Assignment 6 — Due Dec 6, 2018

NAIQING CAI

Ncai5@wisc.edu

1. The following data were collected in a study of the relation between parental socioeconomic status (SES) and the mental health of children. Any systematic variation is of interest. Analyze the data and write a brief report.

2. In the Bradley-Terry model for ranking k competitors, parameters θ1,...,θk representing ‘abilities’ are introduced in such a way that the probability πij that competitor i beats j is a function of the difference in their abilities.

(a) Write out the 21 × 7 model matrix X for the Bradley-Terry model.

(b) Fit the Bradley Terry model to these data to obtain a ranking of the teams. Extend this model by including a home-team advantage effect (equal for all teams). Obtain the likelihood-ratio statistic. Comment briefly on the magnitude of home-field advantage.

(c) Estimate the probability that Detroit beats Boston (i) at Boston, (ii) at Detroit, and (iii) on neutral territory.

(d) Does the extended model fit the data? Comment briefly on any patterns in the residuals.

3. Table 1 gives the mean number of children born per woman, the women being classified by place, education, and years since first marriage. Any systematic variation in the number of children is of interest. See fiji.txt for the dataset.

(a) Fit an appropriate model describing how the number of children varies with marital age, mother’s abode and education. Give a brief synopsis of the arguments justifying your formulation and choice of model, including checks for model adequacy.

**Fit an appropriate model**













In this way, we can analyze the data by fitting the **poisson** model to mean number of children.

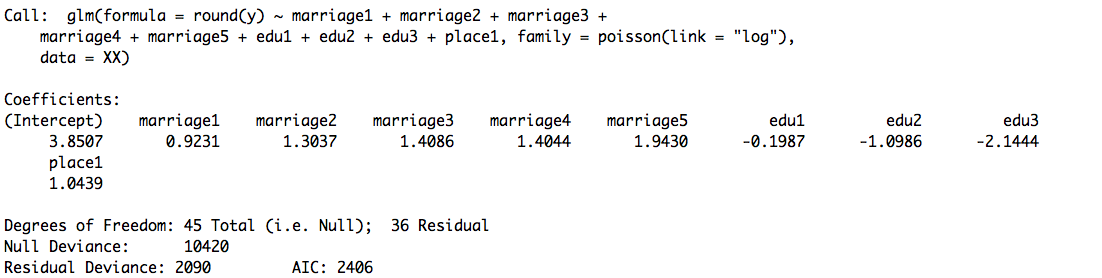


Table 3.1 Log-linear poisson model

Based on the r summary table, we can get the formula as follows:



**Checks for model adequacy**

(b) Explain the meaning of all parameters in your model. Comment on the major factors affecting fertility.

**Meaning of all parameters**

















**Comment**

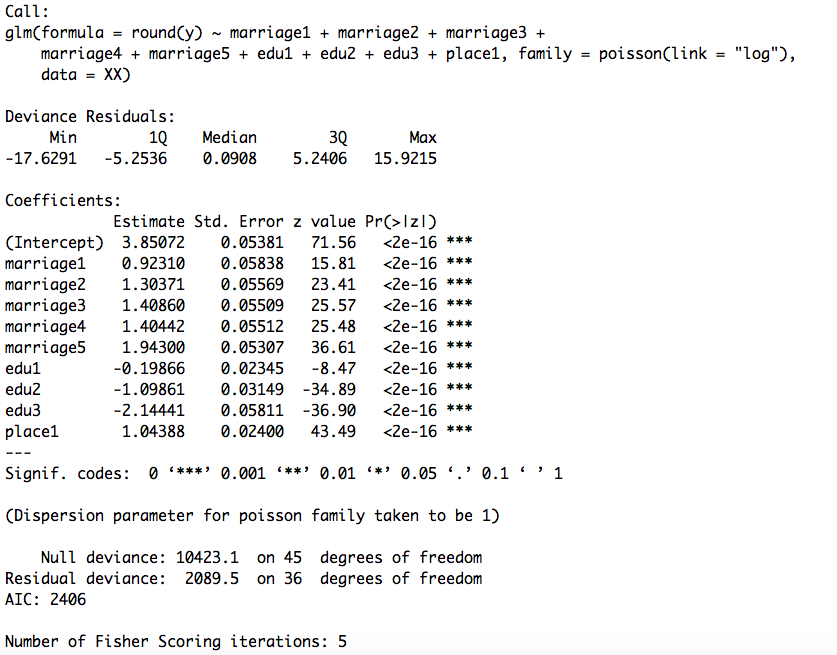


Table 3.2 Estimates for log-linear poisson model

Based on the table above, we can make a comment on the major factors affecting fertility.

major factors are:

Marriage is the most effective factor among the three factors. Also, women with marriage years over 25years will have the highest fertility. That is said, the longer the marriage, the more the children to be born.

Education is the second effective factor. Women with lower elementary level education will have more children than the other three kinds education levels.

Place is the least effective factor. For this factor, women in urban will have more children than those in rural.

(c) Construct a 95% confidence interval for the mean number of children born to an urban woman with upper elementary education after ten years of marriage.

(d) Estimate the lifetime average number of children born to rural women with secondary education. Give 90% confidence limits.

