WORLD FERTILITY SURVEY
TECHNICAL BULLETINS

OCTOBER 1978

NO. 5/TECH. 834

Generalized Linear Models for Cross-Classified Data from the WFS

RODERICK J. A. LITTLE

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office · Director: E. Lunenberg
Prinses Beatrixlaan 428
Voorburg, The Hague,
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Sir Maurice Kendall, Sc. D., F.B.A.
35-37 Grosvenor Gardens
London SW1W OBS, U.K.

The World Fertility Survey is an international research programme whose purpose is to assess the current state of human fertility throughout the world. This is being done principally through promoting and supporting nationally representative, internationally comparable, and scientifically designed and conducted sample surveys of fertility behaviour in as many countries as possible.

The WFS is being undertaken, with the collaboration of the United Nations, by the International Statistical Institute in cooperation with the International Union for the Scientific Study of Population. Financial support is provided principally by the United Nations Fund for Population Activities and the United States Agency for International Development.

This paper is one of a series of Technical Bulletins recommended by the WFS Technical Advisory Committee to supplement the document *Strategies for the Analysis of WFS Data* and which deal with specific methodological problems of analysis beyond the Country Report No. 1. Their circulation is restricted to people involved in the analysis of WFS data, to the WFS depositary libraries and to certain other libraries. Further information and a list of these libraries may be obtained by writing to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, Voorburg, The Hague, Netherlands.

GENERALIZED LINEAR MODELS FOR CROSS-CLASSIFIED DATA FROM THE WFS

WFS/TECH.834

By: RODERICK J.A. LITTLE
WFS Central Staff
International Statistical Institute
35-37 Grosvenor Gardens
LONDON SW1W OBS, U.K.

CONTENTS

ACKNOWLEDGEMENTS 1

ABSTRACT 2

FOREWORD 3

1. INTRODUCTION 6

2. AN EXAMPLE OF A MODEL 13

3. GENERALIZED LINEAR MODELS 19

3.1 LINEAR MODELS 20

3.1.1 The Systematic Component 20

3.1.2 The Normal Error Structure 22

3.2 GENERALIZED LINEAR MODELS 23

3.2.1 The Systematic Component 24

3.2.2 Error Structures other than Normal 25

3.3 SUMMARY 26

4. HIERARCHICAL MODELS: FITTING AND SELECTION 27

4.1 THE SATURATED MODEL 27

4.2 UNSATURATED HIERARCHICAL MODELS 27

4.3 MODEL FITTING 29

4.3.1 Linear Models 29

4.3.2 Log-Linear and Logit-Linear Models 32

4.4 MODEL SELECTION USING THE MEAN DEVIANCE 32

4.4.1 Nested Models 33

4.4.2 F-tests of Relative Fit 33

4.4.3 Chi-Squared Tests of Absolute Fit for Poisson and Binomial Error Structures 35

4.4.4 Chi-Squared Tests of Relative Fit for Poisson and Binomial Error Structures 36

4.5 SUMMARY 37

5. INTERPRETATION OF ESTIMATED EFFECTS 38

5.1 LINEAR MODELS 38

5.2 LOG-LINEAR AND LOGIT-LINEAR ADDITIVE MODELS 39

5.3	STANDARDIZATION OF THE FITTED MEANS	43
5.4	WHY FIT A MODEL	46
5.5	SUMMARY	47
6.	LOGIT LINEAR MODELS FOR PROPORTIONS	47
6.1	THE DATA	47
6.2	MODEL SELECTION	48
6.3	HYPOTHESIS TESTING	48
6.4	INSPECTION OF THE FITTED VALUES: LOGIT PLOTS	50
6.5	SUMMARY	54
REFERENCES		55
APPENDIX I :	ADDITIONAL ANALYSIS BASED ON WITHIN-CELL SAMPLE VARIANCE	56
	NORMAL ERROR	56
	POISSON ERROR	58
	SUMMARY	61
APPENDIX II :	COMPUTING NOTES	62
APPENDIX III:	AN EXAMPLE OF OUTPUT FROM GLIM	66

ACKNOWLEDGEMENTS

Whilst taking full responsibility for errors in this paper, the author would like to acknowledge gratefully the advice and suggestions of Dr. Y. Bishop, Dr. R.O. Carleton, Professor W.J. Dixon, Dr. J.A. Nelder, Professor T.W. Pullum, and numerous members of the World Fertility Survey staff.

ABSTRACT

Generalized linear models for cross classifications of means and proportions are presented with particular reference to data collected by fertility surveys. Besides the standard linear normal models which form the basis of analysis of variance, log-linear and logit-linear models are discussed. Topics covered include the objectives of model fitting, model selection, the estimation and interpretation of parameters and fitted values, standardization and the treatment of within-cell sample variances. The methods are applied to data on fertility and contraceptive use from the Fiji Fertility Survey, using the computer program GLIM.

FOREWORD

The problem discussed in this paper is one of several types occurring in the analysis of classified data. For a population cross-classified by several criteria e.g.,by age, education, social class, and district of residence, it is required to examine how some specified variable (which may be continuous, discontinuous or itself merely classificatory) varies over the different classes and sub-classes.

In some cases the entries in the sub-cells of the classifications are themselves counts of numbers of individuals, that is, the table is a contingency table. A comprehensive study of such cases, among others, will be found in the book by Bishop, Fienberg and Holland (1975). In this paper contingency tables are studied in a more restricted way; a variable taking only two values is formed, for example, the variable Contraceptive Use with two values:1 = Ever Used, 0 = Never Used; then the proportion of respondents taking one of these values is analysed across the cells of the table.

The other situation considered here concerns tables where the cell entries are mean values of a quantitative variable such as parity (mean number of children ever born to the female respondent). This raises some special problems in that, although the number of respondents in each cell is known, the distribution of the dependent variable and, in particular,the variance within cells is not. Recourse could be had to the primary information to resuscitate this information and some suggestions for the treatment of within cell variances are given in Appendix I ; however,many demographic tables are printed out from the computer in the form which is studied here.

The fundamental problem, in this as in many multivariate situations is that the variables are not independent among themselves and it is therefore difficult to assess their relative contribution to the specified variable. The general objective of analysis is then to study how the specified variable varies between sub-classes in the categorization. If, for instance, it were found that the relationship between parity and education was the same for each district of residence and age-group the model of the behaviour of the system could be simplified in that, for such a relationship, district and age could be ignored. The technical way of saying this would be to state that certain interactions were zero, or that behaviour could be accounted for by fewer parameters. One of the objects of model-fitting, in this kind of study, is to find as parsimonious a model as possible; that is to say, one which requires the smallest number of parameters to "explain" it - a version of the old logical principle known as Ockham's razor.

It may be useful to explain two expressions which occur thematically throughout the whole of multivariate contingency analysis. To fix the idea, suppose a specified variable V is studied under categorization by three variables A, B, C. We can then consider how V varies

(a) among the sub-cells ABC (i.e., if there are a categories in A, b in B and c in C, among the abc sub-cells).

(b) among the two-way classifications AB, BC, CA or combinations of them,

(c) among the one-way classification A, B, and C or combinations of them, or

(d) that it does not vary but is a constant.

In short, we can consider a model

$$V = \lambda + \lambda_A + \lambda_B + \lambda_C + \lambda_{AB} + \lambda_{BC} + \lambda_{CA} + \lambda_{ABC}$$

in which any or all of the terms except the constant λ vanish. If only λ_A, λ_B, λ_C survive we have a model in which the categorizations $A, B, C,$ are independently contributing to V in an additive way. If such a model gives an inadequate fit we involve some or all of the terms typified by λ_{AB}. These are known as *first-order interactions* or *two-factor effects.** If these are inadequate we have to involve λ_{ABC}.

These models become complicated for more than three classificatory variables and it is customary to examine the general model *hierarchically*. That is to say, we find out whether we can discard λ_{ABC} as non-contributory. If so we proceed to examine the two-factor effects; and so on. The point of proceeding in this way is that if we retain a particular effect, we automatically retain the effects of lower order which it embodies; for example, if we retain λ_{AB} we retain λ_A and λ_B, and if we retain λ_{AB} and λ_{BC}, we must retain λ_A, λ_B and λ_C. The procedure bears an obvious resemblance to the step-wise backwards fitting of regression analysis.

Sir Maurice Kendall

---

* The term "interaction" arose in the analysis of variances in an agricultural context. If a plant was dressed with a nitrogenous and a phosphoric fertiliser and the two did not affect the yield independently they were said to "interact" and this may have corresponded to a real physical interaction. More generally, the term has come to mean the absence of independence of two or more factors. It may be better, as Bishop *et al* recommend, to avoid the term and speak of one-factor, two-factor, three-factor effects. A p-factor effect is a $(p-1)$ interaction.

1. INTRODUCTION

Data from WFS Surveys are often presented in the form of multi-way classifications, where sample quantities such as counts, percentages or means are classified according to a set of background characteristics such as age, educational level, region. Each characteristic is represented by a *factor*, or *variable*, which has a set of possible levels, and the result is a table of cells with one cell for each combination of levels of the factors.

If the entries in the table are the number of sampled individuals which belong to each cell, then the table is called a *contingency table*. For example, Table 1 is a contingency table with 5 factors defined as follows:

R = Race: 1 = Fijians, 2 = Indians

A = Age (Years): 1 = <25, 2 = 25-29, 3 = 30-39, 4 = 40-49

E = Education: 1 = Lower Primary or less, 2 = Upper Primary or more.

W = Desire for More Children: 1 = Yes, 2 = No.

U = Pattern of Contraceptive Use: 1 = Ever Used, Not Currently Using.
2 = Currently Using.
3 = Never Used.

This table is adapted from Table I15 in the Fiji Fertility Survey Country Report. The number of levels of the variable "Pattern of Contraceptive Use" in that table are reduced from 6 to 3 by dropping the classification according to future intention, and classifying sterilised women as current users.

TABLE 1 Distribution of Currently Married "Fecund"[a] Women by Race, Current Age, Pattern of Contraceptive Use, Desire for More Children and Educational Level

			FIJIANS Age[b]				INDIANS Age			
Educ	Desire for More Children	Pattern[c] of Contraceptive Use	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
LOW	Yes	(1)	12	25	59	11	70	59	43	4
		(2)	6	14	33	6	78	59	35	4
		(3)	41	35	53	24	99	45	57	11
LOW	No	(1)	6	10	55	26	28	37	71	44
		(2)	4	10	80	48	27	106	378	212
		(3)	4	9	22	20	12	17	48	50
HIGH	Yes	(1)	62	73	58	4	71	44	16	2
		(2)	52	54	46	8	113	64	21	0
		(3)	150	82	60	4	97	23	17	1
HIGH	No	(1)	24	54	45	9	20	33	21	6
		(2)	10	27	78	31	22	56	100	32
		(3)	26	11	23	3	8	12	12	9

a Sterilized women are included as currently contracepting.

b Age categories defined as follows:
1 = <25, 2 = 25-29, 3 = 30-39 4 = 40-49.

c Contraceptive Use categories are labelled as follows:
(1) = Ever Used, Not Currently Using (2) = Currently Using,
(3)= Never Used.

