

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Massey University Library. Thesis Copyright Form

Title of thesis: A Review of some Models for the  
Analysis of Contingency Tables

- (1) (a) I give permission for my thesis to be made available to readers in the Massey University Library under conditions determined by the Librarian.
- (b) ~~I do not wish my thesis to be made available to readers without my written consent for \_\_\_\_\_ months.~~
- (2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.
- (b) ~~I do not wish my thesis, or a copy, to be sent to another institution without my written consent for \_\_\_\_\_ months.~~
- (3) (a) I agree that my thesis may be copied for Library use.
- (b) ~~I do not wish my thesis to be copied for Library use for \_\_\_\_\_ months.~~

Signed J.A. Anderson

Date 18/12/87

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME AND ADDRESS

DATE

---

---

---

---

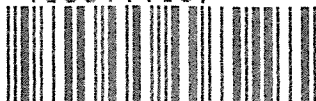
---

---

FOR  
Reference Only

NOT TO BE REMOVED FROM THE LIBRARY

1088111067



A REVIEW OF SOME MODELS FOR THE ANALYSIS  
OF CONTINGENCY TABLES

A thesis presented in partial fulfilment of the requirements  
for the degree of Master of Arts in Statistics  
at Massey University

Julie Anne Anderson

1987

ABSTRACT

Some models proposed for the analysis of contingency tables are reviewed and illustrated with examples.

These include standard loglinear models; models which are suitable for ordinal categorical variables such as ordinal loglinear, log-multiplicative and logit models, and models based on an underlying distribution for the response; and models for incomplete and square tables.

Estimation methods and inference are also discussed.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr Doug Stirling, for his guidance and supervision of this project. Thanks also to Mrs Gail Haydock for the typing of this thesis. Finally, I wish to thank my husband Dean for his help, encouragement and support.

TABLE OF CONTENTS

	<u>Page</u>
Abstract	(i)
Acknowledgements	(ii)
Table of Contents	(iii)
Chapter I: Introduction	1
1.1 Categorical Variables	1
1.2 Two-Dimensional Tables	2
1.3 Three-Dimensional Tables	3
1.4 Sampling Models	4
1.4.1 Poisson	4
1.4.2 Multinomial	5
1.4.3 Product Multinomial	5
1.4.4 Equivalence of Results for Different Sampling Models	6
1.5 Response and Explanatory Variables	6
1.5.1 Three Responses	7
1.5.2 Two Responses	8
1.5.3 One Response	8
1.5.4 Types of Models that can be Fitted	8
1.6 Ordinal Categorical Data	9
1.6.1 Advantages of using Ordinal Methods	10
1.6.2 Odds Ratios for 2x2 Tables	11
1.6.2.1 Incidence of Colds Example	13
1.6.3 Odds Ratios for rxc Tables	13
1.6.3.1 Local Odds Ratios	14
1.6.3.2 Local-Global Odds Ratios	15
1.6.3.3 Global Odds Ratios	16
1.6.3.4 Dumping Severity Example	17

1.7	Estimation	19
1.8	Model testing	20
1.9	Structural and Sampling Zeroes	20
1.9.1	Monkey Example	20
1.10	Loglinear Models	23
1.10.1	Fitting Loglinear Models	24
1.11	Linear Models	24
1.11.1	Linear Models Specified as $A_{\pi} = X\beta$	24
1.11.2	Linear Models Specified in Terms of Constraints	25
1.11.3	Fitting linear models	26
1.12	Other Models	26

	<u>Page</u>
Chapter II: Nominal Loglinear Models	28
2.1 Two Dimensional Tables	28
2.1.1 Multiplicative Form of the Loglinear Model	28
2.1.1.1 Saturated Model	28
2.1.1.2 Independence Model	31
2.1.2 Additive Form	31
2.1.2.1 Saturated Model	31
2.1.2.2 Independence Model	33
2.1.2.2.1 Abortion Attitude Example	33
2.1.2.3 Goodness of Fit Statistics	35
2.1.2.3.1 Abortion Attitude Example	35
2.2 Three Dimensional Tables	36
2.2.1 Associations Between Three Variables	36
2.2.2 Hierarchial Models	38
2.2.3 Estimation	39
2.2.4 Model Testing	39
2.2.5 Abortion Attitude Example	40
2.2.6 Conditional Test Statistics	42
2.2.6.1 Partitioning Chi-Square	43
2.2.6.2 Abortion Attitude Example	43
2.2.6.3 University Admissions Example	
2.3 Higher Order Contingency Tables	44
2.4 Other Loglinear Models	45



	<u>Page</u>
Chapter III: Some General Issues	46
3.1 Model Selection	46
3.1.1 Abortion Attitude Example	46
3.1.2 Residual Analysis	47
3.2 Interpretation	48
3.2.1 Abortion Attitude Example	48
3.3 Collapsing Tables	49
3.3.1 University Admissions Example	50

	<u>Page</u>
Chapter IV: Loglinear Models for Ordinal Variables	53
4.1 Disadvantages of Ignoring Ordinal Nature of Variables	53
4.1.1 Dumping Severity Example	53
4.1.1.1 Odds Ratios	54
4.1.1.2 Residual Analysis	54
4.2 Ordinal-Ordinal Tables	55
4.2.1 Dumping Severity Example	55
4.2.2 Linear by Linear Association Model	57
4.2.3 Estimation	58
4.2.4 Dumping Severity Example	59
4.2.5 Conditional Test of Independence	60
4.3 Ordinal-Nominal Tables	61
4.3.1 Estimation	63
4.3.2 Dumping Severity Example	63
4.3.3 Conditional Test of Independence	65
4.4 Higher Dimensions	67
4.4.1 Odds Ratios	68
4.4.2 All Variables Ordinal	68
4.4.3 Ordinal and Nominal Variables	69
4.4.3.1 One Nominal, Two Ordinal Variables	69
4.4.3.2 Two Nominal, One Ordinal Variable	70
4.4.3.3 Dumping Severity Example	70
4.4.4 Three Factor Interaction Models	74
4.4.4.1 Smoking Example	75

	<u>Page</u>
Chapter V: Log-Multiplicative Models	79
5.1 Estimation	80
5.2 Inference for Log-Multiplicative Models	81
5.3 Dumping Severity Example	81
5.4 Higher Dimensions	84

	<u>Page</u>
Chapter VI: Logit Models	85
6.1 Dichotomous Response	85
6.2 Polytomous Response	87
6.2.1 Adjacent Categories Logits	88
6.2.1.1 Ordinal-Ordinal	87
6.2.1.2 Ordinal-Nominal	89
6.2.2 Continuation Ratio Logits	89
6.2.3 Cumulative Logits	90
6.2.3.1 Heterogeneous Effects	90
6.2.3.1.1 Ordinal-Ordinal Tables	90
6.2.3.1.1.1 Dumping Severity Example	91
6.2.3.1.2 Ordinal-Nominal Tables	95
6.2.3.2 Homogenous Effects	95
6.2.3.2.1 Ordinal-Ordinal Tables	97
6.2.3.2.1.1 Dumping Severity Example	97
6.2.3.2.1.2 Conditional Test of Independence	99
6.2.3.2.2 Ordinal-Nominal Tables	99
6.2.3.2.2.1 Dumping Severity	100
6.2.3.2.2.2 Conditional Test of Independence	102
6.2.4 Cumulative Logit Models for Higher Dimensions	103
6.2.4.1 Homogeneous Linear Logit Effects	104
6.2.4.1.1 Dumping Severity Example	105
6.2.4.2 Higher Order Interaction Models	107
6.2.4.3 Heterogeneous Effects	107

Page

Chapter VII: Models Based on an Underlying Distribution for the Response	108
7.1 Distribution Functions	110
7.1.1 Normal	111
7.1.2 Logistic	111
7.1.3 Extreme Value	113
7.1.4 Estimation	115
7.1.5 Dumping Severity Example	118

	<u>Page</u>
Chapter VIII: Other Models	125
8.1 Mean Response Models	125
8.1.1 Dumping Severity Example	125
8.1.2 Three-Way Dumping Severity Example	126
8.2 Models for Incomplete Tables	127
8.2.1 Definitions	127
8.2.2 Quasi-Independence	127
8.2.3 Estimation	128
8.2.4 Monkey Example	128
8.2.5 Higher-Order Tables	129
8.2.5.1 Health Concerns Example	130
8.3 Models for Square Tables	132
8.3.1 Quasi-Independence	133
8.3.1.1 Social Mobility Example	134
8.3.2 Symmetry	136
8.3.2.1 Eye-Testing Example	136
8.3.3 Quasi-Symmetry	138
8.3.3.1 Eye-Testing Example	139
8.3.4 Marginal Homogeneity	139
8.3.4.1 Estimation	140
8.3.4.2 Eye-Testing Example	140
8.3.5 Multi-Dimensional Tables	142
8.4 Linear Models	142
8.4.1 Drug Example	142
8.5 Summary	145

## Appendices

Appendix 1: Equivalence of MLEs under Poisson, Multinomial and Product-Multinomial Sampling Schemes	148
A.1.1 MLEs	148
A.1.2 Deviance	150
A.1.2.1 Multinomial	150
A.1.2.2 Poisson	150
A.1.2.3 Product-Multinomial	151
Appendix 2: Fitting Loglinear Models	152
A.2.1 Newton-Raphson Algorithm	152
A.2.1.1 Abortion Attitude Example	153
A.2.2 Interactive Proportional Fitting	154
A.2.2.1 Lizard Example	155
Appendix 3: Fitting Linear Models	159
A.3.1 Linear Models Specified in Terms of Constraints	159
A.3.1.1 Wedderburn's Algorithm for Finding MLEs of Generalized Linear Models Specified in Terms of Constraints	160
A.3.2 Linear Models Specified as $A\pi = X\beta$	165
A.3.2.1 Drugs Example	166
Appendix 4: Maximum Likelihood Equations for Loglinear Models for Three-Dimensional Tables	171
Appendix 5: Fitting Mean Response Models	174
A.5.1 Dumping Severity Example	174

	<u>Page</u>
Appendix 6: Quasi-Symmetry	178
A.6.1 One Dummy Variable	178
A.6.1.1 Eye-Testing Example	178
A.6.2 Two Dummy Variables	181
A.6.2.1 Eye-Testing Example	181
Appendix 7: Marginal Homogeneity	182
A.7.1 Method of Solving Simultaneous Equations	182
A.7.1.1 Eye-Testing Example	182
A.7.2 Generalized Linear Models Specified in Terms of Constraints	184
A.7.2.1 Eye-Testing Example	184
Appendix 8: Fitting Models Using Genstat	188
References	273



## CHAPTER I: INTRODUCTION

### 1.1 Categorical Variables

This thesis discusses different types of models that can be used to describe categorical data. A categorical variable differs from a continuous variable in that rather than being able to take on a continuous range of values, it is only classified into a certain number of categories. An example would be marital status, which could have categories such as married, widowed, divorced, or "other". If we classify each member of a sample simultaneously on two or more categorical variables, then we can form a cross-classification table. For example, we might classify 1000 people by their marital status and age (where age has only been measured in categories) such as in Table 1.1.

Table 1.1: Cross-classification table of 1000 people by age and marital status

Age (years)	Marital Status				Total
	Married	Widowed	Divorced	Other	
< 25	100	10	10	180	300
25 - 40	200	50	100	50	400
> 40	120	75	80	25	300
TOTAL	420	135	190	255	1000

A cross-classification table is also referred to as a contingency table or cross-tabulation.

For some variables such as marital status and sex, the only sensible way to measure them is to classify them into categories. However, some other variables, such as age and income, can be measured on a

continuous scale, but it is often more convenient to simply categorize them.

## 1.2 Two-Dimensional Tables

Consider a two-way table of counts with the row variable,  $X$ , having  $r$  categories, and the column variable,  $Y$ , having  $c$  categories - thus there are  $r$  rows and  $c$  columns. We will denote the actual count in the  $i$ th row and  $j$ th column by  $n_{ij}$ , and the corresponding expected count under some model as  $m_{ij}$ . The row and column totals are:

$$n_{i+} = \sum_{j=1}^c n_{ij}$$

$$n_{+j} = \sum_{i=1}^r n_{ij}$$

The total number of observations is

$$\sum_i \sum_j n_{ij} = N$$

Assuming that neither category has fixed marginal totals, the probability that a given individual is classified into cell  $(i, j)$  is:

$$\pi_{ij} = P(X \text{ takes on level } i \text{ and } Y \text{ takes on level } j)$$

where

$$\pi_{ij} = \frac{m_{ij}}{N}$$

$$\sum_i \pi_{ij} = \pi_{+j}$$

$$\sum_j \pi_{ij} = \pi_{i+}$$

$$\sum_{ij} \pi_{ij} = 1$$

If X and Y are independent, then

$$\begin{aligned} \pi_{ij} &= P(X \text{ takes on level } i) \times P(Y \text{ takes on level } j) \\ &= \pi_{i+} \pi_{+j} \end{aligned}$$

Since the expected value of  $n_{ij}$  is

$$m_{ij} = N \pi_{ij}$$

then under the model of independence

$$m_{ij} = N \pi_{i+} \pi_{+j}$$

Later, we will discuss models that allow X and Y to be associated in some way. For these models the expected values depend on more than just the marginal probabilities.

### 1.3 Three-Dimensional Tables

We can extend the notation introduced in Section 1.2 to the case of three-way tables. A three-way table with variables X, Y and Z having  $r$ ,  $c$  and  $\ell$  categories respectively, will be said to have observed counts  $n_{ijk}$  with corresponding expected counts  $m_{ijk}$  and population probabilities  $\pi_{ijk}$ .

An example of a three-way table is Table 1.2 which classifies a sample of 1593 people by their age, religion and frequency of church attendance (Knoke and Burke, 1980, p.68).

Table 1.2: Effect of age and religion on church attendance

Religion	Age	Church Attendance			Total
		Low	Medium	High	
Non-Catholic	Young	322	124	141	587
	Old	250	152	194	596
Catholic	Young	88	45	106	239
	Old	28	24	119	171
TOTAL		688	345	560	1593

Later, we will formulate models that allow various types of association between the variables.

#### 1.4 Sampling Models

There are three common sampling models that are used for the collection of cross-classified data. We will illustrate for the case of an  $r \times c \times l$  table classified by variables  $X$ ,  $Y$  and  $Z$ . These results can be easily generalized to tables of a different dimension.

##### 1.4.1 Poisson

We observe a set of independent Poisson processes, one for each cell in the table over a fixed time period, with no prior knowledge of the total number of observations to be taken. The count  $n_{ijk}$  in each cell will have a Poisson distribution with mean  $m_{ijk}$ , i.e. the probability function for  $n_{ijk}$  has the form

$$f(n_{ijk}) = \frac{m_{ijk}^{n_{ijk}} e^{-m_{ijk}}}{n_{ijk}!}$$

The log likelihood function is

$$\log L(n_{ijk}) = \sum_{i,j,k} n_{ijk} \log m_{ijk} - \sum_{i,j,k} m_{ijk} - \sum_{i,j,k} n_{ijk}!$$

Since the cells contain counts having independent Poisson distributions, the total count in the table,  $N$ , has a Poisson distribution with mean

$$m_{+++} = \sum_{ijk} m_{ijk}$$

#### 1.4.2 Multinomial

We take a fixed sample of size  $N$  and cross-classify each member of the sample according to the categorical variables. The cell counts  $\{n_{ijk}\}$  will have the multinomial distribution specified by the sample size  $N$  and the rcg population probabilities  $\{\pi_{ijk}\}$ . The probability of a particular set of cell counts  $\{n_{ijk}\}$  that sum to  $N$  is the multinomial likelihood

$$L(n_{ijk}) = \frac{N!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} \pi_{ijk}^{n_{ijk}}$$

The log likelihood is

$$\log L(n_{ijk}) = \sum_{i,j,k} n_{ijk} \log \pi_{ijk} + \log N! - \sum_{i,j,k} \log n_{ijk}!$$

The expected value of each  $n_{ijk}$  is  $m_{ijk} = N\pi_{ijk}$ .

#### 1.4.3 Product Multinomial

For each combination of one or more categorical explanatory variables, we take a multinomial sample of fixed size which is classified by the remaining response variable(s). For example, suppose we fix the  $\ell$  layer totals and take a sample of size  $n_{++k}$  for each  $k$ . Let  $\pi_{ij(k)}$  be the probability of an observation falling into the  $i$ th category of  $X$  and the  $j$ th category of  $Y$ , given

that it falls into the  $k$ th category of  $Z$  (i.e.  $\pi_{ijk}/\pi_{++k}$ ). The cell counts within the  $k$ th layer have the multinomial distribution specified by the sample size  $n_{++k}$  and the probabilities  $\{\pi_{ijk} \mid k = 1, \dots, \ell\}$ , and cell counts from different layers are independent. The cell counts in layer  $k$  have the probability function

$$L(n_{ij(k)}) = \frac{n_{++k}!}{\prod_{i,j} n_{ijk}!} \prod_{i,j} \pi_{ij(k)}^{n_{ijk}}$$

and the product of these from the  $\ell$  layers gives the probability function for the whole table (the product multinomial likelihood)

$$L(n_{ijk}) = \prod_k \frac{n_{++k}}{\prod_{i,j} n_{ijk}!} \prod_{i,j} \pi_{ij(k)}^{n_{ijk}}$$

The expected value of each  $n_{ijk}$  is  $m_{ijk} = n_{++k} \pi_{ij(k)}$ .

#### 1.4.4 Equivalence of Results for Different Sampling Models

For the models that will be discussed in this thesis, the maximum likelihood estimates (MLEs) are the same for all sampling schemes. The one condition required is that a term corresponding to the fixed margin(s) in the product multinomial sampling scheme be included in the model (for more details see Appendix 1). Because of this equivalence, generally models will be phrased as though the sampling scheme was multinomial.

### 1.5 Response and Explanatory Variables

Each variable (i.e. margin) in a table can be thought of as either an explanatory variable (factor) which affects others, or as a response variable which depends on other factors.

For three-dimensional tables there are three possible combinations:

- (i) no explanatory, three response variables
- (ii) one explanatory, two response variables
- (iii) two explanatory, one response variable.

Examples of these three types of tables include Tables 1.2, 1.3 and 1.4.

Table 1.3: Occupation (O), Education (E), and Aptitude (A)  
of World War II volunteers

O1 (self employed, business)

O2 (self employed, professional)

	E1	E2	E3	E4		E1	E2	E3	E4
A1	42	55	22	3		1	2	8	19
A2	72	82	60	12		1	2	15	33
A3	90	106	85	25		2	5	25	83
A4	27	48	47	8		2	2	10	45
A5	8	18	19	5		0	0	12	19

O3 (teacher)

O4 (salary employed)

	E1	E2	E3	E4		E1	E2	E3	E4
A1	0	0	1	19		172	151	107	42
A2	0	3	3	60		208	198	206	92
A3	1	4	5	86		279	271	331	191
A4	0	0	2	36		99	126	179	97
A5	0	0	1	14		36	35	99	79

#### 1.5.1 Three Responses

Type (i) tables are only rarely found in practice. However Table 1.3 can be thought of as one. The data, taken from Fienberg (1980, p. 45) refer to the classification of 4353 World War II volunteers into four occupational groups by four levels of education and five

levels of aptitude. Because of the sampling scheme and the way in which the individuals were classified (see Fienberg for further details), all three variables can be thought of as responses.

### 1.5.2 Two Responses

Table 1.4, taken from Fienberg (1980, p.27) is an example of the second type of table. The data refer to the perch heights and diameters of two different species of lizards. Species is an explanatory variable which affects the responses of height and diameter.

Table 1.4: Perch height and diameter of two species of lizards

Perch Diameter	Sagrei Species Perch height		Distichus Species Perch height	
	< 4.0"	> 4.0"	< 4.0"	> 4.0"
>4.75'	32	86	61	73
<4.75'	11	35	41	70

### 1.5.3 One Response

Type (iii) tables are the most common three-dimensional tables. An example is given in Table 1.2 which illustrates the effect of the explanatory variables, religion and age, on the response, frequency of church attendance.

### 1.5.4 Types of Models that can be Fitted

For type (i) tables only Poisson or multinomial sampling schemes are usually appropriate, whereas for types (ii) and (iii) we would also use a product-multinomial model in which the fixed marginal totals correspond to explanatory variables.



The distinction between explanatory and response variables certainly affects the interpretation of the results, but often does not affect the types of models that can be fitted. A sensible approach for the analysis of tables with one or more explanatory variables is to condition on the values of these margins, treating them as fixed even in those cases where they are not. We will discuss this approach more fully later.

### 1.6 Ordinal Categorical Data

When one or more of the variables in a cross-classification is measured on an ordinal scale, we can use models which take account of this to give more powerful tests of association and simpler, more incisive measures of this association than models which simply treat all the variables as nominal.

An illustration of an ordinal variable and the levels of its corresponding scale would be education which might be measured as primary school, high school, or tertiary education.

Other examples would be consumer rating of a new food product as dislike a lot, dislike, indifferent, like, like a lot; or measuring the softness of water as soft, medium or hard.

Ordinal scales commonly occur in many disciplines, such as the social sciences (e.g. for measuring attitudes and opinions), marketing (e.g. for preference scales), medicine (e.g. for describing severity of an injury, or degree of recovery from an illness). In many fields ordinal scales often result when discrete measurement is used with inherently continuous variables such as age, income or social status. Often it is possible to measure a variable perhaps even on a continuous scale, but much quicker and more convenient to simply measure it on an ordinal scale. For instance, the amount of sediment left on a filter pad may be simply classified as none, slight, moderate or excessive by comparing it to a photographic standard rather than drying it and precisely weighing it.

A categorical variable is referred to as "ordinal" rather than "interval" when there is a clear ordering of the categories but the absolute distances among them are unknown. For example, the variable

"education" is ordinal when measured with categories primary school, high school, university, but it is interval when measured with the integer values 0, 1, 2,... representing number of years of education.

An ordinal variable is quantitative because it corresponds to different quantities of a certain characteristic, while qualitative variables which are measured on a nominal scale have no such property. Examples of nominal variables are race, religion or marital status. The order of listing of the categories of a nominal variable is obviously unimportant.

#### 1.6.1 Advantages of Using Ordinal Methods

Most of the well-known methods for analysing categorical data (such as the Pearson chi-squared test of independent or the common loglinear models discussed in Chapter II) treat all variables as nominal, i.e. the results are invariant to permutations of the categories of any of the variables.

Since ordinal variables are inherently quantitative, Agresti (1984) argues that their descriptive measures should be more like those for interval variables than those for nominal variables.

The advantages of using ordinal methods instead of the standard nominal procedures include:

1. Ordinal methods have greater power for detecting particular kinds of association;
2. Ordinal data description is based on measures that are similar to those (e.g. correlations, slopes) used in ordinary regression and analysis of variance for continuous variables;
3. Ordinal analyses can use a greater variety of models, most of which are more parsimonious and have simpler interpretations than the standard models for nominal variables.
4. Interesting ordinal models can be applied in settings where the standard nominal models are trivial or else have too many parameters to be tested for goodness of fit.

In Chapters IV to VII we will discuss particular classes of models that can be used to model ordinal categorical data. These include ordinal loglinear, log-multiplicative and logit models, as well as models based on underlying distributions for the response.

### 1.6.2 Odds Ratios for 2x2 Tables

The odds ratio is a measure that describes the degree of association in a 2x2 table - it is especially important in the study of ordinal models.

Consider the 2x2 population cross-classification with cell probabilities  $\pi_{ij}$ . Within row 1 the odds that variable 2 is in column 2 instead of column 1 is

$$\Omega_1 = \frac{\pi_{12}}{\pi_{11}}$$

Within row 2 the corresponding odds equals

$$\Omega_2 = \frac{\pi_{22}}{\pi_{21}}$$

Each  $\Omega_i$  is nonnegative, with value greater than 1.0 if column 2 is more likely than column 1.

The ratio of these odds

$$\begin{aligned} \theta = \frac{\Omega_2}{\Omega_1} &= \frac{\pi_{22}/\pi_{21}}{\pi_{12}/\pi_{11}} \\ &= \frac{\pi_{11} \pi_{22}}{\pi_{21} \pi_{12}} \end{aligned}$$

is the odds ratio. It is sometimes called the cross product ratio, since it is the ratio of the products  $\pi_{11} \pi_{22}$  and  $\pi_{12} \pi_{21}$  of proportions from cells that are diagonally opposite.

Each odds  $\Omega_i$  can be expressed as

$$\begin{aligned}\Omega_i &= \frac{\pi_{i2}/\pi_{i+}}{\pi_{i1}/\pi_{i+}} \\ &= \frac{\pi_{2(i)}}{\pi_{1(i)}}\end{aligned}$$

so

$$\theta = \frac{\pi_{2(2)}/\pi_{1(2)}}{\pi_{2(1)}/\pi_{1(1)}}$$

The row and column variables are independent if and only if  $\Omega_1 = \Omega_2$  (and so  $\theta = 1.0$ ). If  $1 < \theta < \infty$ , then individuals in row 2 are more likely to be in column 2 than are individuals in row 1, i.e.

$\pi_{2(2)} > \pi_{2(1)}$ . If  $0 < \theta < 1$ , individuals in row 2 are less likely to be in column 2 than are individuals in row 1, i.e.  $\pi_{2(2)} < \pi_{2(1)}$ .

For sample cell frequencies  $\{x_{ij}\}$ , a sample analog of  $\theta$  is

$$\hat{\theta} = \frac{n_{11} n_{22}}{n_{21} n_{12}}$$

The value of  $\theta$  does not change if both cell frequencies within any row are multiplied by a nonzero constant, or if both cell frequencies within any column are multiplied by a constant. So  $\hat{\theta}$  estimates the same characteristic ( $\theta$ ) even if disproportionately large or small samples are selected from the various marginal categories of a variable. In particular, it estimates the same characteristic regardless of whether sampling is full multinomial or independent multinomial. It also takes the same value if the orientation of the table is reversed so that the rows become the columns and the columns become the rows.

If the order of the rows or the order of the columns is reversed, the new value of  $\theta$  is simply the inverse of the original value. So two values of  $\theta$  that are the inverse of one another (such as 3 and  $1/3$ ) represent the same degree of association, but in opposite directions.

The odds ratio is a multiplicative function of the cell proportions. Its logarithm is an additive function, i.e.

$\log \theta = \log \pi_{11} - \log \pi_{12} = \log \pi_{21} + \log \pi_{22}$  and may equal any real number. The log odds ratio is symmetric about the independence value of 0.0 in the sense that a reversal of the two rows or the two columns results in a change of its sign.

#### 1.6.2.1 Incidence of Colds Example

Pauling (1971) describes a double-blind study to evaluate the effect of ascorbic acid (vitamin C) on the common cold. One group of 140 skiers received a placebo, while a second of 139 received 1 g of ascorbic acid per day. The incidence of colds was recorded and is shown in Table 1.5.

Table 1.5: Incidence of common colds

Treatment	No Cold	Cold
Ascorbic acid	122	17
Placebo	109	31

The odds of catching a cold for the ascorbic acid group are  $17/122 = .14$ , while the odds for the placebo group are  $31/109 = 0.28$ . The ratio of these odds is  $0.28/0.14 = (122 \times 31)/(109 \times 17) = 2.04$ . This means that the odds of catching a cold were 2.04 times higher for the placebo group than for the ascorbic acid group. This odds ratio is significantly higher than 1.0, so it is plausible that administration of vitamin C helped to prevent the occurrence of colds.

#### 1.6.3 Odds Ratio for rxc Tables

For the general rxc table odds ratios can be formed using each of  $\binom{r}{2} = r(r-1)/2$  pairs of rows in combination with each of the  $\binom{c}{2} = c(c-1)/2$  pairs of columns. For rows a and b and columns c and d, the odds ratio  $(\pi_{ac} \pi_{bd})/(\pi_{bc} \pi_{ad})$  uses four cells occurring in a

rectangular pattern (see Figure 1), and there are  $\binom{r}{2} \binom{c}{2}$  odds ratios of this type. The independence of the two variables is equivalent to the condition that all these population odds ratios equal 1.0.

However, there is much redundant information when the entire set of these odds ratios is used to characterize the association in a table.

#### 1.6.3.1 Local Odds Ratios

A basic set of  $(r-1)(c-1)$  odds ratios is

$$\theta_{ij} = \frac{\pi_{ij} \pi_{i+1,j+1}}{\pi_{i,j+1} \pi_{i+1,j}}, \quad i=1, \dots, r-1, \\ j=1, \dots, c-1$$

Figure 1.1: General Odds Ratio  $\pi_{ac} \pi_{bd} / \pi_{bc} \pi_{ad}$

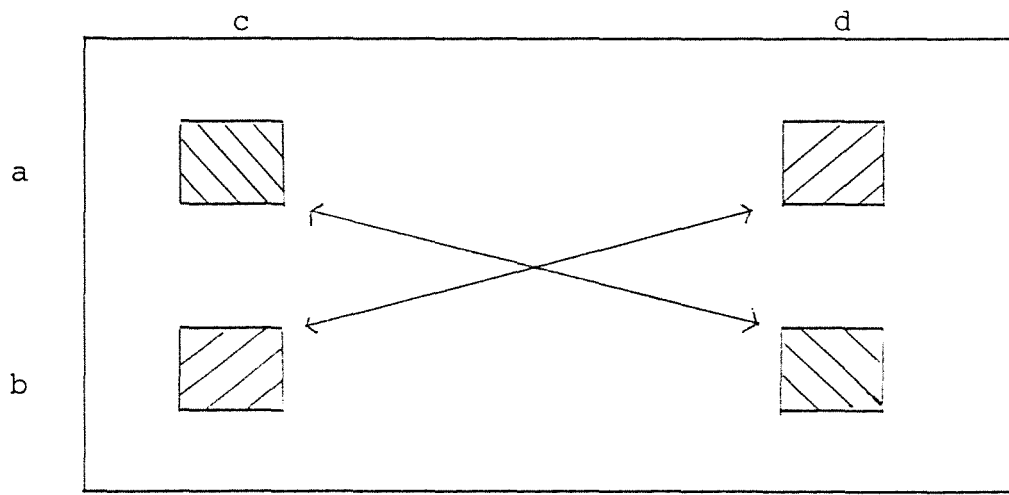
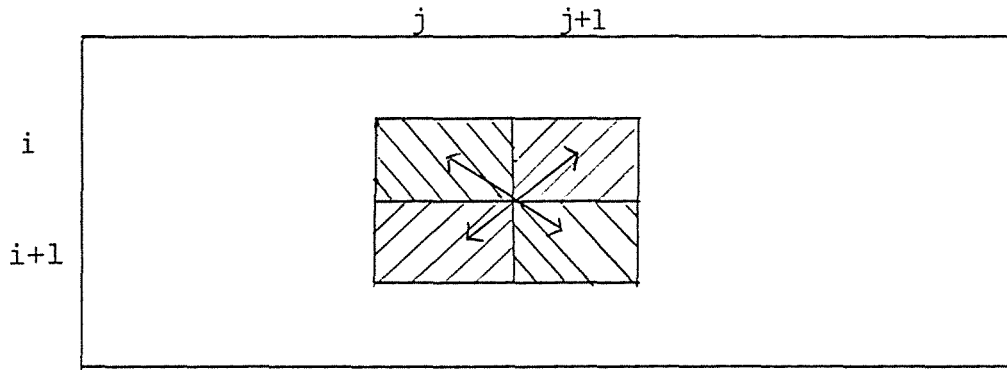


Figure 1.2: Local Odds Ratio  $\theta_{ij}$



This basic set determines all  $\binom{r}{2} \binom{c}{2}$  odds ratios that can be formed from pairs of rows and pairs of columns. Independence of the two variables is therefore also equivalent to the condition that the odds ratios in the basic set are equal to one.

These odds ratios are formed using cells in adjacent rows and adjacent columns, as illustrated in Figure 1.2. Their volumes describe the relative magnitude of "local" associations in the table, so they are called local odds ratios.

#### 1.6.3.2 Local-Global Odds Ratios

Another family of odds ratios is

$$\theta'_{ij} = \frac{\left( \sum_{b \leq j} \pi_{ib} \right) \left( \sum_{b > j} \pi_{i+1,b} \right)}{\left( \sum_{b > j} \pi_{ib} \right) \left( \sum_{b \leq j} \pi_{i+1,b} \right)}$$

$$i = 1, \dots, r-1,$$

$$j = 1, \dots, c-1$$

These odds ratios are local in the row variable but "global" in the column variable, since all  $c$  categories of the column variable are used in each odds ratio (see Figure 1.3). They are particularly meaningful when a distinction is made between response and explanatory variables.

### 1.6.3.3 Global Odds Ratios

A third family of odds ratios is

$$\theta''_{ij} = \frac{\left( \sum_{a \leq j} \sum_{b \leq j} \pi_{ab} \right) \left( \sum_{a > i} \sum_{b > j} \pi_{ab} \right)}{\left( \sum_{a \leq i} \sum_{b > j} \pi_{ab} \right) \left( \sum_{a > i} \sum_{b \leq j} \pi_{ab} \right)}$$

These measures are the regular odds ratios computed for the 2x2 tables corresponding to the  $(r-1)(c-1)$  ways of collapsing the row and column classification into dichotomies. They treat row and column variables alike and describe associations that are global in both variables (see Figure 1.4).

Figure 1.3: Local-Global Odds Ratio  $\theta'_{ij}$

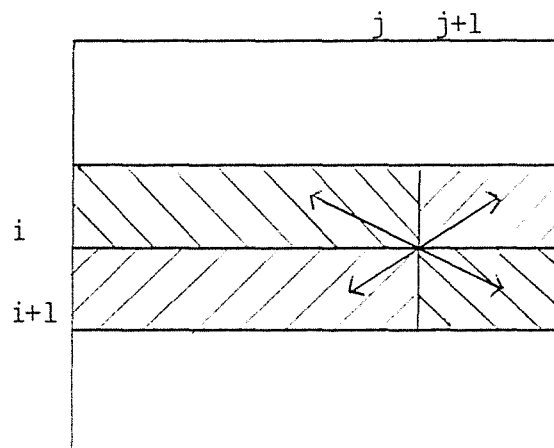
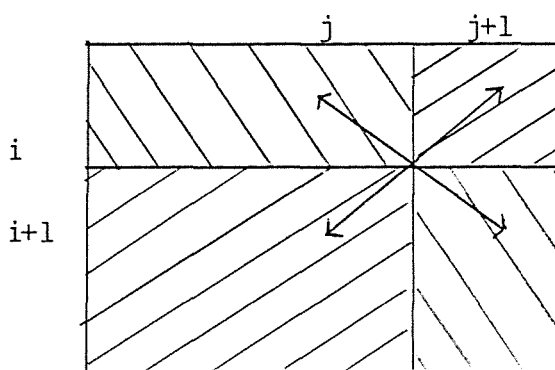




Figure 1.4: Global Odds Ratio  $\theta''_{ij}$



For local, local-global and global odds ratios, independence is equivalent to all log odds ratios equalling zero. An association described by one of these measures is referred to as "positive" or "negative" according to the sign of the log odds ratio.

If all  $\log \theta_{ij} > 0$ , then all  $\log \theta'_{ij} > 0$ . If all  $\log \theta'_{ij} > 0$ , then all  $\log \theta''_{ij} > 0$ . The converses of these statements are not true (Agresti, 1984). The condition that all local log odds ratios be positive is therefore the most stringent of three possible definitions for "uniformly positive association".

The less localized the odds ratio, the more precise its sample value tends to be as an estimation of its population value, since the standard error involves the inverses of larger sample totals. So if all the  $\{\theta_{ij}\}$  are approximately equal, if the  $\{\theta'_{ij}\}$  are approximately equal, and if the  $\{\theta''_{ij}\}$  are approximately equal, the sample estimates of the third set will tend to be smoothest.

#### 1.6.3.4 Dumping Severity Example

We will illustrate these three types of odds ratios for ordinal variables using the data in Table 1.6, from Grizzle, Starmer and Koch (1969). The data refer to a comparison of four different operations for treating duodenal ulcer patients. The operations

correspond to removal of various amounts of the stomach. Operation A is drainage and vagotomy, B is 25% resection and vagotomy, C is 50% resection and vagotomy, and D is 75% resection. The categories of operation are ordered, with A being the least severe operation and D corresponding to the greatest removal of stomach. The variable "dumping severity" describes the extent of a possible undesirable side effect of the operation. The categories of this variable are also ordered, with the response "none" representing the most desirable result.

Table 1.6: Dumping severity and operation

Operation	Dumping Severity			Total
	None	Slight	Moderate	
A	61	28	7	96
B	68	23	13	104
C	58	40	12	110
D	53	38	16	107
TOTAL	240	129	48	417

Table 1.7 contains the sample values  $\{\hat{\theta}_{ij}\}$ ,  $\{\hat{\theta}'_{ij}\}$  and  $\{\hat{\theta}''_{ij}\}$  of the ordinal odds ratios.

To illustrate the calculation of the values in Table 1.7:

$$\hat{\theta}_{12} = \frac{28 \times 13}{23 \times 7} = 2.26$$

$$\hat{\theta}'_{12} = \frac{(61+28) \times 13}{(68+23) \times 7} = 1.82$$

$$\hat{\theta}''_{12} = \frac{(61+28) \times (13 \times 12 \times 16)}{(68+23+58+40+53+38) \times 7} = 1.86$$

Table 1.7: Values of Ordinal Odds Ratios for Dumping Severity Data

	$\hat{\theta}_{ij}$		$\hat{\theta}'_{ij}$		$\hat{\theta}''_{ij}$	
	j 1	2	1	2	1	2
i 1	0.74	2.26	0.92	1.82	1.38	1.86
2	2.04	0.53	1.69	0.86	1.74	1.33
3	1.04	1.40	1.14	1.44	1.55	1.53

The value of  $\hat{\theta}_{12}$  means that the estimated odds that dumping is moderate instead of slight is 2.26 times higher for operation B than for A.

The value of  $\hat{\theta}'_{12}$  means that the estimate odds that dumping is moderate instead of none or slight is 1.82 times higher for operation B than A.

The value of  $\hat{\theta}''_{12}$  means that the estimated odds that dumping is moderate instead of none or slight is 1.86 times higher when some stomach is removed (operations B, C, D) than when none is removed (A).

All three sets of measures indicate a generally positive association, though the  $\{\hat{\theta}''_{ij}\}$  show the most consistency.

### 1.7 Estimation

For all the models discussed in this thesis, the parameters and expected cell counts are estimated by the method of Maximum Likelihood (Lindgren, 1976, p.269).

This well-known statistical principle gives parameter estimates with certain known properties (e.g. asymptotic efficiency, consistency, asymptotic normality with known parameters, etc.) as well as giving

rise to powerful likelihood-ratio tests which can be used to test whether specific models fitted are feasible.

### 1.8 Model Testing

To test the goodness-of-fit of the various models, we can use either of the following two statistics:

$$\chi^2 = \sum_i \frac{(n_i - m_i)^2}{m_i}$$

$$G^2 = 2 \sum_i n_i \log \frac{n_i}{m_i}$$

which are asymptotically equivalent. Under the null hypothesis, both  $\chi^2$  and  $G^2$  are asymptotically distributed as chi-square.  $\chi^2$  is the Pearson chi-square statistic (Pearson, 1900), and  $G^2$  is a likelihood-ratio (LR) statistic, known as the "deviance" in the terminology of generalized linear models. Although both tests usually lead to very similar conclusions, we will use  $G^2$  as the LR statistic is much more useful in testing significance of model terms.

### 1.9 Structural and Sampling Zones

Zero entries in contingency tables are of two types - structural and sampling zeroes. Structural (fixed) zeros occur when it is impossible to observe values for certain combinations of the variables, e.g. males who have had a hysterectomy. Sampling (random) zeroes are due to sampling variation and the relatively small size of the sample when compared with the large number of cells; they disappear when the sample size is increased sufficiently.

When structural zeros occur in a table, it is still possible to analyse the data using models which will be discussed in Section 8.2.

When sampling zeroes are scattered haphazardly throughout the table, there are usually no problems - the appropriate models are fitted in the normal manner.

However, sometimes the zero entries are placed in such a way that when computing estimated values to satisfy the constraints, if one zero entry is given a positive value, then another must be given a negative value. If the extra constraint that all estimated values must be non-negative is applied, then these entries will have estimated values of zero. Table 1.9 gives an example of such a case:

Table 1.9: A table with two-dimensional marginal total equal to zero

	Y1	Y2		Y1	Y2
X1	0	5		6	10
X2	0	12		5	8

The  $n_{+11}$  marginal total is zero. Thus, any model which requires this marginal total to be fitted must necessarily estimate the (1,1,1) and (2,1,1) cells as zero.

It is this circumstance which has given risen to much debate about the "correct" degrees of freedom applying to the deviance in such a case.

There are three views stated in the literature. The first and most widely stated view is that in order to test the goodness-of-fit of a model that uses a set of observed marginal totals with at least one zero entry, the degrees of freedom associated with the test statistic must be reduced (Bishop et al, 1975; Fienberg, 1980; Brown and Fuchs, 1983 and 1984; Aston and Wilson, 1984). This means that if an observed marginal entry is zero, then both the observed and estimated entries for all cells included in that total must be zero, and so the fit of the model for those cells is known to be perfect. As a result, the degrees of freedom associated with the fit of the zero cell values must be deleted. The formula for the degrees of freedom is given as

$$df = (n_C - n_Z) - (n_P - n_N)$$

where

$n_C$  = number of cells in the table,

$n_Z$  = number of cells with estimated values equal to zero,

$n_P$  = number of parameters specified in the model,

$n_H$  = number of parameters that cannot be estimated because of zero marginal totals.

However, in a recent paper, Baker et al (1985) have asserted that such a treatment is incorrect. They state that if a zero occurs in a margin that was fixed prior to the experiment, then by definition the cells in the table contributing to that margin are structural zeros (and are weighted out of all analyses of the table). Therefore, by extension, if a table is analysed "conditional on a margin that was not actually fixed in the experiment, then cells in the table that were not structural zeroes in the experiment will become structural zeroes in the analysis if they contribute to a zero cell in the conditioning margin". However, after having dealt with these "structural" zeroes, if any other zero cells remain which contribute to a margin which is not conditioned on, then no adjustment whatsoever is to be made to the degrees of freedom.

The third view is that of Stirling (1986) who asserts that both the previous two methods are incorrect. He states that to obtain the correct degrees of freedom for any model, one should always use the formula

$df$  = difference between the number of estimable parameters for the model in question and for the saturated model.

He explains that if this formula is used, then "it makes no difference whether or not structural zeroes are kept in the data, whether the margins are classified as responses or explanatory variables, or whether log-linear or logistic models are used when there is a single binary response". However, to correctly apply the formula, "we must correctly identify all estimable parameters. This has been incorrectly done by some previous authors".

It can therefore be seen that the literature on methods for sparse contingency tables is still controversial.

### 1.9.1 Monkey Example

An example of a table containing both sampling and structural zeroes is Table 1.8, which is taken in a slightly modified form from Fienberg (1980, p.146). The table gives the distribution of genital display among six squirrel monkeys (labelled R to W). For each display there is an active and passive participant, but a monkey never displays towards himself. Thus the dashes in the table indicate structural zeroes. There are also several sampling zeroes such as in cell (1,6) where there is no a priori reason to suppose that the event is impossible. We will assume that the opportunity was not available to observe monkey T as an active participant.

Table 1.8: Genital display in a colony of squirrel monkeys

Active Participant	Passive Participant					
	R	S	T	U	V	W
R	-	1	5	8	9	0
S	29	-	14	46	4	0
U	2	3	1	-	38	2
V	0	0	0	0	-	1
W	9	25	4	6	13	-

### 1.10 Loglinear Models

Let  $\underline{n}' = (n_1, \dots, n_I)$  and  $\underline{m}' = (m_1, \dots, m_I)$  denote the observed and expected counts for the  $I$  cells in the table. For simplicity, we will use a single index, though the table may be multi-dimensional.

Loglinear models have the form

$$\log m_i = \underset{\sim}{x}'_i \underset{\sim}{\beta}$$

where  $\underline{\beta}$  is a  $p \times 1$  vector of parameters and  $\underline{x}'_i$  is a row vector of known constants, the choice of which depends on what kind of association one wishes to model. In the nomenclature of analysis of variance  $\underline{x}'_i$  is the  $i$ th row of the  $I \times p$  design matrix  $X$ , i.e.

$$\log \underline{m} = X \underline{\beta}$$

Many of the models discussed in this thesis are simply special cases of the more general category of loglinear models. They can be used to model many kinds of association, and so are probably the most common models used in practice for contingency tables.

#### 1.10.1 Fitting Loglinear Models

Loglinear models can be fitted quite easily using either the Newton-Raphson algorithm or the Iterative Proportional Fitting algorithm. These are discussed in Appendix 2.

#### 1.11 Linear Models

Linear models relate the expected cell count to a linear function of parameters. The two common methods of specifying these models are:

- (i) directly, in a form such as  $A \underline{\pi} = X \underline{\beta}$ , or
- (ii) indirectly, in terms of constraints.

They are not as commonly used as loglinear models, as they usually specify fairly unusual kinds of relationships between the variables of a contingency table. Nevertheless, they form a powerful and useful class of models which can be used to test specific hypotheses that could not normally be tested using loglinear models.

##### 1.11.1 Linear models specified as $A \underline{\pi} = X \underline{\beta}$

Consider the cell counts in a contingency table as making up an  $I \times 1$  vector  $\underline{m}$ . The cell probabilities corresponding to these counts make up an  $I \times 1$  vector  $\underline{\pi}$ . The vector  $\underline{\pi}$  may correspond to



- (i) a Poisson or single multinomial distribution, or to
- (ii) a product-multinomial distribution.

In the former case  $\sum \pi_i = 1$ , while in the product-multinomial case the set of  $I$  cells is comprised of several subsets, each of which corresponds to a separate multinomial sample, and the sum of the elements of  $\pi$  over each subset is unity.

If  $\beta$  is a  $K \times 1$  vector of unknown parameters,  $A$  is a known  $J \times I$  matrix with linearly independent rows, and  $X$  is a known  $J \times K$  matrix, with linearly independent columns, with  $I > J > K$ , then we can write the expected cell probabilities in terms of a linear function of the model parameters as

$$A\pi = X\beta.$$

Haber (1985) discusses linear models which are formulated in this way.

#### 1.11.2 Linear models specified in terms of constraints

Suppose we have some hypothesis about the cell counts which can be specified in terms of  $E$  constraints. We can write the constraints as

$$F\pi = 0$$

where  $F$  is an  $E \times I$  matrix with  $E$  linearly independent rows. Further constraints are imposed by the sampling design. These constraints guarantee that the sum of the probabilities within each sample will be equal to one (or equivalently that the sum of the counts within each sample will be equal to the correct marginal total). They can be written

$$D'\pi = \underline{1}_S$$

where  $S$  is the number of samples ( $S > 1$ ) and  $D = \{d_{is}\}$  is the  $I \times S$  matrix defined by

$$d_{is} = \begin{cases} 1 & \text{if cell } i \text{ belongs to sample } s \\ 0 & \text{otherwise} \end{cases}$$

In terms of the cell counts the constraints can be written

$$D' R \underline{m} = \underline{1}_S$$

where  $R$  is the  $I \times I$  diagonal matrix with diagonal elements  $r_{ii} = 1/m_{i+}$ , where  $m_{i+}$  is the marginal total of the  $i$ th sample ( $i=1, \dots, S$ ), and off-diagonal elements zero.

Thus, the constraints on the cell counts can be written as

$$L \underline{m} = \underline{c}$$

where  $L'$  is the  $I \times (E+S)$  matrix  $L' = (F' : R'D)$  and  $\underline{c}'$  is the  $1 \times (E+S)$  vector consisting of  $E$  zeroes and  $S$  ones, i.e.

$$\underline{c}' = (\underline{0}_E' : \underline{1}_S').$$

A vector  $\underline{a}$  which satisfies these constraints is

$$\underline{a} = R^{-1} \underline{a}^*$$

where  $R^{-1}$  is the  $I \times I$  diagonal matrix with diagonal elements  $r_{ii} = m_{i+}$ , and  $\underline{a}^*$  is the  $I \times 1$  vector with  $i$ th element  $a_i = 1/B_i$  where  $B_i$  is the number of cells in the  $i$ th sample.

### 1.11.3 Fitting Linear Models

Linear models specified in terms of constraints can be easily fitted using the algorithm of Wedderburn (1974). Details of this are given in Appendix 3.

Linear models specified as  $A \underline{\pi} = X \underline{\beta}$  can be most easily fitted by reformulating in terms of constraints so that we can then apply Wedderburn's algorithm. Appendix 3 gives further details.

### 1.12 Other Models

As mentioned previously, most models discussed in this thesis are either linear or loglinear, and so can be fitted using the general

algorithms appropriate for these. Where a model does not fall into one of these two classes, details of estimation methods will be given separately.

## CHAPTER II: NOMINAL LOGLINEAR MODELS

### 2.1 Two-Dimensional Tables

Consider a two-way table of counts, with the row variable  $X$  having  $r$  categories, and the column variable  $Y$  having  $c$  categories.

#### 2.1.1 Multiplicative Form of the Loglinear Model

##### 2.1.1.1 Saturated Model

For the case of only two variables, the most general model that specifies an association between them is called the saturated model. The simplest way to express this model is

$$m_{ij} = \alpha_{ij}^{XY} \quad (2.1).$$

From this expression it can be seen that the expected cell counts can be adjusted to give any pattern. In particular, we can easily arrange for the expected cell counts to be equal to the actual counts.