4. The file byss.txt contains information, obtained from a survey conducted by a large textile company, on the prevalence of byssinosis, a lung disease to which cotton workers are subject. The file lists the observed prevalence of byssinosis (affected, not affected), by race (white = 1; non white = 2), sex (male = 1; female = 2), smoking habits (two levels), length of employment (three levels), and dustiness of the work environment (three levels). In the last three cases, higher-numbered categories denote larger values (more smoking, longer employment and increased dustiness). Parts (a) and (b) are based on the assumption that the main-effects linear logistic model is substantially correct.

(a) Fit the main-effects linear logistic model. Explain how the residual degrees of freedom is calculated for the deviance.

**Fit the model**



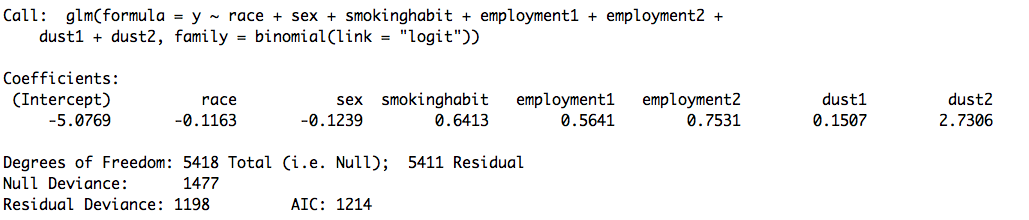


Table 4.1 Linear Logistic Model

Based on the above table, we can find that the linear logistic model is the following form:



**Explanation**

residual degrees of freedom is 5411

residual deviance is 1198

(b) Interpret the coefficient of sex(2). Construct an approximate 90% confidence interval for the odds ratio (males vs females) of contracting byssinosis.

**Interpret the coefficient of sex(2)**

sex(2) means female, coefficient of sex(2) is beta2 in the model.

Beta2 represents the the difference between male and female in prevalence of byssinosis.

**90% confidence interval**



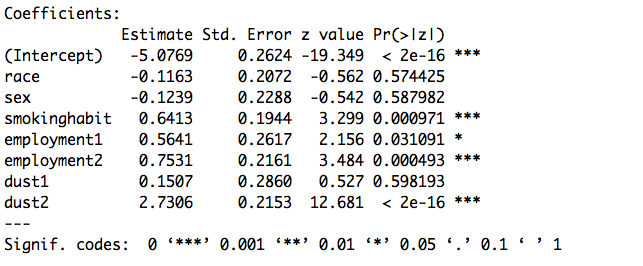




So the 90% confidence interval for the odds ratio (males vs females) of contracting byssinosis is the 90% confidence interval for the coefficient beta2.

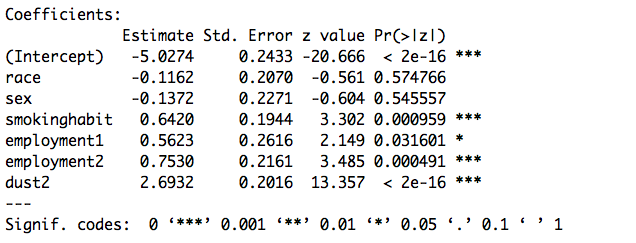
90% confidence interval for beta2: (-0.4973715, 0.2565643)

(c) Drop the least significant factor from the model, proceeding until all the remaining factors are significant at the 5% level. Interpret the reduced model thus obtained.



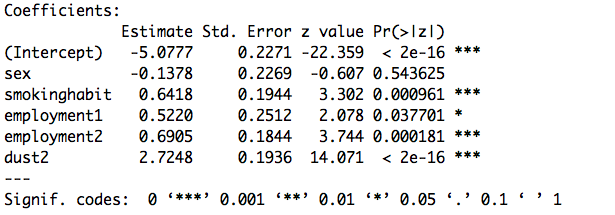
Table

Based on the above table, we find that the least significant factor is dust1. So we drop it from the model.



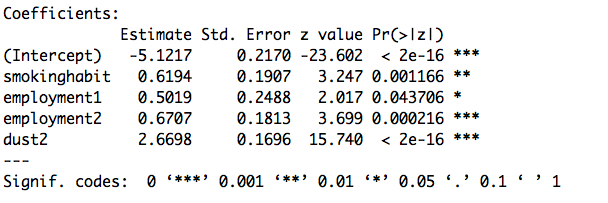
Table

Based on the above table, we find that the least significant factor is race. So we drop it from the model.



Table

Based on the above table, we find that the least significant factor is sex. So we drop it from the model.



Table

(d) Beginning with the complete main-effects model, look for significant interactions by fitting each of the ten models main effects + one interaction. In judging the significance of interactions, you should bear in mind, at least informally, the effects of selection. After detecting the significant interactions, remove insignificant main effects as described in (c), except for those that are included in interactions. Interpret the model thus obtained.

**Fitting each of the ten models main effects + one interaction**

(e) You are required to write a short report giving details of the excess risk associated with cotton dust. How fast does the risk increase with dust level? If necessary, give separate figures for males and females or for smokers and non-smokers.

(f) Does this analysis suggest that the aetiology of byssinosis is related to sex or race? Explain.