TABLE 2a) : Proportion Currently Using An Efficient Contraceptive Method According to Race, Current Age, Desire for More Children and Educational Level

Education	Desire For More Children		FIJIANS Age 1	2	3	4	INDIANS Age 1	2	3	4
LOW	YES	p	.10	.19	.23	.15	.32	.36	.26	.21
		n	59	74	145	41	247	163	135	19
LOW	NO	p	.29	.34	.51	.51	.40	.66	.76	.69
		n	14	29	157	94	67	160	497	306
HIGH	YES	p	.20	.26	.28	.50	.40	.49	.39	.00
		n	264	209	164	16	281	131	54	3
HIGH	NO	p	.17	.29	.53	.72	.44	.55	.75	.68
		n	60	92	146	43	50	101	133	47

p = proportion

n = sample size

TABLE 2b) : Proportion Who Ever Used Efficient Contraceptive Method According to Race, Current Age, Desire For More Children and Educational Level

Education	Desire For More Children		FIJIANS Age 1	2	3	4	INDIANS Age 1	2	3	4
LOW	YES	p	.31	.53	.63	.41	.60	.72	.58	.42
		n	59	74	145	41	247	163	135	19
LOW	NO	p	.71	.69	.86	.79	.82	.89	.90	.84
		n	14	29	157	94	67	160	497	306
HIGH	YES	p	.43	.61	.63	.75	.65	.82	.69	.67
		n	264	209	164	16	281	131	54	3
HIGH	NO	p	.57	.88	.84	.93	.84	.88	.91	.81
		n	60	92	146	43	50	101	133	47

p = proportion

n = sample size

TABLE 3 Mean Number of Children Ever Born for Women of Indian Race, by Marital Duration, Type of Place and Education.

FITTED VALUES[a] FROM LOG LINEAR MODEL, OBSERVED VALUES[b], AND SAMPLE SIZES[c].

TYPE OF PLACE

Years Since First Marriage	SUVA				URBAN				RURAL			
	Education*				Education*				Education*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
<5	0.89[a]	0.91	0.80	0.65	.99	1.02	0.90	0.73	1.03	1.06	.93	0.76
	0.50[b]	1.14	0.90	0.73	1.17	0.85	1.05	0.69	0.97	0.96	0.97	0.74
	8[c]	*21*	*42*	*51*	*12*	*27*	*39*	*51*	*62*	*102*	*107*	*47*
5 - 9	2.41	2.47	2.18	1.77	2.70	2.76	2.44	1.98	2.81	2.87	2.53	2.06
	3.10	2.67	2.04	1.73	4.54	2.65	2.68	2.29	2.44	2.71	2.47	2.24
	10	*30*	*24*	*22*	*13*	*37*	*44*	*21*	*70*	*117*	*81*	*21*
10-14	3.50	3.58	3.16	2.57	3.92	4.01	3.54	2.87	4.07	4.17	3.68	2.99
	4.08	3.67	2.90	2.00	4.17	3.33	3.62	3.33	4.14	4.14	3.94	3.33
	12	*27*	*20*	*12*	*18*	*43*	*29*	*15*	*88*	*132*	*50*	*9*
15-19	4.47	4.57	4.04	3.28	5.00	5.12	4.52	3.67	5.20	5.32	4.69	3.81
	4.21	4.94	3.15	2.75	4.70	5.36	4.60	3.80	5.06	5.59	4.50	2.00
	14	*31*	*13*	*4*	*23*	*42*	*20*	*5*	*114*	*86*	*30*	*1*
20-24	5.30	5.43	4.79	3.89	5.93	6.07	5.36	4.35	6.17	6.31	5.57	4.53
	5.62	5.06	3.92	2.60	5.36	5.88	5.00	5.33	6.46	6.34	5.74	2.50
	21	*18*	*12*	*5*	*22*	*25*	*13*	*3*	*117*	*68*	*23*	*2*
25+	6.42	6.57	5.80	4.71	7.18	7.35	6.49	5.27	7.47	7.64	6.75	5.48
	6.60	6.74	5.38	2.00	6.52	7.51	7.54	-	7.48	7.81	5.80	-
	47	*27*	*8*	*1*	*46*	*45*	*13*	*0*	*195*	*59*	*10*	*0*

SOURCE: FIJI FERTILITY SURVEY

*Categories of Education Level are:
(1) None
(2) Lower Primary
(3) Upper Primary
(4) Secondary or Higher

Note: Figures in italics are Base Frequencies

MODEL: $\hat{m}_{tde} = \hat{\gamma}\cdot\hat{\gamma}_{1t}\cdot\hat{\gamma}_{2d}\cdot\hat{\gamma}_{3e}$

$\hat{\gamma} = 5.479$

$\hat{\gamma}_{11} = 0.860 \quad \hat{\gamma}_{12} = 0.963 \quad \left[\hat{\gamma}_{13} = 1\right]$

$\hat{\gamma}_{21} = 0.139 \quad \hat{\gamma}_{22} = 0.376 \quad \hat{\gamma}_{23} = 0.545 \quad \hat{\gamma}_{24} = 0.696 \quad \hat{\gamma}_{25} = 0.826 \quad \left[\hat{\gamma}_{26} = 1\right]$

$\hat{\gamma}_{31} = 1.363 \quad \hat{\gamma}_{32} = 1.395 \quad \hat{\gamma}_{33} = 1.231 \quad \left[\hat{\gamma}_{34} = 1\right]$

DEVIANCE: $\chi^2 = 70.65$ on 59 degrees of freedom.

The counts in a contingency table describe the joint distribution of the factors. If interest is focussed on how the distribution of one factor varies with respect to the other factors, then the counts are often converted into percentage distributions of the factor of interest, with counts in the margin. For example in Table 1, we may be interested in the effects of R, A, E and W on Contraceptive Use. Hence a new table is constructed, where the percentage distribution of U is cross-classified by the other variables. The marginal counts in this table form a four-way contingency table describing the joint distribution of R, A, E and W.

If the dependent variable (here U) is reduced to two levels, then the proportion p out of n for which it takes one of its two values can be cross-classified, thus effectively reducing the dimensions of the cross-classification by one. For example, in Table 2a) we classify the proportion currently using contraception (U=2), and in Table 2b) we classify the proportion who have ever used contraception (U=2 + U=1) by R, A, E and W. In the analysis of these tables we are usually interested in variation in the proportions p across the cells, and the distribution of the sample sizes is not studied, although the sample sizes do play a role in the analysis since they determine the precision with which the proportions are estimated.

So far we have considered tables of counts. Other tables cross-classify the sample mean $\bar{y}$ (or some other summary statistic such as the median) of the ordinal variable Y for each subsample formed by the joint levels of factors.

As in tables of proportions, two quantities are tabulated, the mean $\bar{y}$ for each cell and the sample size n on which the mean is based, and the primary interest is in the variation of the means across the cells, which describe the effects of the factors on the *response* or *dependent* variable Y.

For example, Table 3 displays a cross-classification of the mean parity (that is, the mean number of live births) for women of Indian race from the Fiji Fertility Survey. There are 3 factors,

T = Type of Place of Residence: 1=Suva, 2=Other Urban, 3=Rural

D = Marital Duration(Years): 1=0-4, 2=5-9, 3=10-14, 4=15-19, 5=20-24, 6=25+

E = Education: 1=No Education, 2=Lower Primary, 3=Upper Primary, 4=Secondary and Higher

The first entry in each cell of the table is described later; the second entry is the mean parity and the third entry is the sample size.

It is worth noting that a cross-tabulation of proportions can be considered a special case of cross-tabulation of means in which the response Y is dichotomous and takes values 1 and 0; for then the mean in each cell is simply the proportion of cases with Y=1.

We shall describe here the formulation of models for cross-tabulations of means or proportions, which describe the effect of a set of factors on the mean of a scalar or dichotomous dependent variable.

Note that we shall not consider models for contingency tables such as the log-linear models formulated by Goodman (see, for example, Goodman, 1970, 1972) and described by Davis (1974) or Bishop, Fienberg and Holland (1975). These models are formally a special case of the log-linear models for cross-classified means discussed here, and they can be calculated using the same computer programs. However they describe the joint distribution of a set of variables, and hence have a very different interpretation from the models given here, which describe the effect of a set of variables on a response*

---

* Analysis of contingency tables is in some respects analogous to correlation analysis of scalar variables whereas the analysis of tables of means is analogous to regression analysis of a scalar response.

Contingency table models have a limited role in the analysis of WFS data, since many contingency tables in the tabulations for the First Country Report are largely descriptive in character, whilst others are usefully analysed by defining one or more dichotomous responses and analysing the cross-classification of proportions by the methods given here. This restricts the scope of the analysis, but has the advantage of reducing the dimensions of the cross-classification by one, a reduction which greatly simplifies the treatment of four or five way tables.

There are two serious limitations of any analysis based solely on a cross-classification of means. The first involves the within cell variance of Y, that is the variability of the response for a group homogeneous with respect to the factors. Often the table of means contains no information about these variances.* Hence either unverifiable assumptions have to be made about them, or additional data are required usually in the form of within cell sample variances. Thus it is important to calculate these sample variances if at all possible; the additional information does not radically affect the analysis of the means described here, but provides checks on the variance assumptions, as illustrated in Appendix 1.

A more fundamental limitation concerns the grouping of continuous variables such as age and marital duration. The analysis of cross-classified data given here does not take into account ordering between the categories of such grouped variables, and the choice of categories is rather arbitrary. In many ways an analysis of the individual level data without grouping is more natural, using regression or analysis of covariance techniques. Although this may be more informative, the results tend to be more abstract and hard to interpret, and cross-classifications at least present a simple picture of multivariate data without the abstract indices such as correlation coefficients required for individual level data.

The general idea behind fitting models is to clarify this picture by distinguishing significant features from insignificant detail. Before launching into a full description of the models we shall illustrate this by describing a simple model for the data in Table 3.

