Department of Statistics

COURSE STATS 330

Assignment 5, 2005

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Tuesday October 18.

Note: No data is supplied for this assignment. You must type your own and construct suitable data frames using the **`rep`** and **`expand.grid`** functions.

Question 1.
The data in table 1 come from a classic British study into the effect of smoking on the incidence of coronary heart disease and lung cancer.

1. Make an appropriate data frame and fit a model to explain the coronary death rate (in deaths per 1000 person years) in terms of age and smoking, using a suitable offset. Make a table of these rates. [10 marks]

2. Does smoking have an effect on coronary death rates? If so, does the effect depend on age? [5 marks]

3. If you find that it does, can you think of an intuitive reason why this might be so? [5 marks]

Table 1. Data for Question 1

	Person-Years		Coronary Deaths	
Age	Nonsmokers	Smokers	Nonsmokers	Smokers
35-44	18793	52407	2	32
45-54	10673	43248	12	104
55-64	5710	28612	28	206
65-74	2585	12663	28	186
75-84	1462	5317	31	102

Question 2
A survey to assess attitudes to AIDS funding posed the following questions; (1) Do you think the Government should pay all health-related costs for AIDs victims? (2) Do you think the government should allocate funds to promote safe sex practices? Table 2 gives the answers of 621 respondents to this questionnaire, classified by gender.

Table 2. Data for Question 2.

			Pay costs	
		Fund program	Yes	No
gender	Male	Yes	76	160
		No	6	25
	Female	Yes	114	181
		No	11	48

1. Fit a suitable log-linear model to these data. Do we require a saturated model or is a smaller model or models adequate? [10 marks]

2. Describe your model(s) in terms of conditional odds ratios and independence. [5 marks]

3. Give a confidence interval for the conditional odds ratio between "gender" and "pay costs", conditional on "Fund program". [5 marks]