However, a more useful parameterization is

$$M_{ij} = \eta \tau_i^X \tau_j^Y \tau_{ij}^{XY} \quad (2.2)$$

The expected count is modelled as a multiplicative function of parameters that depend on  $X$  and  $Y$ . The meaning of the parameters will be explained shortly.

Although this expression is more complicated than (2.1), it is a more useful parameterization for the other types of models that we will meet later.

There are  $rc$  cells in the table. There cannot be  $r$  independent  $\tau_j^Y$  parameters and  $rc$  independent  $\tau_{ij}^{XY}$  parameters, or there would be  $1+r+c+rc$  parameters for only  $rc$  cells. We will

therefore add some identifiability constraints: a convenient set is to restrict the parameters so that

$$\prod_i \tau_i^X = \prod_j \tau_j^Y = \prod_i \tau_{ij}^{XY} = \prod_j \tau_{ij}^{XY} = 1 \quad (2.3)$$

Then we will have only  $r-1$  independent  $\{\tau_i^X\}$ ,  $c-1$  independent  $\{\tau_j^Y\}$  and  $(r-1)(c-1)$  independent  $\{\tau_{ij}^{XY}\}$ . Thus the number of parameters will be  $1+(r-1)+(c-1)+(r-1)(c-1) = rc$ .

Since the number of parameters equals the number of cells in the table, we can see that the estimated expected counts under this model will be equal to the actual counts, i.e.  $m_{ij} = n_{ij}$ . In fact, the model is called "saturated" because it contains all possible effect parameters.

We can express the parameters in terms of the expected counts in the following way:

$$\begin{aligned} \prod_i \prod_j m_{ij} &= \prod_i \prod_j (\eta \tau_i^X \tau_j^Y \tau_{ij}^{XY}) \\ &= (\eta^{rc})(1)(1)(1) \end{aligned}$$

(because of the identifiability constraints), so

$$\eta = \left( \prod_{i,j} m_{ij} \right)^{1/rc} \quad (2.4)$$

The  $\eta$  parameter is thus the geometric mean of the counts. It is analogous to the intercept term in a regression equation. It is a "baseline" or starting point from which effects are measured. Since

$$\prod_i m_{ij} = \eta^I (\tau_j^Y)^r$$

then

$$\tau_j^Y = \frac{(\prod_i m_{ij})^{1/r}}{\eta}$$

$$= \frac{(\prod_i m_{ij})^{1/r}}{(\prod_{i,j} m_{ij})^{1/rc}} \quad (2.5)$$

which is the geometric mean of the  $j$ th column divided by the geometric mean of the whole table.

Similarly

$$\begin{aligned} \tau_j^X &= \frac{(\prod m_{ij})^{1/c}}{n} \\ &= \frac{(\prod m_{ij})^{1/c}}{(\prod_{i,j} m_{ij})^{1/rc}} \end{aligned} \quad (2.6)$$

Also

$$\begin{aligned} \tau_{ij}^{XY} &= \frac{m_{ij}}{n \tau_i^X \tau_j^Y} \\ &= \frac{m_{ij} (\prod_{i,j} m_{ij})^{1/rc}}{(\prod_i m_{ij})^{1/r} (\prod_j m_{ij})^{1/c}} \end{aligned}$$

The  $\tau$  parameters represent effects which the variables have on the cell frequencies. The magnitude of an effect is measured as a departure from the value of unity - effects of 1.0 have no impact since they leave the product unchanged. If every effect was equal to 1.0, then each cell count would be equal to every other cell count, and all would be equal to  $n$ . If an effect parameter is greater than 1.0, then there will be more than the average number of cases expected in that cell, while if the  $\tau$  parameters are less than 1.0, there will be fewer than the average expected.

Since the estimated expected cell counts are equal to the actual cell counts, the saturated model is not very informative.

### 2.1.1.2 Independence Model

The model of independence is unsaturated (i.e. it does not contain all possible effect parameters), and corresponds to setting  $\tau_{ij}^{XY} = 1.0$  for all  $i, j$ , i.e.

$$m_{ij} = \eta \tau_i^X \tau_j^Y.$$

The expected counts depend only on an effect due to  $X$  ( $\tau_j^X$ ) and an effect due to  $Y$  ( $\tau_i^Y$ ). This is because of the definition of independence:

the probability that a given individual is classified into cell  $(i, j)$  is

$$P(X=i \text{ and } Y=j) = P(X=i) \times P(Y=j)$$

if and only if  $X$  and  $Y$  are independent.

In the saturated model of the previous section, the previous counts also depended on the combination of  $X$  and  $Y$  effects ( $\tau_{ij}^{XY}$ ), so that the effect of level  $i$  of  $X$  could differ, depending on whether it was combined with level  $a$  of  $Y$  or say, level  $b$ . The independence model specifies that  $X$  and  $Y$  act quite independently on the cell counts, so that  $X$  and  $Y$  are not associated.

The independence model has  $1+(r-1) + (c-1) = r+c-1$  independent parameters, so that we cannot always adjust the parameters to make the fitted counts equal any observed pattern of cell counts (since  $r \geq 2$  and  $c \geq 2 \Rightarrow r+c-1 < rc$ ) so, in general the fitted counts will not equal the actual counts.

## 2.1.2 Additive Form of the Loglinear Model

### 2.1.2.1 Saturated Model

Taking natural logarithms of all terms in equation (2.2) yields  $\log m_{ij} = \log \eta + \log \tau_i^X + \log \tau_j^Y + \log \tau_{ij}^{XY}$ , which is usually written as

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad (2.7)$$

where  $\mu = \log \eta$ ,  $\lambda_i^X = \log \tau_i^X$ , etc.

The model assumes that the log of the counts is a linear function of the parameters, hence the parameters cannot be interpreted as directly as in the multiplicative form; this additive form corresponds more closely to ordinary normal linear models and we shall see later that the parameters can be more directly estimated in this form.

Equation (2.4) becomes

$$\begin{aligned} \mu &= \log \eta \\ &= \log \left( \prod_{i,j} m_{ij} \right)^{1/rc} \\ &= \frac{1}{rc} \sum_{i,j} \log m_{ij} \end{aligned}$$

while equation (2.6) becomes

$$\begin{aligned} \lambda_i^X &= \log \tau_i^X \\ &= \log \frac{\prod_j m_{ij}^{1/c}}{\eta} \\ &= \frac{1}{c} \sum_j \log m_{ij} - \mu \\ &= \frac{1}{c} \sum_j \log m_{ij} - \frac{1}{rc} \sum_{i,j} \log m_{ij} \end{aligned}$$

and equation (2.5) becomes

$$\begin{aligned} \lambda_{ij}^{XY} &= \log m_{ij} + \frac{1}{rc} \sum_{i,j} \log m_{ij} - \frac{1}{r} \sum_i \log m_{ij} \\ &\quad - \frac{1}{c} \sum_j \log m_{ij} \end{aligned}$$



The identifiability constraints given by equation (2.3) now become

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = 0$$

The additive form of the logarithms is reminiscent of analysis of variance (ANOVA).

Whereas using the multiplicative form, the magnitude of an effect is measured as a departure from 1.0, using the additive form it is measured as a departure from zero.

#### 2.1.2.2 Independence Model

The additive form of the loglinear model of independence is

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y$$

where, as before,

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0$$

This model specifies that X and Y are independent, as defined in Section 2.1.1.2.

This model can be fitted quite easily as described in Section 1.10 and Appendix 2.

The next section gives an example of fitting the independence model to a two-way table.

##### 2.1.2.2.1 Abortion Attitude Example

Table 2.1 is taken from Knoke and Burke (1980, p.72). It is a 4x2 cross-tabulation of religion by attitude to abortion. The "attitude to abortion" variable reflects agreement or disagreement with a questionnaire item asking

whether a woman should be allowed to have a legal abortion because she was too poor to support more children.

Table 2.1: Cross-tabulation of abortion attitude by religion

Religion	Attitude to Abortion		Total
	Favour	Oppose	
Protestant	460	498	958
Catholic	147	240	387
Jew	41	10	51
Other	65	17	82
TOTAL	713	765	1478

Table 2.2 gives the estimated expected counts under the model of independence.

Table 2.2: Estimated expected counts under independence

Religion	Attitude to Abortion		Total
	Favour	Oppose	
Protestant	462.15	495.85	958
Catholic	186.69	200.31	387
Jew	24.60	26.40	51
Other	39.56	42.44	82
TOTAL	713	765	1478

### 2.1.2.3 Goodness-of-Fit Statistics

As mentioned in Section 1.9, to perform a statistical test of the null hypothesis of independence, we can use either of the following two statistics:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

$$G^2 = 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}$$

which are asymptotically equivalent. Under the null hypothesis both  $\chi^2$  and  $G^2$  are asymptotically distributed as chi-square with degrees of freedom (df) equal to the number of parameters in the saturated model minus the number of parameters in the independence model. The number of parameters in the saturated model is  $rc$ , and the number of parameters in the independence model is  $1+(r-1)+(c-1) = r+c-1$ , so the difference is

$$df = rc - (r+c-1) = (r-1)(c-1)$$

#### 2.1.2.3.1 Abortion Attitude Example

For the data of Table 2.1 the goodness-of-fit statistics are

$$\begin{aligned} \chi^2 &= (460 - 462.15)^2 + \dots + (17 - 42.44)^2 \\ &= 69.06 \end{aligned}$$

$$\begin{aligned} G^2 &= 2(460 \log \frac{460}{462.15} + \dots + 17 \log \frac{17}{42.44}) \\ &= 72.44 \end{aligned}$$

Both these statistics have an asymptotic chi-square distribution with 3 df. Since  $\chi^2 = 11.35$

.99(3)

both  $\chi^2$  and  $G^2$  are significant at the 1% level, leading to rejection of the independence hypothesis. So the

respondent's religion seemed to influence his/her attitude to abortion.

## 2.2 Three-Dimensional Tables

We will now generalize loglinear models to three-dimensional tables. Consider a three-way table, with variables  $X$ ,  $Y$  and  $Z$  having  $r$ ,  $c$  and  $\ell$  categories, respectively. The observed count  $n_{ijk}$  has corresponding expected count  $m_{ijk}$  and population probability  $\pi_{ijk}$ .

The general loglinear model has the form

$$\log m_{ijk} = \mu + \lambda_j^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ} \quad (2.7)$$

with identifiability constraints

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_k \lambda_k^Z = 0$$

$$\sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = \sum_i \lambda_{ik}^{XZ} = \sum_k \lambda_{ij}^{XZ} = \sum_j \lambda_{jk}^{YZ} = \sum_k \lambda_{jk}^{YZ} = 0$$

$$\sum_i \lambda_{ijk}^{XYZ} = \sum_j \lambda_{ijk}^{XYZ} = \sum_k \lambda_{ijk}^{XYZ} = 0$$

### 2.2.1 Association Between Three Variables

For simplicity, assume that all variables are responses and that the multinomial sampling scheme applies (although the MLEs are identical for the Poisson and product-multinomial sampling schemes; and the classification of variables as explanatory of response really only affects the interpretation of the models).

There are five basic kinds of relationships that can occur between three variables:

- (1) All three variables related so that the relationship between any two depends on the value of the third. This is the

saturated model and can be represented by equation (2.7). This model imposes no restrictions on the  $\{m_{ijk}\}$  and so the estimated expected counts will be equal to the actual counts.

- (2) Pairwise relations among X, Y and Z, with each two-variable relation being unaffected by the value of the third variable. The representation of this model can be obtained from equation (2.7) by restricting the three-factor interaction to be zero, giving

$$\begin{aligned}\log m_{ijk} = & \mu + \lambda_j^Y + \lambda_k^Z \\ & + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}\end{aligned}$$

- (3) Conditional independence of X and Y given Z. This can be obtained by setting  $\lambda_{ijk}^{XYZ} = 0$ , and also  $\lambda_{ij}^{XY} = 0$ , to give  $\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$ . This model implies that X and Y are independent for each fixed value of Z. For each fixed value of Z we get independence in the corresponding rxc subtable of X and Y.

There are two other versions of this model, corresponding to setting  $\lambda_{ik}^{YZ} = 0$  (conditional independence of X and Y, given Y) and to  $\lambda_{jk}^{YZ} = 0$  (conditional independence of Y and Z, given X).

- (4) Independence of (i) X, and  
(ii) Y and Z jointly.

This can be obtained by setting

$$\lambda_{ijk}^{XYZ} = \lambda_{ij}^{XY} = \lambda_{ik}^{XZ} = 0, \text{ to give}$$

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$$

This means that Y is independent of X, given Z and Z is independent of X, given Y. Y and Z are conditionally dependent, given X.

The two other versions of this model correspond to the sole conditionally dependent pair being either X and Y or X and Z.

- (5) The final model corresponds to all three variables being independent, i.e.

$$\begin{aligned}\pi_{ijk} &= P(X=i, Y=j, Z=k) \\ &= P(X=i) P(Y=j) P(Z=k)\end{aligned}$$

The loglinear representation of this corresponds to setting  $\lambda_{ij}^{XY} = \lambda_{jk}^{YZ} = \lambda_{ijk}^{XYZ} = 0$

giving

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

### 2.2.2 Hierarchical Models

In a hierarchical model higher-order association terms may be included only if the related lower-order terms are included. For example, if  $\lambda_{ijk}^{XYZ}$  is in the model, then so must be  $\lambda_{ij}^{XY}$ ,  $\lambda_{ik}^{XZ}$  and  $\lambda_{jk}^{YZ}$ . Similarly, if  $\lambda_{ij}^{XY}$  is in the model, then  $\lambda_i^X$  and  $\lambda_j^Y$  must also be included. An example of a non-hierarchical model is

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ijk}^{XYZ}.$$

The main reason for using only hierarchical models is ease of interpretation.

In hierarchical models successively higher association terms measure deviations from lower-order terms. The lower-order terms are said to be "marginal" to the higher-order ones, e.g.

$$\lambda_{ij}^{XY} \text{ is marginal to } \lambda_{ijk}^{XYZ}.$$

A useful notation is the abbreviated parentheses notation. A model can be represented by including in parentheses the superscripts for the highest-order term(s) for each variable. For instance, the notation for the model

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

is (X,Y,Z). The notation for

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$$

is (X,Y,XZ). The notation for

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

is (XYZ). This notation uniquely describes all hierarchical loglinear models.

### 2.2.3 Estimation

The cell counts and parameters can be estimated as previously discussed in Section 1.10 and Appendix 2. Appendix 4 gives further details of maximum likelihood estimation for the case of three variables.

### 2.2.4 Model Testing

As discussed in Section 1.9, the two goodness-of-fit statistics are

$$\chi^2 = \sum_{ijk} \frac{(n_{ijk} - \hat{m}_{ijk})^2}{\hat{m}_{ijk}}$$

and

$$G^2 = 2 \sum_{ijk} n_{ijk} \log \frac{n_{ijk}}{\hat{m}_{ijk}}$$

If the model fitted is correct and N is large,  $\chi^2$  and  $G^2$  have approximate chi-square distributions with

$$\begin{aligned} \text{df} &= \# \text{ parameters in saturated model} \\ &\quad - \# \text{ parameters in model under consideration.} \end{aligned}$$

As an example of calculating degrees of freedom, consider the model (XY, XZ). The parameters in the model are  $\mu, \lambda_i^X, \lambda_j^Y, \lambda_k^Z, \lambda_{ij}^{XY}$  and  $\lambda_{ik}^{XZ}$ , i.e. there are  $1+(r-1) + (c-1) + (l-1) + (r-1)(c-1) + (r-1)(l-1)$  parameters. The number of parameters in the saturated model (XYZ) is  $rc\ell$ . Consequently the degrees of freedom associated with the model (XY, XZ) are

$$\begin{aligned}
 df &= rcg - [1 + (r-1) + (c-1) + (g-1) + (r-1)(c-1) \\
 &\quad + (r-1)(g-1)] \\
 &= r(c-1)(g-1)
 \end{aligned}$$

### 2.2.5 Abortion Attitude Example

Consider again the data of Table 2.1. The "attitude to abortion" was actually surveyed in 1972 and 1978. Table 2.1 contains only the 1972 data; the full table is shown in Table 2.3 (taken from Knoke and Burke, 1980, p.72).

Table 2.3: Attitude to abortion by religion and year

Religion	1972 Attitude		1978 Attitude	
	Favour	Oppose	Favour	Oppose
Protestant	460	498	424	501
Catholic	147	240	151	225
Jew	41	10	23	6
Other	65	17	88	30

If all three variables were explanatory, then there would be nine possible hierarchical loglinear models that could be fitted to this data. For the purposes of illustration, they have all been fitted and both goodness-of-fit statistics for the nine models are given in Table 2.4.



Table 2.4: Goodness-of-fit statistics for various loglinear models

Model	df	$\chi^2$	$G^2$
(R,Y,A)	10	140.85***	143.24***
(R,Y,A)	7	124.88***	130.16***
(R,YA)	9	140.79***	143.02***
(Y,RA)	7	14.82*	14.96*
(RY,RA)	4	1.88	1.89
(RY,YA)	6	124.66***	129.94***
(RA,YA)	6	14.60*	14.74*
(RY,YA,RA)	3	1.46	1.47
(RYA)	0	0	0

\*, \*\*\* Denotes values in upper 5% and 0.1% tail respectively, of corresponding  $\chi^2$  distribution, with df as indicated

It can be seen from Table 2.4 that the values of  $\chi^2$  and  $G^2$  are usually quite close, except when both are large and are far out in the tail of the corresponding  $\chi^2$  distributions. The values of the goodness-of-fit statistics decrease as the number of parameters in the models increases.

As mentioned earlier, it is not sensible to fit all these nine models because there is only one response, attitude to abortion, and religion and year are actually explanatory variables which may affect this response.

Because of this we treat the religion by year margins as fixed (irrespective of whether or not they were actually fixed by the sampling design), and so we only consider models which contain the religion by year association term,  $\lambda_{ijRY}$ . This gives us the choice of five models:

(RY,A), (RY,RA), (RY,YA), (RY,YA,RA) and (RYA).

The model (RYA) allows a possible association between religion and year, as explanatory variables, but specifies that neither of these affected attitude. However, this model does not fit at all well, as evidenced by the huge values of  $G^2$  and  $\chi^2$ .

The model (RY,RA) specifies that religion affected attitude (but year did not). This is an example of model (3) of Section 2.2.1, i.e. attitude is conditionally independent of year, given religion. This model is

$$\log m_{ijk} = \mu + \lambda_i^R + \lambda_j^Y + \lambda_k^A + \lambda_{ij}^{RY} + \lambda_{ik}^{RA}$$

so has  $1+3+1+1+3+3 = 12$  parameters, therefore  $df = 16-12 = 4$ . This model fits well and is the simplest model to do so.

The model (RY,YA) states that there was a change in attitude from 1972 to 1978, which was constant for all four religions. This model fits very badly.

The model (RY,YA,RA) specifies that both year and religion affected attitude. However, the absence of the three-factor interaction implies that the effect of religion was the same in 1972 and 1978, and the effect of year was the same for all religions.

This model fits well, but is more complicated than (RY,RA) which also fitted adequately.

The saturated model (RYA) specifies that the effect of religion in attitude varied, depending on the year. Alternatively, this could be thought of in terms of the effect of year on attitude varying, depending on the religion of the person. The saturated model fits perfectly, but provides no simplification of the data whatsoever.

## 2.2.6 Conditional Test Statistics

Consider the case where there are two loglinear models  $M_1$  and  $M_2$ , with estimated counts  $m_{ijk(1)}$  and  $m_{ijk(2)}$ , where model 1 is a special case of model 2 (e.g.  $(X,Y,Z)$  is the special case of  $(XY,Z)$  where  $\lambda_{ij}^{XY}$  is assumed to be zero). Then the difference in deviances,  $G^2(1) - G^2(2)$ , can be used to test whether the

difference between the expected values for the two models is simply due to random variation, given that the true expected values satisfy model 2. This is a property of the likelihood ratio test. The difference in deviance will be

$$\begin{aligned} G^2(1) - G^2(2) &= 2 \sum_{i,j,k} n_{ijk} \log \frac{n_{ijk}}{\hat{m}_{ijk}(1)} - 2 \sum_{i,j,k} n_{ijk} \log \frac{n_{ijk}}{\hat{m}_{ijk}(2)} \\ &= 2 \sum_{i,j,k} n_{ijk} \log \frac{\hat{m}_{ijk}(2)}{\hat{m}_{ijk}(1)} \end{aligned}$$

This conditional test statistic has as asymptotic chi-square distribution with degrees of freedom equal to the difference in the degrees of freedom for the two models (which will also equal the number of parameters of model 2 that are constrained to be zero in model 1).

#### 2.2.6.1 Partitioning Chi-Square

In a hierarchical set of models it is easy to partition the deviance into several additive parts. Take for example the hierarchical set

	<u>Deviance</u>
M1: $\log m_{ijk} = \mu + \lambda X_i + \lambda Y_j + \lambda Z_k$	$G^2(1)$
M2: $\log m_{ijk} = \mu + \lambda X_i + \lambda Y_j + \lambda Z_k + \lambda XY_{ij}$	$G^2(2)$
M3: $\log m_{ijk} = \mu + \lambda X_i + \lambda Y_j + \lambda Z_k + \lambda XY_{ij} + \lambda XZ_{ik}$	$G^3(3)$
M4: $\log m_{ijk} = \mu + \lambda X_i + \lambda Y_j + \lambda Z_k + \lambda XY_{ij} + \lambda XZ_{ik} + \lambda YZ_{jk}$	$G^2(4)$

#### 2.2.6.2 Abortion Attitude Example

We have already seen that the model (RY, RA) provides quite a good fit to the data of Table 2.3. So does the more complicated model (RY, YA, RA) which specifies pairwise associations between

all the variables. The model (RY, RA) assumes that the association term  $\lambda_{jk} = 0$ . To test this assumption we can compare the deviance of model (RY, RA) with that of (RY, YA, RA) which has no restrictions on the  $\lambda_{jk}$  term. The conditional test statistic is  $G^2(\text{RY, RA}) - G^2(\text{RY, YA, RA}) = 1.89 - 1.47 = 0.42$ , which is asymptotically distributed as chi-square with  $4-3 = 1$  df. Thus we have no evidence to reject the null hypothesis that  $\lambda_{jk} = 0$ .

The difference in deviances between model M3 and M4,  $G^2(3) - G^2(4)$ , can be attributed to the  $\lambda_{jk}$  term. This difference is the appropriate test statistic for testing whether  $\lambda_{jk} = 0$ . In a similar manner the  $\lambda_{ik}$  term can be tested by comparing  $G^2(2) - G^2(3)$  to a  $\chi^2$  distribution with  $df = df(2) - df(3)$ . This process will hold true for the whole set because by the very nature of any hierarchical set  $G^2(1) > G^2(2) > G^2(3) > G^2(4)$ .

However, we cannot partition the Pearson goodness-of-fit statistic,  $\chi^2$ , in this manner, because this relationship does not necessarily hold.

### 2.3 Higher-Order Contingency Tables

The models and methods discussed so far can be extended to higher-order contingency tables in a relatively straightforward manner. For example, the general loglinear model for a four-dimensional cross-classification of variables, A, B, C and D would be:

$$\begin{aligned} \log m_{ijkl} = & \mu + \lambda A_i + \lambda B_j + \lambda C_k + \lambda D_\ell + \lambda AB_{ij} \\ & + \lambda AC_{ik} + \lambda AD_{i\ell} + \lambda BC_{jk} + \lambda BD_{j\ell} + \lambda CD_{k\ell} \\ & + \lambda ABC_{ijk} + \lambda ABD_{ij\ell} + \lambda ACD_{ik\ell} + \lambda BCD_{jk\ell} \\ & + \lambda ABCD_{ijkl} \end{aligned}$$

with appropriate identifiability constraints. Models specifying different types of conditional independence can be obtained by setting the appropriate  $\lambda$ -terms to zero.

## 2.4 Other Loglinear Models

There are other non-standard types of loglinear models that have applications to different kinds of data. These will be described in later chapters where they will be contrasted with non-loglinear models for similar kinds of data. For example, Chapter IV deals with loglinear models for ordinal variables, while the models of symmetry and quasi-symmetry are discussed in Chapter VIII.

### CHAPTER III: SOME GENERAL ISSUES

In this chapter we discuss some general issues that affect most of the models discussed in this thesis. These issues can be more easily explained by illustrating with particular examples, so we will use the loglinear models of the previous chapter for illustration.

#### 3.1 Model Selection

Complicated models involving many parameters always fit a set of data more closely than a simpler model that is just a special case of the complicated one. For instance, the deviance of the loglinear model  $(XY,Z)$  must be less than or equal to that of the simpler model  $(X,Y,Z)$ . On the other hand, a simpler model may be more parsimonious and easier to interpret, i.e. we want a model that smooths rather than overfits the data. Thus, there is a balance between goodness-of-fit and simplicity. Just as there are many ways of choosing an appropriate regression equation, there is no general best method of model selection for a contingency table, and the "best" model may well differ between one data analyst and another.

Generally, to balance simplicity and goodness-of-fit, the simplest model possible is accepted as long as its goodness-of-fit statistic,  $G^2$  is not significant, and the more complicated parameters which could be introduced into the model are not significant. In the case of loglinear models, these parameters are the next higher order  $\lambda$ -terms.

##### 3.1.1 Abortion Attitude Example

Table 3.1 gives the five loglinear models that we can consider fitting to the data of Table 2.3:

Table 3.1: Loglinear models

Model	df	G <sup>2</sup>
(RY,A)	7	130.16 ***
(RY,RA)	4	1.89 ns
(RY,YA)	6	129.49 ***
(RY,YA, RA)	3	1.47 ns
(RYA)	0	0

The simplest model is (RY,A); however, when the deviance is compared with an  $\chi^2$  distribution with 7df, it is significant at the 0.1% level, indicating that the model does not provide an adequate explanation of the data. The model (RY,YA) also exhibits significant lack of fit.

The simplest model that fits well is (RY,RA). The terms set to zero in this model are  $\lambda_{jk}$  and  $\lambda_{ijk}$ . In Section 2.2.6.2 we concluded that  $\lambda_{jk}$  was not significantly different from zero. The  $\lambda_{ijk}$  term can be tested by comparing the deviance for (RY,YA,RA) with an  $\chi^2_3$  distribution - again this is not significant.

So the model (RY,RA) seems the best choice as it fits adequately and the terms it sets to zero are non-significant.

### 3.1.2 Residual Analysis

No analysis would be complete without a study of the standardized residuals

$$r_i = \frac{n_i - \hat{m}_i}{\sqrt{\hat{m}_i}}$$

The squared standardized residuals are components of the Pearson chi-square statistic, with  $\sum r_i^2 = \chi^2$ . If the null hypothesis

holds, these residuals are asymptotically normal with mean 0, and the average variance of the residuals equals  $df/\text{number of cells}$ .

### 3.2 Interpretation

After deciding on the best fitting model, it is necessary to interpret it. Often the parameter estimates will shed light on specific hypotheses of interest. The concepts of odds and odds ratios were introduced in Section 1.6.2 and often these provide a very natural way of interpreting a model.

#### 3.2.1 Abortion Attitude Example

Table 3.2 gives the estimated expected counts under the model (RY,RA) fitted to the data of Table 2.3:

Table 3.2: Fitted counts under the model (RY,RA)

Religion	1972 Attitude		1978 Attitude	
	Favour	Oppose	Favour	Oppose
Protestant	449.75	508.25	434.25	490.75
Catholic	151.15	235.85	146.85	229.15
Jew	40.80	10.20	23.20	5.80
Other	62.73	19.27	90.27	27.73

This model can be readily interpreted in terms of odds.

The estimated expected odds for opposing abortion rather than favouring it is  $\hat{m}_{ij2}/\hat{m}_{ij1}$  for religion  $i$  and year  $j$ . For Protestants the estimated expected odds is  $508.25/449.75 = 490.75/434.25 = 1.13$  in both 1972 and 1978. The corresponding odds for the other religions are



Catholic : 1.56  
 Jew : 0.25  
 Other : 0.31

Thus, the model specifies that Protestants and Catholics were more likely to oppose abortion, and Jews and "Others" were more likely to favour abortion, for both the 1972 and 1978 questionnaires.

### 3.3 Collapsing Tables

Table 3.3 is the two-dimensional marginal table of counts  $\{m_{i(+)}k\}$  obtained by collapsing Table 3.2 over year:

Table 3.3: Marginal table of religion by attitude to abortion

Religion	Attitude to Abortion (1972 and 1978)		Odds for Oppose/Favour
	Favour	Oppose	
Protestant	884	999	1.13
Catholic	298	465	1.56
Jew	64	16	0.25
Other	153	47	0.31

The fitted margins  $\{\hat{m}_{i(+)}k\}$  are equal to the actual margins  $\{n_{i(+)}k\}$  because the model (RY,RA) includes the  $\lambda_{ik}$  term. The estimated expected odds that a respondent of a particular religion will oppose rather than favour abortion are:

Protestant : 1.13  
 Catholic : 1.56  
 Jew : 0.25  
 "Other" : 0.31

These are exactly the same values for the odds that were calculated separately for both 1972 and 1978 from the fitted counts of Table 3.2. However, this result will not always hold true; the general rule can be stated as follows. In a three-dimensional table the association between two variables  $X$  and  $Y$  (as given by  $\lambda^{XY}_{ij}$ ) may be measured from the table of sums obtained by collapsing over the third variable  $Z$  if  $Z$  is independent of either  $X$  or  $Y$  or both (i.e.  $\lambda^{XZ}_{jk} = 0$  and/or  $\lambda^{YZ}_{jk} = 0$ ).

Bishop, Fienberg and Holland (1975) incorrectly implied that the converse result is also true. However, Whittemore (1978) gives a counter-example.

Thus, we can see that for the model  $(RY, RA)$  we can collapse over year, because  $\lambda^{YA}_{jk} = 0$ . However, we would not be allowed to collapse over either religion or attitude to measure the  $\lambda^{RY}_{jk}$  term or the  $\lambda^{YA}_{jk}$  term respectively.

So, in general, it can be very misleading to simply examine the three two-dimensional marginal tables resulting from a full three-dimensional cross-classification without first checking whether this is allowed (by testing whether the appropriate  $\lambda$ -terms are zero). A clear illustration of this point is given in the following example.

### 3.3.1 University Admissions Example

Table 3.4 is taken in a slightly modified form from Freedman, Pisani and Purves (1978). The data relate to a study of possible sex bias in graduate admissions to a university. The numbers of admissions of both males and females to four departments are given.

Table 3.4: Admission of males and females to four university department courses

Department	Males Admission		Females Admission	
	No	Yes	No	Yes
A	313	512	19	89
B	207	353	8	17
D	279	138	244	131
F	351	22	317	24

Table 3.5 gives the odds of being admitted, for each sex and department:

Table 3.5: Odds for admission for sex by department

Department	Odds for Admission	
	Males	Females
A	1.64	4.68
B	1.71	2.13
D	.50	.54
F	.06	.08

It can be seen that in each department females had a higher chance of being admitted than males. However, if we simply look at the sex by admission marginal table in Table 3.6, we would be led to the false conclusion that females had less chance of being admitted than males.

Table 3.6: Sex by admission marginal table

Sex	Admission		Odds for Admission
	No	Yes	
Male	1150	1025	.89
Female	588	261	.44

Table 3.7 gives the models that could be fitted to the three-way table:

Table 3.7: Various loglinear models fitted to admissions data

Model	df	Deviance
(D,S,A)	10	1435 ***
(SA,D)	9	1366 ***
(DA,S)	7	651.1 ***
(DS,A)	7	803.9 ***
(DA,SA)	6	582.4 ***
(DA,DS)	4	20.0 ***
(SA,DS)	6	735.3 ***
(DA,SA,DS)	3	11.6 **
(DSA)	0	0

It can be seen that none of the unsaturated models fit the data adequately, and consequently there are no  $\lambda$ -terms that can be assumed to be zero. Because of this we can tell immediately that it would be misleading to look at the data collapsed over any of the variables.

## CHAPTER IV: LOGLINEAR MODELS FOR ORDINAL VARIABLES

### 4.1 Disadvantages of Ignoring Ordinal Nature of Variables

The concept of ordinal categorical variables has already been introduced in Section 1.6.

The ordinary loglinear models of Chapter II treat all variables as nominal, in the sense that parameter estimates and chi-square statistics are invariant to orderings of categories. Thus, these models fail to use all the available information when at least one of the variables is ordinal.

#### 4.1.1 Dumping Severity Example

To illustrate this, let us fit the loglinear model of independence to the dumping severity data of Table 1.6. Fitting the model  $\log m_{ij} = \mu + \lambda_i^O + \lambda_j^D$  gives a deviance (denoted by  $G^2(I)$ ) of 10.88, with 6 df. Since  $\chi^2_{(6).90} = 10.64$  and  $\chi^2_{(6).95} = 12.59$ , there is slight evidence of an association between operation and the severity of the side effect, but we could not reject the model of independence at the 5% level.

The fitted counts under this model are given in Table 4.1:

Table 4.1: Fitted counts under independence dumping severity

Operation	Dumping Severity		
	None	Slight	Moderate
A	55.25	29.70	11.05
B	59.86	32.17	11.97
C	63.31	34.03	12.66
D	61.58	33.10	12.32

#### 4.1.1.1 Odds Ratio

As noted in Section 1.6.2 the model of independence constrains all local odds ratios to be equal to 1.0. This can be illustrated by calculating the six local odds ratios as follows:

$$\hat{\theta}_{11} = \frac{55.25 \times 32.17}{59.86 \times 29.70} = 1.0 \quad \hat{\theta}_{12} = \frac{29.70 \times 11.97}{32.17 \times 11.05} = 1.0$$

$$\hat{\theta}_{21} = \frac{59.86 \times 34.03}{63.31 \times 32.17} = 1.0 \quad \hat{\theta}_{22} = \frac{32.17 \times 12.66}{34.03 \times 11.97} = 1.0$$

$$\hat{\theta}_{31} = \frac{63.31 \times 33.10}{61.58 \times 34.03} = 1.0 \quad \hat{\theta}_{32} = \frac{34.03 \times 12.32}{33.10 \times 12.66} = 1.0$$

#### 4.1.1.2 Residual Analysis

Table 4.2 contains the standardized residuals from the independence model:

Table 4.2: Standardized residuals

Operation	Dumping Severity		
	None	Slight	Moderate
A	0.77	-0.31	-1.22
B	1.05	-1.62	0.30
C	0.67	1.02	-0.19
D	-1.09	0.85	1.05

As discussed in Section 3.1.2, if the model fitted is correct, then these residuals should be normally distributed. However, far from looking "random" there seems to be some sort of trend present in these residuals. Along row 1 (i.e. operation A) the

residuals are scattered about a line with negative slope. Along row 2 the slope is less negative, along row 3 the slope is positive, and along row 4 the slope is even more positive. The pattern to these residuals suggests that for operations with a small amount of stomach removal, there are more people than would be expected with the lesser side-effects, whereas for operations with a large amount of removal, there are more than expected with more severe side-effects.

Thus, since there are systematic departures from independence, the model of independence does not appear to fit as well as might have been thought at first sight.

Since operation and dumping severity have a natural ordering, we might want to ask questions such as "does dumping severity tend to increase when more of the stomach is removed?". To help answer this type of question, we will look at some loglinear models which take into account the ordinality of the variables.

## 4.2 Ordinal-Ordinal Tables

Suppose that both the row and column variables,  $X$  and  $Y$ , of a two-dimensional table are ordinal. A very simple type of association would result from all the local odds ratios being the same fixed constant, say  $\theta_{ij} = \alpha = \exp(\beta)$ . Equivalently the log odds ratio would be  $\log \theta_{ij} = \beta$ . We would call this model the constant local odds ratio model, or the uniform association model.

### 4.2.1 Dumping Severity Example

Table 4.3 gives the estimated expected counts which result from fitting this model to the data of Table 1.6.

The estimated value of the local odds ratio is 1.18, which can be obtained from the fitted counts, e.g.

$$1.18 = \frac{62.51 \times 30.94}{62.84 \times 26.15} = \dots = \frac{35.32 \times 16.72}{36.59 \times 13.72}$$

This means that the odds of having a slight side-effect instead of none (or moderate instead of slight) are 1.18 times higher for operation B instead of A (or C instead of B, or D instead of C).

To illustrate the meaning of the odds ratio, Table 4.3 also shows the estimated odds:

Table 4.3: Estimated counts and odds under uniform association

Operation	Dumping Severity			Odds Slight vs None	Odds Moderate vs Slight
	None	Slight	Moderate		
A	62.51	26.15	7.34	$\frac{26.15}{62.51} = 0.418$	0.281
B	62.84	30.94	10.22	$\frac{30.94}{62.84} = 0.492$	0.330
C	60.97	35.32	13.72	$\frac{35.32}{60.97} = 0.579$	0.388
D	53.69	36.59	16.72	$\frac{36.59}{53.69} = 0.682$	0.457

Consider the odds of slight dumping vs none: the odds for operation B of 0.492 are 1.18 times higher than the odds for C of 0.418. Similarly, the odds for C are 1.18 times higher than those for B, and the odds for D are 1.18 times higher than those for C.

Exactly the same relationship holds for the odds of moderate vs slight dumping.



#### 4.2.2 Linear by Linear Association Model

The uniform association model is a special case of the more general linear by linear association model which can be represented as

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta(u_i - \bar{u})(v_j - \bar{v}) \quad (4.1)$$

where  $\sum \lambda_i^X = \sum \lambda_j^Y = 0$ .

The  $\{u_i\}$  and  $\{v_j\}$  are known scores which we have assigned to the rows and columns respectively, where  $u_1 < u_2 < \dots < u_r$  and  $v_1 < v_2 < \dots < v_c$ . Equally spaced scores result in the simplest interpretation, and in practice the integer scores  $\{u_i = i\}$  and  $\{v_j = j\}$  are the most commonly used.

This model has only one more parameter ( $\beta$ ) than the independence model, so it has  $df = (r-1)(c-1)-1 = rc-r-c$  for testing goodness-of-fit. So this model is unsaturated as long as one of the variables has more than two categories. Unlike the general model for association between  $X$  and  $Y$ ,

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad (2.7)$$

model (4.1) does not require additional association parameters as the number of categories of  $X$  or  $Y$  increases. Model (4.1) is the special case of model (2.7) in which the general association term  $\lambda_{ij}^{XY}$  takes the structured form  $\beta(u_i - \bar{u})(v_j - \bar{v})$ .

The  $\beta$  parameter describes the association between  $X$  and  $Y$ ; the independence model is the special case in which  $\beta = 0$ .

The association term  $\beta(u_i - \bar{u})(v_j - \bar{v})$  reflects a deviation of  $\log m_{ij}$  from the independence model. If  $\beta > 0$  more observations are expected to have (large  $X$ , large  $Y$ ) values or (small  $X$ , small  $Y$ ) values than if  $X$  and  $Y$  are independent. On the other hand, if  $\beta < 0$ , more observations are expected to have (large  $X$ , small  $Y$ ) or (small  $X$ , large  $Y$ ) values than under the independence model. In either case the deviation from independence increases in the directions of the four corner cells of the table.

For an arbitrary pair of rows  $a < b$  and an arbitrary pair of columns  $c < d$ , the log odds ratio formed from the cells  $(ac)$ ,  $(ad)$ ,  $(bc)$  and  $(bd)$  is:

$$\log \frac{m_{ac} m_{bd}}{m_{ad} m_{bc}} = \beta (u_b - u_a)(v_d - v_c)$$

Whenever the row scores are one unit apart and the column scores are one unit apart, the log odds ratio equals  $\beta$ .

Goodman (1979) defined models in terms of the  $(r-1)(c-1)$  local odds ratios

$$\theta_{ij} = \frac{m_{ij} m_{i+1,j+1}}{m_{i,j+1} m_{i+1,j}} \quad \begin{array}{l} 1 \leq i \leq r-1 \\ 1 \leq j \leq c-1 \end{array}$$

defined for adjacent rows and adjacent columns. The uniform association model discussed in the previous section is the special case of (4.1) in which all the  $\theta_{ij}$ 's are equal. In such a model the  $\{u_i\}$  and  $\{v_j\}$  are equal-interval scores, i.e.  $u_2 - u_1 = \dots = u_r - u_{r-1}$  and  $v_2 - v_1 = \dots = v_c - v_{c-1}$ . The use of equal-interval scores assumes that the categories are "equally spaced" in some sense.

The use of integer scores  $\{u_i = i\}$  and  $\{v_j = j\}$  results in all  $\theta_{ij} = \exp(\beta)$  and all  $\log \theta_{ij} = \beta$ , so  $\beta$  can be interpreted simply as the common value of the local log odds ratio.

#### 4.2.3 Estimation

There is no closed-form expression for the maximum likelihood estimates  $\{\hat{m}_{ij}\}$  of the  $\{m_{ij}\}$  in model (4.1). Under the usual sampling assumptions, the estimates satisfy the likelihood equations

$$\begin{aligned} \hat{m}_{i+} &= n_{i+} \quad , \quad i = 1, \dots, r \\ \hat{m}_{+j} &= n_{+j} \quad , \quad j = 1, \dots, c \\ \sum_{ij} \hat{m}_{ij} u_i v_j &= \sum_{ij} n_{ij} u_i v_j \end{aligned}$$

However, the model is no more difficult to fit than those described in Chapter II. We can simply use the methods described in Section 1.10 and Appendix 2.

#### 4.2.4 Dumping Severity Example

Table 4.4 gives a comparison of the observed counts and the fitted counts under the model of independence and also under the uniform

Table 4.4: Observed and fitted counts under independence and uniform association

Operation	Dumping Severity		
	None	Slight	Moderate
A	61	28	7
	55.3	29.7	11.0
	62.5	26.2	7.3
B	68	23	13
	59.9	32.2	12.0
	62.8	30.9	10.2
C	58	40	12
	63.3	34.0	12.7
	61.0	35.3	13.7
D	53	38	16
	61.6	33.1	12.3
	53.7	36.6	16.7

<sup>a</sup> observed counts

<sup>b</sup> fitted counts under independence

<sup>c</sup> fitted counts under uniform association

association (UA) model. The latter model fits much better than the independence model in the corners, where it predicts the greatest departures from independence.

When integer scores are used, the estimate of the association parameter is  $\hat{\beta} = 0.163$ , with standard error 0.065. The positive

value of  $\hat{\beta}$  means that the side-effect tended to be more severe with greater removal of stomach. As previously noted, the estimated uniform odds ratio for adjacent rows and adjacent columns is  $\hat{\theta}_{ij} = \exp(\hat{\beta}) = 1.18$ .

#### 4.2.5 Conditional Test of Independence

The deviance for the UA model is  $G^2 = 4.59$ , with 5 df so it fits well. The significance of the association between dumping severity and operation can be assessed by testing  $H_0: \beta = 0$ . This gives a conditional test of independence, under the assumption that the uniform association model holds. Table 4.5 gives an analysis of deviance table for these nested models.

Table 4.5: Analysis of deviance

Model	$\log m_{ij} =$	$\hat{\theta}_{ij} =$	df	Deviance
Independence	$\mu + \lambda_i^O + \lambda_j^D$	1	6	10.88
Uniform Association	$\mu + \lambda_i^O + \lambda_j^D + \beta(u_i - \bar{u})(v_j - \bar{v})$	$\alpha = e^\beta$	5	4.59
Saturated	$\mu + \lambda_i^O + \lambda_j^D + \lambda_{ij}^{OD}$	$\alpha_i \beta_j = \frac{n_{ij} n_{i+1,j+1}}{n_{i+1,j} n_{i,j+1}}$	0	0

The test statistic is the reduction in deviance denoted by  $G^2(I|U)$ , obtained by adding the association parameter  $\beta$  to the independence model. For this data  $G^2(I|U) = G^2(I) - G^2(U)$   
 $= 10.88 - 4.59 = 6.29$ ,

with  $6 - 5 = 1$  df, which is significant at the 5% level. This is much stronger evidence of association than was obtained with the  $G^2(I)$  statistic.

If the uniform association model holds, the ordinal test based on  $G^2(I|U)$  is asymptotically more powerful than the test based on  $G^2(I)$  for detecting departures from independence. The  $G^2(I)$  statistic cannot use as efficiently a constant pattern in local odds ratios, since it treats the variables as nominal. The  $G^2(I)$  statistic tests for any way in which two variables could be associated, whereas the  $G^2(I|U)$  statistic only tests for a specific type of association, and thus is more powerful at detecting this association. (This is analogous to the ANOVA situation where an overall F test may not pick up a difference in means, but a linear trend is significant.)

However, if in the population some local log odds ratios are positive and some are negative, then the uniform association model may fit poorly, and the parameter that  $\beta$  estimates may equal zero. Thus independence implies that  $\beta = 0$ , but  $\beta = 0$  does not imply independence if the uniform association model does not hold. Since the  $G^2(I)$  statistic is designed to detect any departure from independence, it is better than the  $G^2(I|U)$  statistic at detecting nonmonotonic dependencies for which  $\beta$  is zero or close to zero. The alternative hypothesis in the ordinal test ( $\beta=0$ ) is narrower than the broad alternative of "dependence", so the ordinal test sacrifices power for detecting dependencies that are not well summarized by the uniform association model. However, the uniform association model is still a very useful model in many situations.

#### 4.3 Ordinal-Nominal Tables

Sometimes only one of the variables in a two-way table is ordered. Even if both variables are ordered, we may want to use only the ordinal nature of the response, for instance, if we do not want to specify the "distances" between the ordinal categories of the explanatory variable.

If the row variable,  $X$ , is nominal and the column variable,  $Y$ , is ordinal, we can assign scores  $v_1 < v_2 < \dots < v_C$  (usually integer scores) to the columns and form the row effects model

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \tau_i (v_j - \bar{v}) \quad (4.2)$$

$$\text{where } \sum \lambda_i^X = \sum \lambda_j^Y = \sum \tau_i = 0$$

the  $\{\tau_{ij}\}$  parameters,  $r-1$  of which are linearly independent. Thus the model has

$$\begin{aligned} df &= rc - [1+(r-1)+(c-1)+(r-1)] \\ &= (r-1)(c-2) \end{aligned}$$

and is unsaturated when there are more than two columns.

The independence model is the special case of this model with all  $\tau_i = 0$ . The association term  $\tau_i(v_j - \bar{v})$  reflects the deviation of  $\log m_{ij}$  from the independence model. The  $\{\tau_i\}$  are called the row effects. If a particular row effect  $\tau_i$  is positive, then in row  $i$  the probability of the  $Y$  variable being higher than  $V$  is higher than would be expected if the variables were independent. If  $\tau_i < 0$ , observations in row  $i$  are more likely (compared to the independence case) to fall at the low end of the scale on  $Y$ .

For an arbitrary pair of rows  $a$  and  $b$ , and columns  $j$  and  $j+1$ , the log odds ratio is

$$\frac{\log m_{aj} m_{bj+1}}{m_{bj} m_{aj+1}} = (\tau_b - \tau_a)(v_{j+1} - \bar{v}_j)$$

so in a sense the  $\tau_i$  map the categories onto a linear 1-dimensional scale where the distance between  $\tau_a$  and  $\tau_b$  is directly related to how "similar" the categories are (i.e. how close to 1.0 the odds ratios involving these categories are).

For integer scores the log odds ratio is constant and equals  $\tau_b - \tau_a$  for all pairs of adjacent columns.

The row effects model has the same form and produces the same  $G^2$  value if the rows are permuted.

It can also be applied to ordinal-ordinal tables. This might be done if the departure of  $\log m_{ij}$  from the independence model is not linear across the rows, as is necessary for the linear-by-linear association model is the special case of the row effects model in which  $\tau_i = \beta(u_i - \bar{u})$ .

### 4.3.1 Estimation

The ML estimates of expected counts for the row effects model satisfy the likelihood equations

$$\begin{aligned}\hat{m}_{i+} &= n_{i+}, \quad 1 \leq i \leq r \\ \hat{m}_{+j} &= n_{+j}, \quad 1 \leq j \leq c \\ \sum \hat{m}_{ij} v_j &= \sum n_{ij} v_j, \quad 1 \leq i \leq r\end{aligned}$$

There is no closed form expression for these estimates; however, we can simply use the general algorithm of Section 1.10.

### 4.3.2 Dumping Severity Example

For the dumping severity data of Table 1.6, the fitted values under the row effects model with integer scores are given in Table 4.6:

Table 4.6: Fitted counts under row effects model

Operation	Dumping Severity		
	None	Slight	Moderate
A	61.7	26.6	7.7
B	64.4	30.1	9.4
C	60.2	35.7	14.2
D	53.7	36.6	16.7

The estimates of the row effects parameters are:

$$\begin{aligned}\text{Operation A : } \hat{\tau}_1 &= -0.21 \\ \text{B : } \hat{\tau}_2 &= -0.13 \\ \text{C : } \hat{\tau}_3 &= 0.10 \\ \text{D : } \hat{\tau}_4 &= 0.24\end{aligned}$$

This model predicts constant odds ratios for adjacent columns of dumping, so the difference between the row effect parameters of row 1 and row 2

$$\begin{aligned}\hat{\tau}_2 - \hat{\tau}_1 &= -0.13 - (-0.21) \\ &= 0.08\end{aligned}$$

means that the odds of being classified slight instead of none, or moderate instead of slight are  $\exp(0.08) = 1.08$  times higher for operation B instead of A.