---

* Exceptions to this rule are tables of proportions, where the variance of Y is given by the Binomial formula $p\ (1-p)$.

2. AN EXAMPLE OF A MODEL

There are several paths which might lead to an examination of the data in Table 3. For example, suppose that we are studying the effect of education on fertility. The mean parity for different levels of education seem to indicate that education is strongly related to fertility:

INDIANS	EDUCATION				
	None(1)	L.Prim(2)	U.Prim(3)	Sec+Higher(4)	Mean
Mean Parity	5.2	4.2	2.8	1.5	4.0
(Sample Size)	(895)	(943)	(579)	(271)	(2688)

However, part of these differences is attributable to a compositional effect of marital duration. That is, highly educated women tend to be younger and marry later than less educated women, and hence have shorter exposure to the risk of conception. This compositional effect can be studied by a two-way cross classification of mean parity by education and marital duration. The table (not shown here) indicates that education is still negatively associated with fertility after marital duration is controlled, although the magnitude of the effect is much reduced. Looking further, we may hypothesize that this relationship is merely reflecting urban-rural differentials in fertility. That is, within the urban sector and the rural sector differentials according to educational level do not exist, and the differentials are caused by the concentration of more highly educated women in the urban sector. To investigate this the three-way table, Table 3, is constructed.*

An obvious difficulty in such multiway classification is that the within cell sample sizes are small. In Table 3 twelve of the 72 sample means are based on samples of less than ten and two cells are empty. As a result it is difficult to distinguish real differences from differences caused by random fluctuation.

---

* We do not intend to discuss the merits of marital duration as a demographic control here. For some comments see Pullum (1977). The methods presented here are equally applicable to tables which control age, or age and age at marriage.

It is worth noting that this table is not recommended for the First Country Report, where ten-year rather than five-year marital duration cohorts are used to reduce the number of cells. However, in this case the strong relationship between education and marital duration and between marital duration and parity render this table unsatisfactory, because the effect of marital duration is not properly controlled. Hence we are led to consider the more detailed table.

Using statistical terminology, the problem is that a large number of *parameters*, the 72 population cell means, are being estimated from the sample. Suppose that we try to estimate the cell means from a smaller set of parameters. Specifically we calculate a grand mean ($\hat{\gamma}$), three Type of Place estimates, one for each level of Type of Place ($\hat{\gamma}_{11},\hat{\gamma}_{12},\hat{\gamma}_{13}$), six Marriage Duration estimates ($\hat{\gamma}_{21},\hat{\gamma}_{22},\ldots,\hat{\gamma}_{26}$) and four Education estimates ($\hat{\gamma}_{31},\hat{\gamma}_{32},\hat{\gamma}_{33},\hat{\gamma}_{34}$) and then calculate a fitted value for each cell mean of the form

Fitted value = (Grand mean) x (Type of Place estimate) x (Marriage Duration estimate) x (Education estimate)

In this way estimates of the 72 cell means are produced from 1 + 3 + 6 + 4 = 14 parameters. Let us denote the cell with Type of Place = t, Marriage Duration = d and Education = e by the subscripts tde; let $\bar{y}_{tde}$ be the sample mean parity and $\hat{m}_{tde}$ the fitted value for this cell. Then the fitted values are given by the formula

$$\hat{m}_{tde} = \hat{\gamma}\cdot\hat{\gamma}_{1t}\cdot\hat{\gamma}_{2d}\cdot\hat{\gamma}_{3e}, \text{ for all } t, d \text{ and } e \quad . \qquad (2.1)$$

Loosely speaking, the estimates are chosen so that the fitted values $\hat{m}_{tde}$ resemble as closely as possible the sample means, $\bar{y}_{tde}$. The exact criterion of fitting and the method of calculation are considered later. The estimates giving the best fit for the data in Table 3 are as follows:

Grand mean:	$\hat{\gamma} = 5.479$
Type of Place:	$\hat{\gamma}_{11} = 0.860$, $\hat{\gamma}_{12} = 0.963$, $\hat{\gamma}_{13} = 1$
Marriage Duration:	$\hat{\gamma}_{21} = 0.139$, $\hat{\gamma}_{22} = 0.376$, $\hat{\gamma}_{23} = 0.545$, $\hat{\gamma}_{24} = 0.696$, $\hat{\gamma}_{25} = 0.826$, $\hat{\gamma}_{26} = 1$
Education:	$\hat{\gamma}_{31} = 1.363$, $\hat{\gamma}_{32} = 1.395$, $\hat{\gamma}_{33} = 1.231$, $\hat{\gamma}_{34} = 1$

The fitted values constructed from these estimates are the first entries of each cell in Table 3. For example, consider the cell with t=2, d=3 and e=1, which corresponds to urban women with 10-14 years of marriage duration and no education. The fitted value, $\hat{m}_{231} = 3.92$, is obtained by calculating

$$\hat{m}_{231} = \hat{\gamma}\,\hat{\gamma}_{12}\,\hat{\gamma}_{23}\,\hat{\gamma}_{31} = (5.479)\,(0.963)\,(0.545)\,(1.363) = 3.92\ .$$

The equation (2.1) describes a *multiplicative model* for the fitted values, multiplicative since the estimates on the right hand side are multiplied together. Of course there are many other ways of estimating the cell means from a smaller set of underlying parameters; for example, one might use the *linear model*

$$\hat{m}_{tde} = \hat{\lambda} + \hat{\lambda}_{1t} + \hat{\lambda}_{2d} + \hat{\lambda}_{3e} \quad . \tag{2.2}$$

The $\hat{\lambda}$'s have a similar role to the $\hat{\gamma}$'s in (2.1). This model has the same number of $\hat{\lambda}$'s as the $\hat{\gamma}$'s in (2.1), but the values of the $\hat{\lambda}$'s and the fitted values obtained from them are quite different.

There are two criteria for deciding between models for constructing fitted values. The population cell means should follow approximately the same pattern, and the fitted values should be close to the observed sample means. In this example, let m_{tde} be the population cell mean for the cell with $T = t$, $D = d$ and $E = e$. If these means can be expressed in the multiplicative form

$$m_{tde} = \gamma \ \gamma_{1t} \ \gamma_{2d} \ \gamma_{3e} \quad \text{for all } t, d, \text{ and } e, \tag{2.3}$$

then the fitting equation (2.1) is appropriate; if the population means can be expressed in the linear form, then (2.2) is appropriate. Of course in practice we do not know the population means and do not expect them to follow patterns such as (2.3) exactly; (2.3) is a working hypothesis, which is tested by comparing the fitted values with the sample means. However we shall indicate later that the multiplicative model is theoretically preferable to the linear model for these variables. Also on a more practical level we shall show that the multiplicative model provides a better fit to the data.

Inspection of Table 3 suggests that the fitted values from the multiplicative model succeed very well in reproducing the sample means, that is, the model is indeed a good fit. Note that the correspondence between the fitted values and the observed means is particularly good for cells with a large sample size, where the sample means have a small variance. For example, the cell with t=3, d=4, e=1, with 114 observations, has observed mean 5.06 and fitted value 5.20. For cells with small sample sizes the fitted value can deviate more from the observed value. This is clearly a sensible property for a good model. Indeed a chi-squared test* gives a value of 70.65 on 59 degrees of freedom, indicating that the model fits the data.

The degrees of freedom mentioned here represent the number of parameters saved by fitting the model rather than estimating each cell mean individually**. This reduction in the number of quantities estimated by the data reduces the sampling variation of estimates, and so the fitted values are more stable than the observed means. Also fitted values are provided for the empty cells. These properties imply that the fitted values are useful input for estimating the means of one factor over a standard distribution of the other factors, using the technique of standardization. This idea is taken up in Section 5.

Although the fitted values are useful, perhaps a more compelling reason for fitting a model is that the structure of the model clarifies important relationships between the variables of the cross-classification. All the models considered in this paper have such an associated substantive interpretation.

Consider the present model, (2.1):

$$\hat{m}_{tde} = \hat{\gamma} \; \hat{\gamma}_{1t} \; \hat{\gamma}_{2d} \; \hat{\gamma}_{3e} \quad \text{for all } t, d \text{ and } e.$$

This has the following property: the ratio of the estimated parities for two levels of one factor, with the levels of the other two factors fixed, is the same for all levels of the other two factors. For example, if e and e' are two educational levels, then from (2.1)

$$\hat{m}_{tde}/\hat{m}_{tde'} = (\hat{\gamma}\cdot \hat{\gamma}_{1t} \cdot\hat{\gamma}_{2d} \cdot\hat{\gamma}_{3e})/ (\hat{\gamma}\cdot\hat{\gamma}_{1t} \cdot\hat{\gamma}_{2d} \cdot\hat{\gamma}_{3e'}) = \hat{\gamma}_{3e}/ \hat{\gamma}_{3e'}$$

and this ratio does not depend on t and d. Thus percentage differences in mean parity between education groups, with Marital Duration and Type of Place controlled, are the same for all levels of Marital Duration and Type of Place. Percentage differences between levels of T and D have a similar interpretation.

---

* This test is like the more familiar chi-square test for independence in a two-way table. It is discussed in Section 4.

** The degrees of freedom are calculated as 59 = 70 - 11, where 70 represents the number of non-empty cells and 11 represents the number of distinct parameters in the model. The latter is three less than the original number because only 11 parameters are required to completely determine the fitted values in (2.1). This point is discussed further in Chapter 4.

In general, if differentials in response according to a factor *A* (measured on some scale) are the same for all levels of another factor *B*, then the effects of *A* and *B* are said to be *additive*. If *A* and *B* are not additive then they are said to *interact* in their effects on the response. Thus the model (2.1) has the interpretation that the effects of *T*, *D* and *E* are additive when differences are measured on a percentage scale.

We can proceed to describe the ratios individually. Consider for example the effects of education, represented by the set of ratios $\{\hat{\gamma}_{3e}/\hat{\gamma}_{3e'}\}$ for different values of e and e'. Since $\hat{\gamma}_{34}$ is set equal to one, we have

$$\hat{\gamma}_{31} = \hat{\gamma}_{31}/\hat{\gamma}_{34} = \hat{m}_{td1}/\hat{m}_{td4} \quad ;$$

that is, $\hat{\gamma}_{31} = 1.36$ estimates the ratio of mean parities between E=1 and E=4, and similarly $\hat{\gamma}_{32} = 1.40$ estimates the ratio of mean parities between E=2 and E=4 and $\gamma_{33} = 1.23$ estimates ratio of mean parities between E=3 and E=4. Hence we can say that after controlling for Type of Place and Marital Duration, women with no education (E=1) have an estimated 36 per cent more children than women with secondary or higher education (E=4) and women with upper primary education (E=3) have an estimated 23 per cent more children than women with secondary or higher education. Other comparisons can be made by taking the ratio of the appropriate γ's. For example to compare E=1 and E=2,

$\hat{\gamma}_{31}/\hat{\gamma}_{32} = \hat{m}_{td1}/\hat{m}_{td2} = 1.363/1.395 = 0.98$, so that women with no education have an estimated 2 per cent fewer children than women with lower primary education with the same Type of Place and Marital Duration.

