

## Week 8

**Tutorial Exercise.**

1. The data below come from a 1992 survey in Dayton Ohio on high school students' use of alcohol, tobacco and marijuana. Students were asked if they had ever used each of the drugs.

Alcohol use	Cigarette use	Marijuana use	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

- (a) Calculate the expected cell frequencies for the cell (1, 1, 1) under complete independence.
- (b) Calculate the expected cell frequencies under the model that claims alcohol use and marijuana use are independent given smoking status.
- (c) Test the fit of this model using Pearson's  $X^2$ .
- (d) Express the conditional independence model in log-linear format. Determine the degrees of freedom associated with the deviance for this model.
- (e) Use the log-linear model output and an appropriate model to predict the chance of a randomly selected student having used alcohol but never smoked cigarettes or marijuana.
2. Consider the saturated log-linear model for a 3-way table

$$\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}.$$

Show that if

$$\log(np_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk}.$$

then  $R$  is completely independent of  $S$  and  $T$ .