give the test statistic, its P-value, and your conclusion.)