This last difference is small; indeed in Section 5 we shall indicate that it is not statistically significant. However, the percentage differences between other levels of Education are large and significant. Hence the answer to the question which led us to consider Table 3 is that the relationship between Education and mean parity is not merely reflecting the joint compositional effects of Type of Place and Marital Duration.

In future sections we shall discuss the selection of this particular model and extend further its interpretation. However, the next step is to describe the family of models considered in this paper. The family is general enough to cover most of the tables obtained from WFS Surveys; the substantive meaning of the models may be clarified by reference to the particular examples given here and in later sections.

3. GENERALIZED LINEAR MODELS

In the previous section we discussed a particular procedure for replacing sample means in a cross-tabulation by fitted values calculated from a small set of underlying quantities. Any cross-tabulation of means is a potential candidate for this procedure, but the extent to which the analysis is useful depends upon the underlying structure of the population means and the sampling properties of the data. The statistical models presented in this chapter make explicit assumptions about these aspects of the data. Specifically they consist of two components, an assumption about the structure of the population cell means, represented by a formula such as (2.3), and an assumption about the distribution of the sample cell means about the population cell means. These two parts of the model are called the *systematic component* and the *error structure* respectively, and they are conceptually quite distinct.

The purpose of a statistical model is to define precisely conditions under which the associated analysis is the best possible. Prior knowledge of the data can be compared with the assumptions of the model to decide whether the analysis is worthwhile. Note that for the type of data collected in WFS surveys, where the theoretical structure between the variables is not subject to exact physical laws, the model is always a simplification of the real world and can only be said to be followed approximately. However some models are better than others, and a good model leads to good data analysis.

We describe a large class of models for cross-tabulated means and proportions, taken from the system developed by J.A. Nelder and others at Rothampstead Experimental Station. The statistical methodology is discussed in Nelder and Wedderburn (1972)*, and a computer package called GLIM (Generalized Linear Interactive Modelling) has been developed to fit the models**. Our purpose here is to describe the basic elements with a minimum of technical detail.

---

* Another useful reference here is Nelder (1974).

** For a discussion of available programs for computing models, the reader is referred to Appendix 2. GLIM is obtainable from the Numerical Algorithms Group, 13 Banbury Road, Oxford OX2 6NN, U.K.

3.1 LINEAR MODELS

Until recently nearly all model-based statistical analysis was based on the class of *linear models with Normal (Gaussian) error*. This is defined as follows. Suppose that $\bar{y}_c$ is the sample mean of Y in a typical cell of the table, which we denote by the subscript c. Let the mean value of Y for the population in cell c be m_c; in other words, m_c is the expected value of $\bar{y}_c$. (It is important for later developments to keep clear of the distinction between $\bar{y}_c$, which is observed, and m_c which is not observed). The model consists of two parts:

(i) *a systematic component*, which expresses the value of m_c for each cell as a sum of unknown parameters.

(ii) an *error structure*, which characterizes the distribution of $\bar{y}_c$ about its expected value m_c. The *Normal error structure* states that $\bar{y}_c$ is normal with mean m_c and variance $\sigma^2 k_c/n_c$, where

σ^2 is an unknown positive constant,

k_c is a known multiplier, which may take different values for different cells, and

n_c is the sample size for cell c.

3.1.1 The Systematic Component

Consider the data in Table 3. Here $\bar{y}_{tde}$ is the mean parity for the cell with $T=t$, $D=d$ and $E=e$, and $m_{tde} = E(\bar{y}_{tde})$. Examples of systematic components are

$$m_{tde} = \lambda \quad \text{for all } t, d \text{ and } e; \tag{3.1}$$

$$m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} \quad \text{for all } t, d \text{ and } e; \tag{3.2}$$

$$m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{12td} \quad \text{for all } t, d \text{ and } e; \tag{3.3}$$

$$m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} + \lambda_{12td} + \lambda_{13te} + \lambda_{23de} \quad \text{for all } t, d \text{ and } e. \tag{3.4}$$

The terms on the right hand side are called *parameters*, or *effects*, since they characterize the way in which the mean parity of the population varies for different levels of the factors T, D and E. Thus model (3.1) implies that m_{tde} is constant for all cells, that is, the mean parity is the same for all values of T, D and E. This model is unrealistic since marital duration clearly has an effect on parity. The second model (3.2) states that T, D and E all affect the mean parity, and that these effects are *additive* on a *linear* scale, that is, differences in mean parity according to one factor are the same for all joint levels of the other two factors. Consider, for example, differences according to education; if e and e' are two levels of education, and $T=t$ and $D=d$ are fixed levels of the other factors, then $m_{tde} - m_{tde'}$ represents the difference in mean parity between education groups e and e'. From (3.2), we have

$$m_{tde} - m_{tde'} = (\lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e}) - (\lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e'}) = \lambda_{3e} - \lambda_{3e'}$$

and so this difference is the same for all values of t and d.

This linear additive model forms the basis of a multiple classification analysis of the table. However, the interpretation indicates that it is not appropriate, since we should <u>not</u> expect differences in mean parity according to T or E to be constant for all levels of D: since parity is a cumulative response we expect these differences to increase with D. We shall see later that the data do not fit this linear additive model, confirming our suspicions. Also the linear additive model should be contrasted with the multiplicative model given in Table 3 and discussed in the previous section, where differentials are measured by ratios or percentage differences. The multiplicative model contains the same number of parameters, but appears more realistic since the assumption of additivity when differentials are measured as ratios or percentage differences seems more reasonable. We shall return to this model later.

The model (3.3) omits all effects involving the education subscript e, and hence if e and e' are two levels of education, then $m_{tde} = m_{tde'}$ for all t and d. That is, Education has no effect on mean parity, after controlling for Marital Duration and Type of Place. The inclusion of effects $\{ \lambda_{12td} \}$ involving the subscripts t and d imply that T and D interact in their effect on P, that is, differentials according to T change for different values of D. In symbols,

$$\begin{aligned} m_{tde} - m_{t'de} &= (\lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{12td}) - (\lambda + \lambda_{1t'} + \lambda_{2d} + \lambda_{12t'd}) = \\ &= \lambda_{1t} - \lambda_{1t'} + \lambda_{12td} - \lambda_{12t'd} , \end{aligned}$$

and this differential depends on the value of d. Finally (3.4) includes the effects of T, D and E and all the interactions between pairs of variables.

We shall discuss the choice of systematic component more comprehensively in the next section. However, these examples show how each systematic component involves a hypothesis about how the mean response for the population varies from cell to cell. The error structure, on the other hand, provides the link between this theoretical property and the data by specifying the distribution of the sample means $\{ \bar{y}_c \}$ about their expected values $\{ m_c \}$. We now consider the normal error structure in more detail.

3.1.2 The Normal Error Structure

The variance of the mean for cell c, var $\bar{y}_c = \sigma^2 k_c/n_c$, contains a multiplier k_c which needs to be chosen by the analyst. To motivate this choice, suppose that Y_c is a typical individual parity in cell c. Then if $\bar{y}_c$ is considered the mean of a random sample of n_c values of Y_c, we know from elementary statistics that var $(\bar{y}_c)$ = var $(Y_c)/n_c$. Hence the variance of the mean given above corresponds to a within-cell variance

$$\text{var } Y_c = k_c \, \sigma^2. \qquad (3.5)$$

The standard assumption is that this within-cell variance is equal for all cells (*homoscedasticity*), and from (3.5) this corresponds to setting

$k_c = 1$ for all cells c. However, in Table 3, this assumption is untenable, since without reference to data it is clear that the variance of Y (parity) increases with marital duration. If within-cell sample variances s_c^2 are available, then these can be used to derive estimates of k_c and σ^2, as in Appendix 1; if not, then a reasonable choice seems to be set $k_{tde} = d$, which corresponds to assuming that the variance of Y is proportional to D. This choice of multipliers is used in the linear models fitted in the next section.

In general, the choice of multipliers is not as critical as the choice of model, but it is sensible to make some allowance for obvious departures from the assumption of equal within cell variances.

Linear models with Normal error form the basis of the analysis of variance of cross-classified data. However, a unique decomposition of the variance is only possible for balanced tables, that is tables with an equal sample size in each cell. In observational data from surveys this is rarely the case. The emphasis for unbalanced data is rather on fitting and comparing a set of models and interpreting the model with the best fit.

The fitting process involves estimating the parameters of the model so that the fitted values $\{ \hat{m}_c \}$ found by substituting the parameter estimates into the systematic component are in some sense as close as possible to the means $\{ \bar{y}_c \}$ (See section 4). For linear models, these fitted values can be negative, and this is undesirable for data where the response is inherently non-negative, such as Parity. Similarly, if linear models are fitted to tables of proportions then the fitted proportions $\hat{p}$ can lie outside the range zero to one. These deficiencies lead to the search for suitable generalizations of linear models which give fitted values in the desired range.

3.2 GENERALIZED LINEAR MODELS

The models considered here are more general than the usual linear models in two respects: i) in forming the systematic component, we express a function of the population mean m_c, called the *link function*, as a sum of unknown parameters; ii) in choosing the error structure we consider other distributions apart from the Normal which are more appropriate for mean counts and proportions. We consider each of these elements in turn.

3.2.1 The Systematic Component

The systematic component of a generalized linear model is formed by choosing a link function g and setting $g(m_c)$, the function of the population mean is all c, equal to a sum of parameters *.

Three choices of link function are used here:

(i) The identity function, $g(m_c) = m_c$, which leads to linear models, as before.

(ii) The logarithmic function, $g(m_c) = \log m_c$, $m_c > 0$ which leads to log-linear models.

(iii) The logit (or log-odds) function
$g(m_c) = \text{logit } m_c = \log \{m_c/(1 - m_c)\}$, $\quad 0 < m_c < 1$,
which leads to logit-linear models.