Corresponding odds ratios for the other operations are

$$\begin{aligned}\text{C vs B} &: \exp(0.10 - (0.13)) = \exp(0.23) = 1.26 \\ \text{D vs C} &: \exp(0.24 - (0.10)) = \exp(0.14) = 1.15\end{aligned}$$

We can also obtain these odds ratios from the fitted values:

$$\begin{aligned}\text{B vs A, slight vs none} &: \hat{\theta}_{11} = \frac{61.7 \times 30.1}{64.4 \times 26.6} \\ &= 1.08\end{aligned}$$

$$\begin{aligned}\text{B vs A, moderate vs slight} &: \hat{\theta}_{12} = \frac{26.6 \times 9.4}{30.1 \times 7.7} \\ &= 1.08\end{aligned}$$

$$\begin{aligned}\text{C vs B, slight vs none} &: \hat{\theta}_{21} = \frac{64.4 \times 35.7}{60.2 \times 30.1} \\ &= 1.26\end{aligned}$$

Similarly for any other odds ratios.

If the variable "operation" was actually strictly nominal, we would not be so interested in simply comparing row  $i$  to row  $i+1$ . However, since it is actually ordinal, this comparison is very relevant. Fitting the row effects model allows the odds ratios to be different for every pair of adjacent rows, rather than constraining them to be identical, as in the uniform association



model (which gave an estimated common odds ratio of 1.18). The row effect scores of  $(-0.21, -0.13, 0.10, 0.24)$  seem to indicate that the "distance" (in terms of operation severity) between operations B and C is twice as great as that between A and B, or C and D. This contrasts with using integer scores in the uniform association model, which assumes that all adjacent operations are of equal difference in severity.

#### 4.3.3 Conditional Test of Independence

The deviance for this model is  $G^2 = 4.40$ , with 3 df, so there is no evidence of lack-of-fit.

The significance of the association between the two variables in the row effects model can be assessed by testing

$$H_0 : \tau_1 = \dots = \tau_r = 0$$

If the row effects model holds, this homogeneity of the row effects corresponds to independence. A conditional test of independence can therefore be based on

$$G^2(I|R) = G^2(I) - G^2(R)$$

with  $df = (r-1)(c-1) - (r-1)(c-2) = r-1$ .

Given that the row effects model holds, this test is asymptotically more powerful at detecting an association than the  $G^2(I)$  test, since it concentrates the noncentrality on fewer degrees of freedom. The  $G^2(I)$  test ignores the ordinal nature of  $Y$ , whereas the  $G^2(I|R)$  test focuses on alternatives where the ordinal scaling is utilized through a linear departure of  $\log m_{ij}$  from independence. These alternatives are narrower and usually of greater interest than the general alternative to the null hypothesis of independence.

For the dumping severity data testing

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4$$

$$\begin{aligned}
 \text{gives } G^2(I|R) &= G^2(I) - G^2(R) \\
 &= 10.88 - 4.40 \\
 &= 6.48
 \end{aligned}$$

with  $df = 6-3 = 3$ , which is only significant at the 10% level. Compare this with the  $G^2(I|U)$  statistic of 6.29 with 1 df obtained in Section 4.2.5. Since the X variable in this example (operation) is actually ordinal, and the uniform association model seems to fit better than the row effects model, it is not surprising to note stronger evidence of association is provided by the  $G^2(I|U)$  statistic than by  $G^2(I|R)$ . As before, given that the uniform association model holds, we would expect the  $G^2(I|U)$  test to be asymptotically more powerful at detecting an association than the  $G^2(I|R)$  test, since it concentrates the noncentrality on fewer degrees of freedom.

Since operation is actually ordinal, and the uniform association model is just a special case of the row effects model (in which  $\tau_i = \beta(u_i - \bar{u})$ ), we can see whether we are losing much information by using the uniform association model (which, in a sense "averages" the odds ratios) rather than the row effects model, by comparing the deviances of the two models.

$$\begin{aligned}
 \text{In this case } G^2(U|R) &= G^2(U) - G^2(R) \\
 &= 4.59 - 4.40 \\
 &= 0.19
 \end{aligned}$$

with  $5-3 = 2$  df, which is not significant. So the improved fit due to the two extra parameters in the row effects model is not significant.

The analysis of deviance table for this data is given in Table 4.7.

Table 4.7: Analysis of deviance

Model	$\log m_{ij} = \mu + \lambda_i^O + \lambda_j^D +$	$\sigma_{ij} =$	df	Deviance
Independence	0	1	6	10.88
Uniform association	$\beta(u_i - \bar{u})(v_j - \bar{v})$	$\alpha = \exp(\beta)$	5	4.59
Row effects	$\tau_i(v_j - \bar{v})$	$\alpha_i = \exp(\tau_{i+1} - \tau_i)$	3	4.00
Saturated	$\lambda_{ij}^{OD}$	$\alpha_i \beta_j = \frac{n_{ij} n_{i+1,j+1}}{n_{i+1,j} n_{i,j+1}}$		

#### 4.4 Higher Dimensions

The models in Sections 4.2 and 4.3 can be readily generalized to higher order tables having at least one ordinal variable. We will illustrate for the  $r \times c \times l$  cross-classification of three variables  $X$ ,  $Y$  and  $Z$  having expected frequencies  $\{m_{ijk}\}$ .

The hierarchical models of interest for three dimensions range from the simple mutual independence model

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

to the model that contains all the partial association terms but no three-factor interaction term,

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

The next most complex model beyond this includes the term  $\lambda_{ijk}$  and is of little interest because it is saturated. However, if one or more of the variables is ordinal, there is a richer hierarchy of models that includes partial association models, and unsaturated three-factor interaction models.

#### 4.4.1 Odds Ratios

In interpreting these models, we refer to the odds ratios  $\{\theta_{ij(k)}\}$ ,  $\{\theta_{i(j)k}\}$  and  $\{\theta_{(i)jk}\}$  that describe the local conditional associations between two variables within a fixed level of the third variable. The set of conditional odds ratios  $\{\theta_{ij(k)}\}$ , where

$$\theta_{ij(k)} = \frac{\pi_{ijk} \pi_{i+1,j+1,k}}{\pi_{i,j+1,k} \pi_{i+1,j,k}} \quad \begin{array}{l} 1 \leq i \leq r-1 \\ 1 \leq j \leq c-1 \end{array}$$

is the set of  $(r-1)(c-1)$  local odds ratio at a fixed level of  $Z$ . Similar definitions hold for  $\theta_{i(j)k}$  and  $\theta_{(i)jk}$ . The ratio of odds ratios

$$\theta_{ijk} = \frac{\theta_{ij(k+1)}}{\theta_{ij(k)}} = \frac{\theta_{i(j+1)k}}{\theta_{i(j)k}} = \frac{\theta_{(i+1)jk}}{\theta_{(i)jk}}$$

is used for describing local three-factor interaction.  $\theta_{ijk}$  describes the interaction in a  $2 \times 2 \times 2$  section of the table consisting of adjacent rows, adjacent columns, and adjacent layers. There is an absence of three-factor interaction if all  $(r-1)(c-1)(\ell-1)$  of the  $\theta_{ijk}$  equal 1.0.

#### 4.4.2 All Variables Ordinal

Consider the case in which  $X$ ,  $Y$  and  $Z$  are all ordinal. Once again we have to choose (perhaps somewhat arbitrarily) scores  $\{u_i\}$ ,  $\{v_j\}$  and  $\{w_k\}$  to assign to the levels of  $X$ ,  $Y$  and  $Z$  respectively. A model that utilizes the ordinal nature of the variables is

$$\begin{aligned} \log m_{ijk} = & \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta^{YZ}(u_i - \bar{u})(v_j - \bar{v}) \\ & + \beta^{XZ}(u_i - \bar{u})(w_k - \bar{w}) + \beta^{XY}(v_j - \bar{v})(w_k - \bar{w}) \quad (4.3). \end{aligned}$$

As we will see later, the  $\beta^{ZY}$ ,  $\beta^{XZ}$  and  $\beta^{YZ}$  parameters describe the pairwise partial associations. This model has only three more parameters than the independence model, and  $df = rc\ell - [1 + (r-1) + (c-1) + (\ell-1)] = rc\ell - r - c - \ell + 1$ , so the model is always unsaturated.

The association term for each pair of variables has the same form as the association term in the linear-by-linear association model. There is no three-factor interaction term in the model, and all  $\theta_{ijk}$  equal 1.0. Hence the partial association between each pair of variables is the same for all levels of the third variable.

If integer scores are used for all three variables, the local conditional log odds ratios simplify to

$$\begin{aligned}\log \theta_{ij(k)} &= \beta^{XY} \\ \log \theta_{i(j)k} &= \beta^{XZ} \\ \log \theta_{(i)jk} &= \beta^{YZ}\end{aligned}$$

The local odds ratio is uniform for each pair of variables, and the strength of association is homogeneous across the levels of the third variable, which is why the model is called the homogeneous uniform association model.

If this model holds, and if a particular  $\beta$  equals zero, then there is conditional independence between those variables.

In some cases it is not relevant to utilize the orderings of classifications for all the ordinal variables. This could be because departures from independence are not linear, or because the sampling design dictates fitting certain marginal distributions. Then the variables in question could be treated as nominal, and the models of the following section could be fitted.

#### 4.4.3 Ordinal and Nominal Variables

When there is a mixture of nominal and ordinal variables, the row effects model of Section 4.3 can be generalized.

##### 4.4.3.1 One Nominal, Two Ordinal Variables

If X is nominal and Y and Z are ordinal, a basic model is:

$$\begin{aligned}\log m_{ijk} &= \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \tau_i^{XZ}(w_k - \bar{w}) \\ &\quad + \beta^{YZ}(v_j - \bar{v})(w_k - \bar{w})\end{aligned}\tag{4.4}$$

where  $\sum \lambda_i^X = \sum \lambda_j^Y = \sum \lambda_k^Z = \sum \tau_i^{XY} = \sum \tau_i^{XZ} = 0$

This is usually the simplest model that would be formulated to include all pairwise associations, but no three-factor interaction. The Y-Z association term for the ordinal variables has the same form as the association term for the linear-by-linear association model. The X-Y and X-Z association terms for the pairs of nominal and ordinal variables have the same form as the association term for the row effects model.

For model (4.4) applied with integer scores for the ordinal variables,

$$\begin{aligned}\log \theta_{ij(k)} &= \tau_{i+1}^{XY} - \tau_i^{XY} \\ \log \theta_{i(j)k} &= \tau_{i+1}^{XZ} - \tau_i^{XZ} \\ \log \theta_{(i)jk} &= \beta^{YZ} \\ \log \theta_{ijk} &= 0\end{aligned}$$

#### 4.4.3.2 Two Nominal, One Ordinal Variable

Suppose that X and Y are nominal, and Z is ordinal, then a basic model would be

$$\begin{aligned}\log m_{ijk} &= \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} \\ &\quad + \tau_i^{XZ}(w_u - w) + \tau_j^{YZ}(w_k - w)\end{aligned}\quad (4.5)$$

$$\begin{aligned}\text{where } \sum \lambda_i^X &= \sum \lambda_j^Y = \sum \lambda_k^Z = \sum \tau_i^{XZ} = \sum \tau_j^{YZ} \\ &= \sum \lambda_{ij}^{XY} = \sum \lambda_{ij}^{XY} = 0\end{aligned}$$

A general association term  $\{\lambda_{ij}^{XY}\}$  is used for the association between the nominal variables.

#### 4.4.3.3 Dumping Severity Example

Table 4.8 gives a 4x3x4 cross-classification of operation, dumping severity and hospital. The data used earlier (Table 1.6) is the marginal distribution of this table in which the results are combined for the four hospitals.

Table 4.8: Cross-classification of operation, hospital and dumping severity

Hospital												
Dumping Severity												
Operation	N	S	M	N	S	M	N	S	M	N	S	M
A	23	7	2	18	6	1	8	6	3	12	9	1
B	23	10	5	18	6	2	12	4	4	15	3	2
C	20	13	5	13	13	2	11	6	2	14	8	3
D	24	10	6	9	15	2	7	7	4	13	6	4

Note: N = none, S = slight, M = moderate

A model that takes into account the ordinal nature of operation and dumping is

$$\log m_{ijk} = \mu + \lambda_i^O + \lambda_j^H + \lambda_k^D + \lambda_{ij}^{OH} + \beta^{OD}(u_i - \bar{u})(w_k - \bar{w})$$

According to this model, dumping severity and operation have a uniform association that is the same for each hospital, and dumping severity is independent of hospital for each operation. The general association term  $\lambda_{ij}^{OH}$  is used for the O-H association, rather than the term  $\tau_j^{OH}(u_i - \bar{u})$ , because the O-H marginal distribution is regarded as fixed by the sampling design. When this model is fitted using integer scores, the deviance is  $G^2 = 25.35$ , with 29 df, which is a very good fit.

Given that the model fits, the conditional independence of dumping severity and operation corresponds to  $\beta^{OD} = 0$ . The next model in the hierarchy that does not include the  $\beta^{OD}$  term is  $\log m_{ijk} = \mu + \lambda_i^O + \lambda_j^H + \lambda_k^D + \lambda_{ij}^{OH}$ , which is referred to as (D,OH).

This model has deviance  $G^2 = 31.64$  with 30 df. Thus, the test statistic for testing  $H_0: \beta^{OD} = 0$  is  $G^2 = 31.64 - 25.35 = 6.29$ , with  $30 - 29 = 1$  df, which is significant at the 5% level.

The model (OD,OH), i.e.

$$\log m_{ijk} = \mu + \lambda_i^O + \lambda_j^H + \lambda_k^D + \lambda_{ij}^{OH} + \lambda_{ik}^{OD}$$

contains a general O-D association term, and has deviance  $G^2 = 20.76$  with 24 df. To test whether a general O-D association term,  $\lambda_{ik}^{OD}$ , is necessary (rather than simply a uniform association term,  $\beta^{OD}(u_i - \bar{u})(w_k - \bar{w})$ ), the statistic is  $G^2 = 25.35 - 20.76 = 4.59$ , with  $29 - 24 = 5$  df, which is not significant. So there is not a significant improvement in fit due to the extra five parameters in the more complicated model (OD,OH).

The ML estimate of  $\beta^{OD}$  is  $\hat{\beta}^{OD} = 0.163$ . This means that the estimated odds of moderate instead of slight, or slight instead of no dumping, are  $\exp(0.163) = 1.18$  times higher for each additional 25% of stomach removal.

To illustrate this, consider Table 4.9, which contains the estimated expected frequencies under the O-D uniform association model.

The estimated odds ratios for severity of slight vs none for operation B vs A, for hospitals 3 and 4 are:

$$\text{Hospital 3: } \hat{\theta}_{1(3)1} = \frac{11.07 \times 5.95}{12.09 \times 4.63} = 1.18$$

$$\text{Hospital 4: } \hat{\theta}_{1(4)1} = \frac{14.32 \times 5.95}{12.09 \times 5.99} = 1.18$$



Table 4.9: Estimated expected frequencies under the O-D uniform association model

Hospital												
1				2			3			4		
Dumping Severity												
	N	S	M	N	S	M	N	S	M	N	S	M
Operation												
A	20.84	8.72	2.45	16.28	6.81	1.91	11.07	4.63	1.30	14.32	5.99	1.68
B	22.96	11.31	3.73	15.71	7.74	2.55	12.09	5.95	1.96	12.09	5.95	1.96
C	21.06	12.20	4.74	15.52	8.99	3.49	10.53	6.10	2.37	13.86	8.03	3.12
D	20.07	13.68	6.25	13.05	8.89	4.06	9.03	6.16	2.81	11.54	7.87	3.60

The corresponding estimated odds ratios for adjacent operations and adjacent levels of dumping severity are all 1.18.

Because the model does not contain an H-D association term, the table can be collapsed over hospital, and the O-D association term in the marginal operation by dumping table will be the same as in the full table. This is why the estimate of the  $\beta^{OD}$  parameter is the same as in the uniform association model fitted in Section 4.2.1.

#### 4.4.4 Three-Factor Interaction Models

The models introduced in the previous section can be generalized to include three-factor interaction terms. A simple model for the case when X, Y and Z are ordinal is

$$\begin{aligned} \log m_{ijk} = & \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta^{XY}(u_i - \bar{u})(v_j - \bar{v}) \\ & + \beta^{XZ}(u_i - \bar{u})(w_k - \bar{w}) + \beta^{YZ}(v_j - \bar{v})(w_k - \bar{w}) \\ & + \beta^{XYZ}(u_i - \bar{u})(v_j - \bar{v})(w_k - \bar{w}) \end{aligned} \quad (4.6)$$

This model has  $df = rc\ell - [1 + (r-1) + (c-1) + (\ell-1) + 1 + 1 + 1 + 1] = rc\ell - r - c - \ell - z$ , and is unsaturated whenever there are more than two rows, columns or layers.

When this model is applied with integer scores,

$$\begin{aligned} \log \theta_{ijk} &= \beta^{XYZ} \\ \log \theta_{ij(k)} &= \beta^{XY} + \beta^{XYZ} [k - \frac{(\ell+1)}{2}] \end{aligned}$$

This model is described as a uniform interaction model because the local interaction equals  $\beta^{XYZ}$  for all  $2 \times 2 \times 2$  subtables formed from adjacent rows, adjacent columns and adjacent layers.

Within a particular level of Z, the association between X and Y is uniform with local log odds ratio  $\beta^{XY} + \beta^{XYZ}[k - (\ell+1)/2]$ . Thus, the strength of the X-Y partial association is constant within the levels of Z but changes linearly across the levels of Z. Similar results hold for the X-Z and Y-Z partial associations.

A more general interaction model would be

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{YZ} + \lambda_{jk}^{YZ} \\ + \beta^{XYZ}(u_i - \bar{u})(v_j - \bar{v})(w_k - \bar{w}).$$

This model has  $df = rc\ell - [1 + (r-1) + (c-1) + (\ell-1) + (r-1)(c-1) + (r-1)(\ell-1) + (c-1)(\ell-1) + 1] = (r-1)(c-1)(\ell-1) - 1$ , so is unsaturated whenever there are more than two rows, columns or layers. When this model is applied with integer scores, there is a constant value for  $\log \theta_{ijk}$ , and it is the most general uniform interaction model. Due to the general forms for the association terms, this model does not have uniform association within the partial tables.

Model (4.6) applied with integer scores may be described as a type of heterogeneous uniform association model, since the strength of the uniform partial association for each pair of variables changes across the levels of the third variable.

However, a more general example of a heterogeneous uniform association model would be to assume there is a uniform conditional association between say X and Y which is allowed to change in an unspecified manner across the levels of Z. In this case the interaction is not uniform and Z is treated as nominal. This model is

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \\ + \beta_k^{XY}(u_i - \bar{u})(v_j - \bar{v})$$

For this model with integer scores  $\beta_k^{XY}$  is the constant value of the X-Y local log odds ratio at level k of Z.

#### 4.4.4.1 Smoking Example

Table 4.10 is a cross-classification of age, smoking status and breathing test results of 2289 people (Forthofer and Lehren, 1981, p.21).

Table 4.10: Age by smoking status by breathing test results

Age	Smoking Status	Breathing Test Results		
		Normal	Borderline	Abnormal
<40	Never smoked	577	27	7
	Former smoker	192	20	3
	Current smoker	682	46	11
40-59	Never smoked	164	4	0
	Former smoker	145	15	7
	Current smoker	245	47	27

The standard partial association model (AS, AB, SB) fits poorly, with a deviance of 25.93 based on 4 df, so we consider models having three-factor interaction terms.

The model

$$\log m_{ijk} = \mu + \lambda_i^A + \lambda_j^S + \lambda_k^B + \lambda_{ij}^{AS} + \lambda_{ik}^{AB} + \lambda_{jk}^{SB} + \beta^{ASB}(u_i - \bar{u})(v_j - \bar{v})(w_k - \bar{w})$$

with integer scores has only one additional parameter but fits much better, with  $G^2 = 2.74$  with 3 df. To test  $H_0: \beta^{ASB} = 0$  use  $G^2 = 25.93 - 2.74 = 23.19$  with  $4-3 = 1$  df, which is significant at the 0.1% level. So there is very strong evidence that the association between smoking status and breathing test results depends on age. The estimate of  $\beta^{SBA} = 0.831$  means that the association between smoking status and breathing test result is more positive at the higher age level. Any local odds ratio for the 40-59 age group is estimated to be  $\hat{\theta}_{ijk} = \hat{\theta}_{(2)jk} / \hat{\theta}_{(1)jk} = \exp(0.831) = 2.30$  times higher than the corresponding local odds ratio for the < 40 age group. To illustrate this concept consider Table 4.11, which contains the fitted counts under this uniform interaction model.

Table 4.10: Age by smoking status by breathing test results

Age	Smoking Status	Breathing Test Results		
		Normal	Borderline	Abnormal
<40	Never smoked	577	27	7
	Former smoker	192	20	3
	Current smoker	682	46	11
40-59	Never smoked	164	4	0
	Former smoker	145	15	7
	Current smoker	245	47	27

The standard partial association model (AS, AB, SB) fits poorly, with a deviance of 25.93 based on 4 df, so we consider models having three-factor interaction terms.

The model

$$\log m_{ijk} = \mu + \lambda_i^A + \lambda_j^S + \lambda_k^B + \lambda_{ij}^{AS} + \lambda_{ik}^{AB} + \lambda_{jk}^{SB} + \beta^{ASB}(u_i - \bar{u})(v_j - \bar{v})(w_k - \bar{w})$$

with integer scores has only one additional parameter but fits much better, with  $G^2 = 2.74$  with 3 df. To test  $H_0: \beta^{ASB} = 0$  use  $G^2 = 25.93 - 2.74 = 23.19$  with  $4-3 = 1$  df, which is significant at the 0.1% level. So there is very strong evidence that the association between smoking status and breathing test results depends on age. The estimate of  $\beta^{SBA} = 0.831$  means that the association between smoking status and breathing test result is more positive at the higher age level. Any local odds ratio for the 40-59 age group is estimated to be  $\hat{\theta}_{ijk} = \hat{\theta}_{(2)jk} / \hat{\theta}_{(1)jk} = \exp(0.831) = 2.30$  times higher than the corresponding local odds ratio for the < 40 age group. To illustrate this concept consider Table 4.11, which contains the fitted counts under this uniform interaction model.

Table 4.11: Fitted counts under the uniform interaction model

Age	Smoking Status	Breathing Test Results		
		Normal	Borderline	Abnormal
< 40	Never smoked	577.43	27.09	6.48
	Former smoker	191.76	18.57	4.67
	Current smoker	681.81	47.34	9.85
40-59	Never smoked	163.57	3.91	0.52
	Former smoker	145.24	15.43	5.33
	Current smoker	245.19	45.66	28.15

The odds ratio for abnormal vs borderline breathing test, for former smoker vs never smoked, at age < 40 estimated to be:

$$\text{Age} < 40: \hat{\theta}_{(1)12} = \frac{27.09 \times 4.67}{18.57 \times 6.48} = 1.05$$

The corresponding estimated odds ratio for age 40-59 is:

$$\text{Age } 40-59: \hat{\theta}_{(2)12} = \frac{3.91 \times 5.33}{16.43 \times 0.52} = 2.44$$

So the estimated odds ratio for the 40-59 age group is  $2.44/1.05 = 2.3$  times higher than the estimated odds ratio for the < 40 age group.

The heterogeneous uniform S-B association model,

$$\log m_{ijk} = \mu + \lambda_i^A + \lambda_j^S + \lambda_k^B + \lambda_{ij}^{AS} + \lambda_{ik}^{AB} + \beta_i^{SB}(v_j - \bar{v})(w_k - \bar{w})$$

with integer scores, has a deviance of 10.80 with 6 df, and does not fit quite as well as the uniform interaction model. However, it yields the simple interpretation of constant local

odds ratios  $\exp(\hat{\beta}_1^{SB}) = \exp(0.115) = 1.12$  for the lower age group, and  $\exp(\hat{\beta}_2^{SB}) = \exp(0.781) = 2.18$  for the higher age group. To confirm this, Table 4.12 holds the fitted counts under this model:

Table 4.12: Counts under heterogeneous uniform S-B association model

Age	Smoking Status	Breathing Test Results		
		Normal	Borderline	Abnormal
< 40	Never smoked	572.42	32.25	6.33
	Former smoker	199.61	12.62	2.78
	Current smoker	678.98	48.13	11.89
40-59	Never smoked	160.71	6.44	0.85
	Former smoker	150.08	13.12	3.80
	Current smoker	243.21	46.44	29.35

At age < 40, the odds ratio for abnormal vs borderline results, for former smoker vs never smoked is estimated to be:

$$\text{Age } < 40: \hat{\theta}_{(1)12} = \frac{32.25 \times 2.78}{12.62 \times 6.33} = 1.12$$

The corresponding estimated odds ratio for age 40-59 is

$$\text{Age } 40-59: \hat{\theta}_{(2)12} = \frac{6.44 \times 3.80}{13.12 \times 0.85} = 2.19$$

Since both  $\hat{\beta}_1^{SB}$  and  $\hat{\beta}_2^{SB}$  are positive, breathing test results tend to be more abnormal when an individual's smoking status is more current. The association is estimated to be stronger for the older age group.

## CHAPTER V: LOG-MULTIPLICATIVE MODELS

An obvious disadvantage of the ordinal loglinear models discussed in Chapter IV is the need to assign scores to the categories of ordinal variables. For many cases it is not obvious what scores should be chosen. But the parameter estimates and goodness of fit of the models depend on that choice.

However, we could treat the scores as parameters to estimate, rather than as fixed values.

For two-dimensional tables a model of this type would be

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta \mu_i \nu_j$$

where  $\sum \lambda_i^X = \sum \lambda_j^Y = 0$ . Here the  $\{\mu_i\}$  and  $\{\nu_j\}$  are parameters to be estimated. The model is called log-multiplicative because the log of the expected count is a multiplicative function of the parameters. Goodman (1979) called it the "RC model" because of its multiplicative row and column effects. The basic form of the model is unchanged when the  $\{\mu_i\}$  or  $\{\nu_j\}$  are replaced by linear functions of themselves. So we can assume an arbitrary location and scale, such as

$$\sum \mu_i = \sum \nu_j = 0 \text{ and } \sum \mu_i^2 = \sum \nu_j^2 = 1$$

Because of the constraints on the model parameters,  $r-2$  of the  $\{\mu_i\}$  and  $c-2$  of the  $\{\nu_j\}$  are linearly independent. So for testing goodness of fit

$$\begin{aligned} df &= rc - [1 + (r-1) + (c-1) + 1 + (r-2) + (c-2)] \\ &= (r-2)(c-2) \end{aligned}$$

and so the table must be at least 3x3 for the model to be unsaturated.

The independence model corresponds to  $\beta=0$ . The log-multiplicative model resembles the linear-by-linear association model discussed in Section 4.2.2

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta (u_i - \bar{u})(v_j - \bar{v})$$



The log-multiplicative model can be interpreted like the linear-by-linear association model through the odds ratios if the fixed scores  $\{u_i\}$  and  $\{v_j\}$  are replaced by the parameters  $\{\mu_i\}$  and  $\{v_j\}$ . However, the parameters  $\{\mu_i\}$  and  $\{v_j\}$  need not be monotonic. If the RC model fits well and produces parameter score estimates that are monotonic, then the linear-by-linear association model would also fit well if the fixed scores that were chosen for that model had similar spacings.

The log-multiplicative model is invariant to interchanges of rows and columns. For example, if we interchange rows a and b, then  $\mu_a$  and  $\mu_b$  simply switch places, so the variables are treated as nominal. However, we can describe ordinal characteristics of the data through odds ratios. The local log odds ratio is

$$\log \theta_{ij} = \beta(\mu_{i+1} - \mu_i)(v_{j+1} - v_j)$$

so monotonicity in the scores means that all local associations have the same sign. Lack of monotonicity in the scores indicates that local associations are positive in some locations and negative in others.

The log-multiplicative model also resembles the row effects model

$$\log m_{ij} = \mu + \lambda_i X + \lambda_j Y + \tau_i(v_j - \bar{v})$$

if we treat the  $\{\mu_i\}$  as row effects and the  $\{v_j\}$  as parameter scores.

Generally in the RC model we regard the  $\{\mu_i\}$  as row effects and the  $\{v_j\}$  as column effects.

The RC model can actually be used to give an ordering to the row and column categories, i.e. we could order them in increasing order the  $\mu_i$  or  $v_j$ .

### 5.1 Estimation

Although the RC model is not actually loglinear, it can be fitted using an iterative loglinear fitting procedure.

If one set of parameter scores is treated as fixed, then the model has a loglinear form.

Each cycle of the iterative procedure consists of two steps. First, the column parameters are treated as fixed, and the row scores are estimated as in a loglinear row effects model. Then the estimated row effects are treated as fixed row scores, and column scores are estimated as in a column effects model. Those estimates serve as fixed column scores in the first part of the next cycle.

## 5.2 Inference for Log-Multiplicative Models

Testing for independence corresponds to testing  $H_0: \beta = 0$  in the RC model. The test statistic  $G^2(I) - G^2(RC)$  does not have an asymptotic chi-squared distribution, because the  $\{\mu_i\}$  and  $\{v_j\}$  are undetermined in the RC model if independence holds. Instead, Haberman (1981) showed that the null asymptotic distribution of the statistic is the same as that of the maximum eigenvalue of the  $(r-1)$  by  $(c-1)$  central Wishart matrix with  $df = c-1$ .

## 5.3 Dumping Severity Example

Once again we will use the data from Table 1.6. Fitting the RC model

$$\log m_{ij} = \mu + \lambda_i^O + \lambda_j^D + \beta \mu_i v_j$$

gives a deviance of  $G^2(RC) = 2.85$  with  $df = (4-2)(c-2) = 2$ , which is a good fit. We have already seen that the independence model gives  $G^2(I) = 10.88$ , with 6 df. The test statistic  $G^2(I) - G^2(RC) = 10.88 - 2.85 = 8.03$  can be used to test independence, given that the RC model fits. The null distribution of this statistic is the same as that of the maximum eigenvalue of the  $3 \times 3$  central Wishart matrix, with 2 df. Table 51 of Pearson and Hartley (1972) gives the upper 5% and 1% critical values for this test, which are 10.74 and 14.57 respectively. Thus, the  $G^2(I) - G^2(RC)$  test is significant at the 5% level, so there is evidence that the operation affects the dumping severity.

The parameter estimates for the RC model are:

$$\begin{aligned}
\hat{\mu}_1 &= 0.364 & \hat{\nu}_1 &= -0.797 \\
\hat{\mu}_2 &= -0.613 & \hat{\nu}_2 &= 0.553 \\
\hat{\mu}_3 &= 0.404 & \hat{\nu}_3 &= 0.244 \\
\hat{\mu}_4 &= 0.573 \text{ and } \hat{\beta} & &= 0.442
\end{aligned}$$

The parameter estimates are not monotonic, which indicates that some local associations are positive and some are negative. The estimate of the local odds ratio is

$$\hat{\theta}_{ij} = \exp[\hat{\beta}(\hat{\mu}_{i+1} - \hat{\mu}_i)(\hat{\nu}_{j+1} - \hat{\nu}_j)]$$

which gives the six estimated local odds ratios as

$$\begin{aligned}
\hat{\theta}_{11} &= \exp[0.442(-0.613+0.364)(0.553+0.797)] \\
&= \exp(-0.149) \\
&= 0.86
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}_{21} &= \exp[0.442(0.404+0.613)(0.553+0.797)] \\
&= \exp(0.607) \\
&= 1.83
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}_{31} &= \exp[0.442(0.573-0.404)(0.553+0.797)] \\
&= \exp(0.101) \\
&= 1.11
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}_{12} &= 1.04 \\
\hat{\theta}_{22} &= 0.87 \\
\hat{\theta}_{32} &= 0.98
\end{aligned}$$

These are of course the same as those obtained using the fitted counts in Table 5.1.

Table 5.1: Fitted counts under the RC model

Operation	Dumping Severity		
	None	Slight	Moderate
A	60.28	25.59	10.13
B	68.62	25.10	10.23
C	57.59	38.64	13.77
D	53.50	39.68	13.82

$$\hat{\theta}_{11} = \frac{60.28 \times 25.10}{68.62 \times 25.59} = 0.86$$

$$\hat{\theta}_{21} = \frac{68.62 \times 38.64}{57.59 \times 25.10} = 1.83$$

$$\hat{\theta}_{31} = \frac{57.59 \times 39.68}{53.50 \times 38.64} = 1.11$$

etc.

This RC model states that the odds for heavier dumping severity are sometimes increased by the removal of more stomach and sometimes decreased. The simpler uniform association model

$$\log m_{ij} = \mu + \lambda_i^O + \lambda_j^{D+\beta} (u_i - \bar{u})(v_j - \bar{v})$$

which we fitted in Section 4.2.1, states that the odds of having slight dumping instead of none (or moderate instead of slight) are  $\exp(\beta) = \exp(0.163) = 1.18$  times higher for operation  $i+1$  than operation  $i$ . The uniform association model gives  $G^2(U) = 4.59$ , with 5 df, and is the special case of the RC model where

$$\mu_i = u_i - \bar{u} \text{ and } v_j = v_j - \bar{v}$$

The difference in deviance between the two models,  $G^2(U|RC) = G^2(U) - G^2(RC)$ , provides a test of the null hypothesis that the  $\{\mu_i\}$  and  $\{v_j\}$

are linear transformations of the pre-selected scores for the loglinear model, given that the RC model fits. In this case  $G^2(U/RC) = 4.59 - 2.85 = 1.74$  based on  $5-2 = 3$  df, which is not significant, so the equal interval scores of the uniform association model seem permissible, since the increase in deviance is small compared to using the parameter scores of the RC model. Therefore the simpler uniform association model seems adequate for describing the relationship between operation and dumping severity.

Although the example used here has both ordinal row and column variables, we could of course have used a nominal-ordinal, or nominal-nominal table.

#### 5.4 Higher Dimensions

The log-multiplicative model can be generalized in various ways so that it can be used with multi-dimensional tables. Log-multiplicative models are obtained when parameters are substituted for some or all pairs of sets of fixed scores in the ordinal loglinear models of Chapter IV. For example, a parameter-scores version of the homogeneous linear-effects model (4.3) is

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta_{XY} \mu_i \nu_j + \beta_{XZ} \mu_i \omega_k + \beta_{YZ} \nu_j \omega_k$$

$$\text{where } \sum \lambda_i^X = \sum \lambda_j^Y = \sum \lambda_k^Z = 0$$

$$\sum \mu_i = \sum \nu_j = \sum \omega_k = 0$$

$$\sum \mu_i^2 = \sum \nu_j^2 = \sum \omega_k^2 = 1$$

This model is always unsaturated as it has  $df = rcl - 2(r+c+l) + 5$

The log odds ratios are

$$\log \theta_{ij(k)} = (\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j) \beta_{XY}$$

$$\log \theta_{ijk} = 0$$

so that the model is quite simple to interpret when the parameter scores are monotonic.

## CHAPTER VI: LOGIT MODELS

The models described in the previous chapters are suitable for tables with one or more response variables. In contrast, the logit models of this chapter are only appropriate for the situation where there is only one response that is also ordinal. This response variable is explained by a set of explanatory variables.

### 6.1 Dichotomous Response

A dichotomous response can take one of only two values, e.g. yes/no; success/failure; low/high. It can be trivially treated as ordinal by defining one level to be "high".

Dichotomous logit models can be thought of as simply a different way of writing loglinear models for one dichotomous response.

To illustrate this, consider a cross-classification in which X and Y are categorical explanatory variables and Z is a dichotomous response. We will fit the loglinear model (XY, XZ, YZ) or

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

We define the logit as  $\log m_{ij2}/m_{ij1}$ . This is equal to

$$\begin{aligned} & \log m_{ij2} - \log m_{ij1} \\ &= \mu + \lambda_i^X + \lambda_j^Y + \lambda_2^Z + \lambda_{ij}^{XY} + \lambda_{i2}^{XZ} + \lambda_{j2}^{YZ} \\ &- [\mu + \lambda_i^X + \lambda_j^Y + \lambda_1^Z + \lambda_{ij}^{XY} + \lambda_{i1}^{XZ} + \lambda_{j1}^{YZ}] \\ &= (\lambda_2^Z - \lambda_1^Z) + (\lambda_{i2}^{XZ} - \lambda_{i1}^{XZ}) + (\lambda_{j2}^{YZ} - \lambda_{j1}^{YZ}) \end{aligned}$$

Since Z is dichotomous and  $\sum_k \lambda_k^Z = 0$ ,  $\sum_k \lambda_{ik}^{XZ} = 0$  and  $\sum_k \lambda_{jk}^{YZ} = 0$

$$\begin{aligned} \text{then } \lambda_1^Z &= -\lambda_2^Z \\ \lambda_{i1}^{XZ} &= -\lambda_{i2}^{XZ} \\ \text{and } \lambda_{j1}^{YZ} &= -\lambda_{j2}^{YZ} \end{aligned}$$

so  $\log m_{ij2}/m_{ij1} = 2\lambda_2^Z + 2\lambda_{i2}^{XZ} + 2\lambda_{j2}^{YZ}$  which can be written as

$$\log m_{ij2}/m_{ij1} = \alpha + \tau_i^X + \tau_j^Y$$

where  $\sum_i \tau_i^X = \sum_j \tau_j^Y = 0$

This is the general form of the logit model for nominal X and Y variables.

The logit for Z at level i of X and j of Y is defined to be

$$\begin{aligned} \log \frac{\pi_{2(ij)}}{1-\pi_{2(ij)}} &= \log \frac{\pi_{2(ij)}}{\pi_{1(ij)}} \\ &= \log \frac{\pi_{ij2}}{\pi_{ij1}} \\ &= \log \frac{m_{ij2}}{m_{ij1}} \end{aligned}$$

Logit models always assume that the response is to be explained by a combination of explanatory variables, so the corresponding loglinear model will always contain the most general interaction term for describing associations among the explanatory variables, plus the relevant interaction terms which specify relationships between the explanatory variables and the response (these will depend on exactly which model is being fitted). So, to rewrite a loglinear model in terms of a logit model, we can simply omit any terms that don't involve the response and remove the response subscript from all other terms.

For example, suppose X is nominal but Y is ordinal with monotone scores  $\{v_1, \dots, v_C\}$  assigned to its levels. A loglinear model that specifies that both X and Y affect Z (and also uses the ordinal nature of Y) is

$$\begin{aligned} \log m_{ijk} &= \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} \\ &\quad + \tau_j^{YZ} (v_j - \bar{v}) \end{aligned}$$

The logit equivalent of this model is

$$\log \frac{m_{ij2}}{m_{ij1}} = \alpha + \tau_i^X + \beta^Y (v_j - \bar{v})$$

with  $\sum_i \tau_i^X = 0$

## 6.2 Polytomous Response

Now we will extend these logit models to the case where a response has two or more ordered categories. It makes sense to form logits in a way that takes the category order into account. However, we don't need to use only two categories at a time in forming the logits. Three different types of logits we could use are:

(a) Adjacent-categories logits

$$L_j = \log \frac{\pi_{j+1}}{\pi_j}, \quad j = 1, \dots, c-1$$

(b) Continuation-ratio logits

$$L_j = \log \frac{\pi_{j+1}}{\pi_1 + \dots + \pi_j}, \quad j = 1, \dots, c-1$$

(c) Cumulative logits

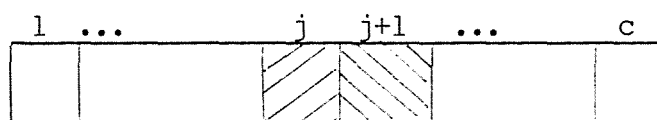
$$L_j = \log \frac{\pi_{j+1} + \dots + \pi_c}{\pi_1 + \dots + \pi_j}, \quad j = 1, \dots, c-1$$

where  $\pi_j$  is the probability of response category  $j$  at a certain combination of levels of explanatory variables. When there are only  $c=2$  categories, all three of these logit types simplify to the standard logit,  $\log(\pi_2/\pi_1)$ .

Figure 6.1 gives an impression of these three types of logits.

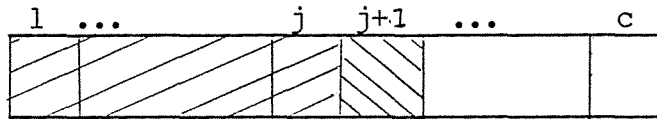
Figure 6.1: Logits for an ordinal response

(a) Adjacent-categories

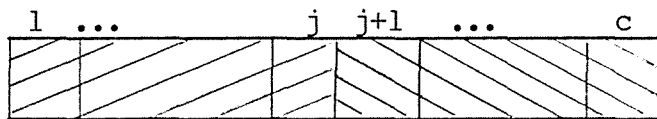




(b) Continuation-ratio



(c) Cumulative



### 6.2.1 Adjacent-Categories Logits

Goodman (1983) presented models using adjacent-categories logits. These models are equivalent to the ordinal loglinear models discussed in Chapter IV, but they have a different emphasis in the sense that they are only appropriate when there is only one (ordinal) response.

#### 6.2.1.1 Ordinal-Ordinal

For a two-way table with ordinal explanatory row variable,  $X$ , which has assigned scores  $\{u_j\}$ , and ordinal response column variable,  $Y$ , a simple logit model would be

$$L_j(i) = \log \frac{m_{ij+1}}{m_{ij}} = \alpha_j + \beta(u_i - \bar{u})$$

$$1 \leq i \leq r,$$

$$1 \leq j \leq c-1$$

Independence corresponds to the case where  $\beta = 0$ .

This model is equivalent to the loglinear uniform association model

$$\log m_{ij} = \mu + \beta(u_i - \bar{u})(v_j - \bar{v})$$

(i.e. with integer scores), and the association parameter  $\beta$  is identical in the two models.

#### 6.2.1.2 Ordinal-Nominal

If the row variable  $X$  is nominal, a simple logit model is

$$L_j(i) = \log \frac{m_{ij+1}}{m_{ij}} = \alpha_j + \tau_i$$

$$1 \leq i \leq r$$

$$1 \leq j \leq c-1$$

where  $\sum \tau_i = 0$ . This is referred to as the parallel odds model by Goodman. Independence corresponds to the case where all the  $\tau_i$  equal zero. The parallel odds model is equivalent to the log linear row effects model with integer scores  $\{v_j = j\}$

$$\log m_{ij} = \mu + \tau_i (v_j - \bar{v})$$

and the  $\{\tau_i\}$  row effect parameters are identical in the two models.

#### 6.2.2 Continuation-Ratio Logits

Continuation-ratio logits have the feature that the results of fitting models for separate logits are independent. Hence the  $C-1$   $G^2$  statistics and their df values can be summed to obtain an overall goodness-of-fit statistic that pertains to the simultaneous fitting of  $C-1$  models, one for each logit. However, if the categories of the ordinal variable are listed in the opposite order and continuation-ratio logits are formed from these (i.e. the logits  $\log [\pi_j/(\pi_{j+1}+\dots+\pi_C)]$  are formed), the results will differ from the original analysis.

Although they do not appear to have been used very much in the literature, models can be formed for continuation-ratio logits in a similar fashion to that described in Section 6.2.1, i.e. we model the logit as a linear function of the explanatory variables. For example, for an ordinal-ordinal table with explanatory row variable

X with scores  $\{u_i\}$ , we could model the logit as a linear function of the scores

$$L_j(i) = \log \frac{m_{ij+1}}{m_{i1} + \dots + m_{ij}} = \alpha_j + \beta (u_i - \bar{u})$$

$$1 \leq i \leq r$$

$$1 \leq j \leq c-1$$

If X is nominal (or if we just wish to treat it as nominal) we could model the logit as

$$L_j(i) = \alpha_j + \tau_i$$

$$1 \leq i \leq r$$

$$1 \leq j \leq c-1$$

$$\sum \tau_i = 0$$

Fienberg (1980, p.114) gives an example of the use of a continuation ratio logits for a four-way table.

### 6.2.3 Cumulative Logits

The cumulative logits

$$L_j = \frac{\pi_{j+1} + \dots + \pi_c}{\pi_1 + \dots + \pi_j}$$

use all c categories for each logit, and satisfy  $L_1 > L_2 \dots > L_{c-1}$

#### 6.2.3.1 Heterogeneous Effects

##### 6.2.3.1.1 Ordinal-Ordinal Tables

For a two-way table with ordinal explanatory row variable, X, which has assigned scores  $\{u_i\}$ , and ordinal response column variable, Y, a simple logit model would be

$$L_j(i) = \log \frac{m_{ij+1} + \dots + m_{ic}}{m_{i1} + \dots + m_{ij}}$$

$$= \alpha_j + \beta_j (u_i - u), i = 1, \dots, r$$

For each fixed cutpoint  $j$  there are  $r$  logits and two parameters, so that  $df = r-2$ . This model is simply the basic logit model of Section 6.1 for a dichotomous response variable and quantitative explanatory variable, applied to all  $c-1$  combinations of adjacent categories. For each cutpoint  $j$ , the first category is obtained by combining the response categories up to and including the  $j$ th, and the second category is obtained by combining the last  $c-j$  categories. Thus we form  $c-1$  separate  $rx2$  tables and fit the basic logit model to each. If the model holds and  $\beta_j = 0$ , then  $X$  and  $Y$  are independent when  $Y$  is collapsed in this manner.

For fixed category cutpoint  $j$  this model, applied with integer scores, is identical to the loglinear uniform association model

$$\log m_{ij} = \mu + \beta(u_i - \bar{u})(v_j - \bar{v})$$

applied to the collapsed  $rx2$  table.

#### 6.2.3.1.1.1 Dumping Severity Example

We will now re-analyse the data of Table 1.6. The first cumulative logit involves comparing "none" to "slight" plus "moderate", i.e.  $(\pi_2 + \pi_3)/\pi_1$ , the second cumulative logit compares "none" plus "slight" to "moderate", i.e.  $\pi_3/(\pi_1 + \pi_2)$ . These two collapsing are given in Tables 6.1 and 6.2.

Table 6.1: Data for first cumulative logit ( $j=1$ )

Operation	Dumping Severity		Total
	None	Slight or Moderate	
A	61	35	96
B	68	36	104
C	58	52	110
D	53	54	107

Table 6.2: Data for second cumulative logit ( $j=2$ )

Operation	Dumping Severity		Total
	None or Slight	Moderate	
A	29	7	96
B	91	13	104
C	98	12	110
D	91	16	107

Assigning integer scores to the operations and fitting the linear logit models

$$L_j(i) = \alpha_j + \beta_j(u_i - \bar{u})$$

gives the fitted counts of Tables 6.3 and 6.4.

Table 6.3: Fitted counts from first collapsing ( $j=1$ )

Operation	Dumping Severity		Total
	None	Slight or Moderate	
A	63.37	32.63	96
B	63.13	40.87	104
C	40.63	49.37	110
D	52.87	54.13	107

Table 6.4: Fitted counts from second collapsing ( $j=2$ )

Operation	Dumping Severity		Total
	None or Slight	Moderate	
A	87.91	8.09	96
B	93.39	10.61	104
C	96.48	13.52	110
D	91.22	15.78	107

The deviances for the logit models are  $G_1^2 = 1.49$  and  $G_2^2 = 0.94$ , both with 2 df, so both cumulative logits are fitted well.

The estimates of the slope parameters are  $\hat{\beta}_1 = 0.229$  and  $\hat{\beta}_2 = 0.211$ , so the odds that there is some side effect (instead of none) is estimated to be  $\exp(\hat{\beta}_1) = \exp(0.229) = 1.26$  times higher for operation  $i+1$  than for operation  $i$ . Similarly the odds that the dumping is moderate, rather than none or slight, is estimated to be  $\exp(\hat{\beta}_2) = \exp(0.211) = 1.23$  times higher for operation  $i+1$  than for operation  $i$ .

Of course, both these odds ratios can be obtained from the tables of fitted counts. For example, the odds ratio for

some dumping instead of none, for operation D vs C is estimated as

$$\hat{\theta}_{31}(1) = \frac{60.63 \times 54.13}{52.87 \times 49.37} = 1.26$$

The odds ratio for moderate dumping instead of slight or none, for operation D vs C, is estimated as

$$\hat{\theta}_{31}(2) = \frac{96.48 \times 15.78}{91.22 \times 13.52} = 1.23$$

To test whether the  $\beta$  parameters are zero, we compare the deviances of the models where they are unconstrained and of the models where they are set to zero.

For the first cumulative logit, fitting the independence model

$$L_1(i) = \alpha_1$$

gave a deviance of  $G^2(I_1) = 7.92$  with 3 df. The unconstrained model

$$L_1(i) = \alpha_1 + \beta_1(u_i - \bar{u})$$

gave a deviance of  $G^2(\beta_1) = 1.49$  with 2 df, so the test statistic for  $H_0: \beta_1 = 0$  is  $G^2(I_1) - G^2(\hat{\beta}_1) = 7.92 - 1.49 = 6.43$ , with  $3-2 = 1$  df, which is significant at the 5% level.

For the second cumulative logit,  $G^2(I_2) = 3.19$  with 3 df, and  $G^2(\hat{\beta}_2) = 0.94$  with 2 df, so the test statistic for  $H_0: \beta_2 = 0$  is  $G^2(I_2) - G^2(\hat{\beta}_2) = 3.19 - 0.94 = 2.25$  with  $3-2 = 1$  df, which is not significant at the 10% level.

### 6.2.3.1.2 Ordinal-Nominal Tables

For the two-way table where X is nominal, this approach is not very enlightening because if we fit

$$L_j(i) = \log \frac{m_{i,j+1} + \dots + m_{ic}}{m_{i1} + \dots + m_{ij}}$$

$$= \alpha_j + \tau_{ij}$$

where  $\sum_i \tau_{ij} = 0$ , then for each fixed cutpoint  $j$  there are

$1+(r-1) = r$  parameters, but only  $r$  logits, and the model is saturated.

