STAT 3014/ 3914

Semester 2	Applied Statistics	2011

Week 7 Solutions

Tutorial Exercise.

1. Consider the log-linear model

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

Given the constraints:

$$\mu = \log P(R=1|S=1)$$
$$\alpha_i = \log(P(R=i|S=1)/P(R=1|S=1))$$
$$\beta_j = \log(P(R=1|S=j)/P(R=1|S=1))$$
$$\lambda_{ij} = \log\left(\frac{P(R=i|S=j)/P(R=1|S=1)}{P(R=i|S=1)P(R=1|S=j)}\right)$$

If $P(R=i|S=j) = P(R=i|S=1)$ for $i = 1,\ldots,r, \quad j = 1,\ldots,s$ (1)
then

$$\begin{aligned} \lambda_{ij} &= \log\left(\frac{P(R=i|S=1)/P(R=1|S=1)}{P(R=i|S=1)P(R=1|S=1)}\right) using(1) \\ &= \log 1 \\ &= 0 \end{aligned}$$

Conversely, if $\lambda_{ij} = 0$ then
$P(R=i|S=j)P(R=1|S=1) = P(R=i|S=1)P(R=1|S=j)$ (2)

Sum over i and recall $\sum_{i=1}^{r} P(R=i|S=j) = 1$ to get

$$P(R=1|S=1) = P(R=1|S=j).$$

Substituting back into (2) gives

$$P(R=i|S=j) = P(R=i|S=1)$$

as required.

2. The expected frequencies under independence are

	Wife			
Husband	Tall	Medium	Short	Totals
Tall	18	28	14	60
	(14.634)	(30.439)	(14.927)	
Medium	20	51	28	99
	(24.146)	(50.224)	(24.629)	
Short	12	25	9	46
	(11.220)	(23.337)	(11.444)	
Totals	50	104	51	205

(a) Test H_0 : Husbands and wives heights are independent.
Pearson's χ^2 is

$$\sum_{ij} \frac{o_{ij}^2}{e_{ij}} - 205 = 2.907.$$

P-value is

$$P(\chi_4^2 \geq 2.907) = 0.5735.$$

Thus the data are consistent with the independence model.

(b) The expected frequency under symmetry in the $(i, j)^{th}$ cell is $(x_{ij} + x_{ji})/2$.
Note that these estimates are the maximum likelihood but we do not need to prove this.
Expected frequencies:

	Wife		
Husband	Tall	Medium	Short
Tall	18	24	13
Medium	24	51	26.5
Short	13	26.5	9

Test $H_0 : p_{ij} = p_{ji}$ for all $i, j = 1, 2, 3$.

Pearson's χ^2 is

$$\sum_{ij} \frac{o_{ij}^2}{e_{ij}} - 205 = 1.657.$$

P-value is

$$P(\chi_3^2 \geq 1.657) = 0.6465.$$

Note the symmetry hypothesis contains 3 parameters (p_{12}, p_{13}, p_{23}) so the statistics has 3 d.f.

From the p-value, we can conclude that a model of symmetry fits this data set.

3. Let the incidence rate for the placebo group be p_p and for the Ascorbic acid group be p_c. Test

$$H_0 : p_p = p_c.$$

	Cold	No Cold	Total
Placebo	31	109	140
	(24.086)	(115.714)	
Ascorbic Acid	17	122	139
	(23.914)	(115.086)	
	48	231	279

$$\chi^2 = \frac{31^2}{24.086} + \ldots + \frac{122^2}{115.086} - 279 = 4.811.$$

$$\begin{aligned} P - value &= P(\chi_1^2 \geq 4.811) \\ &= P(|Z| \geq \sqrt{4.811}) \\ &= P(|Z| \geq 2.19) \\ &= 2(1 - 0.9857) \\ &= 0.0286 \end{aligned}$$

Estimated difference in incidence rate :

$$\frac{31}{140} - \frac{17}{139} = 0.221 - 0.122 = 0.099.$$