Example:

Consider for example a log-linear additive model for Table 3 with systematic component

$$\log m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} \quad \text{for all } t, d \text{ and } e.$$

The base of the logarithm is arbitrary: we shall take (natural) logarithms to base e. If we define a new set of parameters $\gamma = e^{\lambda}$, $\gamma_{1t} = e^{\lambda 1t}$ $\gamma_{2d} = {}^{\lambda 2d}$, $\gamma_{3e} = e^{\lambda 3e}$, then using the properties of logarithms,

$$\begin{aligned}\log m_{tde} &= \log \gamma + \log \gamma_{1t} + \log \gamma_{2d} + \log \gamma_{3e} \\ &= \log (\gamma \cdot \gamma_{1t} \cdot \gamma_{2d} \cdot \gamma_{3e}).\end{aligned}$$

Exponentiating, we obtain

$$m_{tde} = \gamma \cdot \gamma_{1t} \cdot \gamma_{2d} \cdot \gamma_{3e} ,$$

which is precisely the multiplicative model (2.3) in chapter 2. Hence that model is an example of a log-linear additive model. Note that for any log-linear model the fitted values $\hat{m}_{tde}$ are exponents of linear sums, and hence are always positive.

---

* It is important to note that this function is defined on the population means m_c and not the sample means y_c. Hence it is not merely a preliminary transformation of the data.

Just as fitted values of log-linear models are always positive, it is easy to see that fitted values of logit-linear models always lie between zero and one. For if $\hat{p}_c$ is the fitted value for cell c, then

$$\log \{\hat{p}_c/(1-\hat{p}_c)\} = \hat{s},$$

where $\hat{s}$ is a sum of estimated effects. Hence $\hat{p}_c(1-\hat{p}_c) = e^{\hat{s}}$, and so $\hat{p}_c = e^{\hat{s}}/(1 + e^{\hat{s}})$, which always lies between zero and one.

This property suggests that logit-linear models are particularly appropriate for cross-classifications of proportions, that is tables where the response Y, is dichotomous. Then p_c is the expected proportion with $Y = 1$, $p_c/(1-p_c)$ is (in betting language) the odds in favour of $Y=1$, and so $\log \{p_c/(1-p_c)\}$ is the log-odds in favour of $Y=1$. In section 6 we shall fit logit-linear models to the tables of proportions given in Table 2*.

3.2.2 Error Structures Other than Normal

Finally, we consider the error structure for these models. For various reasons, the normal error structure is not the most satisfactory for log-linear and logit-linear models. Two additional error distributions will be considered here:

(i) *Poisson Error*: Assume that $n_c\bar{y}_c$, the total for cell c, has a Poisson distribution with mean $n_c m_c$.

(ii) *Binomial Error* (for tables of proportions): Assume that $n_c\bar{y}_c$, the number of respondents with $Y=1$ in cell c, is Binomial with index n_c and mean $n_c p_c$.

The Poisson distribution is particularly appropriate for response which are accumulated counts (such as parity). The distribution is usually associated with log-linear models. Since the variance of a Poisson distribution equals the mean, we have:

$$\text{var } (n_c\bar{y}_c) = n_c^2 \text{ var}(\bar{y}_c) = n_c m_c,$$

---

* Logit-linear models for proportions are related to log-linear models for contingency tables. As noted in the introduction, a table of proportions can be expressed as a contingency table. Logit-linear models are equivalent to certain log-linear models for this table. See Goodman (1970).

So that the variance of $\bar{y}_c$ is proportional to the mean and inversely proportional to the sample size, that is,

$$\text{var } \bar{y}_c = m_c/n_c.$$

The Binomial distribution is usually appropriate for dichotomous responses, and is associated with the logit-linear model. The variance of the sample proportion is $p_c(1-p_c)/n_c$.

3.3 SUMMARY

A generalized linear model* is characterized by three components

(a) The link function, $g(m)$ (identity, log or logit)

(b) The systematic component (the sum of parameters included in the model)

(c) The error structure (Normal, Poisson or Binomial)

The most common combinations of (a) and (c) are linear Normal, log-linear Poisson and logit-linear Binomial. Thus in what follows we shall identify models by the link function and assume the corresponding error structure, unless the error structure differs from the above.

Logit-linear models are appropriate for cross-classified proportions.

Log-linear models are often appropriate when the response is non-negative, and differentials are to be interpreted on a ratio or percentage scale.

The strategy of analysis is to choose the link function and error structure and then to fit a set of systematic components and choose the model with the best fit. The choice of systematic component, the fitting procedure and an index of goodness-of-fit are discussed in the next section.

---

* As discussed here. The system developed by Nelder includes other link functions and error structures.

4. HIERARCHICAL MODELS: FITTING AND SELECTION

In the previous section, we gave some examples of systematic components for a three-way table (equations (3.1)-(3.4)). In this section, we describe sets of systematic components for two-way and three-way tables; and criteria for choosing between them; the extension to higher way tables involves no new ideas.

4.1 THE SATURATED MODEL

First consider a two-way table with factors T and D, such that $\bar{y}_{td}$ is the sample mean for the cell with $T = t$, $D = d$, and $E(\bar{y}_{td}) = m_{td}$ is the population mean. As an example we shall consider the table derived from Table 3 by summing over the levels of education, so that $Y = P$ = Parity, T = Type of Place and D = Marital Duration. Consider the model with systematic component

$$g(m_{td}) = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{12td} \quad \text{for all } t, d, \tag{4.1}$$

where g is the link function, λ is the grand mean, $\{\lambda_{1t} : t = 1 \text{ to } 3\}$ and $\{\lambda_{2d} : d = 1 \text{ to } 6\}$ are called the *one-factor effects*, or the *main effects*, of T and D respectively on P, and $\{\lambda_{12td}, t = 1 \text{ to } 3, d = 1 \text{ to } 6\}$ are called the *two-factor effects*, or the *interactions* of T and D on P.

Not that (4.1) is a set of 3 x 6 = 18 equations, relating means $\{ m_{td} \}$ to 1 + 3 + 6 + 3 x 6 = 28 parameters $\{\lambda, \lambda_{1t}, \lambda_{2d}, \lambda_{12td}\}$. Hence certain restrictions are required to define the parameters uniquely; these are discussed below. The restrictions reduce the number of distinct parameters to 18. This model is called *saturated*, because it does not reduce the number of quantities to be estimated, and is really no more than a re-expression of the population means*.

4.2 UNSATURATED HIERARCHICAL MODELS

Unsaturated models are formed from (4.1) by setting parameters on the right hand side equal to zero: we say that these parameters are excluded from the model. Two rules are observed in this process.

---

* One could set $\lambda = \lambda_{1t} = \lambda_{2d} = 0$ and $\lambda_{12td} = g(m_{td})$; then the model is clearly vacuous. However, other constraints are more illuminating.

(1) The main or higher order effects of a factor, or set of factors, are either <u>all</u> set to zero or <u>all</u> not set to zero. For example, the parameters $\{\lambda_{11}, \lambda_{12}, \lambda_{13}\}$ are the main effects of T. In any model either λ_{11}, λ_{12} and λ_{13} are all assumed non-zero (included in the model), or all assumed equal to zero (excluded from the model).

(2) If the k-factor effects between k factors $A \times B \times \ldots \times K$ are included in the model, then all 1, 2, ..., $(k-1)$-factor effects between subsets of the factors $\{A, B, \ldots, K\}$ must also be included in the model. For example,

$$g(m_{td}) = m + m_{1t} + m_{12td}$$

is not allowed because the two-factor effects of $T \times D$ are included in the model but the main effects of D are not included in the model.

Models subject to these restrictions are called hierarchical*. There are five hierarchical models for the two-way table, including the saturated model (4.1),

$$(TD) \quad : \quad g(m_{td}) = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{12td}; \text{ for all } t, d; \tag{4.1}$$

$$(T,D) \quad : \quad g(m_{td}) = \lambda + \lambda_{1t} + \lambda_{2d} \quad \text{for all } t, d; \tag{4.2}$$

$$(T) \quad : \quad g(m_{td}) = \lambda + \lambda_{1t} \quad \text{for all } t, d; \tag{4.3}$$

$$(D) \quad : \quad g(m_{td}) = \lambda + \lambda_{2d} \quad \text{for all } t, d; \tag{4.4}$$

$$(\emptyset) \quad : \quad g(m_{td}) = \lambda \quad \text{for all } t, d; \tag{4.5}$$

The models (4.1)-(4.5) are labelled on the left to indicate which effects are included in the model. Thus (4.5) is called the *null model* since no effects of T or D are included. This model is labelled ($\emptyset$). Model (4.2) is called the *additive* model for T and D, and labelled (T,D). The models (4.3) and (4.4) imply that one of the factors has no effect on the response, and are labelled (T) or (D) accordingly. Finally, the saturated model is labelled (TD), since it includes the two-factor effect of T and D. The main effects of T and D are included because the model is hierarchical**.

---

* Non-hierarchical models are sometimes useful in particular situations, but are not considered here.

** The notation is similar to that used by Goodman (1970) to describe models for contingency tables.

For a three-way table with three factors, T, D and E, say, the saturated model becomes

$$(TDE) : g(m_{tde}) = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} + \lambda_{12td} + \lambda_{13te} + \lambda_{23de} + \lambda_{123tde}, \quad (4.6)$$

where, in addition to one-factor and two-factor effects, we have three-factor effects of T, D and E $\{\lambda_{123tde} : t = 1 \text{ to } 3, d = 1 \text{ to } 6, e = 1 \text{ to } 4\}$. By setting parameters in (4.6) equal to zero, we obtain 18 unsaturated hierarchical models for the three-way table. These are listed in the first column of Table 4. The notation is the obvious extension of that for two-way tables. For example, the linear models (3.1), (3.2), (3.3) and (3.4) of Section 3 are labelled (TDE), (T,D,E), (TD) and (TD, TE, DE) respectively; the model (TD, TE, DE) includes all one-factor and two-factor effects but excludes the three factor effects of T, D and E.

4.3 MODEL FITTING

4.3.1 Linear Models

Fitting a model involves estimating the effects included in the systematic component. The estimation procedure depends on the link function and the error structure. For linear models with normal error, the parameters are estimated so that if $\hat{m}_c$ is the resulting fitted value for cell c, then the criterion

$$S = \sum_c n_c (\hat{m}_c - \bar{y}_c)^2/k_c,$$

is minimized, where the summation is over the cells of the table, $\bar{y}_c$ and n_c are the sample mean and sample size respectively for cell c, and k_c is the multiplier in the error variance. This estimation procedure is called *weighted least squares*, since S is a weighted sum of the squares of the deviations of the observed sample means from the fitted values, with weights n_c/k_c. Note that empty cells ($n_c = 0$) are given weight zero, so any value of $\bar{y}_c$ can be substituted in such cells without affecting the estimates of the parameters. The minimum value of S is called the *residual sum of squares* or the *deviance*, and measures the fit of the model.