### 6.2.3.2 Homogeneous Effects

#### 6.2.3.2.1 Ordinal-Ordinal Tables

Consider again the heterogeneous-effects logit model which we applied to the dumping severity data. The estimates of

$$\beta_1 = \alpha_1 + \beta_1(u_i - \bar{u}) \text{ and } L_2(i) = \alpha_2 + \beta_2(u_i - \bar{u})$$

were  $\beta_1 = 0.229$  and  $\beta_2 = 0.211$ . The fact that these estimates are so similar suggests that we might be able to assume  $\beta_1 = \beta_2$  and replace them by a single homogeneous effect parameter  $\beta_1$  and fit the model:

$$L_j(i) = \alpha_j + \beta(u_i - \bar{u}) \quad \begin{array}{l} i = 1, \dots, r \\ j = 1, \dots, c-1 \end{array}$$

An amalgamated logit model of this type for the full  $rc$  table does not correspond to any loglinear model.

In general, there are  $c-1$  logits in each of  $r$  rows, giving a total of  $r(c-1)$  logits. There is one association parameter,  $\beta$ , and  $c-1$  parameters,  $\{\alpha_j\}$ , relating to the various cutpoints for forming the logits. Therefore, the residual  $df = r(c-1) - 1 - (c-1) = rc - r - c$ . This is the same as the residual



df for the linear-by-linear association model; however, the loglinear and cumulative logit models are not equivalent unless  $c=2$ .

Each of the  $c-1$  logits is linearly related to the explanatory variable, with slope  $\beta$  assumed the same for all logits. If the model holds and  $\beta=0$ , then the  $j$ th logit is the same in each row (for all  $j$ ), which implies that  $X$  and  $Y$  are independent.

The difference in logits for adjacent rows

$$\begin{aligned} L_j(i+1) - L_j(i) &= \log \frac{\pi_{i+1,j+1} + \dots + \pi_{i+1,c}}{\pi_{i+1,1} + \dots + \pi_{i+1,j}} - \log \frac{\pi_{i,j+1} + \dots + \pi_{i,c}}{\pi_{i,1} + \dots + \pi_{i,j}} \\ &= \log \frac{(\pi_{i+1,1} + \dots + \pi_{i+1,j}) (\pi_{i,j+1} + \dots + \pi_{i,c})}{(\pi_{i+1,j+1} + \dots + \pi_{i+1,c}) (\pi_{i,1} + \dots + \pi_{i,j})} \end{aligned}$$

which is the log odds ratio for the  $2 \times 2$  table obtained using rows  $i$  and  $i+1$  and the dichotomous response having cutpoint following category  $j$ . This is simply the log of the local-global odds ratio  $\theta_{ij}$  introduced in Section 1.6.3.2. If integer row scores are used, then

$$\begin{aligned} L_j(i+1) - L_j(i) &= \alpha_j + \beta(u_{i+1} - \bar{u}) - \alpha_j - \beta(u_i - \bar{u}) \\ &= \beta(u_{i+1} - u_i) \\ &= \beta \end{aligned}$$

so  $\exp(\beta)$  represents the constant value of the odds ratios  $\{\theta_{ij}\}$  for the  $(r-1)(c-1)$   $2 \times 2$  tables obtained by taking all pairs of adjacent rows and all dichotomous collapsing of the response. We will call this model the logit uniform association model. McCullagh (1980) shows that the whole class of homogeneous logit effect models are proportional odds models. Taking the example of the logit uniform association model, the proportional odds model states that for a given row  $i$ , the odds that a category is less than or equal to  $j$ , is

$$\frac{\pi_{i,1} + \dots + \pi_{i,j}}{\pi_{i,j+1} + \dots + \pi_{i,c}} = \kappa_j \exp(-(\delta_j + \beta(u_i - \bar{u})))$$

where  $\kappa_j$  and  $\delta_j$  are just constants that depend on  $j$ . Taking logs of both sides and multiplying through by  $-1$  gives

$$\begin{aligned} \log \frac{\pi_{i,j+1} + \dots + \pi_{ic}}{\pi_{i1} + \dots + \pi_{ij}} &= -\log \kappa_j + \delta_j + \beta(u_i - \bar{u}) \\ &= \alpha_j + \beta(u_i - \bar{u}) \end{aligned}$$

which is simply the logit uniform association model.

6.2.3.2.1.1 Dumping Severity Example

We will now apply the logit uniform association model to the dumping severity data of Table 1.6. The estimate of the association parameter is  $\beta = 0.225$ , and the deviance is  $G^2 = 4.27$  with 5 df, so the model fits well. The estimate of the local-global odds ratio is  $\hat{\theta}'_{ij} = \exp(\hat{\beta}) = \exp(0.225) = 1.25$ . This means that the odds that dumping severity is some rather than none (or moderate rather than slight or less) is 1.25 times higher for operation  $i+1$  than for operation  $i$ . To illustrate this, consider the fitted counts of Table 6.5:

Table 6.5: Fitted counts under logit uniform association model

Operation	Dumping Severity			Total
	None	Slight	Moderate	
A	63.23	24.87	7.90	96
B	63.08	30.43	10.50	104
C	60.70	35.75	13.56	110
D	53.05	37.94	16.01	107

The first collapsing (corresponding to  $j=1$ ) of these counts is given in Table 6.6:

Table 6.6: Fitted counts, collapsed into dichotomous response ( $j=1$ )

Operation	Dumping Severity		Total
	None	Some (=Slight+ Moderate)	
A	63.23	32.77	96
B	63.08	40.93	104
C	60.70	49.31	110
D	53.05	53.95	107

The odds for some dumping instead of none for operation A is  $32.77/63.23 = 0.52$ ; for operation B it is  $40.93/63.08 = 0.64$ , so the odds that there will be some dumping rather than none is  $0.64/0.52 = 1.25$  times higher for operation B compared to A. This is of course just the local-global odds ratio

$$\hat{\theta}'_{11} = \frac{63.23 \times 40.93}{63.08 \times 32.77} = 1.25$$

The  $\hat{\theta}'_{21}$  and  $\hat{\theta}'_{31}$  are calculated in a similar manner to be 1.25. As always, this means that the odds for operation  $i+1$  is 1.25 times higher than the odds for operation  $i$ .

Operation	Odds for Some vs None
A	$32.77/63.23 = 0.52$
B	$40.93/63.08 = 0.64$
C	$49.31/60.70 = 0.81$
D	$53.91/53.05 = 1.02$

The second collapsing ( $j=2$ ) corresponds to a comparison of moderate dumping vs none plus slight, and the values of the estimated local-global odds ratios,  $\hat{\theta}'_{12}$ ,  $\hat{\theta}'_{22}$ ,  $\hat{\theta}'_{32}$  are also all 1.25.

#### 6.2.3.2.1.2 Conditional Test of Independence

Given that the logit uniform association model holds, one can test independence by testing  $H_0: \beta = 0$ . The test statistic is the difference between the deviances for the independence model and the logit uniform association model:

$$G^2(I|U) = G^2(I) - G^2(U),$$

which has  $df = (r-1)(c-1) - (rc-r-c) = 1$ .

The logit independence model

$$L_j(i) = \alpha_j$$

is equivalent to the loglinear independence model

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y$$

For the dumping severity data, the test of  $H_0: \beta = 0$  is based on  $G^2(I|U) = 10.88 - 4.27 = 6.61$ , with 1 df, which is significant at the 5% level.

#### 6.2.3.2.2 Ordinal-nominal tables

If  $X$  is nominal, we can form a logit analog of the loglinear row effects model:

$$\begin{aligned} L_j(i) &= \log \frac{m_{i,j+1} + \dots + m_{ic}}{m_{i1} + \dots + m_{ij}} \\ &= \alpha_j + \tau_i, \quad \begin{array}{l} 1 \leq i \leq r \\ 1 \leq j \leq c-1 \end{array} \end{aligned}$$

where  $\sum \tau_i = 0$ . This model is also analogous to the one-way ANOVA model for continuous response variables. In this logit row effects model the  $i$ th row effect,  $\tau_i$ , is assumed to be the same for all  $c-1$  ways of forming the cumulative logits and  $(c-1)+(r-1)$  independent parameters, so the residual  $df = (r-1)(c-2)$ . This is the same as the loglinear row effects model, but the two are not equivalent unless  $c=2$ .

For each pair of rows  $a$  and  $b$  the difference in logits

$$L_j(b) - L_j(a) = \tau_b - \tau_a$$

is constant for all  $c-1$  logits, so the log odds ratio for the  $2 \times 2$  table formed by taking rows  $a$  and  $b$  of the table and collapsing the response is constant for all  $c-1$  collapsings.

McCullagh (1980) gives an example of what he calls a proportional odds model, which is the same as this logit row effects model, with one explanatory variable, with two levels, and a response variable with three ordered categories.

#### 6.2.3.2.2.1 Dumping Severity Example

Once again using the data from Table 1.6, we can treat the row variable operation as nominal, and fit the logit row effects model. The model fits well, with  $G^2 = 3.56$  with 3  $df$ . The fitted counts are given in Table 6.7.

The estimates of the row effects parameters are:

$$\begin{aligned} \text{Operation A : } \hat{\tau}_1 &= 0.259 \\ \text{B : } \hat{\tau}_2 &= -0.251 \\ \text{C : } \hat{\tau}_3 &= 0.170 \\ \text{D : } \hat{\tau}_4 &= 0.341 \end{aligned}$$

Table 6.7: Fitted counts under logit row effects model

Operation	Dumping Severity		
	None	Slight	Moderate
A	61.57	25.97	8.45
B	66.51	28.27	9.23
C	59.17	36.62	14.21
D	53.00	37.99	16.01

The model predicts constant log odds ratios for the two collapsings of each pair of rows into 2x2 tables, so the difference between the row effect parameters of rows 3 and 4,  $\hat{\tau}_4 - \hat{\tau}_3 = 0.341 - 0.170 = 0.171$  means that the odds of dumping being some instead of none (or moderate instead of slight or less) are  $\exp(0.171) = 1.19$  times higher for operation D compared to A.

These odds ratios can also be obtained from the fitted values  $\{\hat{m}_{ij}\}$ . For example, Table 6.8 gives the first collapsing of the response, corresponding to  $j=1$ .

The estimated odds ratio for operation D vs C is  $1.02/0.86 = (59.17 \times 54.00)/(53.00 \times 50.83) = 1.19$ . Estimated odds ratios can be calculated similarly for the other operations. The values of the odds ratios obtained for this collapsing are identical to the values obtained for the other collapsing (corresponding to  $j=2$ ).

Table 6.8: Fitted counts, collapsed into dichotomy  
(j=1)

Operation	Dumping Severity		Odds (Some vs None)
	None	Some	
A	61.57	34.43	0.56
B	66.51	37.49	0.56
C	59.17	50.83	0.86
D	53.00	54.00	1.02

Treating operation as nominal and fitting the logit row effects model allows the odds ratios to be different for every pair of adjacent rows, rather than constraining them to be identical, as in the logit uniform association model.

#### 6.2.3.2.2.2 Conditional Test of Independence

Independence is the special case of the logit row effects model in which all the row effect parameters are zero. Given that the row effects model holds, a conditional test of independence can be based on  $G^2(I|R) = G^2(I) - G^2(R)$ , with  $df = (r-1)(c-1) - (r-1)(c-2) = r-1$ .

For the dumping severity data, the test of  $H_0: \tau_1 = \dots = \tau_r = 0$  is based on  $G^2(I|R) = 10.88 - 3.56 = 7.32$  with  $6-3 = 3$  df.

Since operation is actually ordinal and the logit uniform association model is just a special case of the row effects model (in which  $\tau_i = \beta(u_i - \bar{u})$ ), we can see whether we are losing much information by using the logit uniform association model rather than the row effects model, by comparing the deviances of the two models. In this case

$$\begin{aligned}
 G^2(U|R) &= G^2(U) - G^2(R) \\
 &= 4.27 - 3.56 \\
 &= 0.71
 \end{aligned}$$

with  $5-3 = 2$  df, which is not significant, so the improved fit due to the two extra parameters in the logit row effects model is not significant.

The analysis of deviance table for this data is given in Table 6.9:

Table 6.9: Analysis of deviance

Model	$L_j(i) = \log \frac{m_{i,j+1} + \dots + m_{ic}}{m_{i1} + \dots + m_{ij}}$	df	Deviance
Independence	$\alpha_j$	6	10.88
Logit uniform association	$\alpha_j + \beta(u_i - \bar{u})$	5	4.27
Logit row effects	$\alpha_j + \tau_i$	3	3.56

#### 6.2.4 Cumulative Logit Models for Higher Dimensions

The cumulative logit models discussed in Section 6.2.3 can be generalized to higher order tables where there may be both ordinal and nominal explanatory variables. These models resemble multiple regression models for continuous response variables. They are simpler to construct than the analogous loglinear models of Chapter IV, since it is unnecessary to model associations among the explanatory variables.

Anderson and Philips (1981) give an application of the cumulative logit model with multiple explanatory variables to the problem of discriminant analysis with an ordinal classification.



6.2.4.1 Homogeneous Linear Logit Effects

Consider an  $r \times c \times \ell$  table in which  $X$  and  $Y$  are explanatory variables and  $Z$  is an ordinal response. Within each of the  $r \times c$  combinations of  $X$  and  $Y$ , there are  $\ell-1$  cumulative logits.

$$L_k(ij) = \log \frac{m_{ijk+1+} \dots + m_{ij\ell}}{m_{ij1+} \dots + m_{ijk}}$$
$$k = 1, \dots, \ell-1$$

Table 6.10 lists association terms for some simple models that have linear effects of ordinal variables. None of these models allows for three-factor interaction. The parameters have similar interpretations to those given in Section 6.2.3.2, but in terms of partial associations.

Table 6.10: Association terms for cumulative logit models

Scales of Explanatory Variables		Association Terms	
X	Y	X	Y
Ordinal	Ordinal	$\beta^X(u_i - \bar{u})$	$\beta^Y(v_j - \bar{v})$
Nominal	Ordinal	$\tau_i^X$	$\beta^Y(v_j - \bar{v})$
Nominal	Nominal	$\tau_i^X$	$\tau_j^Y$

For example, when  $X$  and  $Y$  are nominal, the model used would be

$$L_k(ij) = \alpha_k + \tau_i^X + \tau_j^Y$$

where  $\sum \tau_i^X = \sum \tau_j^Y = 0$ . Differences between pairs of row effects  $\{\tau_i^X\}$  represent constant log odds ratios for comparing levels of  $X$  (controlling for  $Y$ ) on all  $\ell-1$  ways of collapsing the response  $Z$  into two categories. The  $\{\tau_j^Y\}$  can be interpreted similarly.

#### 6.2.4.1.1 Dumping Severity Example

Consider the three-dimensional version of the dumping severity data, given in Table 4.8. The sampling scheme is assumed to be product multinomial, with an independent multinomial sample being taken of the response dumping severity at each of the 16 combinations of the explanatory variables operation and hospital.

There are two logits at each of the 16 O-H combinations, giving a total of 32 logits. The model

$$L_k(ij) = \alpha_k, \quad k = 1, 2 \quad (6.1)$$

states that dumping severity is jointly independent of operation and hospital. This model is equivalent to the loglinear model

$$\log m_{ijk} = \mu + \lambda_i^O + \lambda_j^H + \lambda_k^D + \lambda_{ij}^{OH}$$

symbolised by (OH, D). Since the logit model has only two parameters, its residual df = 32-2 = 30, and  $G^2 = 31.64$ , so the model fits quite well.

The model

$$L_k(ij) = \alpha_k + \beta^O(u_i - \bar{u}) \quad (6.2)$$

assumes a linear effect of operation on the logit of dumping severity that is the same for both logits ( $k = 1, 2$ ) and the same for each hospital. It also assumes that dumping severity is conditionally independent of hospital for each operation. This model has only one more parameter than the independence model, and yields  $G^2 = 25.03$  based on 29 df when fitted with integer scores. There is a marked improvement in fit compared to the independence model, as  $G^2$  has decreased by  $31.64 - 25.03 = 6.61$  based on 1 df. The estimate of the association parameter is  $\beta^O = 0.225$ . Since this model implies that D is conditionally independent of H, the table can be collapsed over hospital without changing the O-D association if the model holds. This is why these results

regarding the O-D association are the same as those obtained with the logit uniform association model fitted to the marginal O-D table in Section 6.2.3.2.1.

The model

$$L_k(ij) = \alpha_k + \beta^O(u_i - \bar{u}) + \tau_j^H \quad (6.3)$$

where  $\sum \tau_j^H = 0$ , states that the logit of dumping severity is linearly related to operation and is also related to hospital through additive effects. This model assumes that each association parameter is homogeneous for the two ways of forming cumulative logits, and it assumes an absence of three-factor interaction. The model yields  $G^2 = 22.48$  with 26 df, so the improvement in fit over model (6.2) is only  $G^2 = 25.03 - 22.48 = 2.55$ , with  $29 - 26 = 3$  df, which is not significant. So for each operation there is no strong evidence that the distribution of dumping severity differs among the four hospitals.

The analysis of deviance table is given for these three models in Table 6.11:

Table 6.11: Analysis of Deviance

$L_k(ij) =$	df	$G^2$	Difference in df	Difference in $G^2$
$\alpha_k$	30	31.64		
$\alpha_k + \beta^O(u_i - \bar{u})$	29	25.03	1	6.61
$\alpha_k + \beta^O(u_i - \bar{u}) + \tau_j^H$	26	22.48	3	2.25

The loglinear model that is analogous to logit model (6.2) is

$$\log m_{ijk} = \mu + \lambda_i^O + \lambda_j^H + \lambda_k^D + \lambda_{ij}^{OH} + \beta^{OD}(u_i - \bar{u}) (w_k - \bar{w})$$

When fitted using integer scores, this model yields  $G^2 = 25.35$  with 29 df. This is very similar to logit model (6.2), although the models are not equivalent since the number of response categories exceeds two.

#### 6.2.4.2 Higher-Order Interaction Models

It is quite simple to formulate logit models that allow higher-order interactions. For example, suppose that the logit of  $Z$  is linearly related to  $Y$ , but that the slope of the relationship differs across the levels of a nominal variable  $X$ . An appropriate logit model would be

$$L_k(ij) = \alpha_k + \tau_i^X + \beta Y(v_j - \bar{v}) + \xi_i(v_j - \bar{v})$$

where  $\sum \tau_i^X = \sum \xi_i = 0$ . This model has residual df =  $rc - rc - l - 2r + 2$ .

#### 6.2.4.3 Heterogeneous Effects

A feature of the logit models discussed in this Section is the assumption that the effect of each explanatory variable is the same for the different ways of forming the cumulative logits. These models may be generalized to include non-homogeneous logit effects, which means that the effects of the explanatory variables change according to which logit is formed. Williams and Grizzle (1972) give an example of the application of a model that has non-homogeneous logit effects.

Models having homogeneous logit effects are easier to work with, since the effects of explanatory variables are easier to summarize and interpret. If homogeneous effect cumulative logit models fail to fit, it may still be useful to try fitting similar models having another logit form (e.g. adjacent-categories logits).

CHAPTER VII - MODELS BASED ON AN UNDERLYING DISTRIBUTION  
FOR THE RESPONSE

As in the previous three chapters, this chapter discusses the case where there is a single response measured on an ordinal scale, with one or more explanatory factors. The difference here is that we assume there is a continuous random variable,  $Y$ , underlying the observed categories of the response. We can formulate linear models that describe how the distribution of the underlying continuous unobserved response depends on the explanatory variable(s). ordinality is an integral feature of these models and so we do not need to impose arbitrary scores on the response categories.

The  $c$  response categories can be thought of as contiguous intervals on the continuous scale; the points of division or cut points will be denoted by  $\theta_1, \dots, \theta_{c-1}$ . The  $j$ th of the response categories is recorded when  $\theta_{j-1} < Y \leq \theta_j$ , where  $-\infty = \theta_0 < \theta_1 < \dots < \theta_{c-1} < \theta_c = \infty$ . For the extreme case where there is one explanatory variable with only one level, the probability that  $Y$  is less than or equal to a certain value  $\theta_j$  is the cumulative distribution function of  $Y$  evaluated at  $\theta_j$ ,

$$\phi(\theta_j) = \int_{-\infty}^{\theta_j} f(y) dy$$

(where  $f(y)$  is the probability density function of  $y$ ). The difference between  $\phi(\theta_j)$  and  $\phi(\theta_{j-1}) = \int_{-\infty}^{\theta_{j-1}} f(y) dy$

is the probability  $\pi_j$  that  $Y$  falls in category  $j$ , i.e.

$$\pi_j = \phi(\theta_j) - \phi(\theta_{j-1}) \quad j = 1, \dots, c.$$

For the case where there is one explanatory row variable, we model the expected value of the underlying response for the  $i$ th row as

$$\begin{aligned} E(Y_i) &= \underline{x}_i' \underline{\beta} \\ &= (x_{i0}, x_{i1}, \dots, x_{ip}) \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \end{aligned}$$

Now  $\underline{x}_i'\underline{\beta}$  is a general linear function which will depend on the way that the explanatory variable is expected to affect the response.  $\underline{\beta}$  is a vector of unknown parameters, and  $\underline{x}$  is a vector corresponding to the explanatory variable, consisting of dummy variables of 0's and 1's, or of real values.

To illustrate, consider a 3x4 table where the row variable A is purely nominal, and the column variable B is an ordinal response. To model association between A and B we would write

$$E(Y_1) = \beta_0 + \beta_1$$

$$E(Y_2) = \beta_0 + \beta_2$$

$$E(Y_3) = \beta_0 + \beta_3$$

where  $\beta_1 + \beta_2 + \beta_3 = 0$ . So the  $\underline{x}_i'$  contain dummy variables

$$\underline{x}'_1 = (1 \ 1 \ 0 \ 0)$$

$$\underline{x}'_2 = (1 \ 0 \ 1 \ 0)$$

$$\underline{x}'_3 = (1 \ 0 \ 0 \ 1)$$

$$\text{and } \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

but if the row variable is ordinal, with assigned scores -1, 0, 1, and we expect a linear effect of A on Y, then we would model the association as

$$E(Y_1) = \beta_0 - 1 \times \beta_1$$

$$E(Y_2) = \beta_0 + 0 \times \beta_1$$

$$E(Y_3) = \beta_0 + 1 \times \beta_1$$

So the  $\underline{x}_i'$  are

$$\underline{x}'_1 = (1, -1)$$

$$\underline{x}'_2 = (1, 0)$$

$$\underline{x}'_3 = (1, 1)$$

$$\text{and } \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

The probability that for row  $i$  the response falls into category  $j$  is

$$\pi_{ij} = \frac{\exp(\theta_j - \underline{x}'_i \underline{\beta}_j)}{\sum_{j=1}^r \exp(\theta_j - \underline{x}'_i \underline{\beta}_j)}$$

The counts  $(n_{i1}, n_{i2}, \dots, n_{ic})$  recorded in response categories  $(1, 2, \dots, c)$  out of  $n_{i+}$  individuals in row  $i$  are multinomial with probabilities  $\pi_{ij}$ . The log likelihood is

$$\log L(\underline{\beta}', \theta_1, \dots, \theta_{c-1}) = \sum_i \sum_j n_{ij} \log \pi_{ij} + \sum_i \log \binom{n_{i+}}{n_{i1} \dots n_{ic}}$$

and the unknown parameters can be estimated by maximum likelihood.

For multi-dimensional tables with two or more explanatory variables, we can just think of the combinations of the explanatory variables as being arranged together to form a single row variable. Then the expected value of the response for row  $i$  can still be modelled as

$$E(Y_i) = \underline{x}'_i \underline{\beta}$$

where  $\underline{x}'_i$  pertains to the combination of explanatory variables that gives rise to row  $i$ .

## 7.1 Distribution Functions

The shape of the distribution of the underlying response will obviously affect the proportions in the observed ordered categories. The probability density function (pdf) of  $Y_i$  could be either symmetric or skewed. If symmetric, then for a particular row  $i$ , if the expected value of  $Y_i$ ,  $\underline{x}'_i \underline{\beta}$ , is near the middle of the  $c$  ordered categories, the sum of the counts to the left of  $\underline{x}'_i \underline{\beta}$  would be expected to be roughly equal to the sum of the counts to the right. If the pdf of  $Y_i$  is negatively skewed, i.e. there is a sharp drop to the right of the mode, there will be a threshold which  $\underline{x}'_i \underline{\beta}$  must reach before the higher response categories are recorded. A slower drop in the pdf means that increases in  $\underline{x}'_i \underline{\beta}$  are accompanied by more gradual increases in the proportions in higher response categories. The opposite situation will occur with a positively skewed pdf. The tails of the distribution are also important since a long tail to the right implies

that high ordinal response categories are possible even when  $x_i'\beta$  is low and similarly in the left tail. Short tails imply that the counts will be more tightly clustered about  $x_i'\beta$ , and outliers will be very rare.

Some common distributions that are used are the normal, logistic or extreme value distributions. If there are only  $c=2$  response categories, these correspond to probit, logit and complementary-log-log models respectively for the binomial response.

### 7.1.1 Normal

The normal pdf (see Figure 7.1) is

$$f_N(y) = (2\pi)^{-1/2} \exp(-1/2 y^2)$$

with corresponding cumulative distribution function (cdf)

$$\Phi_N(y) = \int_{-\infty}^y (2\pi)^{-1/2} \exp(-1/2 t^2) dt$$

The normal distribution is often chosen because normality is the standard regression assumption for a quantitative response.

However, the short tails of the normal distribution make outliers unlikely, leading to a lack of fit for many tables.

### 7.1.2 Logistic

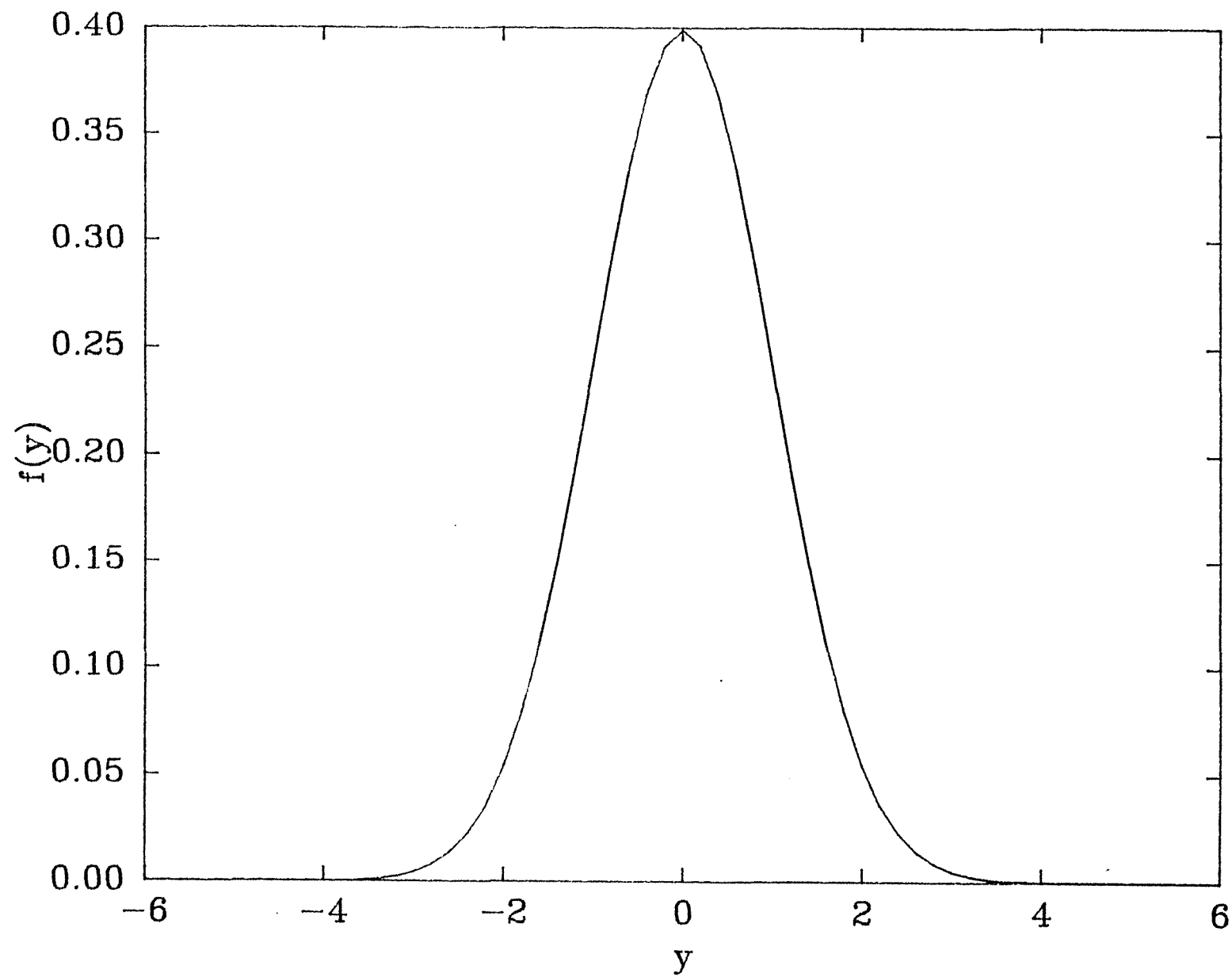
The logistic distribution is also symmetric and has a similar shape to the normal distribution (see Figure 7.2), but has longer tails, so it may fit better than the normal distribution. The logistic pdf is

$$f_L(y) = \frac{\exp(-y)}{(1 + \exp(-y))^2}$$

with cdf



Figure 7.1 Normal Probability Density Function



$$\phi_L(y) = \frac{1}{1 + \exp(-y)}$$

Using the logistic distribution gives exactly the same results as the cumulative logit models discussed in Section 6.2.3 (Agresti, 1984, p.151).

As before, the local-global odds ratio for rows a and b is

$$\begin{aligned} \theta'_{ab} &= \frac{\text{odds that category} > j, \text{ for row } b}{\text{odds that category} > j, \text{ for row } a} \\ &= \frac{P(\text{category} > j \mid \text{row } b)}{P(\text{category} \leq j \mid \text{row } b)} \times \frac{P(\text{category} > j \mid \text{row } a)}{P(\text{category} \leq j \mid \text{row } a)} \\ &= \frac{(\pi_{b,j+1} + \dots + \pi_{bc})}{(\pi_{b1} + \dots + \pi_{bj})} \times \frac{(\pi_{a1} + \dots + \pi_{aj})}{(\pi_{1,j+1} + \dots + \pi_{ac})} \\ &= \exp((\mathbf{x}_b' - \mathbf{x}_a') \boldsymbol{\beta}) \end{aligned}$$

If the row variable is nominal, with parameters  $\beta_a$  and  $\beta_b$  for levels a and b, the odds ratio is simply

$$\exp(\beta_b - \beta_a)$$

So fitting the logistic distribution means the parameters can be interpreted in terms of odds ratios. Anderson and Phillips (1981) and McCullagh (1980) give examples of fitting the logistic distribution.

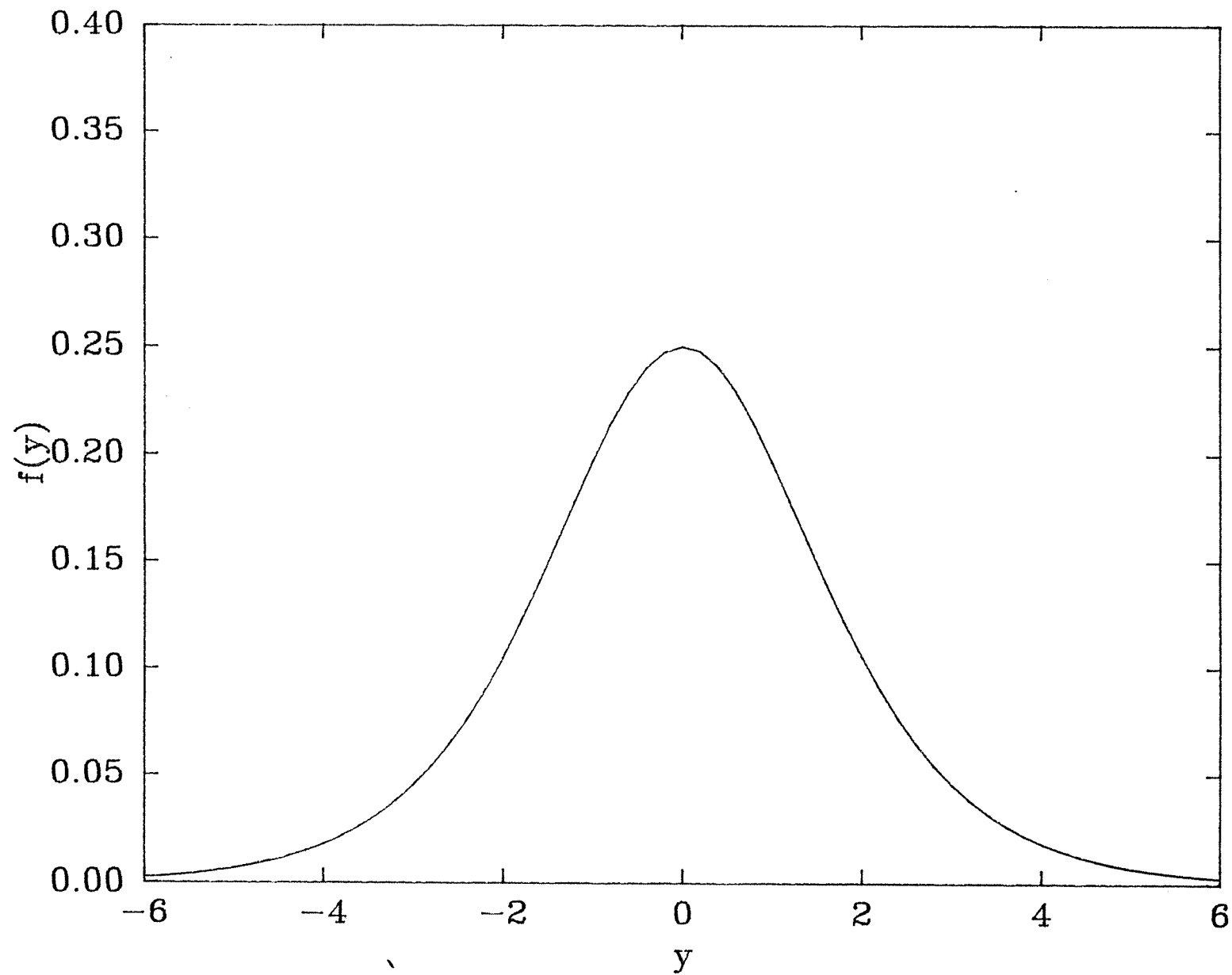
### 7.1.3 Extreme Value

If it is thought that the underlying distribution is skew, then the extreme value distribution (or its mirror image, the negative extreme value distribution) can be tried. The extreme value pdf is

$$f_E(y) = \exp(y - \exp(y))$$

with cdf

Figure 7.2 Logistic Probability Density Function



$$\phi_E(y) = 1 - \exp(-\exp(y))$$

The negative extreme value pdf is

$$f_{NE}(y) = \exp(-y - \exp(-y))$$

with cdf

$$\phi_{NE}(y) = \exp(-\exp(-y))$$

The extreme value distribution is negatively skewed, i.e. there is a very steep drop and short tail to the right of the mode, but a long gradual tail to the left (see Figure 7.3), and the negative extreme value distribution is positively skewed (see Figure 7.4). If  $y_i$  follows the extreme value distribution, then it satisfies a proportional hazards models (McCullagh, 1980), with consequent ease of interpretation of the parameters. The ratio of

$$\begin{aligned} \frac{\log P(\text{category} > j \mid \text{row } b)}{\log P(\text{category} > j \mid \text{row } a)} &= \frac{\log (\pi_{b,j+1} + \dots + \pi_{bc})}{\log (\pi_{a,j+1} + \dots + \pi_{ac})} \\ &= \exp((\mathbf{x}_a' - \mathbf{x}_b')\beta) \end{aligned}$$

which also means that

$$\begin{aligned} P(\text{category} > j \mid \text{row } b) &= \\ P(\text{category} > j \mid \text{row } a) \exp((\mathbf{x}_a' - \mathbf{x}_b')\beta) \end{aligned}$$

#### 7.1.4 Estimation

Stirling (1984) describes an iteratively re-weighted least squares algorithm implemented as a Genstat macro that will fit linear models assuming underlying distributions that are normal, logistic, extreme value, negative extreme value or any mixture of these. He analyses a contingency table with 126 cells using a model with only eight parameters.

Figure 7.3 Extreme Value Probability Density Function

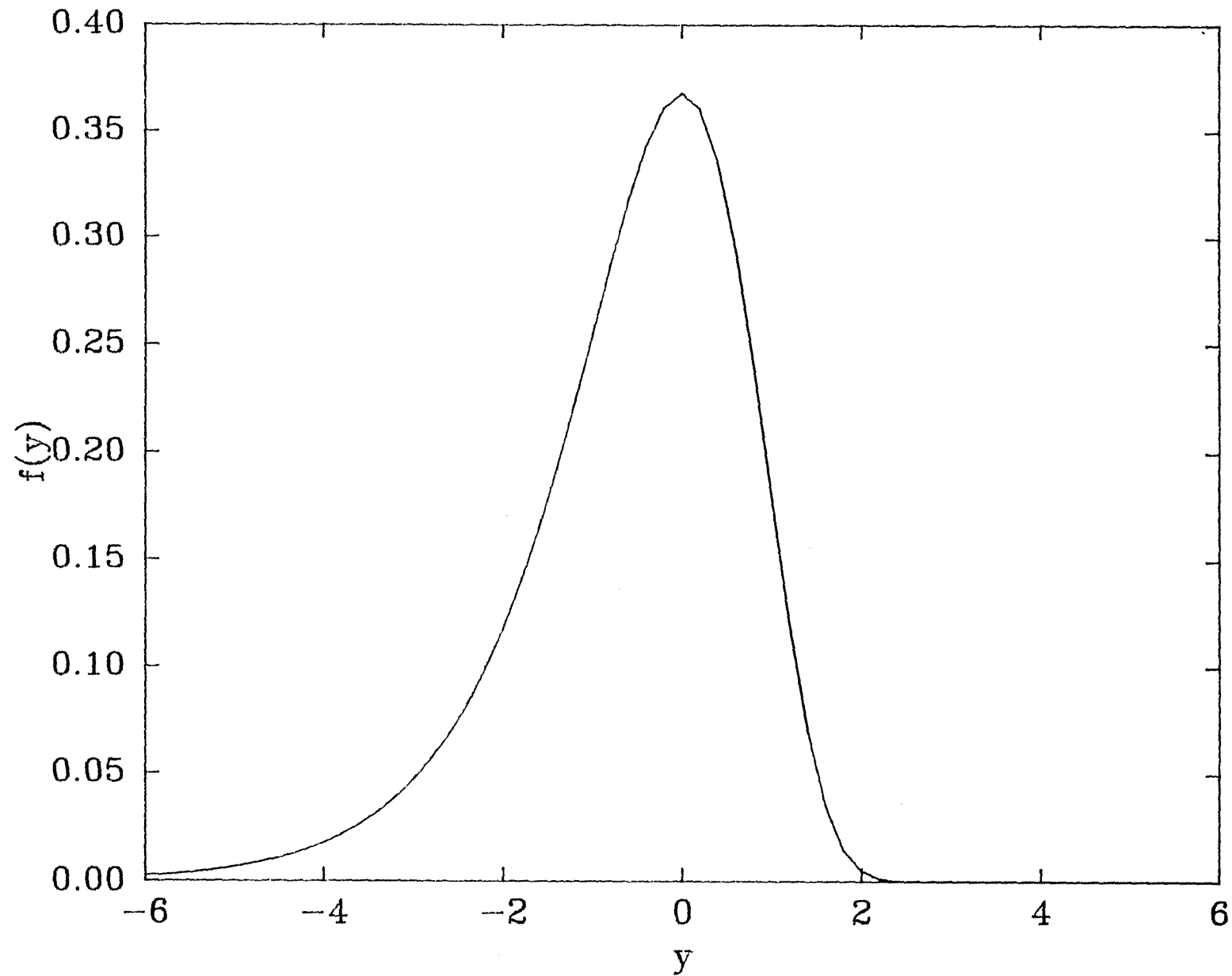
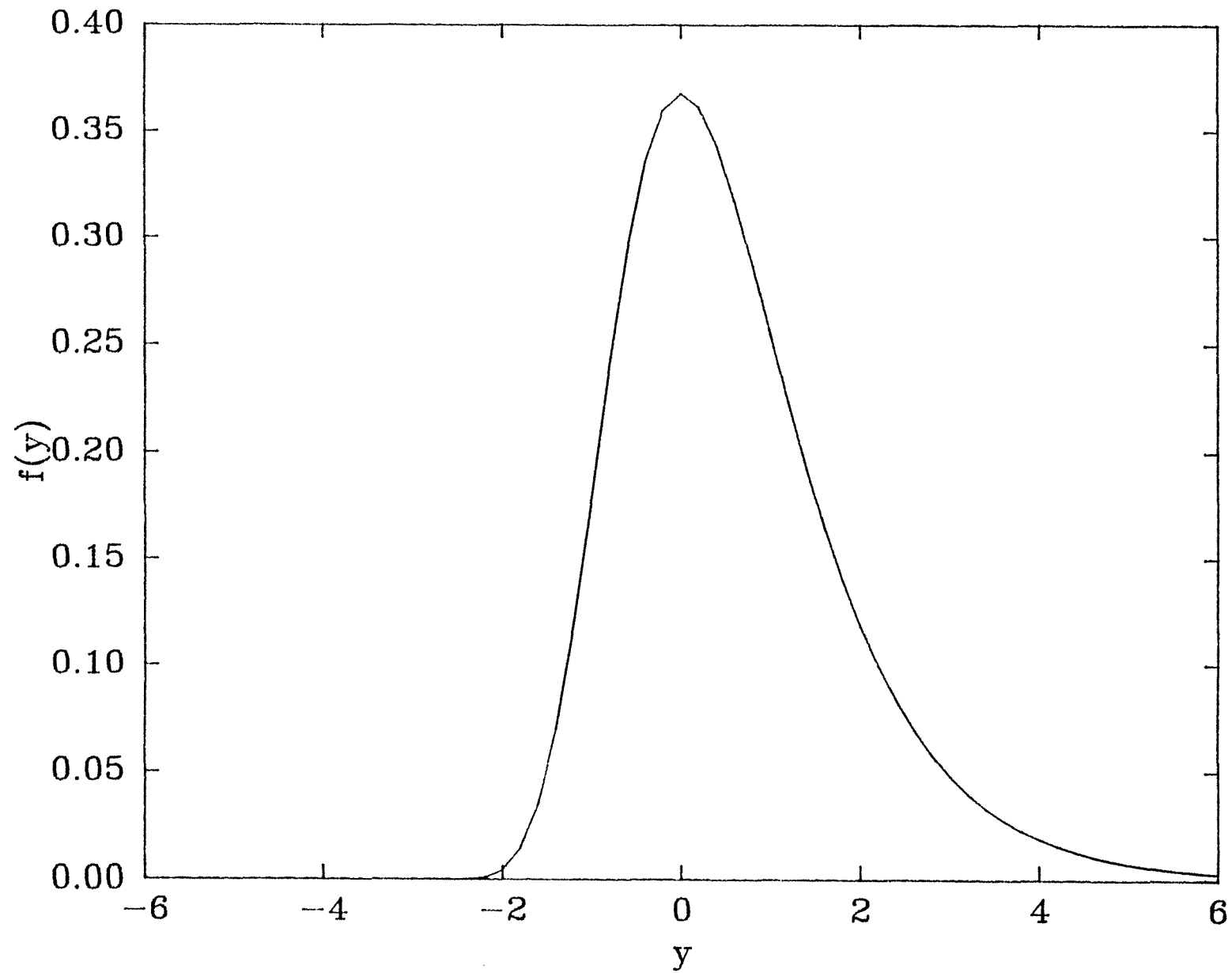


Figure 7.4 Negative Extreme Value Probability Density Function



### 7.1.5 Dumping Severity Example

We will re-analyse the data of Table 1.6 by this method. To choose a suitable distribution, we fit a model where the linear part  $\underline{x}_i'\beta$  is fully parameterized so that each  $\mu_i = \underline{x}_i'\beta$  is allowed to vary freely. This corresponds to treating operation as nominal and fitting

$$E(y_i) = \mu + \tau_i$$

$\sum \tau_i = 0$ , for the four standard distributions. Table 7.1 gives the deviances for these models. There are six independent parameters ( $\mu, \tau_i, \theta_j$ ), three independent constraints (the four row totals), and 12 cells; therefore  $df = 12 - 9 = 3$ .

Table 7.1: Fully parameterized models

Distribution	df	Deviance
Normal	3	4.10 ns
Logistic	3	3.56 ns
Extreme value	3	5.35 ns
Negative extreme value	3	3.17 ns

All these distributions fit quite well.

Treating operation as ordinal, we can fit the model

$$E(y_i) = \mu + \beta(u_i - \bar{u})$$

using integer scores. This gives the deviances in Table 7.2.

Table 7.2: Using integer scores for operation

Distribution	df	Deviance
Normal	5	4.44 ns
Logistic	5	4.27 ns
Extreme value	5	5.37 ns
Negative extreme value	5	4.23 ns

Once again, all distributions seem to fit quite well. In a small table there is often not enough information to choose between different distributions.

Since the negative extreme value distribution gave the smallest deviance in both tables, and has easily interpretable parameters, we will examine the fitted counts for the ordinal model, which are in Table 7.3.

Table 7.3: Fitted counts under the negative extreme value distribution

Operation	Dumping Severity			Total
	Moderate	Slight	None	
A	8.38	24.29	63.32	96
B	10.75	29.98	63.26	104
C	13.45	35.82	60.74	110
D	15.43	38.93	52.63	107

To fit the negative extreme value distribution we can just reverse



the ordering of the response categories and fit the extreme value distribution, so the order of dumping severity in Table 7.3 is the reverse of the usual order.

The parameter estimates are

$$\hat{\mu} = 2.126, \hat{\beta} = 0.178, \hat{\theta}_1 = 0, \hat{\theta}_2 = 1.516.$$

The expected values of  $y$  for operations A to D are:

Operation	$\mu + \beta(u_i - \bar{u}) = \hat{y}_i$
A	$2.126 - 0.178 \times -1.5 = 2.39$
B	$2.126 - 0.178 \times -0.5 = 2.21$
C	$2.126 - 0.178 \times 0.5 = 2.04$
D	$2.126 - 0.178 \times 1.5 = 1.86$

so the underlying extreme value distribution is shifted to the left with increasing removal of stomach (bearing in mind that the category order has been reversed so that the larger  $y$  is the more desirable response, i.e. less side effect (see Figure 7.5). For the extreme value distribution, instead of interpreting the  $\beta$  parameter in terms of odds ratios, as we have done with loglinear and logit models in previous chapters, we interpret it in terms of proportions. If we collapse dumping severity into a dichotomy of moderate vs slight plus none (which corresponds to  $j = 1$ , see Table 7.4), then the log proportions of S+N for each adjacent row are related by  $\exp(-\hat{\beta}) = \exp(+0.178) = 1.95$ , i.e. Propn (S+N) for row  $i+1 = [\text{Propn (S+N) for row } i] \times 1.95$ .

For example, the proportion of patients who had slight or no dumping after operation A is  $87.62/96 = 0.913$ . The corresponding proportion for operation B is  $93.25/104 = 0.897$ . The ratio of the log proportions is  $\log(0.897)/\log(0.913) = -0.109/-0.091 = 1.195$ , and of course  $0.897 = 0.913 \times 1.195$ .

