

442s04 Assignment 11

Do these problems in preparation for the final exam (Test 4). Bring your log and list files to the exam.

The following table shows percentage of applicants admitted to the six largest graduate departments at the University of California at Berkeley in 1973. From this table, you can recover the number of male applicants who were accepted, the number of female applicants who were accepted, the number of male applicants who were rejected, and the number of female applicants who were rejected --- for each department. The result is 24 data lines.

Major	Men		Women	
	Number of applicants	Percent admitted	Number of applicants	Percent admitted
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

Make a SAS program to read the data.

1. First, check your work with `proc freq`. Don't do any significance tests yet.
 - a. Make a two-dimensional table in which the rows are Major field of study and the columns are sex. Your cell frequencies should match the "Number of Applicants" column.
 - b. Make a three-dimensional table, consisting of two Major by Admitted tables. If the rows are Major field of study, your row percentages should match the "Percent Admitted" columns.

2. Now use `proc freq` to make a simple two-dimensional table of sex by whether or not the person was admitted. Use the `chisq` option to test whether these variables are related or not.

- a. What is the independent variable? The dependent variable?
- b. Is this study experimental, observational, or both? Why?
- c. Are the results statistically significant at the 0.05 level?
- d. Give the reduced and full models in the bracket notation of log-linear models.
- e. What is the value of the likelihood ratio test statistic for goodness of fit of your reduced model? The answer is a number on your `proc freq` output.
- d. State your conclusion in plain language. If you just say the variables are "related" without saying who is more likely to be get into graduate school, you lose around half marks.

3. Now use `proc freq` to make a simple two-dimensional table of sex by Department (Major). Use the `chisq` option to test whether these variables are related or not.

- a. What is the independent variable? The dependent variable?
- b. Are the results statistically significant at the 0.05 level?
- c. Give the reduced and full models in the bracket notation of log-linear models.
- d. What is the value of the likelihood ratio test statistic for goodness of fit of your reduced model? The answer is a number on your `proc freq` output.
- e. State your conclusion in plain language. If you just say the variables are "related" without being more specific, you lose around half marks.
- f. Is it right to say that their sex causes students to apply to different departments? Explain.

4. Now test the relationship between sex and getting in to graduate school at Berkeley, *controlling for what department the student applied to*. Please get the answer two ways.

- a. Do it with `proc freq`. Add up likelihood ratio statistics.
 - i) What is the value of your test statistic? The answer is a number.
 - ii) Using the critical value of chi-square with 6 df = 12.59159, is the test significant?
 - iii) If your result is non-significant, stop. But if it is significant, follow up with Bonferroni-corrected tests on each of the six sub-tables separately. What p-value do you need for significance at *joint* level 0.05?
 - iv) Have your conclusions changed? State them in plain language. If you are not at least a little surprised, you are missing something.

b. Do the significance test of part (a) with proc catmod, getting exactly the same value of the test statistic.

5. As a last follow-up to the last result, conduct the test for three-way association among the variables, which I interpret like a sex by department interaction, with admission as the dependent variable.

- i) What is the value of your test statistic? The answer is a number.
- ii) What is the p-value? The answer is a number.
- iii) Are the results statistically significant?
- iv) What do the results mean? Use non-technical language.