TABLE 4 Hierarchical Linear and Log-Linear Models Fitted to Data in Table 3

Model[3]	df	LINEAR MODELS[1] deviance	LINEAR MODELS[1] mean deviance	LOG-LINEAR MODELS[2] deviance	LOG-LINEAR MODELS[2] mean deviance
Ø	69	4906	71.1	3,732	54.1
T	67	4835	72.2	3,659	54.6
D	64	189.1	2.95	165.8	2.59
E	66	3839	58.2	2661	40.3
T, D	62	162.8	2.63	120.7	1.95
T, E	64	3821	59.7	2647	41.4
D, E	61	150.2	2.46	100.0	1.64
T, D, E	59	135.8	2.30	70.65	1.20
TD	52	121.3	2.33	108.8	2.09
TE	58	3812	65.7	2,626	45.3
DE	46	98.95	2.15	84.46	1.84
TD, E	49	89.38	1.82	57.06	1.16
TE, D	53	131.1	2.47	59.89	1.13
DE, T	44	84.49	1.92	54.91	1.25
TD, TE	43	78.93	1.84	44.27	1.03
TD, DE	34	49.02	1.44	42.72	1.26
TE, DE	38	78.30	2.06	44.60	1.17
TD, TE, DE	28	39.68	1.42	30.95	1.11
TDE	0	0.0	-	0.0	-

(1) Linear Models with Normal Error, $k_{tde} = d$ for all t, d, e.
(2) Log-Linear Models with Poisson Error.
(3) For notation, see text.

For example, to fit the linear additive model (T,D,E) to the three-way table of mean parities, we choose parameter estimates

$$\{\hat{\lambda}, \hat{\lambda}_{1t}, \hat{\lambda}_{2d}, \lambda_{3e} : 1 \le t \le 3,\ 1 \le d \le 6,\ 1 \le e \le 4\}$$

so that $\sum_{\text{cells}} n_{tde}\,(\hat{m}_{tde} - \bar{y}_{tde})^2/k_{tde}$ is minimized, where

$$\hat{m}_{tde} = \hat{\lambda} + \hat{\lambda}_{1t} + \hat{\lambda}_{2d} + \hat{\lambda}_{3e} \quad \text{for all } t, d \text{ and } e.$$

As presently defined, this fitting procedure does not uniquely determine the parameter estimates, because different sets of parameters can lead to the same fitted values and hence to the same value of S. For example in (T, D, E) we can replace

$$\hat{\lambda}^1 = \hat{\lambda} + c,\ \hat{\lambda}^1_{1t} = \hat{\lambda}_{1t},\ \hat{\lambda}^1_{2d} = \hat{\lambda}_{2d} + c,\ \hat{\lambda}^1_{3e} = \hat{\lambda}_{3e} - 2c$$

for any constant c, and then $\hat{m}_{tde} = \hat{\lambda} + \hat{\lambda}_{1t} + \hat{\lambda}_{2d} + \hat{\lambda}_{3e} = \hat{\lambda}^1 + \hat{\lambda}^1_{1t} + \hat{\lambda}^1_{2d} + \hat{\lambda}^1_{3e}$

Hence we impose restrictions, or constraints, on the parameters so that each set of fitted values corresponds to a unique set of parameters. One set of restrictions is to set to zero all parameters where one or more of the subscripts t, d or e takes on its maximum value (3, 6 or 4 respectively). Thus for (T, D, E) we set $\lambda_{13} = \lambda_{26} = \lambda_{34} = 0$, and for (TD) we set

$$\lambda_{13} = \lambda_{26} = \lambda_{123d} = \lambda_{12t6} = 0 \text{ for all } d \text{ and } t,$$

a total of 10 restrictions. The choice of restrictions affects the interpretation of individual effects (discussed in section 5), but it does not affect the fitted values or the deviance*.

---

* Restrictions are also required for analysis of variance models for balanced classifications. There it is customary to require that the sum of the parameters over one, or more, of the subscripts is zero. However, there is no agreed way of defining the constraints for classifications with unequal cell sizes, as here.
Similar restrictions are required for log-linear and logit-linear models; for example, for the log-linear model (T,D,E) we may impose $\lambda_{13}=\lambda_{26}=\lambda_{34}=0$. This is equivalent to the constraints $\gamma_{13}=\gamma_{26}=\gamma_{34}=1$ used in defining the estimates in Table 3, since $\lambda_{jk}= \log \gamma_{jk}$ and $\log 1 = 0$.

4.3.2 Log-linear and Logit-linear Models

For log-linear and logit-linear models, the fitting procedure is analogous to weighted least squares, but it involves minimizing other functions of the observed and fitted means. The minimization usually involves an iterative computation, and the resulting minimum value is again called the deviance and measures of the fit of the model. For further details the reader is referred to Nelder and Wedderburn (1972).

Example

All the hierarchical linear and log-linear models for the three-way cross-classification of mean parities (Table 3) were fitted using the computer program GLIM. The deviances are given in columns 3 and 5 of Table 4.

4.4 MODEL SELECTION USING THE MEAN DEVIANCE

It is not possible to compare quantitively the fit of a log-linear model with that of a linear model, but different hierarchical models with the same link function can be compared via the deviances.

However, the deviances cannot be compared directly, since the saturated model always has the smallest deviance, viz., zero, and as successive parameters in the model are set to zero the deviance increases. Hence, if models are chosen according to the smallest deviance then the saturated model always wins. The point is that the introduction of parameters into a model always improves the fit, but since we also prefer models with few parameters, the assessment should take into account the number of parameters in the model. Accordingly the *degrees of freedom* (df) are calculated as

$$\text{df} = \text{number of non-empty cells} - \text{number of parameters} + \text{number of constraints},$$

and the *mean deviance* (the *residual mean square* for linear models) is calculated as the deviance divided by the degrees of freedom.

For example, in Table 3 there are 3 x 6 x 4 = 72 cells and two of these are empty, leaving 70 non-empty cells. For (T,D,E) there are 14 paremeters and 3 constraints, giving df = 70 - 14 + 3 = 59. For (TD) there are 1 + 3 + 6 + 3 x 6 = 28 parameters and 10 constraints, giving df = 70 - 28 + 10 = 52. The degrees of freedom for the models in Table 4 appear in the second column of that table, and the mean deviances for linear and log-linear models appear in columns 4 and 6, respectively.

The model with the best fit, as measured by the lowest mean deviance, is (TD, TE, DE) for both the linear and log-linear cases. However, in general, we prefer simpler models if the increase in mean deviance is slight. Thus the "best" linear model appears to be (TD, DE) and the "best" log-linear model appears to be (T, D, E). Note that models which do not include the effect of marital duration all have a high mean deviance, reflecting the fact that the exclusion of factors which are clearly related to the response leads to models with a poor fit. Also the saturated model (TDE) has no mean deviance, since the deviance and degrees of freedom are both zero. The largest model with a mean deviance is (TD, DE, TE), which assumes that the three factor effects are zero.

4.4.1 Nested Models

For a more formal comparison of mean deviances we need the following definition. Two hierarchical models (m) and (m') are *nested* if (m') is obtained from (m) by setting some non-zero effects in (m) equal to zero. For example, (TD, TE) and (TD, E) are nested since (TD, E) is obtained from (TD, TE) by setting the terms λ_{13te} equal to zero. On the other hand (TD, TE) and (T, DE) are not nested, although the latter model has fewer parameters.

4.4.2 F-Tests of Relative Fit

A formal statistical test can be used to compare the relative fit of two nested models, (m) and (m'). Suppose (m') is obtained from (m) by setting parameters $(\lambda_1,\ldots.,\lambda_r)$ equal to zero. Let de_m, df_m and $de_{m'}$, $df_{m'}$ be the deviances and degrees of freedom of (m) and (m'), respectively. It is always true that $de_m < de_{m'}$ and $df_m < df_{m'}$. If (m) is in fact a true model then (m') is also true if $\lambda_1 = \ldots = \lambda_r = 0$. This hypothesis is tested by comparing the statistic

$$F = \frac{(de_{m'} - de_m)/(df_{m'} - df_m)}{de_m/df_m}$$

with an F distribution with $(df_{m'} - df_m)$ degrees of freedom in the numerator and df_m degrees of freedom in the denominator. Significantly large values of F in a one-tailed F-test suggest that (m) fits better than (m')*.

For example in Table 4 the model (T, D, E) is obtained from (TD, DE) by setting two-factor effects $\{\lambda_{12td}, \lambda_{23de}\}$ equal to zero. To test if (TD, DE) fits better than (T, D, E) we obtain the following F statistics:

(a) linear models: $F_{59-34,\ 34} = \dfrac{(135.8-49.0)/(59-34)}{49.0/34} = 2.39$

(b) log-linear models: $F_{25,\ 34} = \dfrac{(70.65-42.72)/25}{42.72/34} = 0.89$

The tabulated 5 per cent of the $F_{25,34}$ distribution is 1.83. Hence linear (TD, ED) fits significantly better than linear (T, D, E), but log-linear (TD, ED) does not fit better than log-linear (T, D, E). This reflects the fact that the log-linear additive model (T, D, E) makes better sense substantively than the linear additive model (T, D, E). In fact, the best linear model (TD, DE) implies that differences in mean parity by Education and Type of Place are not equal for all levels of Marital Duration, which is precisely the property that linear (T, D, E) does not reflect.

The reader may like to test whether linear (TD, TE, DE) fits significantly better than linear (TD, DE) ($F_{6,28} = 1.10$) and whether log-linear (TD, TE, DE) fits significantly better than log-linear (T, D, E) ($F_{31,28} = 1.16$). In both cases the answer is negative, that is, linear (TD, TE, DE) and log-linear (T, D, E) have the best relative fits.

---

* This test corresponds to the F-test in analysis of variance. Strictly speaking, the test involves an assumption of simple random sampling of individuals which is not valid for the stratified cluster sampling employed in WFS surveys. The effect of this assumption is unknown and the subject of current statistical research.

The fact that the log-linear model with the best relative fit is simpler than the best linear model is a strong reason for preferring log-linear models for this reponse, in addition to the fact that the response is non-negative. In fact, an experienced analyst would probably not fit linear models for this response; we make use of them here for illustrative purposes.