Figure 7.5 Extreme Value pdf for Operations D to A (M,S,N)

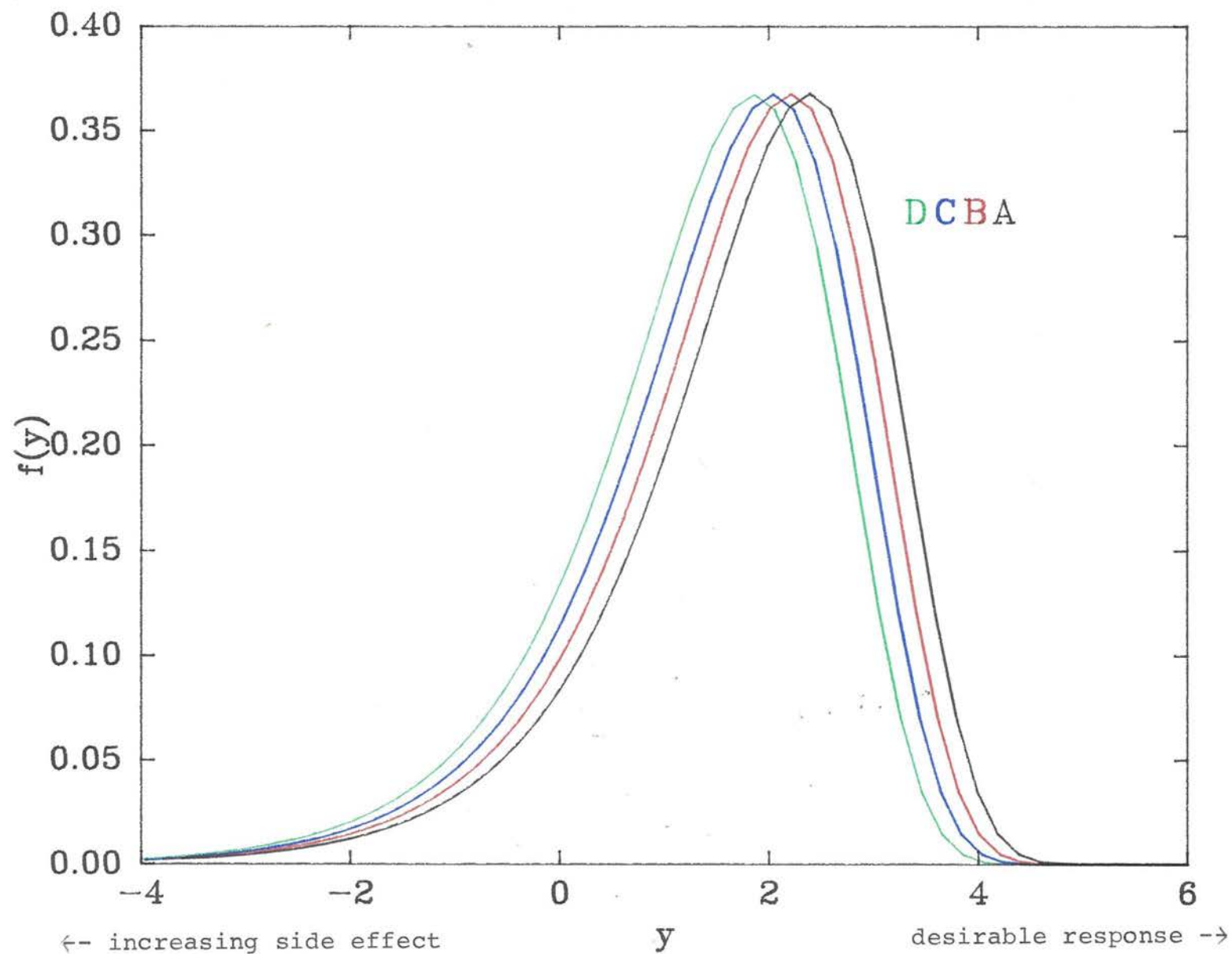


Table 7.4: Fitted counts, collapsed into dichotomy (j=1)

Operation	Dumping Severity		Total	Propn S+N	log (Propn S+N)
	Moderate	Slight or None			
A	8.38	87.62	96	0.913	-0.091
B	10.75	93.25	104	0.897	-0.109
C	13.45	96.55	110	0.878	-0.130
D	15.43	91.57	107	0.856	-0.155

The same relationship holds true for the other collapsing (corresponding to j=2) which compares moderate plus slight vs none (see Table 7.5).

Table 7.5: Fitted counts, collapsed into dichotomy (j=2)

Operation	Dumping Severity		Propn M+S	log (Propn M+S)
	Moderate or Slight	None		
A	32.68	63.32	0.660	-0.416
B	40.74	63.26	0.609	-0.497
C	49.26	60.74	0.552	-0.594
D	29.37	52.63	0.492	-0.710

A similar interpretation holds for fitting the nominal model

$$E(y_i) = \mu + \tau_i$$

The fitted counts under the negative extreme value distribution are in Table 7.6; they are collapsed into a dichotomy of moderate vs slight plus none (corresponding to j=1) in Table 7.7.

Table 7.6: Fitted counts

Operation	Dumping Severity			Total
	Moderate	Slight	None	
A	8.97	25.67	61.36	96
B	9.44	27.19	67.37	104
C	14.21	37.27	58.52	110
D	15.34	38.83	52.83	107

Table 7.7: Fitted counts, collapsed into dichotomy (j=1)

Operation	Dumping Severity		Total	Propn S+N	log (Propn S+N)
	Moderate	Slight or None			
A	8.97	87.03	96	.907	-.098
B	9.44	94.56	104	.909	-.095
C	14.21	95.79	110	.871	-.138
D	15.34	91.66	107	.857	-.154

The  $\tau$  parameters are estimated as

$$\hat{\tau}_A = 0.192$$

$$\hat{\tau}_B = 0.223$$

$$\hat{\tau}_C = -.0152$$

$$\hat{\tau}_D = -0.263$$

The expected log proportions for different rows are related by the exponential of the difference in parameters, e.g. for row D compared to row C

$$\frac{\log \text{propn (S+N) for row D}}{\log \text{propn (S+N) for row C}} = \exp (\hat{\tau}_C - \hat{\tau}_D)$$

$$\text{or propn for D} = (\text{propn for C}) \exp (\hat{\tau}_C - \hat{\tau}_D)$$

$$= 0.871 \exp (0.111)$$

$$= 0.857$$

## CHAPTER VIII - OTHER MODELS

### 8.1 Mean Response Models

Mean response models are appropriate for tables with a single ordinal response. We assign scores to the levels of the response variable and use its mean as the response function in a regression-type model.

For example, consider a two-way table classified by ordinal variables  $X$  and  $Y$  having scores  $\{u_i\}$  and  $\{v_j\}$  respectively. Within level  $i$  of  $X$ , the conditional mean of  $Y$  is

$$\phi = \sum_j \frac{v_j m_{ij}}{n_{i+}}$$

$$i = 1, \dots, r$$

The usual linear regression model is  $\phi_i = \alpha + \beta(u_i - \bar{u})$

The parameter  $\alpha$  is the average of the conditional means, and  $\beta$  is the change in the conditional mean per unit change in  $X$ . There are  $r$  responses and two parameters, so  $df = r-2$ , and we need at least three rows to obtain an unsaturated model. The model for the ordinal nominal table is saturated.

The mean response model is a linear model and can be fitted by the general algorithm described in Section 1.11 and Appendix 3.

#### 8.1.1 Dumping Severity Example

Consider again the 4x3 operation by dumping severity cross-classification of Table 1.6. Treating both variables as ordinal and assigning integer scores  $\{u_i=i\}$  and  $\{v_j=j\}$  (somewhat arbitrarily), we could fit the mean response model

$$\phi_i = \alpha + \beta u_i$$

where  $\phi_i$  is the conditional mean of the dumping severity score,

$$\phi_i = (m_{i1} + 2 m_{i2} + 3 m_{i3}) / m_{i+}$$

$$i = 1, \dots, 4.$$

Table 8.1 gives the fitted counts obtained under this model.

Table 8.1: Fitted counts under mean response model

Operation	Dumping Severity		
	None	Slight	Moderate
A	61.89	27.46	6.65
B	66.19	23.65	14.16
C	58.50	39.75	11.75
D	53.26	37.90	15.84

The estimates of the mean response scores are

$$\hat{\phi}_1 = (61.89 + 2 \times 27.46 + 3 \times 6.65) / 196 = 1.425$$

$$\hat{\phi}_2 = 1.500$$

$$\hat{\phi}_3 = 1.575$$

$$\hat{\phi}_4 = 1.650$$

The parameter estimates are  $\mu = 1.538$  and  $\beta = 0.075$ , which indicate that for each additional 25% of stomach removal, the predicted mean operation score is increased by 0.075. This model has deviance  $G^2 = 0.22$  with a 2 df, which indicates that it fits very well.

### 8.1.2 Three-way Dumping Severity Example

Corresponding models for higher-order tables are easy to construct and interpret. Haber (1985, p.5) gives an analysis of the data of Table 4.8, using main effects for both hospital and operation in a mean response model

$$\begin{aligned}\phi_{ij} &= \mu + \alpha_i + \beta_j \\ i, j &= 1, 2, 3, 4\end{aligned}$$

He shows that operation significantly affects dumping severity, while the four hospitals do not significantly differ with regard to their effect.

## 8.2 Models for Incomplete Tables

The concept of structural zeroes was introduced in Section 1.9. Structural zeroes occur when it is impossible to observe values for certain combinations of the variables. Removing structural zeroes (or other pre-determined values) from the analysis results in a table that is said to be structurally incomplete.

### 8.2.1 Definitions

Let  $S$  be the set of cells in a two-way  $r \times c$  array that remain after the exclusion of missing entries and fixed values. For those cells not in  $S$  we put  $n_{ij} = m_{ij} = 0$ . The general loglinear model can still be obtained by setting

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_{ij}^{XY}$$

for cells  $(i, j) \in S$ , with  $\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0$

$$\text{and } \sum_i \delta_{ij} \lambda_{ij}^{XY} = \sum_j \delta_{ij} \lambda_{ij}^{XY} = 0$$

where  $\delta_{ij} = \begin{cases} 1 & \text{for } (i, j) \in S \\ 0 & \text{otherwise} \end{cases}$

### 8.2.2 Quasi-Independence

The model of quasi-independence can be defined by setting

$$\lambda_{ij}^{XY} = 0 \text{ for } (i, j) \in S$$

so that  $\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y$  for  $(i, j) \in S$ .



In other words,  $X$  and  $Y$  are quasi-independence if we can write the expected counts in the form

$$m_{ij} = \begin{cases} a_i b_j & \text{for } (i,j) \in S \\ 0 & \text{otherwise} \end{cases}$$

so quasi-independence is like independence, so it applies to the non-empty cells of a table.

The general rule for computing degrees of freedom is still applicable here. If there are  $e$  cells that have missing or fixed entries, then the set  $S$  contains  $rc - e$  cells, so the saturated model needs only  $rc - 1$  parameters. The number of parameters in the quasi-independence model is  $1 + (r-1) + (c-1) = r + c - 1$ , so the degrees of freedom are  $rc - e - (r + c - 1) = (r-1)(c-1) - e$ .

### 8.2.3 Estimation

The maximum likelihood estimates under the model of quasi-independence are the same under all three of the usual sampling schemes and are usually unaffected by the presence of sampling zeroes, providing no row or column has an observed zero total. The MLES can be computed by the general log-linear algorithm.

### 8.2.4 Monkey Example

Let us consider the data of Table 1.8 which concerns the genital display of monkeys. Fitting the model of quasi-independence

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^P \\ (i,j) \in S$$

to the data yields the fitted values in Table 8.2. Inspection of this table reveals that the quasi-independence model does not fit well. This is also confirmed by the large value of the deviance which is  $G^2 = 135.17$  with 15 df. There are  $1+4+5 = 10$  parameters in this model, and 25 cells in  $S$ , therefore 25 parameters in the saturated model, which is why the degrees of freedom equals  $25 - 10 = 15$ .

Table 8.2: Fitted counts under the quasi-independence model

Active Participant	Passive Participant					
	R	S	T	U	V	W
R	-	5.26	2.48	8.22	6.65	0.40
S	9.19	-	10.32	34.19	27.66	1.65
U	0.94	12.47	5.88	-	15.77	0.94
V	0.22	0.25	0.12	0.39	-	0.02
W	9.66	11.02	5.20	17.21	13.92	-

Since the quasi-independence model does not fit, we can conclude the various monkeys choose to display themselves more often towards specific members of the colony.

### 8.2.5 Higher-Order Tables

When we deal with multi-dimensional tables, we can consider log-linear models applied only to cells whose values are not structural zeroes or fixed values. We will consider three-way tables; these methods can be easily extended to higher-order tables. Let  $S$  represent the cells in an incomplete  $r \times c \times l$  table with observed values  $n_{ijk}$  and expected values  $m_{ijk}$ . We set  $n_{ijk} = m_{ijk} = 0$  for  $(i,j,k) \notin S$ . The most general loglinear model is

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

where  $(i,j,k) \in S$ . The  $\lambda$ -terms are deviations and sum to zero over each included variable. For example:

$$\begin{aligned} \sum_i \delta_i^{YZ} \lambda_i^X &= \sum_i \delta_{ij}^Z \lambda_{ij}^{XY} \\ &= \sum_i \delta_{ik}^Y \lambda_{ik}^{YZ} = \sum_i \delta_{ijk} \lambda_{ijk}^{XYZ} = 0 \end{aligned}$$

with  $\delta_{ijk} = \begin{cases} 1 & \text{if } (i,j,k) \in S \\ 0 & \text{otherwise} \end{cases}$

$\delta_{ij}^Z = \begin{cases} 1 & \text{if } \delta_{ijk} = 1 \text{ for some } k \\ 0 & \text{otherwise} \end{cases}$

$\delta_{ik}^Y = \begin{cases} 1 & \text{if } \delta_{ijk} = 1 \text{ for some } j \\ 0 & \text{otherwise} \end{cases}$

$\delta_i^{YZ} = \begin{cases} 1 & \text{if } \delta_{ijk}=1 \text{ for some } (j,k) \\ 0 & \text{otherwise} \end{cases}$

We define various unsaturated loglinear models by setting  $\lambda$ -terms equal to zero, and, as in the analysis of complete tables, we restrict our attention to hierarchical models.

The formula for the degrees of freedom of a particular model is once again the difference between the number of parameters in the saturated model and the number of estimable parameters in the model in question. The word "estimable" has been used here to emphasize that when dealing with structural zeroes, sometimes not all parameters in the quasi-loglinear model are table to be estimated. In particular, this will happen when some structural zeroes make up a zero marginal total which is used in a model.

#### 8.2.5.1 Health concerns example

Table 8.3, taken from Fienberg (1980, p.148), gives the results of a survey enquiring into the health concerns of teenagers. The respondents are classified by sex, age and health concern.

Table 8.3: Health concerns of teenagers, cross-classified by age and sex

Health concerns	Males		Females	
	12-15	16-17	12-15	16-17
Sex, reproduction	4	2	9	7
Menstrual problems	-	-	4	8
How healthy I am	42	7	19	10
Nothing	57	20	71	31

Since males do not menstruate, there are two structural zeroes in the table, and a structural zero in the sex by health concern two-dimensional marginal table.

We would normally treat age and sex as explanatory factors which could affect the response health concerns. Thus, any appropriate loglinear models would include the  $\lambda_{ij}^{SA}$  term. Table 8.4 gives the five possible models that could be fitted to this data.

Table 8.4: Models for health concerns data

Model	$G^2$	df
(SA, H)	22.03	7
(SA, SH)	9.43	5
(SA, AH)	13.45	4
(SA, SH, AH)	2.03	2
(SAH)	0	0

The simplest model that fits adequately is (SA, SH), i.e.

$$\log m_{ijk} = \mu + \lambda_i^S + \lambda_j^A + \lambda_k^H + \lambda_{ij}^{SA} + \lambda_{ik}^{SA}$$

for  $(i,j,k) \in S$ .

This model specifies that the age of the respondent does not affect the health concern expressed, but that the sex does. Fitting this model involves the marginal tables  $n_{ij+}$  and  $n_{i+k}$ . The two-dimensional sex by health concerns marginal table is given in Table 8.5. The marginal total  $n_{1+2}$  is a structural zero (because the two cells which it is the total of are both structural zeroes) and so the parameter  $\lambda_{12}^{SH}$  is non-estimable. Thus, the number of estimable parameters is  $1+1+1+3+1+2 = 9$ .

Table 8.5: Sex by health concerns marginal table

Health concerns	Sex	
	Male	Female
Sex, reproduction	6	16
Menstrual problems	-	12
How healthy I am	49	29
Nothing	77	102

There are 14 cells in the table that are not structural zeroes; therefore the saturated model needs only 14 parameters. The df for the model is thus  $14-9 = 5$ .

### 8.3 Models for Square Tables

Square tables are two-dimensional tables where the variable for rows has the same categories as the variable for columns. The following are examples of ways in which such tables might arise:

1. In panel studies where each individual is classified at two points in time;
2. When pairs of matched individuals, such as husbands and wives, are classified, for each member of the pair;
3. When individuals are classified according to two essentially similar variables (e.g. strength of right hand and strength of left hand);
4. In experiments on matched pairs, the members of a pair are subjected to different treatments.

Square tables correspond closely to paired quantitative data.

### 8.3.1 Quasi-Independence

Sometimes we might want to treat the diagonal cells differently from the others. For instance, we may expect independence to be a feasible model for the off-diagonal cells, but not for the diagonal cells.

Consider an  $r \times r$  square table classified by variables  $X$  and  $Y$ . The loglinear formulation of quasi-independence is

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

$$\text{where } \lambda_{ij} = \begin{cases} \lambda_i^* & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

For the off-diagonal cells this is simply the model of independence

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y, \quad i \neq j$$

while for the diagonal cells the model is saturated

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_i^* + \lambda_j^Y, \quad i = j$$

and the estimated expected frequencies for the diagonal cells equal the observed frequencies, i.e.  $\hat{m}_{ii} = n_{ii}$ .

The easiest way to fit this model is just to treat the diagonal cells as structural zeroes and fit the model of quasi-independence as described in Section 8.2.

The number of parameters in this model is  $1 + (r-1) + (r-1) = 3r - 1$ , the number in the saturated model is  $r^2$ , therefore the degrees of freedom equal  $r^2 - 3r + 1$ .

### 8.3.1.1 Social Mobility Example

Table 8.6 presents data on inter-generational social mobility in Britain (Bishop, Fienberg and Holland, 1975, p.206). Because of the nature of social mobility it seems more likely that quasi-independence, rather than independence, will hold in the table, i.e. we would expect a disproportionate number of sons to have the same social status as their fathers, but the status of the sons whose status was different from their fathers might not be affected by their fathers' status.

Table 8.6: British social mobility data

Fathers' status	Son's status				
	1	2	3	4	5
1	50	45	8	18	8
2	28	174	84	154	55
3	11	78	110	223	96
4	14	150	185	714	447
5	0	42	72	320	411

Table 8.7 gives the fitted counts under model of complete independence. This model obviously does not fit well, and was  $G^2 = 811.0$  with 16 df.

The model of quasi-independence also does not fit well, with  $G^2 = 249.4$  with  $25-15+1 = 11$  df; however, it is a vast improvement over the plain independence model. The fitted counts are given in Table 8.8.

Table 8.7: Fitted counts under complete independence

Fathers' status	Son's status				
	1	2	3	4	5
1	3.8	18.0	16.9	52.7	37.5
2	14.6	69.2	65.0	202.3	144.0
3	15.3	72.4	68.0	211.7	150.6
4	44.5	211.2	198.2	617.0	439.1
5	24.9	118.2	110.9	345.3	245.7

Table 8.8: Fitted counts under quasi-independence

Fathers' status	Son's status				
	1	2	3	4	5
1	50.0	9.5	11.0	38.5	20.0
2	6.6	174.0	50.0	174.2	90.3
3	8.5	56.1	110.0	226.2	117.3
4	27.5	181.0	209.0	714.0	378.5
5	10.4	68.4	79.1	276.1	411.0

Bishop, Fienberg and Holland (1975, p.207) discuss more complicated models which involve separate consideration of the triangular sub-tables above and below the diagonal.



### 8.3.2 Symmetry

This model specifies that the table is symmetric about the diagonal, i.e.

$$m_{ij} = m_{ji} \text{ for all } i \neq j$$

The loglinear representation of symmetry is

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^X + \lambda_{ij}^{XY}$$

$$\begin{aligned} \text{where } \lambda_{ij}^{XY} &= \lambda_{ji}^{XY} \\ \text{and } \sum_i \lambda_i^X &= \sum_i \lambda_{ij}^{XY} = 0 \end{aligned}$$

The maximum likelihood estimates are

$$\hat{m}_{ij} = \begin{cases} \frac{n_{ij} + n_{ji}}{2} & i \neq j \\ n_{ii} & i = j \end{cases}$$

There are  $r(r-1)$  off-diagonal cells; therefore there are  $r(r-1)/2$  constraints on the fitted values, so the degrees of freedom are  $r(r-1)/2$ .

#### 8.3.2.1 Eye-Testing Example

Table 8.9 gives the results from eye tests on women. The table is taken from Bishop, Fienberg and Holland (1975, p.284). The model of symmetry specifies that the expected number of women with right eye grade  $i$  and left eye grade  $j$  equals the expected number of women with right eye grade  $j$  and left eye grade  $i$ . In other words, we expect to have the same number of women with, say, excellent right eye vision but terrible left vision, as those with excellent left vision but terrible right vision. The fitted counts under this model are given in Table 8.10. Along the diagonal the fitted counts are equal to the actual counts, while each fitted off-diagonal count,  $\hat{m}_{ij}$ , is equal to the average of  $n_{ij}$  and  $n_{ji}$ , e.g.

$$\begin{aligned}\hat{m}_{41} &= (n_{41} + n_{14}) / 2 \\ &= (36 + 66) / 2 = 51\end{aligned}$$

Table 8.9: Results of eye tests

Right eye Grade	Left Eye Grade				Total
	Highest (1)	Second (2)	Third (3)	Lowest (4)	
Highest (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Lowest (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 8.10: Fitted counts under symmetry model

Right eye Grade	Left Eye Grade				Total
	Highest (1)	Second (2)	Third (3)	Lowest (4)	
Highest (1)	1520	250	120.5	51	1976
Second (2)	250	1512	397	80	2256
Third (3)	120.5	397	1772	192	2456
Lowest (4)	51	80	192	492	789
Total	1907	2222	2507	841	7477

The deviance is calculated in the usual manner as  $G^2 = 19.25$ , with  $df = 4(4-1)/2 = 6$ , which is significant at the 1% level, with indicates that the symmetry model does not fit well.

### 8.3.3 Quasi-Symmetry

The model of quasi-symmetry can be defined by

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

where  $\lambda_{ij}^{XY} = \lambda_{ji}^{XY}$

$$\text{and } \sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_i \lambda_{ij}^{XY} = 0$$

We can see that this is rather like the model of symmetry, except that there are an extra  $(r-1)$  parameters,  $\lambda_j^Y$ , in the loglinear representation. Thus, the degrees of freedom are equal to the  $df$  for the symmetry model minus  $(r-1)$ , which is  $r(r-1)/2 - (r-1) = (r-2)(r-1)/2$ .

The model of quasi-symmetry is often fitted not because of interest in the model itself, but because the difference in  $G^2$  for quasi-symmetry and  $G^2$  for symmetry is a conditional test statistic for the useful model of marginal homogeneity. This will be discussed in more detail in the next section.

The quasi-symmetry model imposes no restrictions on the diagonal cells, so the MLEs of these are

$$m_{ii} = n_{ii},$$

and so we can remove these from the model.

For any  $2 \times 2$  table, the model of quasi-symmetry has  $df = 0$ , and so is saturated.

For  $r=3$ , quasi-symmetry is equivalent to quasi-independence, i.e. if we remove the diagonal cells, or think of them as structural zeroes or fixed values, quasi-independence is independence as it applies to the non-empty cells of the table.

For  $r > 3$ , with the diagonal cells removed, quasi-symmetry is implied by quasi-independence.

The MLEs of the diagonal cells are simply  $\hat{m}_{ii} = n_{ii}$ .

There is no simple closed form expression for the estimates of the off-diagonal cells. However, since quasi-symmetry is a loglinear model, it can be fitted in much the same way as the usual loglinear models.

#### 8.3.3.1 Eye-Testing Example

Fitting the model of quasi-symmetry to the data of Table 8.9 results in the fitted counts given in Table 8.11. The deviance is  $G^2 = 7.27$ , with  $(4-2)(4-1)/2 = 3$  df, which is not quite significant at the 5% level.

Table 8.11: Fitted counts under quasi-symmetry

Right eye grade		Left Eye Grade			
		Highest (1)	Second (2)	Third (3)	Lowest (4)
Highest	(1)	1520	263.4	133.6	59.0
Second	(2)	236.6	1512	419	88.4
Third	(3)	107.4	375	1772	201.6
Lowest	(4)	43.0	71.6	182.4	492

#### 8.3.4 Marginal Homogeneity

The model of marginal homogeneity specifies that

$$m_{i+} = m_{+i}$$

for  $i = 1, \dots, r$ , i.e. corresponding marginal totals are equal. Since there are  $r-1$  constraints on the cell counts, the df for the model equals  $r-1$ .

Marginal homogeneity is implied by, but does not imply, symmetry.

The constraints of marginal homogeneity plus quasi-symmetry give the model of symmetry (see Bishop, Fienberg and Holland, 1975, p.286 for a proof). Thus a conditional test for marginal homogeneity is given by the difference in deviances between the models of symmetry and quasi-symmetry, i.e.  $G^2(MH|QS) = G^2(S) - G^2(QS)$ .

The degrees of freedom of this conditional test statistic are equal to the difference in df for the models of symmetry and quasi-symmetry, i.e.  $df = r(r-1)/2 - (r-1)(r-2)/2 = r-1$ . Stirling (1986) has pointed out that if the model of quasi-symmetry does not fit well, then a test which is conditional on quasi-symmetry does not seem very sensible. He suggests testing the hypothesis of marginal homogeneity directly by fitting the model and calculating the deviance in the usual way.

#### 8.3.4.1 Estimation

The marginal homogeneity model is a linear model because it specifies linear constraints on the cell counts:

$$m_{11} + m_{12} + \dots + m_{1r} = m_{1i} + m_{2i} + \dots + m_{ri}$$

The MLEs of the diagonal cells  $\{m_{ii}\}$  are equal to the observed values in these cells,  $\{n_{ii}\}$ .

The MLEs of the off-diagonal cells can be obtained using either the method of solving simultaneous equations (Stirling, 1984), or generalized linear models specified in terms of constraints (Wedderburn, 1974). For details see Appendix 7.

#### 8.3.4.2 Eye-testing example

We will consider again the data of Table 8.9. The model of marginal homogeneity specifies that  $m_{i+} = m_{+i}$ , i.e.  $m_{1+} = m_{+1}$ ,

$m_{2+} = m_{+2}$ , etc. This means that the expected total number of women with left eye grade of  $i$  equals the expected total number with right eye grade of  $i$ .

We have already seen that the models of quasi-symmetry and symmetry fitted to this data gave deviance of 7.27, with 3 df, and 19.25, with 6 df, respectively. The test statistic for testing the null hypothesis of marginal homogeneity, conditional on the quasi-symmetry model holding, is  $G^2 (MH|QS) = 19.25 - 7.27 = 11.98$ , with  $6-3 = 3$  df.

Fitting the marginal homogeneity model directly gives the estimated expected counts of Table 8.12. The direct goodness-of-fit test gives  $G^2 = 11.99$  with 3 df which is very close to the conditional test statistic. Both test statistics are significant at the 1% level, indicating that the model does not appear to describe the data adequately.

Table 8.12: Estimated expected counts under the model of marginal homogeneity

Right eye grade		Left eye grade				Total
		Highest (1)	Second (2)	Third (3)	Lowest (4)	
Highest	(1)	1520	252.49	111.84	56.95	1941.28
Second	(2)	247.24	1512	409.43	70.59	2239.25
Third	(3)	131.27	383.14	1772	195.27	2481.67
Lowest	(4)	42.77	91.62	188.40	492	814.80
Total		1941.28	2239.25	2481.67	814.80	7477.00

### 8.3.5 Multi-dimensional tables

The models of symmetry, quasi-symmetry and marginal homogeneity can be extended to higher-order tables. For further details see Bishop, Fienberg and Holland, 1975, p.299.

## 8.4 Linear Models

Many of the models that we have considered so far have been loglinear models, i.e. the log of the expected cell count is assumed to be a linear function of the model parameters. However, there are cases where linear models are more appropriate. For instance, the marginal homogeneity model of Section 8.3.4, which is specified by  $m_{i+} = m_{+i}$ , is a linear model. Many linear models are more easily specified in terms of constraints, rather than formally writing the specified cell counts in terms of parameters. Often linear models are used to characterize the margins of a table. They can also be used to specify non-standard hypotheses about a table that often could be difficult to specify any other way.

Estimation is much easier when linear models are specified in terms of constraints. For details see Section 1.11 and Appendix 3.

### 8.4.1 Drug example

Table 8.13, taken from Haber (1984, p.4) deals with the comparison of the reaction (favourable/unfavourable) to three drugs, A, B and C. Seven samples of subjects were used, with each sample corresponding to a subset of the three drugs. Thus the 46 subjects in the first sample received each of the three drugs; the 28 subjects in the second sample received only drugs A and B, etc. Let  $\theta_A$ ,  $\theta_B$  and  $\theta_C$  denote the probabilities of favourable reactions to drugs A, B and C, respectively. The hypothesis that these probabilities are the same for all seven samples can be formulated in terms of the constraints

$$\theta_{A1} = \theta_{A2} = \theta_{A3} = \theta_{A5} = \theta_A$$

$$\theta_{B1} = \theta_{B2} = \theta_{B4} = \theta_{B6} = \theta_B$$

$$\theta_{C1} = \theta_{C2} = \theta_{C3} = \theta_{C7} = \theta_C$$

where  $\theta_{Ai}$ ,  $\theta_{bi}$  and  $\theta_{Ci}$  are the probabilities of favourable reactions to drugs A, B and C, respectively, in sample  $i$ . For instance

$$\theta_{A1} = (m_1 + m_2 + m_3 + m_4) / (m_1 + m_2 + m_3 + m_4 + m_5 + m_6 + m_7 + m_8)$$

$$\theta_{A2} = (m_9 + m_{10}) / (m_9 + m_{10} + m_{11} + m_{12})$$

$$\theta_{A3} = (m_{13} + m_{14}) / (m_{13} + m_{14} + m_{15} + m_{16})$$

$$\theta_{A5} = m_{21} / (m_{21} + m_{22})$$

The last column of Table 8.13 gives the maximum likelihood estimates under the null hypothesis. The deviance is  $G^2 = 1.25$  with 9 df which indicates that the model fits well. Thus it seems reasonable to conclude that each drug worked just as well, whether it was used individually, or in combination with the others.

Another important hypothesis would be that the three drugs worked equally well, i.e.  $\theta_A = \theta_B = \theta_C$ . Haber (1974, p.5) discusses testing this more restrictive hypothesis.



Table 8.13: Responses to drugs A, B and C

Sample	Drugs Used	Pattern of Response *			Observed Frequency	MLE of Expected Frequency
		A	B	C		
1	A,B,C (N <sub>1</sub> =46)	1	1	1	6	6.11
		1	1	0	16	16.03
		1	0	1	2	1.93
		1	0	0	4	3.80
		0	1	1	2	2.11
		0	1	0	4	4.14
		0	0	1	6	5.98
		0	0	0	6	5.89
2	A,B (N <sub>2</sub> =28)	1	1	-	12	13.23
		1	0	-	4	3.74
		0	1	-	4	4.06
		0	0	-	8	6.97
3	A,C (N <sub>3</sub> =25)	1	-	1	5	4.92
		1	-	0	10	10.23
		0	-	1	4	3.85
		0	-	0	6	6.00
4	B,C (N <sub>4</sub> =26)	-	1	1	4	4.08
		-	1	0	12	11.97
		-	0	1	5	5.03
		-	0	0	5	4.92
5	A (N <sub>5</sub> =16)	1	-	-	10	9.70
		0	-	-	6	6.30
6	B (N <sub>6</sub> =15)	-	1	-	11	9.26
		-	0	-	4	5.74
7	C (N <sub>7</sub> =14)	-	-	1	5	4.91
		-	-	0	9	9.09

\* 1 denotes favourable response, 0 denotes unfavourable response, - denotes that the drug was not received

## 8.5 Summary

The preceding chapters have shown that there are many different kinds of models that can be fitted to contingency table data. In general, if there is more than one response variable, we would have a choice of loglinear or log-multiplicative models. If there is only one response, then we have a choice of six classes of models:

- (1) loglinear, which is equivalent to the adjacent categories logit
- (2) continuation ratio logit
- (3) cumulative logit, which is equivalent to assuming an underlying logistic distribution for the response
- (4) underlying normal distribution
- (5) underlying extreme value distribution
- (6) log-multiplicative (RC).

The first three types are equivalent if the response has only two levels.

Loglinear or adjacent categories logit models are well known, easy to fit and have readily interpretable parameters. They are usually interpreted in terms of odds ratios. Unfortunately, if ordinal models are fitted, they are reliant on the choice of arbitrary scores for the response and/or explanatory variables.

Continuation ratio logits are simple to fit, and can be thought of as a series of logits, adding up to the whole model. However, they suffer from the disconcerting feature that different results will be obtained if the ordering of the response category is reversed.

Cumulative logit models can be fairly easily fitted and also have parameters that are easily interpreted in terms of odds ratios. However, they are also reliant on the choice of suitable scores when using ordinal explanatory variables.

Using models which assume the underlying response has a normal distribution seem reasonable, as normality is the standard regression

assumption for a quantitative response. However, the short tails of the normal distribution makes outliers unlikely, leading to a lack of fit for many tables.

Assuming an underlying extreme value distribution is appropriate if the distribution is assumed to be skewed, the models are easily interpreted in terms of proportions rather than odds ratios.

RC models are more difficult to fit than loglinear models, and test statistics are more complicated. However, they eliminate the problem of choosing scores as they are estimated in the course of fitting these models. They are also appropriate for both nominal and ordinal variables.

The choice between these models can sometimes be made on logical grounds, i.e. for a given set of data, perhaps it might seem more reasonable to use a proportional hazards model (such as assuming an extreme value distribution) rather than assuming normality, or if the analysis is to be performed on a logit scale, we may choose to use cumulative rather than adjacent-categories logits because it might seem more reasonable to assume that several categories collapsed together will "behave better" (i.e. more like binomial) rather than assuming we can compare each category to every other category.

Sometimes there are grounds for choosing one type of model over another as a matter of simplicity. For example, although the RC model may estimate scores for the categories of a variable, it may be more parsimonious and easier to explain to a lay-person (as well as perhaps not fitting much more badly) to simply use equal-interval scores.

Often there may be very little difference in the results of the analysis, no matter which model is chosen. This is especially true for small tables where there is just not enough data to differentiate between the goodness-of-fit of various models. Take, for example, the dumping severity data which was re-analysed repeatedly in the previous chapters. Most of the different models fitted resulted in very similar conclusions, and almost any of them could be justified on one ground or another.

The other models discussed (for incomplete or square tables and linear models) are appropriate for certain kinds of tables and it is usually

fairly obvious when to use these.

## APPENDICES

### APPENDIX 1: EQUIVALENCE OF MLEs UNDER POISSON, MULTINOMIAL AND PRODUCT MULTINOMIAL SAMPLING SCHEMES

#### Al.1 MLEs

Consider an  $r \times c$  table classified by variables  $X$  and  $Y$ . The Poisson probability density function (pdf) is

$$f(n_{ij}) = \frac{m_{ij}^{n_{ij}} e^{-m_{ij}}}{n_{ij}!}$$

with corresponding log likelihood for the whole table of counts

$$\log L = \sum_{i,j} n_{ij} \log m_{ij} - \sum_{i,j} m_{ij} + \sum_{i,j} \log (n_{ij}!)$$

The multinomial log likelihood is

$$\begin{aligned} \log L &= \log \left[ \frac{N!}{\prod_{i,j} (n_{ij}!)} \prod_{i,j} \pi_{ij}^{n_{ij}} \right] \\ &= \sum_{i,j} (n_{ij} \log \pi_{ij}) + \log N! + \sum_{i,j} \log (n_{ij}!) \end{aligned}$$

If we reparameterize in terms of the expected values in each cell,  $m_{ij} = N \pi_{ij}$ , the log likelihood becomes

$$\begin{aligned} \log L &= \sum_{i,j} (n_{ij} \log \frac{m_{ij}}{N}) + \log N! + \sum_{i,j} \log (n_{ij}!) \\ &= \sum_{i,j} (n_{ij} \log m_{ij}) - \sum_{i,j} (n_{ij} \log N) + \log N! + \sum_{i,j} \log (n_{ij}!) \end{aligned}$$

We can treat the last three terms as constants because the derivative of any of them with respect to  $m_{ij}$  is zero.

The maximum of  $\log L$  with respect to  $m_{ij}$ , subject to the constraint  $\sum_{i,j} m_{ij} = N$ , is found using the Lagrange multiplier  $\phi$  by maximizing

$$M = \sum_{i,j} n_{ij} \log m_{ij} + \phi (\sum_{i,j} m_{ij} - N)$$

with respect to  $\phi$  and  $m_{ij}$ . If we assume a loglinear model for the expected cell counts, i.e.  $m_{ij} = \exp(\eta_{ij}) = \exp(\alpha + x_{ij}'\beta)$  then we maximize

$$Q = \sum_{i,j} n_{ij} \eta_{ij} + \phi (\sum_{i,j} \exp(\eta_{ij}) - N)$$

First differentiating with respect to  $\alpha$  gives

$$\frac{\partial M}{\partial \alpha} = \sum_{i,j} n_{ij} + \phi \sum_{i,j} \exp(\eta_{ij})$$

$$= \sum_{i,j} n_{ij} + \phi \sum_{i,j} M_{ij}$$

$$= N + \phi N$$

Setting this to zero for a maximum gives  $\phi = -1$ , so that we can equivalently maximize

$$\sum_{i,j} n_{ij} \log m_{ij} - \sum_{i,j} m_{ij}$$

which is identical to what is maximized when using the Poisson model. Therefore, the maximum likelihood estimates of  $\alpha$  and  $\beta$  will be identical for both sampling schemes.

It can be similarly shown that the product-multinomial sampling scheme gives the same likelihood function to be maximized as the Poisson and multinomial models, as long as the  $\lambda$ -terms corresponding to the fixed margins are included in the loglinear model.

## Al.2 Deviance

The deviance is defined as  $2 (\log L_{\text{sat}} - \log L_0)$  where  $\log L_{\text{sat}}$  is the maximum log likelihood for the saturated model and  $\log L_0$  is the maximum log likelihood for the model under consideration.

### Al.2.1 Multinomial

For the multinomial sampling scheme

$$\begin{aligned}\log L_0 &= \sum_{i,j} n_{ij} \log \hat{\pi}_{ij} + \log N! + \sum \log n_{ij} \\ &= \sum n_{ij} \log \hat{m}_{ij} - N \log N + \log N! + \sum \log n_{ij}!\end{aligned}$$

For the saturated model  $\hat{m}_{ij} = n_{ij}$  so  
 $\log L_{\text{sat}} = \sum n_{ij} - N \log N + \log N! + \sum \log n_{ij}!$ ,  
 so the deviance is

$$\begin{aligned}G^2 &= 2 (\log L_{\text{sat}} - \log L_0) \\ &= 2 [\sum (n_{ij} \log n_{ij}) - \sum (n_{ij} \log \hat{m}_{ij})] \\ &= 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}\end{aligned}$$

### Al.2.2 Poisson

For the Poisson sampling scheme  $\log L_0 = \sum n_{ij} \log \hat{m}_{ij} - \sum \hat{m}_{ij} - \sum \log n_{ij}!$ ,

$$\log L_{\text{sat}} = \sum n_{ij} \log n_{ij} - \sum n_{ij} - \sum \log n_{ij}!$$

$$\begin{aligned}G^2 &= 2 (\log L_{\text{sat}} - \log L_0) \\ &= 2 [\sum (n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}) + \sum \hat{m}_{ij} - \sum n_{ij}]\end{aligned}$$

and, since  $\sum n_{ij} = \sum \hat{m}_{ij} = N$

$$G^2 = 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}.$$

which is identical to the deviance obtained in the multinomial model.

### Al.2.3 Product-multinomial

For the product-multinomial sampling scheme where we fix say the  $r$  row totals and take a sample of size  $n_{i+}$  for each  $i$ , the likelihood function under some model is

$$L_0 = \prod_i \left[ \frac{n_{i+}!}{\prod_j n_{ij}!} \prod_j \left( \frac{\hat{\pi}_{ij}}{\hat{\pi}_{i+}} \right)^{n_{ij}} \right]$$

$$\log L_0 = \sum_i \log n_{i+}! - \sum_{i,j} \log n_{ij}! + \sum_{i,j} n_{ij} \log \frac{\hat{\pi}_{ij}}{\hat{\pi}_{i+}}$$

$$= \sum_i \log n_{i+}! - \sum_{i,j} \log n_{ij}! + \sum_{i,j} n_{ij} \log \frac{\hat{m}_{ij}}{\hat{m}_{i+}}$$

$$\log L_{\text{sat}} = \sum_i \log n_{i+}! - \sum_{i,j} \log n_{ij}! + \sum_{i,j} n_{ij} \log \frac{n_{ij}}{n_{i+}}$$

$$\begin{aligned} G^2 &= 2 (L_{\text{sat}} - L_0) \\ &= 2 \left( \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}} - \sum_{i,j} n_{ij} \log n_{i+} + \sum_{i,j} n_{ij} \log \hat{m}_{i+} \right) \end{aligned}$$

if we have the term  $\lambda_i X$  in the loglinear model, then  $\hat{m}_{i+} = n_{i+}$ , so the last two terms cancel out and

$$G^2 = 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{m}_{ij}}$$

which is identical to the deviance obtained with the multinomial and Poisson models.

These results can be easily generalized to multi-dimensional contingency tables.



## APPENDIX 2: FITTING LOGLINEAR MODELS

### A2.1 Newton-Raphson Algorithm

Let  $\underline{n}' = (n_1, \dots, n_I)$  and  $\underline{m}' = (m_1, \dots, m_I)$  denote the observed and expected counts for the  $I$  cells in the table. For simplicity we will use a single index, though the table may be multi-dimensional. Under the Poisson sampling model the log likelihood is

$$\log L(\underline{m}) = \sum n_i \log m_i - \sum m_i - \sum \log n_i!$$

Loglinear models have the form

$$\log m_i = \underline{x}'_i \underline{\beta}$$

where  $\underline{\beta}$  is a  $p \times 1$  vector of parameters and  $\underline{x}'_i$  is the  $i$ th row of the  $I \times p$  design matrix  $X$ , i.e.

$$\log \underline{m} = X\underline{\beta}$$

The MLE  $\hat{\underline{\beta}}$  maximizes the log likelihood expressed as a function of the parameter  $\underline{\beta}$ , namely

$$\log L(\underline{\beta}) = \sum n_i \underline{x}'_i \underline{\beta} - \sum \exp(\underline{x}'_i \underline{\beta}) - \sum \log n_i!$$

We will use the Newton-Raphson (NR) technique to maximize  $\log L(\underline{\beta})$ . This is an iterative procedure that replaces an approximation to the optimum,  $\underline{\beta}^{(t)}$ , by  $\hat{\underline{\beta}}^{(t+1)}$  at each iteration, where

$$\hat{\underline{\beta}}^{(t+1)} = \hat{\underline{\beta}}^{(t)} - \left[ \frac{\partial^2 \log L(\underline{\beta})}{\partial \underline{\beta}^2} \right]^{(t)-1} \left[ \frac{\partial \log L(\underline{\beta})}{\partial \underline{\beta}} \right]^{(t)}$$

This can be expressed as

$$\hat{\underline{\beta}}^{(t+1)} = (X'WX)^{-1}X'W\underline{z}$$

$$\text{where } W = \begin{bmatrix} \hat{m}_1(t) & & & \\ & \hat{m}_2(t) & & \\ & & \ddots & \\ & & & \hat{m}_n(t) \end{bmatrix}$$

$$\text{and } y_i = \log m_i(t) + \frac{n_i - \hat{m}_i(t)}{\hat{m}_i(t)}$$

which is weighted least squares (WLS) with a transformed response. In other words,  $\hat{\beta}^{(t+1)}$  is the WLS solution to the model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

where the  $\{\varepsilon_i\}$  are uncorrelated with variances  $\{1/\hat{m}_i(t)\}$ . Thus the ML estimate may be regarded as the limit of a sequence of WLS estimates.

We can begin the iterative process by setting all  $\hat{m}_i^{(0)} = n_i$ , or all  $m_i^{(0)} = n_i + 0.5$  if any  $n_i = 0$ . The  $\hat{m}^{(t)}$  and  $\hat{\beta}^{(t)}$  usually converge rapidly to the ML estimates  $\hat{m}$  and  $\hat{\beta}$ . The estimated covariance matrix of  $\hat{\beta}$  is  $(X'WX)^{-1}$  (where  $W$  is evaluated at  $\hat{\beta}$ ).

The computer packages GLIM (Baker and Nelder, 1978) or Genstat (Alvey et al, 1983) use the N-R technique for fitting loglinear models to contingency tables as part of the wider framework of generalized linear models (Nelder and Wedderburn, 1972).

#### A2.1.1 Abortion Attitude Example

To illustrate the  $\log \underline{m} = \underline{X}\underline{\beta}$  representation needed for fitting a loglinear model using the Newton-Raphson method, consider the 4x2 table of religion and abortion attitude of Table 2.1.

The model of independence

$$\log m_{ij} = \mu + \lambda_i^R + \lambda_j^A$$

can be expressed as

$$\log \underline{m} = \underline{X}\underline{\beta}$$

with

$$\begin{aligned} \log \underline{m}' &= (\log m_{11}, \log m_{12}, \log m_{21}, \dots, \log m_{42}) \\ \underline{\beta}' &= (\mu, \lambda_1^R, \lambda_2^R, \lambda_3^R, \lambda_1^A) \end{aligned}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

Here we use the fact that  $\lambda_4^R = -\lambda_1^R - \lambda_2^R - \lambda_3^R$  and  $\lambda_2^A = -\lambda_1^A$

## A2.2 Iterative Proportional Fitting Algorithm

To illustrate the Iterative Proportional Fitting (IPF) algorithm (Deming and Stephan, 1940) we will compute the estimated expected cell counts for the no 3-factor interaction model (XY, XZ, YZ), i.e.

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

The  $\{m_{ijk}\}$  must satisfy the ML equations

$$\begin{aligned} \hat{m}_{ij+} &= n_{ij+} \\ \hat{m}_{i+k} &= n_{i+k} \\ \hat{m}_{+jk} &= n_{+jk} \end{aligned}$$

We cannot write out the MLEs in closed form - the following IPF procedure yields the MLEs.

Stage 1: Set  $\hat{m}_{ijk}^{(0)} = 1$  for all  $i, j, k$

Stage 2: Successively adjust the  $\{\hat{m}_{ijk}\}$  so that they satisfy the marginal constraints. The first cycle of the adjustment process has the following three steps:

$$\hat{m}_{ijk}^{(1)} = \hat{m}_{ijk}^{(0)} \frac{n_{ij+}}{\hat{m}^{(0)}_{ij+}}$$

$$\hat{m}_{ijk}^{(2)} = \hat{m}_{ijk}^{(1)} \frac{n_{i+k}}{\hat{m}^{(1)}_{i+k}}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

Here we use the fact that  $\lambda_4^R = -\lambda_1^R - \lambda_2^R - \lambda_3^R$  and  $\lambda_2^A = -\lambda_1^A$

## A2.2 Iterative Proportional Fitting Algorithm

To illustrate the Iterative Proportional Fitting (IPF) algorithm (Deming and Stephan, 1940) we will compute the estimated expected cell counts for the no 3-factor interaction model (XY, XZ, YZ), i.e.

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

The  $\{m_{ijk}\}$  must satisfy the ML equations

$$\begin{aligned} \hat{m}_{ij+} &= n_{ij+} \\ \hat{m}_{i+k} &= n_{i+k} \\ \hat{m}_{+jk} &= n_{+jk} \end{aligned}$$

We cannot write out the MLEs in closed form - the following IPF procedure yields the MLEs.

Stage 1: Set  $\hat{m}_{ijk}^{(0)} = 1$  for all  $i, j, k$

Stage 2: Successively adjust the  $\{\hat{m}_{ijk}\}$  so that they satisfy the marginal constraints. The first cycle of the adjustment process has the following three steps:

$$\hat{m}_{ijk}^{(1)} = \hat{m}_{ijk}^{(0)} \frac{n_{ij+}}{\hat{m}^{(0)}_{ij+}}$$

$$\hat{m}_{ijk}^{(2)} = \hat{m}_{ijk}^{(1)} \frac{n_{i+k}}{\hat{m}^{(1)}_{i+k}}$$

$$\hat{m}_{ijk}^{(3)} = \hat{m}_{ijk}^{(2)} \frac{n_{+jk}}{\hat{m}_{+jk}^{(2)}}$$

Summing both sides of the first expression over  $k$ , we note that after the first step, all  $\hat{m}_{ij+}^{(1)} = n_{ij+}$ , so that the estimated expected frequencies satisfy the first set of marginal constraints. After the second step all  $\hat{m}_{i+k}^{(2)} = n_{i+k}$ , as well as  $m_{ij+} = n_{ij+}$  for all  $i$  and  $j$ . After the third step all  $\hat{m}_{+jk}^{(3)} = n_{+jk}$ , but the first two sets of constraints no longer hold.

Stage 3: Repeat stage 2 until the change in the  $\{\hat{m}_{ijk}\}$  from one cycle to the next is sufficiently small.

#### A2.2.1 Lizard Example

We will use the IPF procedure to fit the model (SD, SH, DH) to the data of Table 1.4. The observed marginal totals are

Species		Diameter	
i	j	$n_{ij+}$	
1	1	32 + 86 = 118	
1	2	11 + 35 = 46	
2	1	61 + 73 = 134	
2	2	41 + 70 = 111	

Species		Height	
i	k	$n_{i+k}$	
1	1	43	
1	2	121	
2	1	102	
2	2	143	

Diameter i	Height k	$n_{+jk}$
1	1	93
1	2	159
2	1	52
2	2	105

Table A2.1 shows various stages in the IPF procedure. We begin with initial cell values equal to 1, i.e.  $\hat{m}_{ijk}^{(0)} = 1$ . Each marginal total  $\hat{m}_{ij+}^{(0)}$  is obtained by summing the  $\{\hat{m}_{ijk}^{(0)}\}$  over  $k$ , e.g.  $\hat{m}_{12+}^{(0)} = \hat{m}_{121}^{(0)} + \hat{m}_{122}^{(0)} = 2$ .

