MATH 523 - ASSIGNMENT 2

To be handed in not later than 5pm, Friday 9th March 2012.
Please hand in during lectures, to Burnside 1225, or to the Mathematics Office Burnside 1005

1 In a Poisson log-linear model with factor predictors, particular linear predictor constructions correspond to specific independence and conditional independence assumptions for the effects of each of the factors. For example, the model where, for $j = 1, \ldots, J, k = 1, \ldots, K$

$$\log \mu_{jk} = \log \mu + \alpha_j + \beta_k$$

(with specified identifiability constraints) can be rewritten

$$\mu_{jk} = \mu e^{\alpha_j} e^{\beta_k} = \lambda_j^{(1)} \lambda_k^{(2)} \tag{1}$$

say, where the superscripts (1) and (2) identify effects associated with Factor 1 (X_1) and Factor 2 (X_2). The above model corresponds to the main effects only model $1+X_1+X_2$. The multiplicative formula (1) indicates that the two factors act **independently** to influence the expected response.

(a) Write down constraints that render the parameters given in (1) identifiable.

2 MARKS

(b) Consider now a model with three factor predictors X_1, X_2, X_3 with J, K and L levels respectively. Explain the independence structures implied by the two models

$$1 + X_1 + X_2 + X_3 + X_1.X_2 \qquad\qquad 1 + X_1 + X_2 + X_3 + X_2.X_3$$

Give the numbers of parameters, p, that each of these two models contains.

6 MARKS

(c) Consider the following data with 3 factors. The response data are the counts y_{jkl} in a $2 \times 2 \times 2$ table obtained using the cross classification of a certain study population.

- **Factor 1:** Personality Type
  - A - highly stressed personality
  - B - relaxed personality
- **Factor 2:** Cholesterol Level
  - Normal
  - High
- **Factor 3:** Diastolic Blood Pressure
  - Normal
  - High

		X_3	
X_1	X_2	Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

Using a Poisson model with a log link function, and standard analysis of deviance tests, find the most appropriate model to explain the observed variation in response, and comment on the implied independence structures of the selected model.

6 MARKS

2 The following data set represents the counts of numbers of subjects who suffered/did not suffer from a thromboembolism, where the study group comprised users of a particular medication who were either smokers or non smokers. The study design was a case-control study that selected 58 thromboembolism sufferers, and 116 control subjects who were known to have not suffered the condition.

		Smoker		Non-smoker	
		Medication		Medication	
		Y	N	Y	N
Thromboembolism	Y	14	7	12	25
	N	2	22	8	84

Using a Poisson model with a log link function on the eight counts with the three binary factors as predictors, and standard analysis of deviance tests, find the most appropriate model to explain the observed variation in response. Explain the implication of your chosen model in terms of which factors affect response, and how they affect the response.

6 MARKS

Note: These data could also be analyzed using a binomial response model. For the purposes of this Assignment, you should use a Poisson response model.