Question 37

Do not write in this margin

The data shown below arose in a study of the causes of the lung disease byssinosis. They relate to 5419 workers in the US cotton industry, and record the dustiness of their workplace as well as whether or not they had byssinosis.

	Byssinosis		
Dustiness	yes	no	Total
High	105	564	669
Medium	18	1282	1300
Low	42	3408	3450
Total	165	5254	5419

Calculate the contribution of the top left cell of the table to a chi-squared statistic in a test of the hypothesis that byssinosis is independent of the dustiness of the workplace. The overall chi-squared statistic has value 413.8; conduct a test of this hypothesis, stating clearly the number of degrees of freedom involved and your conclusion. [3]

$105 - \frac{165 \times 669}{5419} = 351.61$

$E_i = \frac{165 \times 669}{5419} = 20.37$

$\sum \frac{O_i - E_i}{E_i} = \frac{(105-20.37)^2}{20.37} = 351.6$

$413.8 \sim \chi^2(2)$

SP ≈ 0 The null hypothesis of no association is rejected - Byssinosis is related to dustiness

Question 38

A biologist observed an insect larva once every 15 minutes for a total of 30 observations. Each time the larva was classified as 'resting' (R) or 'feeding' (F). The matrix of transition frequencies was given by

$$N = \begin{matrix} R \\ F \end{matrix} \begin{bmatrix} 3 & 12 \\ 12 & 2 \end{bmatrix}$$

(a) Assuming a Markov chain model for the changes of activity, use these data to estimate the activity transition probabilities.

$\alpha = 0.8 \quad \beta = 6/7$

$1-\alpha = 0.2 \quad 1-\beta = 1/7$

$\begin{bmatrix} 1/5 & 4/5 \\ 6/7 & 1/7 \end{bmatrix}$

(b) Over the period, the number of runs expected in a Bernoulli model is 16; in fact, the observed number of runs was 23, giving a z-score of 3.3. What conclusions can you draw from these results? [5]

E(R) - The SP for the null hypothesis is very small. The possibility that the sequence arose from a Bernoulli process is very small. The sequence arose probably from a Markov chain with α, β both between 0.5 and 1