To form the  $\{\hat{m}_{ijk}^{(1)}\}$  we multiply each  $\hat{m}_{ijk}^{(0)}$  by  $n_{ij+}/\hat{m}_{ij+}^{(0)}$ , e.g.

$$\begin{aligned}\hat{m}_{122}^{(1)} &= \hat{m}_{122}^{(0)} \frac{n_{12+}}{\hat{m}_{12+}^{(0)}} \\ &= 1 \times \frac{46}{2} \\ &= 23\end{aligned}$$

Then we compute the  $\{\hat{m}_{ijk}^{(2)}\}$  by  $\hat{m}_{ijk}^{(2)} = \hat{m}_{ijk}^{(1)} \frac{n_{i+k}}{\hat{m}_{i+k}^{(1)}}$

$$\begin{aligned}\text{e.g. } \hat{m}_{122}^{(2)} &= 23 \times \frac{121}{82} \\ &= 33.939\end{aligned}$$

The next step is to calculate the  $\{\hat{m}_{ijk}^{(3)}\}$  by

$$\begin{aligned}\hat{m}_{ijk}^{(3)} &= \hat{m}_{ijk}^{(2)} \frac{n_{+kj}}{\hat{m}_{+jk}^{(2)}} \\ \text{e.g. } \hat{m}_{122}^{(3)} &= 33.939 \times \frac{105}{98.727} \\ &= 36.095\end{aligned}$$

Then we repeat the whole cycle again, until the estimates have converged sufficiently.

Table A2.1: Various stages in the IPF procedure

Cell (i,j,k)	$n_{ijk}$	(0) $\hat{m}_{ijk}$	(1) $\hat{m}_{ijk}$	(2) $\hat{m}_{ijk}$	(3) $\hat{m}_{ijk}$	(4) $\hat{m}_{ijk}$	(12) $\hat{m}_{ijk}$
(1,1,1)	32	1	59	30.939	33.177	33.479	32.80
(1,1,2)	86	1	59	87.061	83.757	84.521	85.20
(1,2,1)	11	1	23	12.061	10.763	10.566	10.20
(1,2,2)	35	1	23	33.939	36.095	35.434	35.80
(2,1,1)	61	1	67	55.788	59.823	59.351	60.20
(2,1,2)	73	1	67	78.212	75.243	74.649	73.80
(2,2,1)	41	1	55.5	46.212	41.237	41.648	41.80
(2,2,2)	70	1	55.5	64.788	68.905	69.592	69.20
		(0) $i \ j \ \hat{m}_{ij+}$	(1) $i \ k \ \hat{m}_{i+k}$	(2) $j \ k \ \hat{m}_{+jk}$	(3) $i \ j \ \hat{m}_{ij+}$	(4) $i \ k \ \hat{m}_{i+k}^{(4)}$	
		1 1 2	1 1 82	1 1 86.727	1 1 116.934	1 1 44.045	
		1 2 2	1 2 82	1 2 165.273	1 2 46.858	1 2 119.955	
		2 1 2	2 1 122.5	2 1 58.273	2 1 135.066	2 1 100.999	
		2 2 2	2 2 122.5	2 2 98.727	2 2 109.905	2 2 144.241	
			(1) $\hat{m}_{ij+} = n_{ij+}$	(2) $\hat{m}_{i+k} = n_{i+k}$	(3) $\hat{m}_{+jk} = n_{+jk}$	(4) $\hat{m}_{ij+} = n_{ij+}$	(12) $\hat{m}_{ij+} = n_{ij+}$
				(2) $m_{ij+} = n_{ij+}$			(12) $\hat{m}_{i+k} = n_{i+k}$
							(12) $\hat{m}_{+jk} = n_{+jk}$

With only a slight modification, the iterative method can also be used to compute estimated expected values for the other models. For example, consider the model (XY, XZ). In this model each cycle of the iterative method would have only two steps. The first adjusts for the marginal totals  $\{n_{ij+}\}$  and the second for the totals  $\{n_{i+k}\}$ . The third set of marginal totals is not needed here. Suppose the table is of size  $2 \times 2 \times 2$ . The  $\hat{m}_{ijk}^{(0)} = 1$  and  $\hat{m}_{ijk}^{(1)} = n_{ij+}/2$ ,

$$\hat{m}_{i+k}^{(1)} = \frac{n_{i1+}}{2} + \frac{n_{i2+}}{2} = \frac{n_{i++}}{2}$$

$$\text{so } \hat{m}_{ijk}^{(2)} = \frac{n_{ij+}}{2} \times \frac{n_{i+k}}{n_{i++}/2}$$

$$= \frac{n_{ij+} m_{i+k}}{n_{i++}}$$

This model could have been fitted directly using the observed marginal totals  $n_{ij+}$ ,  $n_{i+k}$  and  $n_{i++}$ ; however, the IPF method gives a systematic way of estimating the expected cell counts for all the usual loglinear models and we don't have to worry about whether direct estimation is possible or not.



### APPENDIX 3: FITTING LINEAR MODELS

#### A3.1 Linear models specified in terms of constraints

Let  $\underline{m}$  be an  $I \times 1$  vector of cell counts, with corresponding vector of probabilities  $\underline{\pi}$ . The  $I \times 1$  vector  $\underline{\pi}$  may correspond to

- (i) a Poisson, or to
- (ii) a product-multinomial distribution.

In the former case  $\sum \pi_i = 1$ , while in the product-multinomial case, the set of  $I$  cells is comprised of several subsets, each of which corresponds to a separate multinomial sample, and the sum of the elements of  $\pi$  over each subset is unit.

If we have a hypothesis about the cell counts which can be specified in terms of  $E$  constraints, we can write the constraints as

$$\underline{F} \underline{m} = \underline{0}$$

where  $F$  is an  $E \times I$  matrix with  $E$  linearly independent rows. Further constraints are imposed by the sampling design. These constraints guarantee that the sum of the probabilities within each sample will be equal to one (or equivalently that the sum of the counts within each sample will be equal to the correct marginal total). They can be written

$$\underline{D}' \underline{\pi} = \underline{1}_S$$

where  $S$  is the number of samples ( $S > 1$ ) and  $D = \{d_{is}\}$  is the  $I \times S$  matrix defined by

$$d_{is} = \begin{cases} 1 & \text{if cell } i \text{ belongs to sample } S \\ 0 & \text{otherwise} \end{cases}$$

In terms of the cell counts the constraints can be written

$$\underline{D}' \underline{R} \underline{M} = \underline{1}_S$$

where  $R$  is the  $I \times I$  diagonal matrix with diagonal elements  $r_{ii} = 1/m_{i+}$ , where  $m_{i+}$  is the marginal total of the  $i$ th sample ( $i=1, \dots, S$ ), and off-diagonal elements are zero.

Thus, the constraints on the cell counts can be written as

$$\tilde{L} \tilde{m} = \tilde{c}$$

where  $\tilde{L}'$  is the  $I \times (E+S)$  matrix  $\tilde{L}' = (\tilde{F}' : \tilde{R}'D)$  and  $\tilde{c}'$  is the  $1 \times (E+S)$  vector consisting of  $E$  zeroes and  $S$  ones, i.e.

$$\tilde{c}' = (0'E : 1'S)$$

A vector  $\tilde{a}$  which satisfies these constraints is

$$\tilde{a} = R^{-1} \tilde{a}^*$$

where  $R^{-1}$  is the  $I \times I$  diagonal matrix with diagonal elements  $r_{ii} = m_{i+}$ , and  $\tilde{a}^*$  is the  $I \times 1$  vector with  $i$ th element  $a_i = 1/B_i$  where  $B_i$  is the number of cells in the  $i$ th sample.

We can easily calculate the MLEs using the algorithm of Wedderburn (1974) which is explained in the following section.

#### A3.1.1 Wedderburn's Algorithm for finding MLEs of generalized linear models specified in terms of constraints

Wedderburn (1974) gives an iterative method for finding the MLEs of generalized linear models specified in terms of constraints.

If we assume the cell counts,  $n_i$ , are distributed according to the Poisson distribution with mean,  $m_i$ , which is a function of  $\eta_i$ , where  $\eta_i$  is a linear function of the parameters associated with the model we wish to fit, i.e.  $\eta_i = \tilde{x}_i' \tilde{\beta}$ , the pdf of  $n_i$  is

$$p(n_i) = \frac{\exp(-m_i) m_i^{n_i}}{n_i!}$$

We define  $F_i = \log p(n_i)$ ,  
 $F_i' = \frac{\partial F_i}{\partial \eta_i}$

and

$$F_i'' = \frac{\partial^2 F_i}{\partial \eta_i^2}$$

So  $F_i = -m_i + n_i \log m_i - \log(m_i!)$   
 $= -\eta_i + n_i \log m_i - \log(m_i!)$

$$F_i = \frac{n_i}{\eta_i} - 1$$

$$F_i'' = \frac{-n_i}{\eta_i^2}$$

$$E[F_i''] = -1/\eta_i$$

Wedderburn shows that if a model is expressed as  $E(\underline{\eta}) = \underline{m} = \underline{\lambda}$

where  $\lambda_i = f(\eta_i)$

with constraints  $L \underline{\eta} = 0$

then instead of iteratively fitting  $E(\underline{\eta}) = X\beta$

we may fit  $E(\underline{\eta}) = W^{-1} L' \underline{\gamma}$

$$\text{where } W = \begin{bmatrix} -E[F_1''] & & \\ & \ddots & \\ & & -E[F_n''] \end{bmatrix}$$

$$= \begin{bmatrix} 1/\eta_1 & & \\ & \ddots & \\ & & 1/\eta_n \end{bmatrix}$$

and then use the residuals as the new value for  $\underline{\eta}$ .

The algorithm takes the following form:

(a) Set  $\underline{\eta} = \underline{n}$ . Set  $\underline{z} = \underline{\eta}$

- (b) Calculate  $W = \begin{bmatrix} 1/n_1 & \dots & 1/n_n \end{bmatrix}$
- (c) Regress  $\underline{z}$  on  $W^{-1}L$ , using  $W$  as the weight matrix Set  $\underline{\eta} = \underline{z} - \hat{\underline{z}}$  (residuals).  
If the process has gone far enough, stop. Otherwise go to step (b).

If we have inhomogeneous constraints of the form

$$L \underline{\eta} = \underline{c}$$

then we simply choose any vector  $\underline{a}$  such that  $L \underline{a} = \underline{c}$

The algorithm then becomes

- (a) Set  $\underline{\eta} = \underline{n}$ . Calculate  $\underline{z} = \underline{n} - \underline{a}$
- (b) Calculate  $W = \begin{bmatrix} 1/n_1 & \dots & 1/n_n \end{bmatrix}$
- (c) Regress  $\underline{z}$  on  $W^{-1}L'$ , using  $W$  as weight matrix.  
Set  $\underline{\eta} = \underline{\underline{z}} + \underline{a} - \hat{\underline{z}}$ .  
If the process has gone far enough, stop. Otherwise go to step (b).

The MLEs,  $\hat{m}_i$ , will be the values of  $n_i$  at the conclusion of the algorithm.

The deviance,  $G^2$ , is calculated in the usual way, with associated degrees of freedom equal to the rank of  $F$ , i.e.  $df = E$ .

### A3.1.2 Drugs example

For the data of Table 8.13 the hypothesis of equally favourable reactions to, say, drug A, for all seven samples can be written as

$$\theta_A = \theta_{A1} = \theta_{A2} = \theta_{A3} = \theta_{A5}$$

where  $\theta_{A1} = \pi_1 + \pi_2 + \pi_3 + \pi_4$

$$\theta_{A2} = \pi_9 + \pi_{10}$$

$$\theta_{A3} = \pi_{13} + \pi_{14}$$

$$\theta_{A5} = \pi_{21}$$

This implies that

$$\pi_1 + \pi_2 + \pi_4 - \pi_9 - \pi_{10} = 0$$

$$\pi_9 + \pi_{10} - \pi_{13} - \pi_{14} = 0$$

$$\pi_{13} + \pi_{14} - \pi_{21} = 0$$

The constraints relating to drugs B and C can be written in a similar manner, giving

$$F^* \underline{\pi} = 0$$

where  $F^*$  is the  $9 \times 26$  matrix

$$F^* = \begin{bmatrix} 11110000 & -1-100 & & & & & & \\ & 1 \ 100 & -1-100 & & & & & \\ & & 1 \ 100 & 0000 & -10 & & & \\ \hline 11001100 & -10-10 & & & & & & \\ & 1010 & 0000 & -1-100 & & & & \\ & & & 1 \ 100 & 00 & -10 & & \\ \hline 10101010 & 0000 & -10-10 & & & & & \\ & & 1010 & -10-10 & & & & \\ & & & 1010 & 00 & 00 & -10 & \end{bmatrix}$$

and  $\underline{\pi}' = (\pi_1, \pi_2, \dots, \pi_{26})$



$$D' R \underline{m} = \underline{17}.$$

Thus the constraints on the cell counts can be written as

$$L \underline{m} = L R^{-1} R \underline{m} = L^* R \underline{m} = \underline{c}$$

where  $L^* = (F^* : D)$ .

A vector  $\underline{a}$  which satisfies these constraints is

$$\underline{a} = R^{-1} \underline{a}^*$$

where  $\underline{a}^* = (1/8, \dots, 1/8, 1/4, \dots, 1/4, 1/2, \dots, 1/2)$ .

We can implement Wedderburn's algorithm using  $n$ ,  $L$  and  $\underline{a}$  as given above.

The deviance is calculated in the usual way as  $G^2 = 1.25$ , with 9 df.

### A3.2 Linear Models Specified as $A\underline{\pi} = X\underline{\beta}$

Haber (1985) discusses linear models which are formulated as

$$A\underline{\pi} = X\underline{\beta}$$

where  $A$  is a known  $I \times I$  matrix,  $\underline{\pi}$  is an  $I \times 1$  vector of cell probabilities,  $X$  is a known  $J \times K$  matrix, and  $\underline{\beta}$  is a  $K \times 1$  vector of unknown parameters, with  $I > J > K$ . We assume that the rows of  $A$ , as well as the columns of  $X$ , are linearly independent. We will now formulate the linear model in terms of constraints, so that we can apply Wedderburn's (1974) algorithm as discussed in Section (a).

We find a  $J \times (J-K)$  matrix  $W$  with independent columns (i.e. of rank  $J-K$ ) which when pre-multiplied by the  $J \times J$  matrix  $I - X(X'X)^{-1}X'$  gives a  $J \times (J-K)$  matrix  $U$  (of rank  $J-K$ ).  $U$  is orthogonal to  $X$  since  $X'U = X'(I - X(X'X)^{-1}X')W = (X' - X'X(X'X)^{-1}X')W = 0$ . Thus

$$U'A\underline{\pi} = \underline{0}.$$

Now, since  $\underline{\pi} = R\underline{m}$ , where  $R$  is as defined in Section (a), the equation becomes

$$U'AR_{\sim} = 0$$

Writing  $L = U'AR$  gives the set of  $J-K = E$  linear constraints

$$F_{\sim} m = 0$$

in the same notation as Section (a).

The constraints imposed by the sampling design can once again be written as

$$D'R_{\sim} = 15$$

Thus the constraints on the cell counts can be written as

$$L_{\sim} = c$$

with  $L' = (R'A'U : R'D)$ , and Wedderburn's algorithm can be used to find the MLEs.

The degrees of freedom associated with the deviance is equal to the rank of  $U'A$ , which is  $E = J-K$ .

### A3.2.1 Drugs example

For the data of Table 8.13 the linear model which specifies that  $\theta_A$  is the same for all seven samples, and similar restrictions on  $\theta_\beta$  and  $\theta_C$ , can be written

$$A_{\sim} \pi = X_{\sim} \beta$$

where  $A$  is the  $12 \times 26$  matrix



$$A = \begin{bmatrix} 11110000 & 1100 & 1100 & 10 & & \\ 11001100 & 1010 & 1100 & & 10 & \\ 10101010 & & 1010 & 1010 & & 10 \end{bmatrix}$$

$$\pi' = (\pi_1, \pi_2, \dots, \pi_{26})$$

X is the 12x3 matrix

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{and } \beta' = (\theta_A, \theta_B, \theta_C)$$

Now to write this model in the form

$$F^* \pi = U' A \pi = 0$$

we chose the 12x9 matrix W with 9 linearly independent columns:

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Pre-multiplying W by  $I - X(X'X)^{-1}X'$ , we obtain the 12x9 matrix U which is orthogonal to X,

$$4U = \begin{bmatrix} 3 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}$$

We can then obtain the 9x26 matrix  $F^* = U'A$ ,

$$4F^{*'} = \begin{bmatrix} 3 & -1 & -1 & 3 & -1 & -1 & 3 & -1 & -1 \\ 3 & -1 & -1 & 3 & -1 & -1 & 0 & 0 & 0 \\ 3 & -1 & -1 & 0 & 0 & 0 & 3 & -1 & -1 \\ 3 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -1 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 3 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 3 & -1 & -1 & 3 & -1 & 0 & 0 & 0 \\ 1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1 & -1 & 3 & 0 & 0 & 0 & -1 & 3 & -1 \\ -1 & -1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 & -1 & 3 & -1 & -1 & 3 \\ 0 & 0 & 0 & -1 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We can then proceed to write the model as

$$\underset{\sim}{L} \underset{\sim}{m} = \underset{\sim}{c}$$

as explained in Section A3.1.2, and use Wedderburn's algorithm to find the MLEs.

APPENDIX 4: MAXIMUM LIKELIHOOD EQUATIONS FOR LOGLINEAR MODELS  
FOR THREE-DIMENSIONAL TABLES

Consider an  $r \times c \times l$  table, classified by variables  $X$ ,  $Y$  and  $Z$ . The Poisson log likelihood function for this table is

$$\log L = \sum_{i,j,k} n_{ijk} \log m_{ijk} - \sum_{i,j,k} m_{ijk} + \sum_{i,j,k} \log n_{ijk}!$$

Under the model of independence  $(X,Y,Z)$

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

$$\begin{aligned} \text{so } \log L &= \sum_{i,j,k} n_{ijk} (\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z) \\ &\quad - \sum_{i,j,k} \exp (\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z) \\ &\quad + \sum_{i,j,k} \log n_{ijk}! \end{aligned}$$

To find the maximum of  $\log L$  we can differentiate with respect to the model parameters and equate to zero, e.g.

$$\begin{aligned} \frac{\partial \log L}{\partial \mu} &= \sum_{i,j,k} n_{ijk} - \sum_{i,j,k} \exp (\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z) \\ &= \sum_{i,j,k} n_{ijk} - \sum_{i,j,k} m_{ijk} \\ &= 0 \text{ at T.P.} \end{aligned}$$

$$\begin{aligned} \frac{\partial \log L}{\partial \lambda_i^X} &= \sum_{j,k} n_{ijk} - \sum_{j,k} \exp (\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z) \\ &= 0 \text{ at T.P.} \end{aligned}$$

which gives the maximum likelihood (ML) equations

$$\begin{aligned}\hat{m}_{+++} &= N \\ \hat{m}_{i++} &= n_{i++} \\ \hat{m}_{+j+} &= n_{+j+} \\ \hat{m}_{++k} &= n_{++k}\end{aligned}$$

$$\text{Now } m_{i++} = \sum_{j,k} m_{ijk}$$

$$= \sum_{j,k} \exp (\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z)$$

$$= \exp (\mu + \lambda_i^X) \sum_{j,k} \exp (\lambda_j^Y + \lambda_k^Z)$$

$$= \exp (\mu + \lambda_i^X) \sum_j \exp (\lambda_j^Y) \sum_k \exp (\lambda_k^Z)$$

$$m_{+j+} = \exp (\mu + \lambda_j^Y) \sum_i \exp (\lambda_i^X) \sum_k \exp (\lambda_k^Z)$$

$$m_{++k} = \exp (\mu + \lambda_k^Z) \sum_i \exp (\lambda_i^X) \sum_j \exp (\lambda_j^Y)$$

$$m_{+++} = \exp (\mu) \sum_i \exp (\lambda_i^X) \sum_j \exp (\lambda_j^Y) \sum_k \exp (\lambda_k^Z)$$

Combining these equations, we find that

$$m_{ijk} = \frac{m_{i++} m_{+j+} m_{++k}}{(m_{+++})^2}$$

so the maximum likelihood estimate (MLE) of  $m_{ijk}$  is

$$\hat{m}_{ijk} = \frac{n_{i++} n_{+j+} n_{++k}}{N^2}$$

Thus the MLEs of the  $\{m_{ijk}\}$  depend on the cell counts only through certain sufficient statistics. For the model of independence, the

minimal sufficient statistics are the one-dimensional marginal frequencies  $\{n_{i++}\}$ ,  $\{n_{+j+}\}$  and  $\{n_{++k}\}$ .

The sufficient statistics for the other loglinear models can be found in a similar manner. For example, for the model  $(XY, Z)$  or

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$$

$$\frac{\partial \log L}{\partial \mu} = \sum_{i,j,k} n_{ijk} - \sum_{i,j,k} m_{ijk}$$

$$\Rightarrow \hat{m}_{+++} = N$$

$$\frac{\partial \log L}{\partial \lambda_{ij}^{XY}} = \sum_k n_{ijk} - \sum_k m_{ijk}$$

$$= \hat{m}_{ij+} = n_{ij+}$$

$$\frac{\partial \log L}{\partial \lambda_k^Z} = \sum_{i,j} n_{ijk} - \sum_{i,j} m_{ijk}$$

$$= \hat{m}_{++k} = n_{++k}$$

The minimal sufficient statistics are  $\{n_{++k}\}$  and  $\{n_{ij+}\}$ , and the MLE of  $m_{ijk}$  is

$$\hat{m}_{ijk} = \frac{n_{ij+} n_{++k}}{N}$$

Table A4.1 gives the minimal sufficient statistics and the expressions for the MLEs for various models.

Table A4.1: Minimal sufficient statistics and expected frequency estimates for various models

Model	Minimal Sufficient Statistics	$m_{ijk} =$
$(X, Y, Z)$	$\{n_{i++}\}, \{n_{+j+}\}, \{n_{++k}\}$	$n_{i++} n_{+j+} n_{++k} / N^2$
$(XY, Z)$	$\{n_{ij+}\}, \{n_{++k}\}$	$n_{ij+} n_{++k} / n$
$(XY, YZ)$	$\{n_{ij+}\}, \{n_{+jk}\}$	$n_{ij+} n_{+jk} / n_{+j+}$
$(XY, XZ, YZ)$	$\{n_{ij+}\}, \{n_{i+k}\}, \{n_{+jk}\}$	no closed form exists



## APPENDIX 5: FITTING MEAN RESPONSE MODELS

Mean response models can be thought of as simply special cases of linear models, and can be fitted using the same techniques.

Using the same definitions of  $\underline{m}$ ,  $\underline{\pi}$ ,  $A$ ,  $X$ ,  $\beta$ ,  $L$ ,  $L^*$ ,  $\underline{c}$ , etc as in Appendix 3, we can write mean response models either in the form

$$A \underline{\pi} = X \beta$$

or in terms of constraints

$$L^* \underline{\pi} = \underline{c}$$

We can then use Wedderburn's algorithm to calculate the MLEs.

### A5.1 Dumping Severity Example

For the data of Table 1.6 the mean response model

$$\phi_i = \mu + \beta u_i$$

where  $\phi_i = \pi_{i1} + 2\pi_{i2} + 3\pi_{i3}$ , and  $u_i = i$   
 $i = 1, \dots, 4$

can be written  $A \underline{\pi} = X \beta$

where  $A$  is the  $4 \times 12$  matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & & & & & & & & & \\ & & & 1 & 2 & 3 & & & & & & \\ & & & & & & 1 & 2 & 3 & & & \\ & & & & & & & & & 1 & 2 & 3 \end{bmatrix}$$

$$\underline{\pi}' = (\pi_1, \pi_2, \dots, \pi_{12})$$

X is the 4x2 matrix

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

and  $\underline{\beta}' = (\mu, \beta)$

To write this model in the form  $F^* \underline{\pi} = U' A \underline{\pi} = \underline{0}$ , we choose the 4x2 matrix W with two linearly independent columns:

$$W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Pre-multiplying W by  $I - X(X'X)^{-1}X'$ , we obtain the 4x2 matrix U, which is orthogonal to X

$$10 U = \begin{bmatrix} 3 & -4 \\ -4 & 7 \\ -1 & -2 \\ 2 & -1 \end{bmatrix}$$

We then obtain the 2x12 matrix  $F^* = U' A$

$$10 F^* = \begin{bmatrix} 3 & 6 & 9 & -4 & -4 & -12 & -1 & -2 & -3 & 2 & 4 & 6 \\ -4 & -1 & -12 & 7 & 14 & 21 & -2 & -4 & -6 & -1 & -2 & -3 \end{bmatrix}$$

We can write the constraints in terms of the cell counts as

$$F^* \underline{\pi} = F^* R \underline{m} = F \underline{m} = \underline{0}$$

where R is the 12x12 diagonal matrix with diagonal elements  $r_{ii}=1/m_i$

$$R = \begin{bmatrix} 1/96 & & & \\ & 1/104 & & \\ & & 1/110 & \\ & & & 1/107 \end{bmatrix}$$

and  $\underline{m}' = (m_1, m_2, \dots, m_{12})$ .

The constraints that the sum of the probabilities within each of the four operations will equal one can be written

$$D' \underline{\pi} = \underline{1}_4$$

where  $D'$  is the 4x12 matrix

$$D' = \begin{bmatrix} 111 & & & \\ & 111 & & \\ & & 111 & \\ & & & 111 \end{bmatrix}$$

or, equivalently, in terms of cell counts,  $D' R \underline{m} = \underline{1}_4$ .

Thus, the constraints on the cell counts can be written as

$$L \underline{m} = L^* R \underline{m} = \underline{c}$$

where  $L^* = (F^* : D)$ .

A vector  $\underline{a}$  which satisfies these constraints is

$$\underline{a} = R^{-1} \underline{a}^*$$

where  $\underline{a}^* = (1/3, 1/3, \dots, 1/3)$ .

We can implement Wedderburn's algorithm using  $\underline{n}$ ,  $L$  and  $\underline{a}$  as given above.

The deviance is  $G^2 = 0.22$  with 2 df.

Alternatively, we can use the constraints implicit in the model

$$\phi_i = \mu + \beta u_i$$

The  $\phi_i$  are constrained to lie on a straight line:

$$\phi_2 - \phi_1 = \phi_3 - \phi_2 = \phi_4 - \phi_3 = \beta$$

so we can use any two sets of constraints:

$$\begin{aligned} (1) \quad & \phi_2 - \phi_1 = \phi_3 - \phi_2 \\ & \Rightarrow -\phi_1 + 2\phi_2 - \phi_3 = 0 \end{aligned}$$

$$\begin{aligned} (2) \quad & \phi_4 - \phi_3 = \phi_3 - \phi_2 \\ & \Rightarrow \phi_2 - 2\phi_3 + \phi_4 = 0 \end{aligned}$$

So the constraints can be written  $F^* \underline{\pi} = \underline{0}$

$$\text{where } F^* = \begin{bmatrix} -1 & -2 & -3 & 2 & 4 & 6 & -1 & -2 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & -2 & -4 & -6 & 1 & 2 & 3 \end{bmatrix}$$

To use Wedderburn's algorithm we can write the constraints as

$$L \underline{m} = \underline{c}$$

as illustrated previously.

## APPENDIX 6: QUASI-SYMMETRY

For a two-way table the model of quasi-symmetry is

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

where  $\lambda_{ij}^{XY} = \lambda_{ji}^{XY}$ .

### A6.1 One Dummy Variable

In this method we use one dummy variable to force the  $\lambda_{ij}^{XY}$  terms to obey the constraints. The dummy variable takes the value 1 for all cells where  $i=1$  or  $j=1$  (i.e. the first row and the first column). The dummy variable takes the values of the next consecutive  $r-1$  integers for the remainder of the diagonal cells. Then, for every pair of off-diagonal cells, of which neither cell is in the first row or first column, the next consecutive integer is assigned to both members of the pair.

The log-linear model involving  $X$ ,  $Y$  and the dummy variable is then fitted, i.e.

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^D$$

#### A6.1.1 Eye-testing example

We will illustrate this method using the data of Table 8.9, and will call the classifying variables "Right" and "Left", and the dummy variable "Dummy". Table A6.1 gives the values of the dummy variable corresponding to the levels of the classifying variables, and Table A6.2 gives the values of the classifying and dummy variables and the observed cell counts that would be used to fit the quasi-symmetry model:

Table A6.1: Values of the dummy variable

Right	Left			
	1	2	3	4
1	1	1	1	1
2	1	2	5	6
3	1	5	3	7
4	1	6	7	4

Table A6.2: Variables needed to fit quasi-symmetry model

Right	Left	Dummy	Count
1	1	1	1520
1	2	1	266
1	3	1	124
1	4	1	66
2	1	1	234
2	2	2	1512
2	3	5	432
2	4	6	78
3	1	1	117
3	2	5	362
3	3	3	1772
4	1	1	36
4	2	6	82
4	3	7	179
4	4	4	492

## A6.2 Two Dummy Variables

This method involves using two dummy variables. The first dummy variable takes on the value of the larger of  $i$  and  $j$  for the  $ij$ th cell, while the second dummy variable takes on the smaller of  $i$  and  $j$ . Then the log-linear model involving all the classifying and dummy variables, plus the interaction between the two dummy variables, is fitted, i.e.

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_i^{D1} + \lambda_j^{D2} + \lambda_{ij}^{D1D2}$$

The fitted values are then the MLEs  $\hat{m}_{ij}$ .

### A6.2.1 Eye-testing example

Table A6.3 gives the values of the classifying and dummy variables, and the observed cell counts that would be necessary to fit the quasi-symmetry model to the data of Table 8.9.

The dummy variables have been named "High" and "Low".

Table A6.3: Variables needed to fit quasi-symmetry model

Right	Left	High	Low	Count
1	1	1	1	1520
1	2	2	1	266
1	3	3	1	124
1	4	4	1	66
2	1	2	1	234
2	2	2	2	1512
2	3	3	2	432
2	4	4	2	78
3	1	3	1	117
3	3	3	2	362
3	3	3	3	205
4	1	4	1	36
4	2	4	2	82
4	3	4	3	179
4	4	4	4	492

## APPENDIX 7: MARGINAL HOMOGENEITY

For an  $r \times r$  table, the marginal homogeneity model specifies

$$m_{i+} = m_{+i} \\ \text{for } i=1, \dots, r.$$

The following sections explain two methods of estimation.

### A7.1 Method of Solving Simultaneous Equations

For an  $r \times r$  table, there are  $r$  constraints of the form  $m_{i+} = m_{+i}$ , of which  $r-1$  are independent. There is also one constraint of the form

$$\sum_{i \neq j} m_{ij} = N^* \quad (\text{where } N^* = N - \sum m_{ii}).$$

We write down the equations which define these constraints, then express  $r$  of the  $r(r-1)$  off-diagonal cells  $\{m_{ij}, i \neq j\}$  in terms of  $N^*$ , and the other  $r(r-2)$  off-diagonal cells. We can then write these equations as

$$\underline{\tilde{Y}} = \underline{\tilde{X}}\underline{\tilde{\beta}} + \underline{\tilde{O}} + \underline{\tilde{\epsilon}}$$

where  $\underline{\tilde{Y}}$  is the vector of cell counts,  $\underline{\tilde{X}}$  is the matrix containing the coefficients of the expected cell counts in the equations,  $\underline{\tilde{\beta}}$  is the vector of "parameters" (in this case  $r(r-2)$  of the off-diagonal counts),  $\underline{\tilde{O}}$  is an off-set vector (containing zeroes, diagonal counts and appropriate multiples of  $N^*$ ) and  $\underline{\tilde{\epsilon}}$  is a vector of Poisson errors. We can obtain the MLEs by using a Poisson error and an identity link function in a package like Genstat.

#### A7.1.1 Eye-Testing Example

We will illustrate using the data of Table 8.9. There are four constraints of the form  $m_{i+} = m_{+i}$ , of which only three are independent:

- (1)  $m_{1+} = m_{+1} \Rightarrow m_{12} + m_{13} + m_{14} = m_{21} + m_{31} + m_{41}$
- (2)  $m_{2+} = m_{+2} \Rightarrow m_{21} + m_{23} + m_{24} = m_{12} + m_{32} + m_{42}$
- (3)  $m_{3+} = m_{+3} \Rightarrow m_{31} + m_{32} + m_{34} = m_{13} + m_{23} + m_{43}$



The constraint that the cell counts must sum to the total can be expressed as

$$(4) \quad m_{21} + m_{31} + m_{41} + m_{12} + m_{42} + m_{13} + m_{23} + m_{43} + m_{14} + m_{24} + m_{34} = N^*$$

We can write  $m_{21}$ ,  $m_{42}$ ,  $m_{31}$  and  $m_{23}$  in terms of  $N^*$ ,  $m_{41}$ ,  $m_{12}$ ,  $m_{32}$ ,  $m_{13}$ ,  $m_{43}$ ,  $m_{14}$ ,  $m_{24}$  and  $m_{34}$  to give

$$m_{21} = N^* + 3m_{12} + 2m_{13} + 4m_{14} - 2m_{41} + 2m_{24} - m_{43} + 3m_{34} + 2m_{32}$$

$$m_{31} = N^* - 2m_{12} - m_{13} - 3m_{14} + m_{41} - 2m_{24} + m_{43} - 3m_{34} - 2m_{32}$$

$$m_{42} = m_{14} - m_{41} + m_{24} - m_{43} + m_{34}$$

$$m_{23} = N^* - 2m_{12} - 2m_{13} - 3m_{14} + m_{41} - 2m_{24} - 2m_{34} - m_{32}$$

We can obtain the MLEs by writing these equations in the form

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{0} + \underline{\varepsilon}, \text{ i.e.}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \\ Y_9 \\ Y_{10} \\ Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{15} \\ Y_{16} \end{bmatrix} = \begin{bmatrix} n_{11} \\ n_{21} \\ n_{31} \\ n_{41} \\ n_{12} \\ n_{22} \\ n_{32} \\ n_{42} \\ n_{13} \\ n_{23} \\ n_{33} \\ n_{43} \\ n_{14} \\ n_{24} \\ n_{34} \\ n_{44} \end{bmatrix} = \begin{bmatrix} 1520 \\ 234 \\ 117 \\ 36 \\ 266 \\ 1512 \\ 362 \\ 82 \\ 124 \\ 432 \\ 1772 \\ 179 \\ 66 \\ 78 \\ 205 \\ 492 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 4 & -2 & 2 & -1 & 3 & 2 \\ -2 & -1 & -3 & 1 & -2 & 1 & -3 & -2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & -2 & -3 & 1 & -2 & 0 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} m_{12} \\ m_{13} \\ m_{14} \\ m_{41} \\ m_{24} \\ m_{43} \\ m_{34} \\ m_{32} \end{bmatrix} + \begin{bmatrix} 1520 \\ -2181 \\ 2181 \\ 0 \\ 0 \\ 1512 \\ 0 \\ 0 \\ 0 \\ 2181 \\ 1772 \\ 0 \\ 0 \\ 0 \\ 0 \\ 492 \end{bmatrix} + \underline{\varepsilon}$$

## A7.2 Wedderburn's Method - Generalized Linear Models Specified in Terms of Constraints

This method was explained in Appendix A3.1.

Basically, we consider the  $r^2$  cell counts as forming an  $n \times 1$  vector  $\underline{n}$ .

If we express the model as

$$E(\underline{n}) = \underline{\lambda}$$

where  $\lambda_i = f(n_i)$ , with constraints  $L \underline{n} = \underline{0}$

then we may fit  $E(\underline{n}) = W^{-1} L' \underline{y}$

$$\text{where } W = \begin{bmatrix} -E[F_1''] & & \\ & \ddots & \\ & & -E[F_n''] \end{bmatrix} = \begin{bmatrix} 1/n_1 & & \\ & \ddots & \\ & & 1/n_n \end{bmatrix}$$

and then we use the residuals as the new value for  $\underline{n}$ .

## A7.2 Generalized Linear Models Specified in Terms of Constraints

We can simply use the Wedderburn's algorithm described in Appendix 3, Section A3.1.1.

### A7.2.1 Eye-testing example

For the eye-testing data of Table 8.9, the constraints  $m_{i+} = m_{+i}$  imply that

$$-m_{12} - m_{13} - m_{14} + m_{21} + m_{31} + m_{41} = 0$$

$$-m_{21} - m_{23} - m_{24} + m_{12} + m_{32} + m_{42} = 0$$

$$-m_{31} - m_{32} - m_{34} + m_{13} + m_{23} + m_{43} = 0$$

So  $L \underline{n} = L \underline{m} =$

$$\begin{bmatrix}
 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
 0 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\
 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & 0
 \end{bmatrix}
 \begin{bmatrix}
 m_{11} \\
 m_{21} \\
 m_{31} \\
 m_{41} \\
 m_{12} \\
 m_{22} \\
 m_{32} \\
 m_{42} \\
 m_{13} \\
 m_{23} \\
 m_{33} \\
 m_{43} \\
 m_{14} \\
 m_{24} \\
 m_{34} \\
 m_{44}
 \end{bmatrix}
 = \underset{\sim}{0}$$

The first estimate of  $\underset{\sim}{\eta}$  would be  $\underset{\sim}{\hat{y}} = \underset{\sim}{\hat{z}} =$

$$\begin{bmatrix}
 1520 \\
 234 \\
 117 \\
 36 \\
 266 \\
 1512 \\
 362 \\
 82 \\
 124 \\
 432 \\
 1772 \\
 179 \\
 66 \\
 78 \\
 205 \\
 492
 \end{bmatrix}$$

The first estimate of  $W$  would be

$$\begin{bmatrix} 1/1520 & & & \\ & 1/234 & & \\ & & \ddots & \\ & & & 1/492 \end{bmatrix}$$

so  $W^{-1} L'$  would be

$$\begin{bmatrix} 0 & 0 & 0 \\ 234 & -234 & 0 \\ 117 & 0 & -117 \\ 36 & 0 & 0 \\ -266 & 266 & 0 \\ 0 & 0 & 0 \\ 0 & 362 & -362 \\ 0 & 82 & 0 \\ -124 & 0 & 0 \\ 0 & -432 & 432 \\ 0 & 0 & 0 \\ 0 & 0 & 179 \\ -66 & 0 & 0 \\ 0 & -78 & 0 \\ 0 & 0 & -205 \\ 0 & 0 & 0 \end{bmatrix}$$

We would regress  $\underline{z}$  on  $W^{-1}L'$ , using  $W$  as the weight matrix, and then set the new estimate of  $\underline{\eta}$  to be the residuals,  $\underline{z} - \hat{\underline{z}}$ . We would then re-calculate  $W$  and cycle through the procedure until the estimated expected values,  $\hat{\underline{\eta}}$ , and the deviance converged.

If we decide to also use the constraint  $\sum_{i,j} m_{ij} = N$ , then the

constraint matrix would have an extra column of 1's and the constraints would be  $L \underline{\eta} = L \underline{m} =$

$$\begin{bmatrix} 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tilde{m} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ N \end{bmatrix}$$

A vector which satisfies these constraints can be formed from a table with symmetry

$$\tilde{a} = \begin{bmatrix} 1520 \\ 250 \\ 120.5 \\ 51 \\ 250 \\ 1512 \\ 397 \\ 80 \\ 120.5 \\ 397 \\ 1772 \\ 192 \\ 51 \\ 80 \\ 192 \\ 492 \end{bmatrix}$$

## APPENDIX 8: FITTING MODELS USING GENSTAT

Genstat (Alvey *et al.*, 1983) is probably the most powerful computer package commonly available for analyzing contingency tables. The GLM directive fits generalized linear models (Nelder and Wedderburn, 1972) using the Newton-Raphson algorithm. Loglinear models can be easily fitted by specifying independent Poisson sampling for the cell counts. Genstat automatically selects the log link function and uses  $\log m_{ij}$  as the response in the linear model. Logit models for dichotomous response variables can be fitted by specifying a binomial sampling distribution for which the logit,  $\log [\pi(1-\pi)]$ , is the default link function. Other types of models can be fitted using Genstat's versatile programming features.

The remarks made here also apply to GLIM (Baker and Nelder, 1978), which is virtually just a subset of Genstat with some differences in syntax.

The following examples illustrate how Genstat can be used to fit the models discussed in this thesis.

Table A8.1 contains the input for fitting the models (RY,A) and (RY,RA) to the data of Table 3.2. The 'REFERENCE' directive simply indicates the job-name, while 'UNITS' specifies there are 16 observations. RELIGION, YEAR and ATTITUDE are defined as factors, while COUNT is defined as a variate. The 'TERMS' statement contains all variables which might possibly be included in the model. The next statement specifies that the response variable, COUNT, has a Poisson distribution. The first 'FIT' statement fits the model (RY,A), i.e.

$$\log m_{ijk} = \mu + \lambda_i^R + \lambda_k^A + \lambda_{ij}^{RY}$$

while the second one fits the model (RY,RA). RELIGION \* YEAR means include the main effects for RELIGION and YEAR, plus their interaction, RELIGION, YEAR. The default output from the 'FIT' statement is the regression coefficients, plus the deviance. To obtain the estimated expected values as well, we could use

'FIT/PRIN = CAU'

Table A8.2 contains the Genstat output. Instead of the usual identifiability constraint that the sum of the parameter estimates equals zero, Genstat uses the constraint that the parameter estimate for the first level of any factor or interaction is zero. For instance, for the model RELIGION \* YEAR + ATTITUDE, the estimates of RELIGION1, YEAR1, RELIGION1.YEAR1, RELIGION2.YEAR1, RELIGION3, YEAR1, RELIGION4.YEAR1, and RELIGION1.YEAR2 are all zero. This makes no difference whatsoever to the fitted values, etc, and the parameter estimates such as  $\lambda_{ij}^{RY}$  subject to the constraint

$$\sum_i \lambda_{ij}^{RY} = \sum_j \lambda_{ij}^{RY} = 0$$

can be calculated using a little algebra.

The remainder of this Appendix contains Genstat programmes which can be used to fit the models discussed in this thesis. They should be self-explanatory if read in conjunction with the appropriate Sections in the text, and also the previous Appendices in which estimation methods are discussed.

'REFERENCE' RELIGION

..

SEE APPENDIX 2, SECTION A2.1.1  
 FITTING LOGLINEAR MODEL OF INDEPENDENCE  
 AND THE SATURATED MODEL  
 TO TABLE 2.1 - RELIGION BY ATTITUDE TO ABORTION  
 FROM KNOKE & BURKE (1980)

..

'UNITS' \$ 8

'FACTOR' RELIGION \$4 = 1,1,2,2,3,3,4,4

: ATTITUDE \$2 = 1,2,1,2,1,2,1,2

'VARIATE' COUNT

'READ/PRIN=DEM' COUNT

'RUN'

460 498

147 240

41 10

65 17

'EOD'

'TABLE' T1 \$ RELIGION, ATTITUDE

'TABULATE' COUNT ; T1

'PRINT' T1 \$ 10

'RUN'

'TERMS' COUNT+RELIGION\*ATTITUDE

'Y/ERROR=POISSON' COUNT

'CAPTION' '' (R,A) ''

'FIT/PRIN=CAU' RELIGION+ATTITUDE ; FVAL = FIT ; RES = RESID

'RUN'

'TABULATE' FIT ; T1

'PRINT' T1 \$ 10.3

'RUN'

'FIT/PRIN=CAU' RELIGION\*ATTITUDE ; FVAL = FIT ; RES = RESID

'RUN'

'TABULATE' FIT ; T1

'PRINT' T1 \$ 10.3

'RUN'

'CLOSE'

'STOP'



GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1  'REFERENCE' RELIGION
-2  ''
-3      SEE APPENDIX 2, SECTION A2.1.1
-4      FITTING LOGLINEAR MODELS OF INDEPENDENCE
-5      AND THE SATURATED MODEL
-6      TO TABLE 2.1 - RELIGION BY ATTITUDE TO ABORTION
-7      FROM KNOKE & BURKE (1980)
8      ''
9  'UNITS' $ 8
10 'FACTOR' RELIGION $4 = 1,1,2,2,3,3,4,4
11      : ATTITUDE $2 = 1,2,1,2,1,2,1,2
12 'VARIATE' COUNT
13 'READ/PRIN=DEM' COUNT
14 'RUN'

```

```

15 460 498
16 147 240
17 41 10
18 65 17
19 'EOD'

```

IDENTIFIER	MINIMUM	MEAN	MAXIMUM	VALUES	MISSING
COUNT	10.0	184.8	498.0	8	0

```

20 'TABLE' T1 $ RELIGION, ATTITUDE
21 'TABULATE' COUNT ; T1
22 'PRINT' T1 $ 10
23 'RUN'

```

	T1	
ATTITUDE	1	2
RELIGION		
1	460	498
2	147	240
3	41	10
4	65	17

```

24 'TERMS' COUNT+RELIGION*ATTITUDE
25 'Y/ERROR=POISSON' COUNT
26 'CAPTION' '' (R,A) ''

```

27 'FIT/PRIN=CAU' RELIGION+ATTITUDE ; FVAL = FIT ; RES = RESID  
 28 'RUN'

(R,A)

27.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
 Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	6.1359	0.0421	145.85
RELIGION 2	-0.9064	0.0602	-15.05
RELIGION 3	-2.933	0.144	-20.41
RELIGION 4	-2.458	0.115	-21.37
ATTITUDE 2	0.0704	0.0521	1.35

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	72.44	24.15

\*\*\* OBSERVED AND FITTED VALUES \*\*\*

	OBSERVED	FITTED	RESIDUAL
1	460.0	462.1	-0.1
2	498.0	495.9	0.1
3	147.0	186.7	-2.9
4	240.0	200.3	2.8
5	41.0	24.6	3.3
6	10.0	26.4	-3.2
7	65.0	39.6	4.0
8	17.0	42.4	-3.9

29 'TABULATE' FIT ; T1  
 30 'PRINT' T1 \$ 10.3  
 31 'RUN'

ATTITUDE RELIGION	T1	
	1	2
1	462.148	495.853
2	186.692	200.308
3	24.603	26.397
4	39.558	42.442

32 'FIT/PRIN=CAU' RELIGION\*ATTITUDE ; FVAL = FIT ; RES = RESID  
33 'RUN'

32.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	6.1312	0.0466	131.50
RELIGION 2	-1.1408	0.0947	-12.04
RELIGION 3	-2.418	0.163	-14.83
RELIGION 4	-1.957	0.133	-14.77
ATTITUDE 2	0.0794	0.0647	1.23
RELIGION 2 .ATTITUDE 2	0.411	0.123	3.34
RELIGION 3 .ATTITUDE 2	-1.490	0.358	-4.16
RELIGION 4 .ATTITUDE 2	-1.421	0.280	-5.07

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	0	0	*

\*\*\* OBSERVED AND FITTED VALUES \*\*\*

	OBSERVED	FITTED	RESIDUAL
1	460.00	460.00	0.00
2	498.00	498.00	0.00
3	147.00	147.00	0.00
4	240.00	240.00	0.00

5	41.00	41.00	0.00
6	10.00	10.00	-0.00
7	65.00	65.00	0.00
8	17.00	17.00	0.00

```

34 'TABULATE'  FIT ; T1
35 'PRINT'     T1 $ 10.3
36 'RUN'

```

	T1	
ATTITUDE	1	2
RELIGION		
1	460.000	498.000
2	147.000	240.000
3	41.000	10.000
4	65.000	17.000

```

38 'CLOSE'

```

```

***** END OF RELIGION.  MAXIMUM OF 1252 DATA UNITS USED AT LINE 32
(31516 LEFT)

```

```

'REFER' DRUG
      ' ' APPENDIX 3, SECTION A3.1.2
          FITTING THE LINEAR MODEL OF NO
          INTERACTION BETWEEN DRUGS TO
          TABLE 8.13 USING WEDDERBURN'S METHOD
'UNITS' $ 26
'VARIATE' COUNT
      : ASTAR
      : A
'FACTOR' DRUGS $ 7 = 8(1),4(2),4(3),4(4),2(5),2(6),2(7)
'SCALAR' DEV
'READ/PRIN=Z' COUNT
'RUN'
6 16 2 4 2 4 6 6
12 4 4 8
5 10 4 6
4 12 5 5
10 6
11 4
5 9
'EOD'
'TABLE/M' T1 $ DRUGS
'READ/PRIN=Z' L1STAR,L2STAR,L3STAR,L4STAR,L5STAR,L6STAR,
              L7STAR,L8STAR,L9STAR
'RUN'
1 0 0 1 0 0 1 0 0
1 0 0 1 0 0 0 0 0
1 0 0 0 0 0 1 0 0
1 0 0 0 0 0 0 0 0
0 0 0 1 0 0 1 0 0
0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0
-1 1 0 -1 1 0 0 0 0
-1 1 0 0 0 0 0 0 0
0 0 0 -1 1 0 0 0 0
0 0 0 0 0 0 0 0 0
0 -1 1 0 0 0 -1 1 0
0 -1 1 0 0 0 0 0 0
0 0 0 0 0 0 -1 1 0
0 0 0 0 0 0 0 0 0
0 0 0 0 -1 1 0 -1 1
0 0 0 0 -1 1 0 0 0
0 0 0 0 0 0 0 -1 1
0 0 -1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0
0 0 0 0 0 -1 0 0 0
0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 -1
0 0 0 0 0 0 0 0 0
'EOD'
'READ/P,PRIN=Z' L10STAR,L11STAR,L12STAR,L13STAR,
                L14STAR,L15STAR,L16STAR

```

```

'RUN'
1 0 0 0 0 0 0
1 0 0 0 0 0 0
1 0 0 0 0 0 0
1 0 0 0 0 0 0
1 0 0 0 0 0 0
1 0 0 0 0 0 0
1 0 0 0 0 0 0
1 0 0 0 0 0 0

0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0
0 1 0 0 0 0 0

0 0 1 0 0 0 0
0 0 1 0 0 0 0
0 0 1 0 0 0 0
0 0 1 0 0 0 0

0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0
0 0 0 1 0 0 0

0 0 0 0 1 0 0
0 0 0 0 1 0 0

0 0 0 0 0 1 0
0 0 0 0 0 1 0

0 0 0 0 0 0 1
0 0 0 0 0 0 1
'EOD'
'READ/PRIN=Z'    T
'RUN'
46 46 46 46 46 46 46 46 28 28 28 28 25 25 25 25
 26 26 26 26 16 16 15 15 14 14
'EOD'
'READ/PRIN=Z'    AHASH
'RUN'
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
'EOD'
'READ/PRIN=Z'    NO
'RUN'
8 8 8 8 8 8 8 8 4 4 4 4 4 4 4 4 4 4 2 2 2 2 2 2
'EOD'
'CALC'    ASTAR = AHASH/NO
'CALC'    L1 = L1STAR/T
'CALC'    L2 = L2STAR/T
'CALC'    L3 = L3STAR/T
'CALC'    L4 = L4STAR/T
'CALC'    L5 = L5STAR/T
'CALC'    L6 = L6STAR/T

```