4.4.3 Chi-Squared Tests of Absolute Fit for Poisson and Binomial Error Structures

So far, we have considered the relative fit of different models, that is, we have compared the fit of all models with the largest model for which the fit can be measured, that is (TD,TE,DE). If this model is wrong, that is there are significant three-factor effects, then the F-tests no longer apply (although the comparison of mean deviances still has some value). The question remains whether absolute measures of fit can be derived. Here there lies an important distinction between Normal error and Poisson or Binomial, error structures.

Recall that for Normal error the variance of Y in each cell is assumed a known multiple of a constant σ^2, which fixes the level of unexplained random variation. When a linear model is fitted the mean deviance estimates σ^2. If the model is wrong, the mean deviance is inflated by significant effects which are assumed zero in the model, leading to an overestimate of the true variance. For this reason an absolute measure of fit requires an independent estimate of σ^2. This can be formed from the within-cell sample variances if they are available. (See Appendix 1).

For Poisson and Binomial Error, var $(\bar{y})$ is a known function of the mean and there is no independent parameter σ^2 associated with the variance. For this reason, the absolute fit for log-linear and logit-linear models can be assessed directly via chi-squared test on the deviances. If de_m, df_m are the deviance and degrees of freedom for model (m), then if model

(m) is true, de_m is approximately* distributed as chi-squared with df_m degrees of freedom, that is $de_m \sim X^2_{df_m}$. Significantly large values of chi-squared indicate that the model does not fit. For example, log-linear (T,D,E) in Table 4 gives a chi-squared value of 70.65 on 59 degrees of freedom, which is not significant at the 5 per cent level, thus indicating a satisfactory fit.

4.4.4 Chi-Squared Tests of Relative Fit for Poisson and Binomial Error Structures

The chi-squared test also provides an alternative to the F-test for comparing nested log-linear or logit-linear models: If (m') is obtained from (m) by setting parameters to zero, and the model (m') is true, then $de_{m'}-de_m$ is approximately chi-squared with $df_{m'}-df_m$ degrees of freedom, that is $de_{m'}- de_m \sim X^2_{df_{m'}- df_m}$ Significantly large values of chi-squared indicate that (m) fits significantly better than (m'). For example, a test for the effect of education on parity, controlling for T and D, may be achieved by comparing models (TD) and (TD,E). The chi-squared test for these log-linear models gives (from Table 4)

$$X^2_{52-49} = 70.65 - 57.06 = 13.59,$$

significant at the one per cent level. That is, (TD,E) fits significantly better than TD, so that an effect of education is established.

Thus we have two alternative methods for comparing log-linear and logit-linear models. The chi-squared test is more powerful than the F-test, particularly when df_m is small, but the F-test remains valid under somewhat weaker assumptions about the error structure**.

We have described fitting and comparing hierarchical models for three-way tables; the same procedures can be used to fit tables of four or more dimensions. However, the number of hierarchical models increases rapidly (there are 166 hierarchical models for 4-way tables). Consequently it becomes impractical to fit all models, and the possibility of more than

---

* In large samples. Also, simple random sampling is assumed (see previous footnote).

** Namely that $n_c\bar{y}_c$ is <u>proportional</u> to a Poisson (Binomial) variate, with unknown constant of proportionality.

one model giving satisfactory fit is stronger. some literature exists on automatic stepwise procedures for selecting models. However, in general, it is simpler to use substantive knowledge to limit the number of models to be fitted, or to reduce the dimensions of the table by disaggregation, as illustrated in Section 6.

4.5 SUMMARY

We have constructed a class hierarchical models for the systematic component of a generalized linear model. The fitting procedures for these models yield indices, the deviance and the mean deviance, for measuring the goodness-of-fit. The mean deviances, which correspond to residual mean squares for linear models, can be used to obtain F-tests to compare the relative fit of models. In addition for log-linear and logit-linear models, the deviances provide chi-squared tests of absolute and relative fit.

5. INTERPRETATION OF ESTIMATED EFFECTS

We have seen how models are formulated in terms of a set of parameters or effects, which we have denoted by the Greek symbol λ with suitable subscripts. When the models are fitted, estimates of these effects are calculated; also, in some computer programs (including GLIM) associated standard errors can be obtained. In this section, we discuss how to interpret these results.

5.1 LINEAR MODELS

The interpretation for a particular effect depends on the choice of link function, the systematic component and the set of constraints imposed to define the effects uniquely. Consider for example, the best linear model for the data of Table 3, that is linear (*TD*, *ED*). This has systematic component

$$m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} + \lambda_{12td} + \lambda_{23de} \quad \text{for all } t,d,e. \qquad (5.1)$$

Suppose, as before, that in order to define the effects uniquely, we set to zero all effects involving Rural Type of Place (T=3), Marital Duration over 25 years (D=6) or Secondary and Higher Education (E=4), that is, effects with the highest value of any subscript. This gives

$$\lambda_{13} = \lambda_{26} = \lambda_{34} = \lambda_{12t6} = \lambda_{123d} = \lambda_{23d4} = \lambda_{236e} = 0 \text{ for all } t,\ d \text{ and } e.$$

Then it is easily established from (5.1) that, in terms of the cell means,

$$\lambda_{1t} = m_{t6e} - m_{36e} \quad \text{for all } t \text{ and } e, \qquad (5.2)$$

$$\lambda_{2} = m_{3d4} - m_{364} \quad \text{for all } d, \qquad (5.3)$$

$$\lambda_{12td} = (m_{tde} - m_{3de}) - (m_{t6e} - m_{36e}) \text{ for all } t,\ d \text{ and } e \qquad (5.4)$$

$$\lambda_{23de} = (m_{tde} - m_{td4}) - (m_{t6e} - m_{t64}) \text{ for all } t,\ d \text{ and } e \qquad (5.5)$$

Hence, all the main effects can be identified as differences, and the two-factor effects as differences of differences, in the cell means. For example from (5.2), λ_1 measures the difference in parity between Type of Place t and Rural (T=t and T=3), for women with 25 or more years of marriage (D=6) and the same level of education (E=e). Note that this difference is specific to D=6 and is not assumed the same for other values of marital duration; this reflects the fact that the model (TD,ED) includes the two-factor effects of T and D. Indeed, from (5.4) we see that λ_{12td} measures the change in this difference between marriage duration level d, $(m_{tde} - m_{3de})$ and marital duration level 6 $(m_{t6e} - m_{36e})$.

In contrast, consider the model (T,DE) which includes the main effects of T, D and E and the two-factor effects of D and E. For this model, the systematic component is

$$m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} + \lambda_{23de} \quad \text{for all } t, d \text{ and } e,$$

and hence:

$$\lambda_{1t} = m_{tde} - m_{3de} \quad \text{for all } t, d \text{ and } e.$$

That is, the effect λ_{1t} is no longer specific to women married over 25 years, but represents the difference in mean parity between Type of Place t and Rural (T=t and T=3) for all levels of Education and Marital Duration, reflecting the absence of interaction effects involving Type of Place in the model. This generality and simplicity of interpretation is reserved for the main effects of variables which are *additive* in the model, that is which do not appear in two-factor or higher order effects in the model. For this reason, we shall limit the discussion here to the interpretation of additive variables. The interpretation of variables which appear non-additively is probably best achieved by considering the fitted means themselves rather than the effects which underly them (see Section 6).

5.2 LOG-LINEAR AND LOGIT-LINEAR ADDITIVE MODELS

However, we shall not limit ourselves to linear models. The main effects of log-linear models clearly measure differences in the log-means, and the main effects of logit-linear models measure differences in the logit-means, log $(m/(1-m))$. For example, the log-linear additive model (T,D,E) has systematic component

$$\log m_{tde} = \lambda + \lambda_{1t} + \lambda_{2d} + \lambda_{3e} \quad \text{for all } t, d \text{ and } e,$$

and hence with the restriction $\lambda_{34} = 0$ we find that

$$\lambda_{3e} = \log m_{tde} - \log m_{td} ,$$

so that λ_{3e} measures the difference in the log mean parity between any educational level and secondary and higher ($E=e$ and $E=4$), for fixed T and D.

As before, we can re-express this as a ratio of means by exponentiating:

$$\log m_{tde} - \log m_{td4} = \log (m_{tde}/m_{td4}), \text{ so that}$$

$$\gamma_{3e} = e^{\lambda_{3e}} = m_{tde}/m_{td4}$$

That is, γ_{3e} measures the ratio of the mean parities between any educational level e and secondary and higher for fixed T and D. This parallels the interpretation of estimated effects given in Section 2.

Example:

For the log-linear model (T,D,E) fitted to Table 3, the estimated main effects of education were as follows (standard errors in brackets):

$$\hat{\lambda}_{31} = 0.310\ (.055),\ \hat{\lambda}_{32} = 0.333\ (.054),\ \hat{\lambda}_{33} = 0.208\ (.056),\ \hat{\lambda}_{34} = 0(0),$$

the values for $\hat{\lambda}_{34}$ being a consequence of the imposed constraint $\hat{\lambda}_{34} = 0$. In view of this the other estimates should be interpreted as differences in the log-mean parity between category e and category 4, (for e = 1,2 and 3), and the standard errors refer to these estimated differences.

From these values, the differences in log-mean parity between any two levels of education can be estimated. For example $\hat{\lambda}_{31}$ estimates the difference between categories 1 and 4, and $\hat{\lambda}_{31} - \hat{\lambda}_{32}$ estimates the difference between categories 1 and 2 (since 1 vs. 4 minus 2 vs. 4 equals 1 vs. 2).

The complete set of comparisons are given in the lower triangle of the matrix in Table 5a). For example, the entry in row 3, column 2, .125, estimates the difference in log mean parity between educational levels 2 and 3. As previously noted, it is easier to interpret these estimates by exponentiating; thus the ratio of mean parities between levels 2 and 3 is $e^{0.125} = 1.13$. In other words, lower primary educated women are estimated to have 13 per cent more children than upper primary educated women with the same levels of Type of Place and Marital Duration.

In addition, the standard errors for these comparisons, displayed in the upper triangle of Table 5a), can be used to calculate statistical tests and confidence intervals. In the comparison of levels 2 and 3 for example, the difference in log mean parities of 0.125 has standard error .030, and is therefore significant at the 5 per cent level.

A 95 per cent confidence interval* for the difference is given by

$$0.125 \pm 2\ (.030) = (.065,\ .185)$$

This can itself be interpreted by exponentiating to obtain a confidence interval for the ratio of mean parities. In this case we obtain

$$(e^{.065},\ e^{.185}) = (1.07,\ 1.20)$$

In other words, lower primary educated Indian women have on the average somewhere between 7 per cent and 20 per cent more children than upper primary educated Indians of comparable Marital Duration and the same Type of Place. Note that a difference has been uncovered, but the magnitude of the difference is not at all well determined, a typical finding from this kind of data.

