

Week 7

Tutorial Exercise.

1. Consider the log-linear model

$$\log P(R = i|S = j) = \mu + \alpha_i + \beta_j + \lambda_{ij},$$

with $\alpha_1 = \beta_1 = 0$ and $\lambda_{1j} = \lambda_{i1} = 0$ for all $i = 1, \dots, r$ and $j = 1, \dots, s$. Prove that $\lambda_{ij} = 0$ for $i = 1, \dots, r, j = 1, \dots, s$ if and only if $P(R = i|S = j) = P(R = i|S = 1)$ for $i = 1, \dots, r, j = 1, \dots, s$.

2. Galton's height data. The following data for 205 married people were reported in Yule (1900).

Husband	Wife			Totals
	Tall	Medium	Short	
Tall	18	28	14	60
Medium	20	51	28	99
Short	12	25	9	46
Totals	50	104	51	205

- (a) Test the hypothesis that heights of husbands and wives are independent.
 - (b) Test the hypothesis of symmetry using this data set by determining the expected frequencies under a model of symmetry ($p_{ij} = p_{ji}$ for all i and j) and using the appropriate X^2 statistic to assess fit. Explain how you calculate the degrees of freedom.
3. Incidence of common colds in a double blind study involving 279 French skiers (Pauling (1971)).

	Cold	No Cold	Total
Placebo	31	109	140
Ascorbic Acid	17	122	139

Test the hypothesis that the incidence of colds is the same in the 140 randomly chosen to get the placebo and in the 139 chosen to receive the vitamin C treatment. What is the p -value in this case? Estimate the difference in the incidence rate if there is a difference.

Computer Exercise.

The following data from Wermuth (1976) gives the gestation period and perinatal mortality rates for a group of German women many of whom were pregnant for the first time or had complications with previous pregnancies. The following data refer to premature births with gestation periods of 30 to 38 weeks.

Mother's age	Cigarettes smoked per day	Baby died	Baby survived
less than 30	≤ 5	50	315
	> 5	9	40
30 or more	≤ 5	41	147
	> 5	4	11

- Fit all log-linear models to these data containing the intercept and all combinations of models containing up to a maximum of one interaction term, e.g. intercept + outcome + mother's age + smoking or intercept + outcome + mother's age + interaction between outcome and mother's age. There should be 8 models in total. Carefully justify your choice of which model is best.
- Use your model to estimate the probability of a premature baby surviving if born to a mother aged 31 who is a heavy smoker. How does the model based estimate compare to the empirical estimate?
- Use your model to estimate the probability of a premature baby surviving if born to a mother aged 31 who is a non-smoker. Comment.