```

'CALC'   L7 = L7STAR/T
'CALC'   L8 = L8STAR/T
'CALC'   L9 = L9STAR/T
:        L10 = L10STAR/T
:        L11 = L11STAR/T
:        L12 = L12STAR/T
:        L13 = L13STAR/T
:        L14 = L14STAR/T
:        L15 = L15STAR/T
:        L16 = L16STAR/T
'CALC'   A  = ASTAR*T

'CALC'   ETA = COUNT
'CALC'   Z  = COUNT - A
'PRINT/P' COUNT,ETA,Z  $ 10.4
'RUN'
'TABULATE/PRIN=T'   ETA ; T1
'RUN'
'FOR' I = 1...3
  'CALC' W = 1/ETA
  'CALC' P1 = ETA*L1
  'CALC' P2 = ETA*L2
  'CALC' P3 = ETA*L3
  'CALC' P4 = ETA*L4
  'CALC' P5 = ETA*L5
  'CALC' P6 = ETA*L6
  'CALC' P7 = ETA*L7
  'CALC' P8 = ETA*L8
  'CALC' P9 = ETA*L9
  'CALC' P10 = ETA*L10
  'CALC' P11 = ETA*L11
  'CALC' P12 = ETA*L12
  'CALC' P13 = ETA*L13
  'CALC' P14 = ETA*L14
  'CALC' P15 = ETA*L15
  'CALC' P16 = ETA*L16
  'TERMS/WT=W'  Z+P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12
                +P13+P14+P15+P16
  'Y'          Z
  'FIT/INT=N,PRIN=Z'  P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11
                    +P12+P13+P14+P15+P16      ; FVAL = PRED
  'CALC' ETA = COUNT - PRED
  'PRINT/S' ETA $10.4
  'CALC' DEV = 2*(SUM(ETA-COUNT) + SUM(COUNT*LOG(COUNT/ETA)))
  'PRINT' DEV $ 10.4
'REPEAT'
'RUN'
'CLOSE'
'STOP'

```

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1  'REFER' DRUG
-2      '' APPENDIX 3, SECTION A3.1.2
-3      FITTING THE LINEAR MODEL OF NO
-4      INTERACTION BETWEEN DRUGS TO
5      TABLE 8.13 USING WEDDERBURN'S METHOD
''

6  'UNITS' $ 26
7  'VARIATE' COUNT
8      : ASTAR
9      : A
10 'FACTOR' DRUGS $ 7 = 8(1),4(2),4(3),4(4),2(5),2(6),2(7)
11 'SCALAR' DEV
12 'READ/PRIN=Z' COUNT
13 'RUN'

22 'TABLE/M' T1 $ DRUGS
23 'READ/PRIN=Z' L1STAR,L2STAR,L3STAR,L4STAR,L5STAR,L6STAR,
24 L7STAR,L8STAR,L9STAR
25 'RUN'

53 'READ/P,PRIN=Z' L10STAR,L11STAR,L12STAR,L13STAR,
54 L14STAR,L15STAR,L16STAR
55 'RUN'

89 'READ/PRIN=Z' T
90 'RUN'

94 'READ/PRIN=Z' AHASH
95 'RUN'

98 'READ/PRIN=Z' NO
99 'RUN'

102 'CALC' ASTAR = AHASH/NO
103 'CALC' L1 = L1STAR/T
104 'CALC' L2 = L2STAR/T

```



```

105 'CALC'    L3 = L3STAR/T
106 'CALC'    L4 = L4STAR/T
107 'CALC'    L5 = L5STAR/T
108 'CALC'    L6 = L6STAR/T
109 'CALC'    L7 = L7STAR/T
110 'CALC'    L8 = L8STAR/T
111 'CALC'    L9 = L9STAR/T
112      :    L10 = L10STAR/T
113      :    L11 = L11STAR/T
114      :    L12 = L12STAR/T
115      :    L13 = L13STAR/T
116      :    L14 = L14STAR/T
117      :    L15 = L15STAR/T
118      :    L16 = L16STAR/T
119 'CALC'    A  = ASTAR*T
121 'CALC'    ETA = COUNT
122 'CALC'    Z  = COUNT - A
123 'PRINT/P'  COUNT,ETA,Z  $ 10.4
124 'RUN'

```

COUNT	ETA	Z
6.0000	6.0000	0.2500
16.0000	16.0000	10.2500
2.0000	2.0000	-3.7500
4.0000	4.0000	-1.7500
2.0000	2.0000	-3.7500
4.0000	4.0000	-1.7500
6.0000	6.0000	0.2500
6.0000	6.0000	0.2500
12.0000	12.0000	5.0000
4.0000	4.0000	-3.0000
4.0000	4.0000	-3.0000
8.0000	8.0000	1.0000
5.0000	5.0000	-1.2500
10.0000	10.0000	3.7500
4.0000	4.0000	-2.2500
6.0000	6.0000	-0.2500
4.0000	4.0000	-2.5000
12.0000	12.0000	5.5000
5.0000	5.0000	-1.5000
5.0000	5.0000	-1.5000
10.0000	10.0000	2.0000
6.0000	6.0000	-2.0000
11.0000	11.0000	3.5000
4.0000	4.0000	-3.5000
5.0000	5.0000	-2.0000
9.0000	9.0000	2.0000

```

125 'TABULATE/PRIN=T'  ETA ; T1
126 'RUN'

```

	T1
DRUGS	
1	46.00
2	28.00
3	25.00
4	26.00
5	16.00
6	15.00
7	14.00
MARGIN	170.00

```

127 'FOR' I = 1...3
128 'CALC' W = 1/ETA
129 'CALC' P1 = ETA*L1
130 'CALC' P2 = ETA*L2
131 'CALC' P3 = ETA*L3
132 'CALC' P4 = ETA*L4
133 'CALC' P5 = ETA*L5
134 'CALC' P6 = ETA*L6
135 'CALC' P7 = ETA*L7
136 'CALC' P8 = ETA*L8
137 'CALC' P9 = ETA*L9
138 'CALC' P10 = ETA*L10
139 'CALC' P11 = ETA*L11
140 'CALC' P12 = ETA*L12
141 'CALC' P13 = ETA*L13
142 'CALC' P14 = ETA*L14
143 'CALC' P15 = ETA*L15
144 'CALC' P16 = ETA*L16
145 'TERMS/WT=W' Z+P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12
146                +P13+P14+P15+P16
147 'Y' Z
148 'FIT/INT=N,PRIN=Z' P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11
149                +P12+P13+P14+P15+P16 ; FVAL = PRED
150 'CALC' ETA = COUNT - PRED
151 'PRINT/S' ETA $10.4
152 'CALC' DEV = 2*(SUM(ETA-COUNT) + SUM(COUNT*LOG(COUNT/ETA)))

153 'PRINT' DEV $ 10.4
154 'REPEAT'
155 'RUN'

```

ETA

6.1441  
16.1209  
1.9111  
3.7564  
2.1171  
4.1683  
5.9405  
5.8417  
13.2604  
3.7420  
4.1180  
6.8796  
4.9259  
10.2548  
3.8310  
5.9883  
4.1115  
12.0257  
4.9957  
4.8671  
9.7157  
6.2843  
9.3099  
5.6901  
4.9039  
9.0961

DEV 1.2564

ETA  
6.1119  
16.0299  
1.9315  
3.8038  
2.1082  
4.1449  
5.9825  
5.8872  
13.2288  
3.7399  
4.0551  
6.9763  
4.9187  
10.2319  
3.8499  
5.9996  
4.0855  
11.9638  
5.0337  
4.9170  
9.6964  
6.3036

9.2592  
5.7408  
4.9104  
9.0896

DEV 1.2495

ETA  
6.1123  
16.0326  
1.9314  
3.8028  
2.1075  
4.1437  
5.9828  
5.8869  
13.2270  
3.7429  
4.0576  
6.9725  
4.9194  
10.2323  
3.8491  
5.9993  
4.0853  
11.9647  
5.0340  
4.9161  
9.6971  
6.3029  
9.2596  
5.7404  
4.9104  
9.0896

DEV 1.2495

156 'CLOSE'

\*\*\*\*\* END OF DRUG. MAXIMUM OF 3942 DATA UNITS USED AT LINE 145 (2  
8826 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFER' DRUG
-3      '' APPENDIX 3, SECTION A3.2.1
-4      FITTING THE LINEAR MODEL OF NO
-5      INTERACTION BETWEEN DRUGS TO
-6      TABLE 8.13 USING WEDDERBURN'S METHOD
7      USING A DIFFERENT FORM OF THE CONSTRAINTS ''
8  'UNITS' $ 26
9  'VARIATE' COUNT
10     : ASTAR
11     : A
12  'FACTOR' DRUGS $ 7 = 8(1),4(2),4(3),4(4),2(5),2(6),2(7)
13  'SCALAR' DEV
14  'READ/PRIN=Z' COUNT
15  'RUN'

```

```

24  'TABLE/M' T1 $ DRUGS
25  'READ/PRIN=DEM' L1STAR,L2STAR,L3STAR,L4STAR,L5STAR,L6STAR,
26                      L7STAR,L8STAR,L9STAR
27  'RUN'

```

```

28      3 -1 -1 3 -1 -1 3 -1 -1
29      3 -1 -1 3 -1 -1 0 0 0
30      3 -1 -1 0 0 0 3 -1 -1
31      3 -1 -1 0 0 0 0 0 0
32      0 0 0 3 -1 -1 3 -1 -1
33      0 0 0 3 -1 -1 0 0 0
34      0 0 0 0 0 0 3 -1 -1
35      0 0 0 0 0 0 0 0 0
36     -1 3 -1 -1 3 -1 0 0 0
37     -1 3 -1 0 0 0 0 0 0
38      0 0 0 -1 3 -1 0 0 0
39      0 0 0 0 0 0 0 0 0
40     -1 -1 3 0 0 0 -1 3 -1
41     -1 -1 3 0 0 0 0 0 0
42      0 0 0 0 0 0 -1 3 -1
43      0 0 0 0 0 0 0 0 0
44      0 0 0 -1 -1 3 -1 -1 3
45      0 0 0 -1 -1 3 0 0 0
46      0 0 0 0 0 0 -1 -1 3
47      0 0 0 0 0 0 0 0 0
48     -1 -1 -1 0 0 0 0 0 0
49      0 0 0 0 0 0 0 0 0
50      0 0 0 -1 -1 -1 0 0 0
51      0 0 0 0 0 0 0 0 0

```

```

52      0  0  0  0  0  0  -1 -1 -1
53      0  0  0  0  0  0  0  0  0
54  'EOD'

```

IDENTIFIER	MINIMUM	MEAN	MAXIMUM	VALUES	MISSING
L1STAR	-1.0000	0.2692	3.0000	26	0
L2STAR	-1.00000	-0.03846	3.00000	26	0 SKEW
L3STAR	-1.00000	-0.03846	3.00000	26	0 SKEW
L4STAR	-1.0000	0.2692	3.0000	26	0
L5STAR	-1.00000	-0.03846	3.00000	26	0 SKEW
L6STAR	-1.00000	-0.03846	3.00000	26	0 SKEW
L7STAR	-1.0000	0.2692	3.0000	26	0
L8STAR	-1.00000	-0.03846	3.00000	26	0 SKEW
L9STAR	-1.00000	-0.03846	3.00000	26	0 SKEW

```

55  'READ/P,PRIN=Z'      L10STAR,L11STAR,L12STAR,L13STAR,
56                          L14STAR,L15STAR,L16STAR
57  'RUN'

```

```

91  'READ/PRIN=Z'      T
92  'RUN'

```

```

96  'READ/PRIN=Z'      AHASH
97  'RUN'

```

```

100 'READ/PRIN=Z'      NO
101 'RUN'

```

```

104 'CALC'      ASTAR = AHASH/NO
105 'CALC'      L1 = L1STAR/T
106 'CALC'      L2 = L2STAR/T
107 'CALC'      L3 = L3STAR/T
108 'CALC'      L4 = L4STAR/T
109 'CALC'      L5 = L5STAR/T
110 'CALC'      L6 = L6STAR/T
111 'CALC'      L7 = L7STAR/T
112 'CALC'      L8 = L8STAR/T
113 'CALC'      L9 = L9STAR/T
114      :      L10 = L10STAR/T

```

```

115      :      L11 = L11STAR/T
116      :      L12 = L12STAR/T
117      :      L13 = L13STAR/T
118      :      L14 = L14STAR/T
119      :      L15 = L15STAR/T
120      :      L16 = L16STAR/T
121 'CALC'  A  = ASTAR*T
122 'RUN'

```

```

123 'CALC'  ETA = COUNT
124 'CALC'  Z  = COUNT - A
125 'PRINT/P' COUNT,ETA,Z  $ 10.4
126 'RUN'

```

COUNT	ETA	Z
6.0000	6.0000	0.2500
16.0000	16.0000	10.2500
2.0000	2.0000	-3.7500
4.0000	4.0000	-1.7500
2.0000	2.0000	-3.7500
4.0000	4.0000	-1.7500
6.0000	6.0000	0.2500
6.0000	6.0000	0.2500
12.0000	12.0000	5.0000
4.0000	4.0000	-3.0000
4.0000	4.0000	-3.0000
8.0000	8.0000	1.0000
5.0000	5.0000	-1.2500
10.0000	10.0000	3.7500
4.0000	4.0000	-2.2500
6.0000	6.0000	-0.2500
4.0000	4.0000	-2.5000
12.0000	12.0000	5.5000
5.0000	5.0000	-1.5000
5.0000	5.0000	-1.5000
10.0000	10.0000	2.0000
6.0000	6.0000	-2.0000
11.0000	11.0000	3.5000
4.0000	4.0000	-3.5000
5.0000	5.0000	-2.0000
9.0000	9.0000	2.0000

```

127 'TABULATE/PRIN=T'  ETA ; T1
128 'RUN'

```

	T1
DRUGS	
1	46.00
2	28.00
3	25.00
4	26.00
5	16.00
6	15.00
7	14.00
MARGIN	170.00

```

129 'FOR' I = 1...3
130   'CALC' W = 1/ETA
131   'CALC' P1 = ETA*L1
132   'CALC' P2 = ETA*L2
133   'CALC' P3 = ETA*L3
134   'CALC' P4 = ETA*L4
135   'CALC' P5 = ETA*L5
136   'CALC' P6 = ETA*L6
137   'CALC' P7 = ETA*L7
138   'CALC' P8 = ETA*L8
139   'CALC' P9 = ETA*L9
140   'CALC' P10 = ETA*L10
141   'CALC' P11 = ETA*L11
142   'CALC' P12 = ETA*L12
143   'CALC' P13 = ETA*L13
144   'CALC' P14 = ETA*L14
145   'CALC' P15 = ETA*L15
146   'CALC' P16 = ETA*L16
147   'CALC' SL1ETA = SUM(P1)
148   'CALC' SL2ETA = SUM(P2)
149   'CALC' SL3ETA = SUM(P3)
150   'CALC' SL4ETA = SUM(P4)
151   'CALC' SL5ETA = SUM(P5)
152   'CALC' SL6ETA = SUM(P6)
153   'CALC' SL7ETA = SUM(P7)
154   'CALC' SL8ETA = SUM(P8)
155   'CALC' SL9ETA = SUM(P9)
156   'CALC' SL10ETA = SUM(P10)
157   'CALC' SL11ETA = SUM(P11)
158   'CALC' SL12ETA = SUM(P12)
159   'CALC' SL13ETA = SUM(P13)
160   'CALC' SL14ETA = SUM(P14)
161   'CALC' SL15ETA = SUM(P15)
162   'CALC' SL16ETA = SUM(P16)
163   'TERMS/WT=W' Z+P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11
164                   +P12+P13+P14+P15+P16
165   'Y' Z
166   'FIT/INT=N,PRIN=Z' P1+P2+P3+P4+P5+P6+P7+P8+P9+P10
167                   +P11+P12+P13+P14+P15+P16 ; FVAL = PRED

```



```

168      'CALC'  ETA = COUNT - PRED
169      'PRINT/S'  ETA  $10.4
170      'CALC'  DEV = 2*(SUM(ETA-COUNT) + SUM(COUNT*LOG(COUNT/ETA)))

171      'PRINT'  DEV  $ 10.4
172      'REPEAT'
173      'RUN'

```

```

      ETA
6.1441
16.1209
1.9111
3.7564
2.1171
4.1683
5.9405
5.8417
13.2604
3.7420
4.1180
6.8796
4.9259
10.2548
3.8310
5.9883
4.1115
12.0257
4.9957
4.8671
9.7157
6.2843
9.3099
5.6901
4.9039
9.0961

```

```

      DEV      1.2564

```

```

      ETA
6.1119
16.0299
1.9315
3.8038
2.1082
4.1449
5.9825
5.8872
13.2288

```

3.7399  
4.0551  
6.9763  
4.9187  
10.2319  
3.8499  
5.9996  
4.0855  
11.9638  
5.0337  
4.9170  
9.6964  
6.3036  
9.2592  
5.7408  
4.9104  
9.0896

DEV 1.2495

ETA  
6.1123  
16.0326  
1.9314  
3.8028  
2.1075  
4.1437  
5.9828  
5.8869  
13.2270  
3.7429  
4.0576  
6.9725  
4.9194  
10.2323  
3.8491  
5.9993  
4.0853  
11.9647  
5.0340  
4.9161  
9.6971  
6.3029  
9.2596  
5.7404  
4.9104  
9.0896

DEV 1.2495

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFERENCE'  RELIGION
-3      ''  SEE SECTION 2.2.5
-4      FITTING VARIOUS LOGLINEAR MODELS
-5      TO RELIGION BY ATTITUDE TO ABORTION
-6      TABLE 2.3
-7      FROM KNOKE & BURKE (1980, P72)
8
9  'UNITS'  $ 16
10 'FACTOR'  RELIGION  $4 = 4(1...4)
11      :      YEAR      $2 = 2(1,2)4
12      :      ATTITUDE  $2 = (1,2)8
13 'VARIATE'  COUNT
14 'READ/PRIN=D' COUNT
15 'RUN'

```

```

16 460 498 424 501
17 147 240 151 225
18 41 10 23 6
19 65 17 88 30
20 'EOD'

```

```

21 'TABLE'      T1 $ RELIGION,  YEAR,  ATTITUDE
22 'TABULATE'   COUNT ; T1
23 'PRINT'      T1 $ 10
24 'RUN'

```

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	460	498
	2	424	501
2	1	147	240
	2	151	225
3	1	41	10
	2	23	6
4	1	65	17
	2	88	30

```

25 'TERMS' COUNT+RELIGION*YEAR*ATTITUDE

```

```

26 'Y/ERROR=POISSON'      COUNT
27 'CAPTION'  ''          (R,T,A)  ''
28 'FIT/PRIN=A' RELIGION+YEAR+ATTITUDE ; FVAL = PRED
29 'TABULATE'  PRED ; T1
30 'PRINT'      T1 $ 10.2
31 'RUN'

```

(R,T,A)

```

28.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	10	143.2	14.32

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	454.77	496.38
	2	445.54	486.31
2	1	184.28	201.14
	2	180.54	197.05
3	1	19.32	21.09
	2	18.93	20.66
4	1	48.30	52.72
	2	47.32	51.65

```

32 'CAPTION'  ''          (RT,A)  ''
33 'FIT/PRIN=A' RELIGION*YEAR+ATTITUDE; FVAL = PRED
34 'TABULATE'  PRED ; T1
35 'PRINT'      T1 $ 10.2
36 'RUN'

```

(RT,A)

```

33.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	7	130.2	18.59

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	458.05	499.95
	2	442.27	482.73
2	1	185.04	201.96
	2	179.78	196.22
3	1	24.38	26.62
	2	13.87	15.13
4	1	39.21	42.79
	2	56.42	61.58

```
37 'CAPTION'    ''            (R,TA)    ''  
38 'FIT/PRIN=A' RELIGION+YEAR*ATTITUDE; FVAL = PRED  
39 'TABULATE'   PRED ; T1  
40 'PRINT'       T1 $ 10.2  
41 'RUN'
```

(R,TA)

38.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	9	143.0	15.89

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	458.84	492.31
	2	441.47	490.38
2	1	185.93	199.49
	2	178.89	198.70
3	1	19.49	20.92
	2	18.76	20.83
4	1	48.74	52.29
	2	46.89	52.08

```

42 'CAPTION'      ''      (T,RA)  ''
43 'FIT/PRIN=A' RELIGION+YEAR+ATTITUDE
44               +RELIGION.ATTITUDE; FVAL = PRED
45 'TABULATE'    PRED ; T1
46 'PRINT'       T1 $ 10.2
47 'RUN'

```

(T,RA)

43.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	7	14.96	2.137

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	446.53	504.62
	2	437.47	494.38
2	1	150.53	234.88
	2	147.47	230.12
3	1	32.33	8.08
	2	31.67	7.92
4	1	77.28	23.74
	2	75.72	23.26

```

48 'CAPTION'  ''          (RT,RA)  ''
49 'FIT/PRIN=A' RELIGION*YEAR+ATTITUDE
50           +RELIGION.ATTITUDE; FVAL = PRED
51 'TABULATE'  PRED ; T1
52 'PRINT'     T1 $ 10.2
53 'RUN'

```

(RT,RA)

```

49.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	4	1.888	0.4720

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	449.75	508.25
	2	434.25	490.75
2	1	151.15	235.85
	2	146.85	229.15
3	1	40.80	10.20
	2	23.20	5.80
4	1	62.73	19.27
	2	90.27	27.73

```

54 'CAPTION'  ''          (RT,TA)  ''
55 'FIT/PRIN=A' RELIGION*YEAR+ATTITUDE
56           +YEAR.ATTITUDE; FVAL = PRED
57 'TABULATE'  PRED ; T1
58 'PRINT'     T1 $ 10.2
59 'RUN'

```

(RT,TA)

55.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
 Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
 SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	6	129.9	21.66

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	462.15	495.85
	2	438.23	486.77
2	1	186.69	200.31
	2	178.13	197.87
3	1	24.60	26.40
	2	13.74	15.26
4	1	39.56	42.44
	2	55.90	62.10

```

60 'CAPTION' ' ' (RA,TA) ' '
61 'FIT/PRIN=A' RELIGION+YEAR+ATTITUDE+RELIGION.ATTITUDE
62           +YEAR.ATTITUDE ; FVAL = PRED
63 'TABULATE' PRED ; T1
64 'PRINT'    T1 $ 10.2
65 'RUN'

```

(RA,TA)

61.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
 Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
 SCALE PARAMETER IS    1.000



	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	6	14.74	2.457

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	450.53	500.48
	2	433.47	498.52
2	1	151.88	232.96
	2	146.12	232.04
3	1	32.62	8.02
	2	31.38	7.98
4	1	77.98	23.55
	2	75.02	23.45

```

66 'CAPTION' ' ' (RT,TA,RA) ' '
67 'FIT/PRIN=A' RELIGION+YEAR*ATTITUDE
68 +RELIGION.ATTITUDE+RELIGION.YEAR ; FVAL = PRED
69 'TABULATE' PRED ; T1
70 'PRINT' T1 $ 10.2
71 'RUN'

```

(RT,TA,RA)

67.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	1.466	0.4886

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	455.51	502.49
	2	428.49	496.51

2	1	153.38	233.62
	2	144.62	231.38
3	1	40.95	10.05
	2	23.05	5.95
4	1	63.16	18.84
	2	89.84	28.16

```
72 'CAPTION' '' (RTA) ''
73 'FIT/PRIN=A' RELIGION*YEAR*ATTITUDE ; FVAL = PRED
74 'TABULATE' PRED ; T1
75 'PRINT' T1 $ 10.2
76 'RUN'
```

(RTA)

73.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	0	0	*

RELIGION	ATTITUDE YEAR	T1	
		1	2
1	1	460.00	498.00
	2	424.00	501.00
2	1	147.00	240.00
	2	151.00	225.00
3	1	41.00	10.00
	2	23.00	6.00
4	1	65.00	17.00
	2	88.00	30.00

```
78 'CLOSE'
```

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

```

2  'REFERENCE' DUMPING
-3      '' SEE SECTIONS 4.1.1, 4.2.1, 4.3.2
-4      FITTING THE LOGLINEAR MODELS OF INDEPENDENCE,
-5      UNIFORM ASSOCIATION, AND ROW EFFECTS TO THE
6      DUMPING SEVERITY DATA OF TABLE 1.6 ''
8  'UNITS' $ 12
9  'FACTOR' OPERAT $4
10     : DUMP $3
11  'VARIATE' COUNT
12     : V
13     : U
14     : UV
15  'READ/PRIN=D' OPERAT,DUMP,U,V,COUNT
16  'RUN'

```

```

17 1 1 -1.5 -1 61
18 1 2 -1.5 0 28
19 1 3 -1.5 1 7
20 2 1 -0.5 -1 68
21 2 2 -0.5 0 23
22 2 3 -0.5 1 13
23 3 1 0.5 -1 58
24 3 2 0.5 0 40
25 3 3 0.5 1 12
26 4 1 1.5 -1 53
27 4 2 1.5 0 38
28 4 3 1.5 1 16
29 'EOD'

```

```

30 'TABLE' T1 $ OPERAT, DUMP
31 'TABULATE' COUNT; T1
32 'PRINT' T1 $10.2
33 'RUN'

```

	T1		
DUMP	1	2	3
OPERAT			
1	61.00	28.00	7.00
2	68.00	23.00	13.00
3	58.00	40.00	12.00
4	53.00	38.00	16.00

```

34 'CALC'    UV = U*V
35 'TERMS' COUNT+OPERAT+DUMP+UV+OPERAT.V
36 'Y/ERROR=POISSON'    COUNT
37 ''    MODEL OF INDEPENDENCE  ''
38 'FIT/PRIN=A' OPERAT+DUMP ; FVAL = PRED ; RES = RESID
39 'RUN'

```

```

38.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	6	10.88	1.813

```

40 'TABULATE'    PRED; T1
41 'PRINT'    T1  $10.2
42 'RUN'

```

	T1		
DUMP	1	2	3
OPERAT			
1	55.25	29.70	11.05
2	59.86	32.17	11.97
3	63.31	34.03	12.66
4	61.58	33.10	12.32

```

43 'TABULATE'    RESID; T1
44 'PRINT'    T1  $10.2
45 'RUN'

```

	T1		
DUMP	1	2	3

OPERAT				
1	0.77	-0.31	-1.22	
2	1.05	-1.62	0.30	
3	-0.67	1.02	-0.19	
4	-1.09	0.85	1.05	

```
47  ''  UNIFORM ASSOCIATION MODEL  ''
48  'FIT'  OPERAT+DUMP+UV ; FVAL = PRED
49  'RUN'
```

48.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	3.891	0.120	32.30
OPERAT 2	0.168	0.147	1.14
OPERAT 3	0.300	0.158	1.91
OPERAT 4	0.336	0.169	1.98
DUMP 2	-0.627	0.110	-5.73
DUMP 3	-1.654	0.162	-10.22
UV	0.1626	0.0656	2.48

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	5	4.590	0.9180

```
50  'TABULATE'  PRED; T1
51  'PRINT'  T1  $10.2
52  'RUN'
```

	T1		
DUMP	1	2	3
OPERAT			

1	62.51	26.15	7.34
2	62.84	30.94	10.22
3	60.97	35.32	13.72
4	53.69	36.59	16.72

```

54  '' ROW EFFECTS MODEL ''
55  'FIT'  OPERAT+DUMP+OPERAT.V ; FVAL = PRED
56  'RUN'

```

```

55.....
.....

```

```

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT
    1 PARAMETER(S) OF THIS TERM ARE ALIASED.
***** REGRESSION ANALYSIS *****

```

```

ERROR DISTRIBUTION: POISSON   LINK FUNCTION: LOG
Y-VARIATE: COUNT

```

```

*** REGRESSION COEFFICIENTS ***

```

	ESTIMATE	S.E.	T
CONSTANT	3.665	0.192	19.07
OPERAT 2	0.124	0.185	0.67
OPERAT 3	0.293	0.176	1.67
OPERAT 4	0.318	0.173	1.84
DUMP 2	-0.384	0.161	-2.38
DUMP 3	-1.168	0.269	-4.35
V.OPERAT 1	-0.457	0.208	-2.19
V.OPERAT 2	-0.377	0.200	-1.88
V.OPERAT 3	-0.139	0.189	-0.74
V.OPERAT 4	0	*	*

```

* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

```

```

*** RESIDUAL DEVIANCE ***
SCALE PARAMETER IS 1.000

```

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	4.403	1.468

```

57  'TABULATE'  PRED; T1
58  'PRINT'  T1  $10.2
59  'RUN'

```

221.

	T1		
DUMP	1	2	3
OPERAT			
1	61.69	26.61	7.69
2	64.44	30.13	9.44
3	60.17	35.67	14.17
4	53.70	36.59	16.70

60 'CLOSE'

\*\*\*\*\* END OF DUMPING. MAXIMUM OF 1446 DATA UNITS USED AT LINE 55  
(31322 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFERENCE' DUMPING
-3      ''
-4      SEE SECTION 4.4.3.3
-5      FITTING VARIOUS LOGLINEAR MODELS
-6      TO DUMPING SEVERITY DATA OF TABLE 4.8
-7      FROM AGRESTI P67
8
9  'UNITS' $ 48
10 'FACTOR' OPERAT $4 = 12(1...4)
11      : HOSPITAL $4 = 3(1...4)4
12      : DUMP $3 = (1...3)16
13 'VARIATE' COUNT
14 'READ/PRIN=D' COUNT
15 'RUN'

```

```

16 23 7 2 18 6 1 8 6 3 12 9 1
17 23 10 5 18 6 2 12 4 4 15 3 2
18 20 13 5 13 13 2 11 6 2 14 8 3
19 24 10 6 9 15 2 7 7 4 13 6 4
20 'EOD'

```

```

21 'TERMS' COUNT+OPERAT*HOSPITAL*DUMP
22 'Y/ERROR=POISSON' COUNT
23 'CAPTION' '' (O,D,H) ''
24 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP
25 'RUN'

```

(O,D,H)

24.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
 Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
 SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	39	32.61	0.8362



```

26 'CAPTION'  ''          (OD,H)  ''
27 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+OPERAT.DUMP
28 'RUN'

```

(OD,H)

27.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	33	21.73	0.6586

```

29 'CAPTION'  ''          (HD,0)  ''
30 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+HOSPITAL.DUMP
31 'RUN'

```

(HD,0)

30.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	33	24.51	0.7427

```

32 'CAPTION'  ''          (OH,D)  ''
33 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+OPERAT.HOSPITAL
34 'RUN'

```

(OH,D)

33.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	30	31.64	1.055

35 'CAPTION' '' (HD,OD) ''  
36 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+HOSPITAL.DUMP+OPERAT.DUMP  
37 'RUN'  
  
(HD,OD)

36.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	27	13.63	0.5049

38 'CAPTION' '' (OD,OH) ''  
39 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+OPERAT.DUMP  
40 +OPERAT.HOSPITAL  
41 'RUN'  
  
(OD,OH)

39.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

DF	DEVIANCE	MEAN DEVIANCE
----	----------	---------------

RESIDUAL	24	20.76	0.8650
----------	----	-------	--------

```

42 'CAPTION'  ''      (HD,OH)  ''
43 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+HOSPITAL.DUMP
44                +OPERAT.HOSPITAL
45 'RUN'

```

(HD,OH)

43.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	24	23.54	0.9807

```

46 'CAPTION'  ''      (HD,OD,OH)  ''
47 'FIT/PRIN=A' OPERAT+HOSPITAL+DUMP+HOSPITAL.DUMP
48                +OPERAT.DUMP+OPERAT.HOSPITAL
49 'RUN'

```

(HD,OD,OH)

47.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	18	12.50	0.6946

50 'CLOSE'

\*\*\*\*\* END OF DUMPING.    MAXIMUM OF 4348 DATA UNITS USED AT LINE 47  
(28420 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFERENCE' DUMPING
-3      ''
-4      SEE SECTION 4.4.3.3
-5      FITTING THE 0-D UNIFORM ASSOCIATION MODEL
-6      TO DUMPING SEVERITY DATA OF TABLE 4.8
-7      FROM AGRESTI P67
8      ''
9  'UNITS' $ 48
10 'FACTOR' OPERAT $4 = 12(1...4)
11      : HOSPITAL $4 = 3(1...4)4
12      : DUMP $3 = (1...3)16
13 'VARIATE' COUNT
14      : U = 12(-1.5,-0.5,0.5,1.5)
15      : W = (-1,0,1)16
16 'READ/PRIN=Z' COUNT
17 'RUN'

```

```

23 'CALC' UW = U*W
25 'TABLE' T1 $ OPERAT, HOSPITAL, DUMP
26 'TABULATE' COUNT ; T1
27 'PRINT' T1 $ 10
28 'RUN'

```

OPERAT	DUMP	T1		
		1	2	3
1	1	23	7	2
	2	18	6	1
	3	8	6	3
	4	12	9	1
2	1	23	10	5
	2	18	6	2
	3	12	4	4
	4	15	3	2
3	1	20	13	5
	2	13	13	2
	3	11	6	2
	4	14	8	3
4	1	24	10	6
	2	9	15	2
	3	7	7	4
	4	13	6	4

```

30 'TERMS' COUNT+OPERAT+HOSPITAL+DUMP+OPERAT.HOSPITAL+UW
31 'Y/ERROR=POISSON' COUNT
33 'FIT' OPERAT+HOSPITAL+DUMP+OPERAT.HOSPITAL+UW ; FVAL = FIT
34 'RUN'

```

```

33.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	2.793	0.188	14.86
OPERAT 2	0.260	0.243	1.07
OPERAT 3	0.336	0.251	1.34
OPERAT 4	0.450	0.255	1.77
HOSPITAL 2	-0.247	0.267	-0.93
HOSPITAL 3	-0.633	0.300	-2.11
HOSPITAL 4	-0.375	0.277	-1.35
DUMP 2	-0.627	0.109	-5.73
DUMP 3	-1.654	0.162	-10.24
OPERAT 2 .HOSPITAL 2	-0.133	0.369	-0.36
OPERAT 2 .HOSPITAL 3	-0.009	0.407	-0.02
OPERAT 2 .HOSPITAL 4	-0.267	0.391	-0.68
OPERAT 3 .HOSPITAL 2	-0.059	0.365	-0.16
OPERAT 3 .HOSPITAL 3	-0.061	0.411	-0.15
OPERAT 3 .HOSPITAL 4	-0.044	0.378	-0.12
OPERAT 4 .HOSPITAL 2	-0.184	0.367	-0.50
OPERAT 4 .HOSPITAL 3	-0.166	0.413	-0.40
OPERAT 4 .HOSPITAL 4	-0.179	0.381	-0.47
UW	0.1626	0.0655	2.48

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	29	25.35	0.8741

```

35 'TABULATE' FIT ; T1
36 'PRINT' T1 $ 10.2

```

37 'RUN'

		T1		
	DUMP	1	2	3
OPERAT	HOSPITAL			
1	1	20.84	8.72	2.45
	2	16.28	6.81	1.91
	3	11.07	4.63	1.30
	4	14.32	5.99	1.68
2	1	22.96	11.31	3.73
	2	15.71	7.74	2.55
	3	12.09	5.95	1.96
	4	12.09	5.95	1.96
3	1	21.06	12.20	4.74
	2	15.52	8.99	3.49
	3	10.53	6.10	2.37
	4	13.86	8.03	3.12
4	1	20.07	13.68	6.25
	2	13.05	8.89	4.06
	3	9.03	6.16	2.81
	4	11.54	7.87	3.60

38 'CLOSE'

\*\*\*\*\* END OF DUMPING. MAXIMUM OF 2226 DATA UNITS USED AT LINE 33  
(30542 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFERENCE' DUMPING_ITERATIVE
-3      '' SEE SECTION 5.3
-4      FITTING A LOG-MULTIPLICATIVE (RC) MODEL
-5      TO DUMPING SEVERITY DATA FROM TABLE 1.6
6      (AGRESTI, 1984, P81)
7  'UNITS' $ 12
8  'FACTOR' OPERAT $4 = 3(1...4)
9      : DUMP $3 = (1...3)4
10 'VARIATE' COUNT
11      : U = 3(-1.5,-0.5,0.5,1.5)
12      : V = (-1,0,1)4
13      : USTD $ 4
14      : VSTD $ 3
15 'INTEGER' I7_10 = 3(7...10)
16      : I7_9 = (7...9)4
17      : I14710 = 1,4,7,10
18      : I123 = 1,2,3
19 'SCALAR' USSUM
20      : VSSUM
21 'READ/PRIN=D' COUNT
22 'RUN'

23      61  28  7
24      68  23  13
25      58  40  12
26      53  38  16
27 'EOD'

29 'TABLE' T1 $ OPERAT, DUMP
31 'TERMS' COUNT+OPERAT+DUMP+OPERAT.V+DUMP.U
32 'Y/ERROR=POISSON' COUNT
34 'FOR' I = 1...8
35      'FIT/PRIN=Z' OPERAT+DUMP+OPERAT.V ; COEF = CF1 ; DEV = DV
36      'COPY' U $ = CF1 $ I7_10
37      'CALC' U = U - MEAN(U)
38      'COPY' USTD $ = U $ I14710
39      'CALC' USSUM = SQRT(SUM(USTD*USTD))
40      : U = U/USSUM
41      'PRINT' DV $10.6
44      'FIT/PRIN=Z' OPERAT+DUMP+DUMP.U ; COEF = CF2 ; DEV = DV ; F
VAL = PRED
45      'COPY' V $ = CF2 $ I7_9
46      'CALC' V = V - MEAN(V)
47      'COPY' VSTD $ = V $ I123
48      'CALC' VSSUM = SQRT(SUM(VSTD*VSTD))
49      : V = V/VSSUM

```

50 'PRINT' DV \$10.6  
52 'REPEAT'  
53 'RUN'

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 4.403389

44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 3.498055

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 3.089726

44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.938355

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.885235

44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.866411

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.859711



44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.857305

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.856437

44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.856123

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.856010

44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.855968

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.855953

44.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
DV 2.855948

35.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.

DV 2.855946

44.....  
 .....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
 1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
 DV 2.855945

```

54      'FIT/PRIN=CA' OPERAT+DUMP+OPERAT.V ; COEF = CF1 ; DEV = DV
55      'COPY'      U $      = CF1 $ I7_10
56      'CALC'      U = U - MEAN(U)
57      'COPY'      USTD $ = U $ I14710
58      'CALC'      USSUM = SQRT(SUM(USTD*USTD))
59      :           U = U/USSUM
60      'PRINT'     DV $10.6
63      'FIT/PRIN=CA' OPERAT+DUMP+DUMP.U ; COEF = CF2 ; DEV = DV ;
FVAL = PRED
64      'COPY'      V $      = CF2 $ I7_9
65      'CALC'      V = V - MEAN(V)
66      'COPY'      VSTD $ = V $ I123
67      'CALC'      VSSUM = SQRT(SUM(VSTD*VSTD))
68      :           V = V/VSSUM
69      'PRINT'     DV $10.6
71      'PRINT/P'   U,V $10.6
72      'CAPTION'   ''      ACTUAL COUNTS''
73      'TABULATE'  COUNT ; T1
74      'PRINT'     T1 $ 9
75      'CAPTION'   ''      FITTED COUNTS''
76      'TABULATE'  PRED ; T1
77      'PRINT'     T1 $ 10.3
78      'RUN'

```

54.....  
 .....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM V.OPERAT  
 1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
 \*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
 Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	3.770	0.167	22.60
OPERAT 2	0.042	0.163	0.26
OPERAT 3	0.225	0.153	1.47
OPERAT 4	0.210	0.153	1.38
DUMP 2	-0.299	0.205	-1.46

DUMP 3	-1.353	0.210	-6.46
V.OPERAT 1	-0.414	0.224	-1.85
V.OPERAT 2	-0.524	0.222	-2.36
V.OPERAT 3	-0.074	0.212	-0.35
V.OPERAT 4	0	*	*

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	2.856	0.9520

DV 2.855945

63.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM U.DUMP  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	3.932	0.148	26.59
OPERAT 2	0.015	0.158	0.09
OPERAT 3	0.307	0.253	1.22
OPERAT 4	0.311	0.292	1.07
DUMP 2	-0.640	0.111	-5.77
DUMP 3	-1.617	0.160	-10.13
U.DUMP 1	-0.460	0.319	-1.44
U.DUMP 2	0.136	0.343	0.40
U.DUMP 3	0	*	*

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	4	2.856	0.7140

DV 2.855945

U	V
-0.363673	-0.796780
-0.363673	0.552843
-0.363673	0.243937
-0.613369	-0.796780
-0.613369	0.552843
-0.613369	0.243937
0.404214	-0.796780
0.404214	0.552843
0.404214	0.243937
0.572828	-0.796780
0.572828	0.552843
0.572828	0.243937

ACTUAL COUNTS

	T1		
DUMP	1	2	3
OPERAT			
1	61	28	7
2	68	23	13
3	58	40	12
4	53	38	16

## FITTED COUNTS

	T1		
DUMP	1	2	3
OPERAT			
1	60.285	25.587	10.128
2	68.622	25.099	10.279
3	57.595	38.634	13.771
4	53.498	39.680	13.822

79 'CLOSE'

\*\*\*\*\* END OF DUMPING\_. MAXIMUM OF 2210 DATA UNITS USED AT LINE 63  
(30558 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

3  'REFERENCE' DUMPING_ITERATIVE
-4      '' SEE SECTION 5.3
-5      ESTIMATING THE BETA PARAMETER FROM
-6      A LOG-MULTIPLICATIVE (RC) MODEL
-7      DUMPING SEVERITY DATA FROM TABLE 1.6
8      (AGRESTI, 1984, P81) ''
9  'UNITS' $ 12
10 'FACTOR' OPERAT $4 = 3(1...4)
11      : DUMP $3 = (1...3)4
12 'VARIATE' COUNT
13      : U = 3(-.3636, -.6134, .4042, .5728)
14      : V = (-.7968, .5529, .2439)4
15 'READ/PRIN=Z' COUNT
16 'RUN'
```

```

23 'TABLE' T1 $ OPERAT, DUMP *
24 'CAPTION' '' ACTUAL COUNTS''
25 'TABULATE' COUNT ; T1
26 'PRINT' T1 $ 9
27 'RUN'
```

#### ACTUAL COUNTS

	T1		
DUMP	1	2	3
OPERAT			
1	61	28	7
2	68	23	13
3	58	40	12
4	53	38	16

```

28 'CALC' UV = U*V
29 'TERMS' COUNT+OPERAT+DUMP+UV
30 'Y/ERROR=POISSON' COUNT
31 'FIT/PRIN=CA' OPERAT+DUMP+UV ; FVAL = PRED
32 'RUN'
```

31.....  
 .....

## \*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

## \*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	3.971	0.110	36.18
OPERAT 2	0.042	0.143	0.29
OPERAT 3	0.225	0.144	1.56
OPERAT 4	0.210	0.145	1.45
DUMP 2	-0.640	0.111	-5.79
DUMP 3	-1.617	0.159	-10.20
UV	0.442	0.157	2.81

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

## \*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	5	2.856	0.5712

```

33 'CAPTION'  ''  FITTED COUNTS'
34 'TABULATE' PRED ; T1
35 'PRINT' T1 $ 10.3
36 'RUN'

```

## FITTED COUNTS

	T1		
DUMP	1	2	3
OPERAT			
1	60.284	25.588	10.128
2	68.623	25.098	10.279
3	57.595	38.634	13.771
4	53.498	39.680	13.822

```

37 'CLOSE'

```

\*\*\*\*\* END OF DUMPING\_. MAXIMUM OF 1286 DATA UNITS USED AT LINE 31  
(31482 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1  'REFERENCE' DUMPING_LOGIT1
-2      ''SEE SECTION 6.2.3.1.1.1
-3      FITTING SOME HETEROGENEOUS LOGIT MODELS
4      TO DUMPING SEVERITY DATA OF TABLE 1.6  ''
5  'UNITS' $ 4
6  'FACTOR' OPERAT $4
7  'VARIATE' COUNT
8      : TOT
9  'READ/PRIN=Z' OPERAT,U,COUNT,TOT
10 'RUN'

```

```

16 'CALC' PROP = COUNT/TOT
-17 '' IN THIS CASE A "SUCCESS" IS DUMPING = SLIGHT OR MODERATE,
18     AND "FAILURE" IS DUMPING = NONE ''
19 'PRINT/P' OPERAT,U,COUNT,TOT,PROP $ 9, 9.1, 9, 9, 9.3
20 'RUN'

```

OPERAT	U	COUNT	TOT	PROP
1	-1.5	35	96	0.365
2	-0.5	36	104	0.346
3	0.5	52	110	0.473
4	1.5	54	107	0.505

```

21 'TERMS/TOTAL=TOT' COUNT+U
22 'Y/ERROR=BINOMIAL' COUNT
-23 '' THIS MODEL ASSUMES OPERATION HAS NO EFFECT
24     ON FIRST DUMPING LOGIT''
25 'FIT/PRIN=A'
26 'RUN'

```

25.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: BINOMIAL LINK FUNCTION: LOGIT  
 Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
 SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	7.920	2.640

```

-27  ''      THIS MODEL ASSUMES OPERATION HAS AN EFFECT
28      ON FIRST DUMPING LOGIT''
29  'FIT/PRIN=CAU'  U
30  'RUN'

```

```

29.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: BINOMIAL LINK FUNCTION: LOGIT  
Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.320	0.100	-3.20
U	0.2291	0.0910	2.52

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	2	1.485	0.7426

\*\*\* OBSERVED AND FITTED VALUES \*\*\*

	OBSERVED	FITTED	RESIDUAL	TOTAL
1	35	32.63	0.51	96
2	36	40.87	-0.98	104
3	52	49.37	0.50	110
4	54	54.13	-0.03	107

```

31  'CLOSE'

```

\*\*\*\*\* END OF DUMPING\_. MAXIMUM OF 1012 DATA UNITS USED AT LINE 21  
(31756 LEFT)



GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFERENCE' DUMPING_LOGIT2
-3      ''SEE SECTION 6.2.3.1.1.1
-4      FITTING SOME HETEROGENEOUS LOGIT MODELS
5      TO DUMPING SEVERITY DATA OF TABLE 1.6 ''
6  'UNITS' $ 4
7  'FACTOR' OPERAT $4
8  'VARIATE' COUNT
9      : TOT
10 'READ/PRIN=Z' OPERAT,U,COUNT,TOT
11 'RUN'

```

```

17 'CALC' PROP = COUNT/TOT
-18 '' IN THIS CASE "SUCCESS" IS DUMPING = MODERATE
19     AND "FAILURE" IS DUMPING = NONE OR SOME ''
20 'PRINT/P' OPERAT,U,COUNT,TOT,PROP $ 9, 9.1, 9, 9, 9.3
21 'TERMS/TOTAL=TOT' COUNT+U
22 'Y/ERROR=BINOMIAL' COUNT
23 ''OPERATION HAS NO EFFECT ON SECOND DUMPING LOGIT''
24 'FIT/PRIN=A'
25 'RUN'

```

OPERAT	U	COUNT	TOT	PROP
1	-1.5	7	96	0.073
2	-0.5	13	104	0.125
3	0.5	12	110	0.109
4	1.5	16	107	0.150

24.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: BINOMIAL LINK FUNCTION: LOGIT  
 Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
 SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	3.187	1.062

```

26  ''OPERATION HAS AN EFFECT ON SECOND DUMPING LOGIT''
27  'FIT/PRIN=CAU'  U
28  'RUN'

```

```

27.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: BINOMIAL LINK FUNCTION: LOGIT  
Y-VARIATE: COUNT

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-2.070	0.158	-13.13
U	0.211	0.142	1.49

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	2	0.9384	0.4692

\*\*\* OBSERVED AND FITTED VALUES \*\*\*

	OBSERVED	FITTED	RESIDUAL	TOTAL
1	7	8.09	-0.40	96
2	13	10.61	0.78	104
3	12	13.52	-0.44	110
4	16	15.78	0.06	107

```

29  'CLOSE'

```

\*\*\*\*\* END OF DUMPING\_. MAXIMUM OF 1032 DATA UNITS USED AT LINE 21  
(31736 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2 'REFERENCE/NUNN=100' DUMPING_LINEAR
4 'MACRO' ORDINAL $
-116 'SEE SECTION 6.2.3.2.1.1
-117 FITTING A UNIFORM ASSOCIATION MODEL
-118 TO DUMPING SEVERITY DATA OF TABLE 1.6
119 USING MACRO OF SECTION 7.1.4 (STIRLING, 1984)''
121 'SCALAR' NL=3 : NROWS=4 : MAXIT=10 : NCELLS=12
122 'UNIT' $ NCELLS
123 'FACTOR' LEV $NL =(1...3)4
124 'FACTOR' OPERAT $4 = 3(1...4)
125 'VARIATE' N
126 : OPERAT_L = 3(-1.5,-0.5,0.5,1.5)
127 'SCALAR' P1=1.0 : P2=0.0 : P3=0.0
128 'READ' N
129 'RUN'
```

IDENTIFIER	MINIMUM	MEAN	MAXIMUM	VALUES	MISSING
N	7.00	34.75	68.00	12	0

```

135 'SET/LIST=M' MODEL2=DUM(1)
136 'SET/LIST=M' MODEL1=OPERAT_L
137 'USE' ORDINAL $
```

```
138 'RUN'
```

FITTING A LINEAR MODEL TO UNDERLYING RESPONSES  
 PROPORTIONS OF LOGISTIC, NORMAL AND EXTREME VALUE DISTNS ARE

P1 1.00000

P2 0.00000

P3 0.00000

INITIAL ESTIMATES FOR MEAN & CLASS BOUNDARIES

MN -0.3045

BOUNDS  
 -14.1200  
 0.0000  
 1.7351  
 13.4407

DEV 10.8782

137.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.3149	0.0992	-3.18
DUM(1)	1.735	0.143	12.09
OPERAT_L	0.2236	0.0870	2.57

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 4.2952 DEG 5

137.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.320	0.100	-3.20
DUM(1)	1.754	0.143	12.24
OPERAT_L	0.2247	0.0881	2.55

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 4.2701 DEG 5

137.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.320	0.100	-3.20
DUM(1)	1.754	0.145	12.11
OPERAT_L	0.2247	0.0882	2.55

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 4.2701 DEG 5  
ACTUAL AND FITTED CELL COUNTS AND RESIDUALS

LEV ROW	T3		
	1	2	3
1	61	28	7
2	68	23	13
3	58	40	12
4	53	38	16

LEV ROW	T3		
	1	2	3
1	63.230961	24.867994	7.901047
2	63.075539	30.426111	10.498355
3	60.696751	35.746326	13.556925
4	53.050480	37.936993	16.012526

LEV ROW	T3		
	1	2	3
1	-0.2822	0.6155	-0.3270
2	0.6122	-1.4075	0.7441
3	-0.3488	0.6980	-0.4314
4	-0.0073	0.0103	-0.0028

139 'CLOSE'

\*\*\*\*\* END OF DUMPING\_. MAXIMUM OF 4096 DATA UNITS USED AT LINE 13  
7 (28672 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2 'REFERENCE/NUNN=100' DUMPING
4 'MACRO' ORDINAL $
-116 'SEE SECTION 6.2.3.2.2.1
-117 FITTING A LOGIT ROW EFFECTS MODEL
-118 TO DUMPING SEVERITY DATA OF TABLE 1.6
119 USING MACRO OF SECTION 7.1.4 (STIRLING, 1984)''
120 'SCALAR' NL=3 : NROWS=4 : MAXIT=10 : NCELLS=12
121 'UNIT' $ NCELLS
122 'FACTOR' LEV $NL =(1...3)4
123 'FACTOR' OPERAT $4 = 3(1...4)
124 'VARIATE' N
125 : OPERAT_L = 3(-1.5,-0.5,0.5,1.5)
126 'SCALAR' P1=1.0 : P2=0.0 : P3=0.0
127 'READ' N
128 'RUN'

```

IDENTIFIER	MINIMUM	MEAN	MAXIMUM	VALUES	MISSING
N	7.00	34.75	68.00	12	0

```

134 'SET/LIST=M' MODEL2=DUM(1)
135 'SET/LIST=M' MODEL1=OPERAT
136 'USE' ORDINAL $

```

```

137 'RUN'

```

FITTING A LINEAR MODEL TO UNDERLYING RESPONSES  
 PROPORTIONS OF LOGISTIC, NORMAL AND EXTREME VALUE DISTNS ARE

P1 1.00000

P2 0.00000

P3 0.00000

INITIAL ESTIMATES FOR MEAN & CLASS BOUNDARIES

MN -0.3045

BOUNDS  
 -14.1200  
 0.0000  
 1.7351  
 13.4407

DEV 10.8782

136.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.578	0.202	-2.86
DUM(1)	1.735	0.143	12.09
OPERAT 2	0.022	0.278	0.08
OPERAT 3	0.428	0.274	1.56
OPERAT 4	0.603	0.276	2.19

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 3.6013 DEG 3

136.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.581	0.210	-2.77
DUM(1)	1.756	0.143	12.25
OPERAT 2	0.009	0.289	0.03
OPERAT 3	0.429	0.278	1.54
OPERAT 4	0.600	0.278	2.15

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 3.5637 DEG 3

136.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

ESTIMATE	S.E.	T
----------	------	---

CONSTANT	-0.581	0.210	-2.77
DUM(1)	1.756	0.145	12.10
OPERAT 2	0.008	0.290	0.03
OPERAT 3	0.429	0.279	1.54
OPERAT 4	0.600	0.279	2.15

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 3.5637 DEG 3  
ACTUAL AND FITTED CELL COUNTS AND RESIDUALS

	T3		
LEV	1	2	3
ROW			
1	61	28	7
2	68	23	13
3	58	40	12
4	53	38	16

	T3		
LEV	1	2	3
ROW			
1	61.571697	25.974493	8.453816
2	66.506371	28.266640	9.226990
3	59.172386	36.617966	14.209652
4	53.004574	37.987713	16.007710

	T3		
LEV	1	2	3
ROW			
1	-0.0729	0.3924	-0.5155
2	0.1825	-1.0240	1.1693
3	-0.1529	0.5506	-0.6025
4	0.0000	0.0000	-0.0020

138 'CLOSE'

\*\*\*\*\* END OF DUMPING. MAXIMUM OF 4132 DATA UNITS USED AT LINE 136  
(28636 LEFT)



GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1 'REFERENCE/NUNN=100' DUMPING
2 'MACRO' ORDINAL $
-115 'SEE SECTION 6.2.4.1.1
-116 FITTING A THREE-DIMENSIONAL LOGIT MODEL
-117 TO DUMPING SEVERITY DATA OF TABLE 4.8
118 USING MACRO OF SECTION 7.1.4 (STIRLING, 1984)''
120 'SCALAR' NL=3 : NROWS=16 : MAXIT=10 : NCELLS=48
121 'UNIT' $ NCELLS
122 'FACTOR' LEV $NL =(1...3)16
123 'FACTOR' OPERAT $4 = 12(1...4)
124 : HOSPITAL $4 = 3(1...4)4
125 'VARIATE' N
126 : OP_LIN = 12(-1.5,-0.5,0.5,1.5)
127 'SCALAR' P1=1.0 : P2=0.0 : P3=0.0
128 'READ' N
129 'RUN'

```

IDENTIFIER	MINIMUM	MEAN	MAXIMUM	VALUES	MISSING
N	1.000	8.688	24.000	48	0

```

135 'SET/LIST=M' MODEL2=DUM(1)
136 'SET/LIST=M' MODEL1=OP_LIN+HOSPITAL
137 'USE' ORDINAL $
138 'RUN'

```

FITTING A LINEAR MODEL TO UNDERLYING RESPONSES  
 PROPORTIONS OF LOGISTIC, NORMAL AND EXTREME VALUE DISTNS ARE

P1 1.00000

P2 0.00000

P3 0.00000

INITIAL ESTIMATES FOR MEAN & CLASS BOUNDARIES

MN -0.3045

BOUNDS  
 -14.1200  
 0.0000  
 1.7351  
 13.4407

DEV 31.6381

137.....  
 .....

## \*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

## \*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.423	0.163	-2.59
DUM(1)	1.735	0.143	12.09
OP_LIN	0.2262	0.0870	2.60
HOSPITAL 2	0.113	0.251	0.45
HOSPITAL 3	0.426	0.280	1.53
HOSPITAL 4	0.019	0.262	0.07

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 22.5287 DEG 26

137.....  
 .....

## \*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

## \*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.429	0.166	-2.58
DUM(1)	1.762	0.143	12.29
OP_LIN	0.2264	0.0883	2.56
HOSPITAL 2	0.102	0.253	0.40
HOSPITAL 3	0.429	0.277	1.55
HOSPITAL 4	0.027	0.267	0.10

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 22.4782 DEG 26

137.....  
 .....

## \*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
 WEIGHT VARIATE: W

## \*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	-0.429	0.166	-2.58
DUM(1)	1.762	0.145	12.12
OP_LIN	0.2267	0.0884	2.56
HOSPITAL 2	0.104	0.254	0.41
HOSPITAL 3	0.429	0.277	1.55
HOSPITAL 4	0.027	0.267	0.10

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 22.4781      DEG 26  
 ACTUAL AND FITTED CELL COUNTS AND RESIDUALS

	T3		
LEV ROW	1	2	3
1	23	7	2
2	18	6	1
3	8	6	3
4	12	9	1
5	23	10	5
6	18	6	2
7	12	4	4
8	15	3	2
9	20	13	5
10	13	13	2
11	11	6	2
12	14	8	3
13	24	10	6
14	9	15	2
15	7	7	4
16	13	6	4

	T3		
LEV ROW	1	2	3
1	21.864841	7.775896	2.359262
2	16.509319	6.462050	2.028632
3	9.931291	5.217273	1.851436
4	14.903255	5.433831	1.662914
5	24.028255	10.522032	3.449714
6	15.804072	7.602905	2.593021
7	10.566076	6.775257	2.658669
8	12.520808	5.618610	1.860582
9	21.972786	11.797505	4.229710
10	15.475719	9.108009	3.416274

11	8.962210	6.973101	3.064690
12	14.291278	7.858732	2.849991
13	20.887819	13.680888	5.431294
14	12.902023	9.238433	3.859544
15	7.484497	7.016993	3.498509
16	11.855894	7.947727	3.196381

	T3		
LEV	1	2	3
ROW			
1	0.2407	-0.2831	-0.2402
2	0.3616	-0.1840	-0.8016
3	-0.6345	0.3346	0.7738
4	-0.7787	1.3965	-0.5556
5	-0.2113	-0.1623	0.7816
6	0.5403	-0.6038	-0.3838
7	0.4317	-1.1553	0.7649
8	0.6792	-1.2134	0.1010
9	-0.4274	0.3444	0.3640
10	-0.6473	1.2111	-0.8312
11	0.6571	-0.3776	-0.6498
12	-0.0773	0.0502	0.0881
13	0.6650	-1.0457	0.2399
14	-1.1494	1.7370	-1.0438
15	-0.1791	-0.0065	0.2621
16	0.3271	-0.7225	0.4324

139 'CLOSE'

\*\*\*\*\* END OF DUMPING. MAXIMUM OF 5218 DATA UNITS USED AT LINE 137  
(27550 LEFT)

```

'REFERENCE/NUNN=100' DUMPING
  '' SEE SECTION 7.1.5
    FITTING A MODEL BASED ON THE
    NEGATIVE EXTREME VALUE DISTRIBUTION
    TO DUMPING SEVERITY DATA OF TABLE 1.6  ''
'MACRO' ORDINAL $
'' STIRLING, 1984''
'' Macro to fit linear models to ordinal responses
  Input parameters :
    MODEL1 - model formula with explanatory variables
    MODEL2 - DUM(1)+DUM(2)+ ..... +DUM(no of response levels-2)
    LEV     - factor with response levels
    N       - variate with numbers in each cell of table. They
              must be ordered in rows (i.e. with all response
              levels for the first combination of
              explanatory variables, then those for the
              second, etc.
    NROWS   - scalar with no of rows in the table
    MAXIT   - scalar with maximum no of iterations
              (10 should be enough)
    P1      - Scalars giving the proportions of logistic, normal
    P2      - and extreme value distributions (Make sure they
    P3      - add up to 1.0)
  Output :
    BETA     - parameter estimates
    Z        - variate with residuals
    FIT      - variate with fitted values''

'CAPTION' ''FITTING A LINEAR MODEL TO UNDERLYING RESPONSES
          PROPORTIONS OF LOGISTIC, NORMAL AND EXTREME VALUE DISTNS ARE'
,
'PRIN' P1,P2,P3 $ 8.5
'SET' JLIST= DEV,NTOT,NLEVP1,NLEVM2,MN,NLEVS,OLDDEV,LPL,LPU,PDFL,
          PDFU,W,ROWTOT,T1,T2,T3,BOUNDS,CONV,DEG,ROW
'LOCAL' JLIST
'SCALAR' DEV,NTOT,NLEVP1,NLEVM2,MN,NLEVS,OLDDEV,DEG
'START'
'CALC' NLEVS=NLEV(LEV) : NLEVM2=NLEVS-2 : NLEVP1=NLEVS+1
'RUN'
'FACTOR' ROW $NROWS = NLEVS!(1...NROWS)
'SET' ILIST=1...NLEVM2
'VARIATE' LPL,LPU,FIT,PDFL,PDFU,W,Z,DUM(ILIST),ROWTOT
'TABLE/N' T1 $LEV : T2 $ROW
'TABULATE' N ; T1 : N ; T2
'VARIATE' BOUNDS $NLEVP1
'EQUATE' BOUNDS = 0,T1
'EQUATE' Z $NROWS = T2
'CALC' ROWTOT=ELEM(Z ; ROW)
      : NTOT=SUM(BOUNDS)
      : FIT=ELEM(BOUNDS ; LEV+1)/NTOT
      : BOUNDS=CUM(BOUNDS)
      : BOUNDS= BOUNDS/NTOT+
        ((BOUNDS.EQ.0)-(BOUNDS.EQ.NTOT))*0.000001

```

```

:   BOUNDS= P1*LOG(1.0/BOUNDS-1.0)-P2*NED(BOUNDS)-
:   P3*LOG(-LOG(1-BOUNDS))
:   LPL=ELEM(BOUNDS ; LEV)
:   LPU=ELEM(BOUNDS ; LEV+1)
:   DEV=2*SUM(N*LOG(N/ROWTOT/FIT + (N.LT.0.5)))
:   MN= ELEM(BOUNDS ; 2)
:   BOUNDS= MN-BOUNDS
'CAPTION' '' INITIAL ESTIMATES FOR MEAN & CLASS BOUNDARIES''
'PRIN' MN,BOUNDS,DEV    $ 10.4

'FOR' I=1...MAXIT
  'CALC' OLDDEV=DEV
  :   Z=EXP(-LPL)
  :   PDFL=P1*Z/(1+Z)**2 + P2*0.39894228*EXP(-0.5*LPL*LPL)
  :   PDFL=(LEV.GT.1.5)*(PDFL + P3*Z*EXP(-Z*(Z.LT.100))*
  :   (Z.LT.100))
  :   Z=EXP(-LPU)
  :   PDFU=P1*Z/(1+Z)**2 + P2*0.39894228*EXP(-0.5*LPU*LPU)
  :   PDFU=(LEV.LT.NLEVS-0.5)*(PDFU+P3*Z*EXP(-Z*(Z.LT.100))*
  :   (Z.LT.100))
  :   W=ROWTOT*(PDFL-PDFU)**2/FIT
  :   Z=(PDFL*LPL-PDFU*LPU+N/ROWTOT-FIT)/(PDFL-PDFU)
  'FOR' J=ILIST ; DUMJ=DUM(ILIST)
    'CALC' DUMJ=((LEV.EQ.J+1)*PDFU-(LEV.EQ.J+2)*PDFL)
    :   /(PDFL-PDFU)
  'REPEAT'
  'TERMS/WT=W' MODEL2+MODEL1+Z
  'Y/SCALE=1.0' Z
  'FIT/PRIN=C' MODEL2+MODEL1 ; FVAL=FIT ; COEFF=BETA ; DF=DEG
  'CALC' BOUNDS = 0.0
  'EQUATE' BOUNDS $ X,X,NLEVM2 = BETA $ X,NLEVM2
  'CALC' LPU=ELEM(BOUNDS ; LEV+1)-ELEM(BOUNDS ; LEV)
  :   LPL=FIT-PDFU*LPU/(PDFL-PDFU)
  :   LPU=FIT-PDFL*LPU/(PDFL-PDFU)
  :   PDFL=P1/(1.0+EXP(-LPL)) + P2*NPI(LPL) +
  :   P3*EXP(-EXP(-LPL))
  :   PDFU=P1/(1.0+EXP(-LPU)) + P2*NPI(LPU) +
  :   P3*EXP(-EXP(-LPU))
  :   FIT=(LEV.LT.1.5)+(LEV.GT.1.5)*PDFL-
  :   (LEV.LT.NLEVS-0.5)*PDFU
  'CALC' DEV=2*SUM(N*LOG(N/ROWTOT/FIT+(N.LT.0.5)))
  :   DEG=DEG-NROWS
  'PRIN/P' DEV,DEG $ 10.4,8
  'JUMP' CONV*(OLDDEV-DEV.LT.1.0E-4)
  'REPEAT'

'CAPTION' '' TOO MANY ITERATIONS !!!!! IF THE PARAMETER
ESTIMATE FOR A FACTOR LEVEL IS TENDING TO INFINITY, CHECK
WHETHER ALL CELLS FOR THAT LEVEL ARE ZERO APART FROM THE
HIGHEST OR LOWEST RESPONSE LEVEL; IF SO OMIT ALL OBSERVATIONS
AT THAT FACTOR LEVEL AND REFIT''
'LABEL' CONV
'CAPTION' '' ACTUAL AND FITTED CELL COUNTS AND RESIDUALS''
'CALC' FIT=FIT*ROWTOT

```

```

      :      Z=SQRT(2*(N*LOG(N/FIT+(N.LT.0.5))+FIT-N))*(2*(N.GT.FIT)-1)
'TABLE/N' T3 $ROW,LEV
'TABULATE' N ; T3
'PRIN' T3 $ 10
'TABULATE' FIT ; T3
'PRIN' T3 $14.6
'TABULATE' Z ; T3
'PRIN' T3 $ 10.4
'DEVALUE' JLIST
'ENDMACRO'

'SCALAR' NL=3      :   NROWS=4      :   MAXIT=10      :   NCELLS=12
'UNIT' $ NCELLS
'FACTOR' LEV $NL =(1...3)4
'FACTOR' OPERAT $4 = 3(1...4)
'VARIATE' N
      :      OPERAT_L = 3(-1.5,-0.5,0.5,1.5)
'SCALAR' P1=0.0    :   P2=0.0      :   P3=1.0
'READ' N
'' NOTE TABLE IN REVERSE ORDER TO
FIT NEGATIVE EXTREME VALUE DISTRIBUTION ''
'RUN'
  7 28 61
 13 23 68
 12 40 58
 16 38 53
'EOD'
'SET/LIST=M' MODEL2=DUM(1)
'SET/LIST=M' MODEL1=OPERAT
'USE' ORDINAL $
'RUN'
'CLOSE'
'STOP'

```

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1 'REFERENCE/NUNN=100' DUMPING
2 'MACRO' ORDINAL $
-114 '' SEE SECTION 7.1.5
-115 FITTING A MODEL BASED ON THE
-116 NEGATIVE EXTREME VALUE DISTRIBUTION
117 TO DUMPING SEVERITY DATA OF TABLE 1.6 ''
119 'SCALAR' NL=3 : NROWS=4 : MAXIT=10 : NCELLS=12
120 'UNIT' $ NCELLS
121 'FACTOR' LEV $NL =(1...3)4
122 'FACTOR' OPERAT $4 = 3(1...4)
123 'VARIATE' N
124 : OPERAT_L = 3(-1.5,-0.5,0.5,1.5)
125 'SCALAR' P1=0.0 : P2=0.0 : P3=1.0
126 'READ' N
-127 '' NOTE TABLE IN REVERSE ORDER TO
128 FIT NEGATIVE EXTREME VALUE DISTRIBUTION ''
129 'RUN'

```

IDENTIFIER	MINIMUM	MEAN	MAXIMUM	VALUES	MISSING
N	7.00	34.75	68.00	12	0

```

135 'SET/LIST=M' MODEL2=DUM(1)
136 'SET/LIST=M' MODEL1=OPERAT
137 'USE' ORDINAL $
138 'RUN'

```

FITTING A LINEAR MODEL TO UNDERLYING RESPONSES  
 PROPORTIONS OF LOGISTIC, NORMAL AND EXTREME VALUE DISTNS ARE

P1 0.00000

P2 0.00000

P3 1.00000

INITIAL ESTIMATES FOR MEAN & CLASS BOUNDARIES

MN 2.1014

BOUNDS  
 -11.7010  
 0.0000  
 1.5080  
 4.7262



DEV 10.8782

137.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	2.303	0.200	11.52
DUM(1)	1.508	0.132	11.43
OPERAT 2	0.022	0.218	0.10
OPERAT 3	-0.340	0.216	-1.58
OPERAT 4	-0.459	0.217	-2.11

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 3.2027 DEG 3

137.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
WEIGHT VARIATE: W

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	2.321	0.210	11.03
DUM(1)	1.517	0.131	11.57
OPERAT 2	0.030	0.237	0.13
OPERAT 3	-0.344	0.220	-1.56
OPERAT 4	-0.455	0.218	-2.09

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 3.1729 DEG 3

137.....  
.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

Y-VARIATE: Z  
WEIGHT VARIATE: W

## \*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
CONSTANT	2.321	0.212	10.97
DUM(1)	1.518	0.132	11.46
OPERAT 2	0.031	0.238	0.13
OPERAT 3	-0.344	0.221	-1.55
OPERAT 4	-0.455	0.219	-2.08

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

DEV 3.1729 DEG 3  
 ACTUAL AND FITTED CELL COUNTS AND RESIDUALS

	T3		
LEV	1	2	3
ROW			
1	7	28	61
2	13	23	68
3	12	40	58
4	16	38	53

	T3		
LEV	1	2	3
ROW			
1	8.973919	25.668829	61.357254
2	9.443013	27.185537	67.371445
3	14.214483	37.266232	58.519287
4	15.337992	38.829819	52.832195

	T3		
LEV	1	2	3
ROW			
1	-0.6856	0.4534	-0.0456
2	1.0943	-0.8248	0.0764
3	-0.6037	0.4425	-0.0680
4	0.1678	-0.1336	0.0231

139 'CLOSE'

\*\*\*\*\* END OF DUMPING. MAXIMUM OF 4138 DATA UNITS USED AT LINE 137  
 (28630 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1  'REFER' DUMP2WAY
-2      '' SEE SECTION 8.1.1
-3      FITTING A MEAN RESPONSE MODEL TO
-4      OPERATION BY DUMPING DATA OF TABLE 1.6
5      USING WEDDERBURN'S METHOD ''
6  'UNITS' $ 12
7  'VARIATE' COUNT
8      : ASTAR
9      : A
10 'FACTOR' OPERATION $ 4 = 3(1...4)
11      : DUMP $ 3 = (1,2,3)4
12 'SCALAR' DEV
13      : SL1ETA
14      : SL2ETA
15      : SL3ETA
16      : SL4ETA
17      : SL5ETA
18      : SL6ETA
19 'READ/PRIN=Z' COUNT
20 'RUN'

26 'TABLE/M' T1 $ OPERATION, DUMP
27 'READ/PRIN=Z' L1STAR,L2STAR,L3STAR,L4STAR,L5STAR,L6STAR
28 'RUN'

42 'READ/PRIN=Z' T
43 'RUN'

46 'READ/PRIN=Z' AHASH
47 'RUN'

50 'CALC' ASTAR = AHASH/3
51 'PRINT/P' L1STAR, L2STAR,L3STAR,L4STAR,L5STAR,
52      L6STAR,T,AHASH,ASTAR $ 8(7.0),6.4
53 'RUN'

```

L1STAR	L2STAR	L3STAR	L4STAR	L5STAR	L6STAR	T	AHASH	ASTAR
-1	0	1	0	0	0	96	1	0.3333
-2	0	1	0	0	0	96	1	0.3333
-3	0	1	0	0	0	96	1	0.3333
2	1	0	1	0	0	104	1	0.3333
4	2	0	1	0	0	104	1	0.3333
6	3	0	1	0	0	104	1	0.3333
-1	-2	0	0	1	0	110	1	0.3333
-2	-4	0	0	1	0	110	1	0.3333
-3	-6	0	0	1	0	110	1	0.3333
0	1	0	0	0	1	107	1	0.3333
0	2	0	0	0	1	107	1	0.3333
0	3	0	0	0	1	107	1	0.3333

```

54 'CALC'    L1 = L1STAR/T
55 'CALC'    L2 = L2STAR/T
56 'CALC'    L3 = L3STAR/T
57 'CALC'    L4 = L4STAR/T
58 'CALC'    L5 = L5STAR/T
59 'CALC'    L6 = L6STAR/T
60 'CALC'    A  = ASTAR*T
62 'PRINT/P'  T,A,L1, L2,L3,L4,L5,L6  $ 5, 5, 2(8.4), 4(7.4)
63 'RUN'

```

T	A	L1	L2	L3	L4	L5	L6
96	32	-0.0104	0.0000	0.0104	0.0000	0.0000	0.0000
96	32	-0.0208	0.0000	0.0104	0.0000	0.0000	0.0000
96	32	-0.0313	0.0000	0.0104	0.0000	0.0000	0.0000
104	35	0.0192	0.0096	0.0000	0.0096	0.0000	0.0000
104	35	0.0385	0.0192	0.0000	0.0096	0.0000	0.0000
104	35	0.0577	0.0288	0.0000	0.0096	0.0000	0.0000
110	37	-0.0091	-0.0182	0.0000	0.0000	0.0091	0.0000
110	37	-0.0182	-0.0364	0.0000	0.0000	0.0091	0.0000
110	37	-0.0273	-0.0545	0.0000	0.0000	0.0091	0.0000
107	36	0.0000	0.0093	0.0000	0.0000	0.0000	0.0093
107	36	0.0000	0.0187	0.0000	0.0000	0.0000	0.0093
107	36	0.0000	0.0280	0.0000	0.0000	0.0000	0.0093

```

64 'CALC'    ETA = COUNT
65 'CALC'    Z  = COUNT - A
66 'PRINT/P'  COUNT,ETA,Z  $ 7, 7.0, 10.4
67 'RUN'

```

COUNT	ETA	Z
61	61	29.0000

28	28	-4.0000
7	7	-25.0000
68	68	33.3333
23	23	-11.6667
13	13	-21.6667
58	58	21.3333
40	40	3.3333
12	12	-24.6667
53	53	17.3333
38	38	2.3333
16	16	-19.6667

```
68 'TABULATE/PRIN=T'   ETA ; T1
69 'RUN'
```

	T1			
DUMP	1	2	3	MARGIN
OPERATIO				
1	61.00	28.00	7.00	96.00
2	68.00	23.00	13.00	104.00
3	58.00	40.00	12.00	110.00
4	53.00	38.00	16.00	107.00
MARGIN	240.00	129.00	48.00	417.00

```
70 'FOR' I = 1...3
71   'CALC' W = 1/ETA
72   'CALC' P1 = ETA*L1
73   'CALC' P2 = ETA*L2
74   'CALC' P3 = ETA*L3
75   'CALC' P4 = ETA*L4
76   'CALC' P5 = ETA*L5
77   'CALC' P6 = ETA*L6
78   'CALC' SL1ETA = SUM(P1)
79   'CALC' SL2ETA = SUM(P2)
80   'CALC' SL3ETA = SUM(P3)
81   'CALC' SL4ETA = SUM(P4)
82   'CALC' SL5ETA = SUM(P5)
83   'CALC' SL6ETA = SUM(P6)
84   'TERMS/WT=W' Z+P1+P2+P3+P4+P5+P6
85   'Y' Z
86   'FIT/INT=N,PRIN=Z' P1+P2+P3+P4+P5+P6; FVAL = PRED
87   'CALC' ETA = COUNT - PRED
88   'TABULATE/PRIN=T'   ETA ; T1
89   'CALC' DEV = 2*(SUM(ETA-COUNT) + SUM(COUNT*LOG(COUNT/ETA)))
90   'PRINT' DEV $ 10.4
```

91 'REPEAT'  
 92 'RUN'

	T1			
DUMP	1	2	3	MARGIN
OPERATIO				
1	61.93	27.45	6.62	96.00
2	66.19	23.69	14.12	104.00
3	58.52	39.74	11.74	110.00
4	53.27	37.90	15.83	107.00
MARGIN	239.91	128.78	48.31	417.00

DEV 0.2220

	T1			
DUMP	1	2	3	MARGIN
OPERATIO				
1	61.89	27.46	6.65	96.00
2	66.19	23.65	14.16	104.00
3	58.50	39.75	11.75	110.00
4	53.26	37.90	15.84	107.00
MARGIN	239.84	128.76	48.40	417.00

DEV 0.2216

	T1			
DUMP	1	2	3	MARGIN
OPERATIO				
1	61.89	27.46	6.65	96.00
2	66.19	23.65	14.16	104.00
3	58.50	39.75	11.75	110.00
4	53.26	37.90	15.84	107.00
MARGIN	239.84	128.76	48.40	417.00

DEV 0.2216

93 'CLOSE'

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

3  'REFERENCE' MONKEY
-4      ''SEE SECTION 8.2.4
-5      FITTING THE LOGLINEAR MODEL OF
-6      QUASI-INDEPENDENCE TO SQUIRREL MONKEY
7      TABLE WITH FIXED ZEROES (TABLE 1.8)  ''
8  'UNITS' $ 30
9  'FACTOR' ACTIVE $ 5 = 6(1...5)
10      : PASSIVE $ 6 = (1...6)5
11  'VARIATE' COUNT
12  'READ/PRIN=D' COUNT
13  'RUN'

```

```

14  * 1 5 8 9 0
15  29 * 14 46 4 0
16  2 3 1 * 38 2
17  0 0 0 0 * 1
18  9 25 4 6 13 *
19  'EOD'

```

```

20  'TABLE' T1 $ ACTIVE, PASSIVE
21  'TABULATE' COUNT ; T1
22  'PRINT' T1 $ 10
23  'RUN'

```

	T1					
PASSIVE	1	2	3	4	5	6
ACTIVE						
1	0	1	5	8	9	0
2	29	0	14	46	4	0
3	2	3	1	0	38	2
4	0	0	0	0	0	1
5	9	25	4	6	13	0

```

24  'TERMS' COUNT+ACTIVE+PASSIVE

```

```

25 'Y/ERROR=POISSON'      COUNT
26 'CAPTION'  ''          (A,P)  ''
27 'FIT/PRIN=A' ACTIVE+PASSIVE ; FVAL = PRED
28 'RUN'

```

(A,P)

```

27.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	15	135.2	9.011

```

29 'TABULATE'  PRED ; T1
30 'PRINT'     T1 $ 10.3
31 'RUN'

```

	T1					
PASSIVE	1	2	3	4	5	6
ACTIVE						
1	0.000	5.260	2.481	8.216	6.648	0.396
2	19.186	0.000	10.322	34.185	27.661	1.647
3	10.936	12.474	5.884	0.000	15.767	0.939
4	0.220	0.251	0.118	0.392	0.000	0.019
5	9.658	11.016	5.196	17.208	13.924	0.000

```

32 'CLOSE'

```

\*\*\*\*\* END OF MONKEY.    MAXIMUM OF 1416 DATA UNITS USED AT LINE 27 ( 31352 LEFT)



GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

1  'REFER' HEALTH
-2      '' SEE SECTION 8.2.5.1
-3      FITTING THE LOGLINEAR MODEL (SA, SH)
-4      TO TABLE 8.3
-5      SEX X AGE X HEALTH CONCERNS
6      (WITH STRUCTURAL ZEROES) ''
7  'UNITS' $ 16
8  'VARIATE' COUNT $ 16
9      : PRED $ 16
11 'FACTOR' CONCERN $ 4 = 4(1...4)
12      : SEX $ 2 = 2(1,2)4
13      : AGE $ 2 = (1,2)8
15 'READ/PRIN=D' COUNT
17 'RUN'

```

```

18 4 2 9 7
19 * * 4 8
20 42 7 19 10
21 57 20 71 31
22 'EOD'

```

```

23 'TABLE' TAB1 $ CONCERN, SEX, AGE
24 'TABULATE' COUNT ; TAB1
25 'PRINT' TAB1 $ 9
26 'RUN'

```

		TAB1	
		1	2
CONCERN	SEX		
1	1	4	2
	2	9	7
2	1	0	0
	2	4	8
3	1	42	7
	2	19	10
4	1	57	20
	2	71	31

```

27 'TERMS' CONCERN*SEX*AGE+COUNT
28 'Y/ERROR=POISSON' COUNT

```

```
29 'FIT/PRIN=A'      SEX*AGE+CONCERN+SEX.CONCERN ; FVAL = PRED
30 'RUN'
```

29.....  
.....

\*\*\* LINEAR DEPENDENCE DETECTED WHILE FITTING TERM CONCERN.SEX  
1 PARAMETER(S) OF THIS TERM ARE ALIASED.  
\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	5	9.426	1.885

```
31 'TABULATE'      PRED ; TAB1
32 'PRINT'         TAB1 $ 10.2
33 'RUN'
```

CONCERN	AGE SEX	TAB1	
		1	2
1	1	4.68	1.32
	2	10.36	5.64
2	1	0.00	0.00
	2	7.77	4.23
3	1	38.23	10.77
	2	18.79	10.21
4	1	60.08	16.92
	2	66.08	35.92

```
34 'CLOSE'
```

\*\*\*\*\* END OF HEALTH. MAXIMUM OF 1648 DATA UNITS USED AT LINE 29 ( 31120 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

3  'REFERENCE' MOBILITY
-4      ''SEE SECTION 8.3.1.1
-5      FITTING THE LOGLINEAR MODEL OF QUASI-INDEPENDENCE
6      TO BRITISH SOCIAL MOBILITY DATA OF TABLE 8.6  ''
7  'UNITS' $ 25
8  'FACTOR' FATHER $ 5 = 5(1...5)
9      : SON $ 5 = (1...5)5
10 'VARIATE' COUNT
11 'READ/PRIN=D' COUNT
12 'RUN'

```

```

13 * 45 8 18 8
14 28 * 84 154 55
15 11 78 * 223 96
16 14 150 185 * 447
17 0 42 72 320 *
18 'EOD'

```

```

19 'TABLE' T1 $ FATHER, SON
20 'TABULATE' COUNT ; T1
21 'PRINT' T1 $ 10
22 'RUN'

```

	T1				
SON	1	2	3	4	5
FATHER					
1	0	45	8	18	8
2	28	0	84	154	55
3	11	78	0	223	96
4	14	150	185	0	447
5	0	42	72	320	0

```

23 'TERMS' COUNT+FATHER+SON
24 'Y/ERROR=POISSON' COUNT
25 'CAPTION' '' QUASI-INDEPENDENCE ''
26 'FIT/PRIN=A' FATHER+SON ; FVAL = PRED
27 'RUN'

```

QUASI-INDEPENDENCE

26.....  
 .....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: LOG  
 Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
 SCALE PARAMETER IS    1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	11	249.4	22.68

28 'TABULATE'    PRED ; T1  
 29 'PRINT'        T1 \$ 10.3  
 30 'RUN'

	T1				
SON	1	2	3	4	5
FATHER					
1	0.000	9.540	11.020	38.488	19.952
2	6.565	0.000	49.885	174.231	90.319
3	8.522	56.059	0.000	226.173	117.246
4	27.510	180.965	209.042	0.000	378.483
5	10.403	68.436	79.053	276.107	0.000

31 'CLOSE'

\*\*\*\*\* END OF MOBILITY.    MAXIMUM OF 1352 DATA UNITS USED AT LINE 26  
 (31416 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

```

1  'REFER' EYE
-2      '' SEE SECTION 8.3.3.1 AND APPENDIX 6
-3      FITTING THE MODEL OF QUASI-SYMMETRY TO
4      EYE-TESTING DATA OF TABLE 8.9      ''
5  'UNITS' $ 16
6  'VARIATE' COUNT      $ 16
7      :      PRED      $ 16
9  'FACTOR' LEFT $ 4
10     :      RIGHT $ 4
11     :      DUMMY $ 7
13 'READ/PRIN=D' RIGHT,LEFT,COUNT,DUMMY
15 'RUN'

```

```

16 1 1 1520 1
17 1 2 266 1
18 1 3 124 1
19 1 4 66 1
20 2 1 234 1
21 2 2 1512 2
22 2 3 432 5
23 2 4 78 6
24 3 1 117 1
25 3 2 362 5
26 3 3 1772 3
27 3 4 205 7
28 4 1 36 1
29 4 2 82 6
30 4 3 179 7
31 4 4 492 4
32 'EOD'

```

```

34 'TABLE'      TAB1 $ RIGHT, LEFT
35 'TABULATE'   COUNT ; TAB1
36 'PRINT'      TAB1 $ 9
37 'RUN'

```

LEFT RIGHT	TAB1			
	1	2	3	4
1	1520	266	124	66
2	234	1512	432	78
3	117	362	1772	205
4	36	82	179	492

```

39 'TERMS' LEFT+RIGHT+DUMMY+COUNT
40 'Y/ERROR=POISSON' COUNT
42 'FIT/PRIN=A' LEFT+RIGHT+DUMMY ; FVAL = PRED
43 'RUN'

```

```

42.....
.....

```

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG  
Y-VARIATE: COUNT

\*\*\* RESIDUAL DEVIANCE \*\*\*  
SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	7.271	2.424

```

44 'TABULATE' PRED ; TAB1
45 'PRINT' TAB1 $ 10.2
46 'RUN'

```

	TAB1			
LEFT RIGHT	1	2	3	4
1	1520.00	263.38	133.58	59.04
2	236.62	1512.00	418.99	88.39
3	107.42	375.01	1772.00	201.57
4	42.96	71.61	182.43	492.00

```

47 'CLOSE'

```

\*\*\*\*\* END OF EYE. MAXIMUM OF 1430 DATA UNITS USED AT LINE 42 (313  
38 LEFT)

GENSTAT V RELEASE 4.04B  
 COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
 ION)

```

2  'REFER' EYE
-3      '' SEE SECTION 8.3.3.1 AND APPENDIX 6
-4      FITTING THE MODEL OF QUASI-SYMMETRY TO
5      EYE-TESTING DATA OF TABLE 8.9      ''
6  'UNITS' $ 16
7  'VARIATE' COUNT      $ 16
8      :      PRED      $ 16
10 'FACTOR' LEFT $ 4
11      :      RIGHT $ 4
12      :      HIGH $ 4
13      :      LOW $ 4
15 'READ/PRIN=D' RIGHT,LEFT,COUNT,HIGH,LOW
17 'RUN'

```

```

18 1 1 1520 1 1
19 1 2 266 2 1
20 1 3 124 3 1
21 1 4 66 4 1
22 2 1 234 2 1
23 2 2 1512 2 2
24 2 3 432 3 2
25 2 4 78 4 2
26 3 1 117 3 1
27 3 2 362 3 2
28 3 3 1772 3 3
29 3 4 205 4 3
30 4 1 36 4 1
31 4 2 82 4 2
32 4 3 179 4 3
33 4 4 492 4 4
34 'EOD'

```

```

36 'TABLE'      TAB1 $ RIGHT, LEFT
37 'TABULATE'   COUNT ; TAB1
38 'PRINT'      TAB1 $ 9
39 'RUN'

```

	TAB1			
LEFT	1	2	3	4
RIGHT				
1	1520	266	124	66
2	234	1512	432	78
3	117	362	1772	205
4	36	82	179	492

```

41 'TERMS' LEFT+RIGHT+HIGH*LOW+COUNT
42 'Y/ERROR=POISSON' COUNT
43 'FIT/PRIN=A' LEFT+RIGHT+HIGH*LOW ; FVAL = PRED
44 'RUN'

```

```

43.....
.....

```

```

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM LOW
    ALL PARAMETER(S) OF THIS TERM ARE ALIASED.
*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM HIGH.LOW
    6 PARAMETER(S) OF THIS TERM ARE ALIASED.
***** REGRESSION ANALYSIS *****

```

```

ERROR DISTRIBUTION: POISSON LINK FUNCTION: LOG
Y-VARIATE: COUNT

```

```

*** RESIDUAL DEVIANCE ***
SCALE PARAMETER IS 1.000

```

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	3	7.271	2.424

```

46 'TABULATE' PRED ; TAB1
47 'PRINT' TAB1 $ 10.2
48 'RUN'

```

	TAB1			
LEFT	1	2	3	4
RIGHT				
1	1520.00	263.38	133.58	59.04
2	236.62	1512.00	418.99	88.39
3	107.42	375.01	1772.00	201.57
4	42.96	71.61	182.43	492.00

```

49 'CLOSE'

```

```

***** END OF EYE. MAXIMUM OF 1884 DATA UNITS USED AT LINE 43 (308
84 LEFT)

```



GENSTAT V      RELEASE 4.04B  
COPYRIGHT 1984 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STAT  
ION)

```
2  'REFER' EYE

-3      '' SEE SECTION 8.3.4.2 AND APPENDIX 7
-4      FITTING THE MODEL OF MARGINAL HOMOGENEITY TO
-5      EYE-TESTING DATA OF TABLE 8.9
6      USING THE METHOD OF SOLVING SIMULTANEOUS EQUATIONS ''

7  'UNITS' $ 16
8  'VARIATE'  COUNT      $ 16
9  'FACTOR'   RIGHT      $ 4
10     :      LEFT       $ 4
11  'SCALAR'   DEV
12  'READ/PRIN=D' RIGHT,LEFT,COUNT,P1,P2,P3,P4,P5,P6,P7,P8,OFF
13  'RUN'

14  1 1 1520  0  0  0  0  0  0  0  0  1520
15  2 1 234   3  2  4 -2  2 -1  3  2 -2181
16  3 1 117  -2 -1 -3  1 -2  1 -3 -2 2181
17  4 1 36    0  0  0  1  0  0  0  0  0
18  1 2 266   1  0  0  0  0  0  0  0  0
19  2 2 1512  0  0  0  0  0  0  0  0 1512
20  3 2 362   0  0  0  0  0  0  0  1  0
21  4 2 82    0  0  1 -1  1 -1  1  0  0
22  1 3 124   0  1  0  0  0  0  0  0  0
23  2 3 432  -2 -2 -3  1 -2  0 -2 -1 2181
24  3 3 1772  0  0  0  0  0  0  0  0 1772
25  4 3 179   0  0  0  0  0  1  0  0  0
26  1 4 66    0  0  1  0  0  0  0  0  0
27  2 4 78    0  0  0  0  1  0  0  0  0
28  3 4 205   0  0  0  0  0  0  1  0  0
29  4 4 492   0  0  0  0  0  0  0  0 492
30  'EOD'

31  'RUN'
32  'TABLE/M'  PREDTABLE $ RIGHT, LEFT
33  'TABULATE/PRIN=T'  COUNT ; PREDTABLE
34  'RUN'
```

	PREDTABL				
LEFT	1	2	3	4	MARGIN
RIGHT					
1	1520.00	266.00	124.00	66.00	1976.00
2	234.00	1512.00	432.00	78.00	2256.00
3	117.00	362.00	1772.00	205.00	2456.00
4	36.00	82.00	179.00	492.00	789.00
MARGIN	1907.00	2222.00	2507.00	841.00	7477.00

```

35 'TERMS/OFFSET=OFF'          P1+P2+P3+P4+P5+P6+P7+P8+COUNT
36 'Y/ERROR=POISSON, LINK=IDENTITY'      COUNT
37 'FIT/INT=N'          P1+P2+P3+P4+P5+P6+P7+P8 ; FVAL = PRED
38 'RUN'

```

37.....

\*\*\*\*\* REGRESSION ANALYSIS \*\*\*\*\*

ERROR DISTRIBUTION: POISSON    LINK FUNCTION: IDENTITY  
Y-VARIATE: COUNT  
OFFSET VARIATE: OFF

\*\*\* REGRESSION COEFFICIENTS \*\*\*

	ESTIMATE	S.E.	T
P1	252.5	11.8	21.47
P2	111.84	9.29	12.04
P3	56.95	6.89	8.26
P4	42.77	6.11	7.00
P5	70.59	7.65	9.23
P6	188.4	10.6	17.79
P7	195.3	10.6	18.34
P8	383.1	13.4	28.56

\* STANDARD ERRORS BASED ON SCALE PARAMETER WITH VALUE 1.000

\*\*\* RESIDUAL DEVIANCE \*\*\*

SCALE PARAMETER IS 1.000

	DF	DEVIANCE	MEAN DEVIANCE
RESIDUAL	8	11.99	1.498

```

39 'TABULATE/PRIN=T'      PRED ; PREDTABLE
40 'RUN'

```

	PREDTABL				
LEFT	1	2	3	4	MARGIN
RIGHT					
1	1520.00	252.49	111.84	56.95	1941.28
2	247.24	1512.00	409.43	70.59	2239.25
3	131.27	383.14	1772.00	195.27	2481.67
4	42.77	91.62	188.40	492.00	814.80
MARGIN	1941.28	2239.25	2481.67	814.80	7477.00

# REFERENCES

- Agresti, A. (1983) A survey of strategies for modelling cross-classifications having ordinal variables.  
Journal of the American Statistical Association 78, 184-198.
- Agresti, A. (1984) Analysis of ordinal categorical data.  
New York : Wiley.
- Alvey, N G et al. (1983) GENSTAT : a general statistical package.  
Rothamsted Experimental Station.
- Anderson, J A & Phillips, P R. (1981) Regression, discrimination and measurement models for ordered categorical variables.  
Applied Statistics 30, 22-31.
- Aston, C E & Wilson, S R. (1984) Comment on M B Brown and C Fuchs, "On maximum likelihood estimation in sparse contingency tables".  
Computational Statistics & Data Analysis 2, 71-77.
- Baker, R J and Nelder, J A. (1978) The GLIM Manual : Release 3.  
Oxford : Numerical Algorithms Group.
- Bishop, Y M M, Fienberg, S E and Holland, P W. (1975) Discrete multivariate analysis : theory and practice.  
Cambridge, Mass. : The MIT Press.
- Brown, M B and Fuchs, C. (1983) On maximum likelihood estimation in sparse contingency tables.  
Computational Statistics & Data Analysis 1, 3-15.
- Brown, M B and Fuchs, C. (1984) Rejoinder on comment by C E Aston and S R Wilson.  
Computational Statistics & Data Analysis 2, 79-80.
- Cornfield, J. (1962) Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure : a discriminant function analysis.  
Federation Proc. 21, 58-61.

- Edwards, A W F. 1963. The measure of association in a 2x2 table.  
J. Roy. Statist. Soc. A 126, 109-114.
- Fienberg, S E. (1980) The analysis of cross-classified data, 2nd ed.  
Cambridge, MA : MIT Press.
- Fleiss, J L. 1979. "Confidence intervals for the odds ratio in case-control studies : the state of the art".  
J. Chron. Dis. 32, 69-77.
- Forthofer, R N & Lehnen, R G. (1981) Public program analysis, a new categorical data approach.  
Belmont, Calif. : Lifetime Learning Publications.
- Freedman, D, Pisani, R and Purves, R. (1978). Statistics.  
New York : W W Norton.
- Gart, J J and Zweifel, J R. 1967. "On the bias of various estimators of the logit and its variance with application to Quantal bioassay".  
Biometrika 54, 181-187.
- Goodman, L A. (1979) Simple models for the analysis of association in cross-classifications having ordered categories.  
Journal of the American Statistical Association 74, 537-552.
- Goodman, L A. (1983) The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds.  
Biometrics 39, 149-160.
- Grizzle, J E, Starmer C F and Koch, G G. (1969). Analysis of categorical data by linear models.  
Biometrics 25, 489-504.
- Haber, M. (1985) Maximum likelihood methods for linear and log-linear models in categorical data.  
Computational Statistics & Data Analysis 3, 1-10.
- Haberman, S J. (1981) Tests for interaction in two-way contingency tables based on caronical correlation and on linear-by-linear interaction.  
Annals of Statistics 9, 1178-1186.

- Knoke, D and Burke, P J. (1980) Log-linear models.  
Sage University Paper series on quantitative applications in the  
social sciences, 07-001.  
London : Sage Publications.
- Ku, H H and Kullback, S. (1974) Log-linear models in contingency  
table analysis.  
American Statistician 28, 115-122.
- Lindgren, B W. (1976) Statistical theory.  
New York : Macmillan.
- McCullagh, P. (1980) Regression models for ordinal data.  
Journal of the Royal Statistical Society, Series B 42, 109-142.
- Nelder, J A & Wedderburn, R W M. (1972) Generalized linear models.  
Journal of the Royal Statistical Society, Series A 135, 370-384.
- Pauling, L. 1971. "The Significance of the evidence about ascorbic  
acid and the common cold".  
Proc. Nat. Acad. Sci. (USA) 68, 2678-2681.
- Pearson, E S and Hartley, H O. (1972) Biometrika tables for  
statisticians, vol. II.  
Cambridge : Cambridge University Press.
- Pearson, K. (1900) On a criterion that a given system of deviations  
from the probable in the case of a correlated system of variables  
is such that it can be reasonably supposed to have arisen from  
random sampling.  
Philos. Mag. 50, No.5, 157-175.
- Stirling, W D. (1984a) Fitting linear models to ordinal responses.  
British Journal of Mathematical and Statistical Psychology 37,  
263-270.
- Stirling, W D. (1984b) Further analysis of categories containing  
overlapping responses.  
The New Zealand Statistician 19, no.1, 49-53.
- Stirling, W D. (1986) Testing linear hypotheses in contingency  
tables with zero cell counts.  
Computational Statistics & Data Analysis 4, 13-25.

Stirling, W D. (1986) A note on degrees of freedom on sparse contingency tables.

Computational Statistics & Data Analysis, 4, 26-28.

Wedderburn, R W M. (1974) Generalized linear models specified in terms of constraints.

Journal of the Royal Statistical Society B 36, 449-454.

Whittemore, A S. (1978) Collapsibility of multidimensional contingency tables.

Journal of the Royal Statistical Society, Series B 40, 328-340.

Williams, O D and Grizzle, J E. (1972) Analysis of contingency tables having ordered response categories.

Journal of the American Statistical Association 67, 55-63.