---

* These tests and confidence intervals are approximate; appropriate standard distributions are the t distribution with degrees of freedom as for the fitted model (here df=59), or if the degrees of freedom are large, the Normal distribution. No allowance has been made for multiple comparisons although this can be done in the usual way if desired.

TABLE 5 Effects of Education and Type of Place of Residence on Log_e (Mean Parity), based on log-linear additive model

a) Education

		Education Level 1	2	3	4
	1	-	(.023)	(.031)	(.055)
Education	2	-.023	-	(.030)	(.054)
Level	3	.102	.125	-	(.056)
	4	.310	.333	.208	-

b) Type of Place

		Type of Place: Suva	Urban	Rural
	Suva	-	(.032)	(.028)
Type of	Urban	-.112	-	(.025)
Place	Rural	-.151	-.039	-

Key : Lower triangle gives estimates in difference in $\log_e$ (mean parity) between column level j and row level i. Upper triangle gives corresponding standard errors. For example, in a) -0.023 is the estimated difference in $\log_e$ (mean parity) between education level 1 (no education) and education level 2 (lower primary education), and this has standard error .023. Standard errors are based on a mean deviance of one.

The differentials according to Type of Place are presented in Table 5b). These indicate that when Education and Marital Duration are controlled, there is no evidence of a difference in mean parity between rural and other urban Indians, but the Suvan Indians have significantly fewer children than these groups. From the sample the estimated mean parities of Suvan Indians were 14 per cent and 11 per cent lower than Rural and Other Urban Indians, respectively, but these figures are subject to a standard error of 3 per cent and hence the differentials are not well determined.

5.3 STANDARDIZATION OF THE FITTED MEANS

An alternative way of presenting the results is to use the technique of standardization on the fitted means $\{\hat{m}_{tde}\}$, that is to present the fitted means of one factor averaged over a standard distribution of the other factors.

For example the standardized mean $\tilde{m}_e$ for educational level e is formed as weighted average of $\hat{m}_{tde}$ over the cells of T and D, that is

$$\tilde{m}_e = \sum_t \sum_d w_{td} \hat{m}_{tde}, \tag{5.6}$$

where the w_{td} are non-negative weights which sum to unity:

$$\sum_t \sum_d w_{td} = 1 .$$

The choice of standard (that is, of weights w_{td}) is somewhat arbitrary. We shall (following Pullum, 1977) use the marginal observed distribution of counts, so that

$$w_{td} = n_{td+}/n_{+++} , \tag{5.7}$$

where "+" denotes summation over the corresponding subscript. This leads to the following standardized means for our example:

	Education (controlling for T and D)					
	(1)	(2)	(3)	(4)	Mean (weighted)	
Mean Parity	4.05	4.14	3.66	2.97	3.89	(5.8)
(Sample Size)	(895)	(943)	(579)	(271)	(2688)	

These are interpreted as the expected parities for each education group if they had the distribution of T and D given by (5.6) and (5.7). However, an important property of the standardized means from this log-linear model is that they preserve the ratios between mean parities estimated by the model, for any choice of standard distribution. That is, if e and e' are two educational levels, then for any choice of weights, the ratio of the standardized means is

$$\tilde{m}_e / \tilde{m}_{e'} = \hat{\gamma}_{3e} / \hat{\gamma}_{3e'} \qquad (5.9)$$

which does not depend on the choice of weights used in (5.6).

This is easily proved algebraically by substituting $\hat{m}_{tde} = \hat{\gamma}\,\hat{\gamma}_{1t}\hat{\gamma}_{2d}\hat{\gamma}_{3e}$ in (5.6) and (5.9). For example, from (5.8) the ratio of standardized means parities for levels 1 and 4 is

$$\tilde{m}_1 / \tilde{m}_4 = 4.05/2.97 = 1.36,$$

and this equals the ratio of γ_{34} ($e^{0.31} = 1.36$) to γ_{34} ($e^0 = 1$) for this model from the foot of Table 3. Thus the choice of standard distribution affects the overall level of the standardized means, but does not affect the ratios between pairs of means*.

In general, this invariance property between ratios holds if the variable which is being standardized is additive in the fitted log-linear model. Here the variable education is additive in the log-linear model (T,D,E).

* This property also reduces the calculation involved, since only one standardized mean need be calculated from (5.6), the others being derived from (5.9).

In these cases we recommend re-expressing the standardized means (5.8) in terms of percentage deviations from the overall standardized mean, here 3.89. This gives

	Education Level (standardized for T and D)					
	(1)	(2)	(3)	(4)	Mean	
Parity:						
Per cent deviation from Mean	+4	+6	-6	-24	3.89	(5.10)

The advantage of this form of presentation is that the percentage deviations do not depend on the choice of standard (although the overall mean does).

This way of presenting effects does not readily allow the inclusion of standard errors, as in Table 5, and hence it is less detailed than that table. However, it has the advantage of being easier to comprehend and it allows direct comparison with unstandardized results. For example, the unstandardized means for education (in the beginning of section 2) can be written in the same way:

	Educational Level				
	(1)	(2)	(3)	(4)	Mean
Parity:					
Per cent deviation from Mean	+30	+5	-30	-62	4.0 (2688)

By comparison with (5.10), we see that deviations between educational levels are considerably reduced when Type of Place and Marital Duration are controlled.

The results from linear additive models can also be presented by standardizing the fitted values. However, in that case the *raw differences* between the standardized means give the estimated effects from the model and are the same for any choice of standard. Consequently, the results are best expressed as raw deviations from the overall standardized mean.

5.4 WHY FIT A MODEL?

The introduction of standardization as a method of presenting results suggests an intriguing thought. Since the fitted values from a good model should be close to the original means $\{\bar{Y}_c\}$, a similar result should be achieved by simply standardizing the sample means, without fitting a model at all. Hence why fit models?

The answer lies in the following difficulties associated with direct standardization of the observed table.

(1) A procedure is necessary to deal with cells with no observations. These can be avoided by combining or omitting levels of a factor; alternatively, estimates of missing cell means must be substituted. The best estimates are, naturally enough, the fitted values from a suitable model, although simpler fitted values may be satisfactory.

(2) Standardization of the observed means is statistically inefficient, in that estimates of differences have a larger variance than corresponding estimates from the fitted means from the correct model. For example, cells with small numbers of observations are often given too much weight when the observed means are standardized. Also, standard errors of estimated differences are not provided.

(3) Perhaps most importantly, the model-fitting process indicates when standardization is appropriate for summarizing comparisons, and on what scale the comparison should be made (raw differences for linear additive models, log-differences or ratios for log-linear additive models). If models are not fitted then this can only be decided by a subjective assessment of the data (See Pullum, 1977).

Standardization of the observed means remains a useful method, particularly when the apparatus of model-fitting is not available. However, even when the superior methods are not used, the theory of additive models provides valuable insight into the uses and limitations of standardization as a technique.

5.5 SUMMARY

The presentation and interpretation of effects from additive models and associated standard errors has been illustrated using the log-linear (T,D,E) model for data on mean parity. The presentation led to a discussion of standardization as a method of summarizing comparisons.

6. LOGIT LINEAR MODELS FOR PROPORTIONS

A large amount of the information from fertility surveys can be represented by cross-classifying the proportion $\bar{y}$ of the sample n with a particular attribute; for example, the proportion who ever used or are currently using contraception, the proportion currently married, or the proportion who want no more children. We have noted that logit-linear models with Binomial error are often appropriate for analysing these data. In this section, we illustrate this with applications to data on contraceptive use.

6.1 THE DATA

We shall analyse the data given in Table 2, where the proportion who ever used contraception and the proportion currently using contraception are classified by Race (R), Current Age (A), Desire for More Children (W) and Education (E). The levels of these factors are described in the beginning of Section 1; note, in particular, that in this example education is grouped into not four but two levels; Lower Primary and below (E=1) and Upper Primary and above (E=2).

It would be possible to fit models to these two four-way tables with factors R, A, W and E, but to reduce the number of models we have analysed the results for Fijians (R=1) and Indians (R=2) separately, giving four three-way tables with factors A, W and E.

6.2 MODEL SELECTION

The hierarchical models for these tables are listed in the first column of Table 6, with associated degrees of freedom in the second column. The deviances from fitting these logit-linear hierarchical models to the four tables appear in the next four columns, with blanks when a model was omitted. As noted in section 4, these deviances can be interpreted as chi-squared values.

Consider the column for Ever Use among Indians. In absolute terms, seven models give a satisfactory fit in the sense of a corresponding chi-squared value which is not significant at the 5 per cent level: the additive model (A,E,W) and six models involving two factor effects, viz. (AW), (A,EW), (E,AW), (AW,EW), (AE,AW) and (AE,AW,EW). To choose between these we also consider their relative fit. Consider, for example, the two models (A,EW) and (AW,EW), which are nested since the former is obtained from the latter by setting the two-factor effects of A and W equal to zero. The inclusion of the AW effect in the model (A,EW) results in a reduction in the deviance of 8.9 (from 10.7 to 1.8) for the addition of 3 degrees of freedom (from 6 to 9). Since 8.9 is greater than the 95 per cent point of a chi-squared deviate on three degrees of freedom (viz. 7.8) we conclude that this is a significant reduction and therefore model (AW,EW) fits better than (A,EW). Further comparisons with other models lead to the conclusion that (AW,EW) is the best model. The best model for each response and race are indicated by asterisks in Table 6.

6.3 HYPOTHESIS TESTING

As before, the deviances in Table 6 help us to test general hypotheses about effects. For example, consider the hypothesis that education is not related to current use of contraception after the effects of age and desire for more children have been controlled. (Controlling the latter variable means that the question is directed towards differentials in the link between desire and contraceptive practice rather than differentials in desire itself). If this hypothesis is true then the

TABLE 6 Deviances from fitting hierarchical logit-linear models to data in Table 2

		FIJIANS		INDIANS	
Model	df	Ever Use	Current Use	Ever Use	Current Use
1	15	215.4	166	208	336
A	12			138	196
E	11			207	330
W	14	80.6	74.1	47.0	84.4
A,E	11			136	195
E,W	13			41.9	82.2
A,W	11	25.4	36.9	20.0	67.8
A,E,W	10	19.3	29.9	15.4	64.5
AE	8			135	16.7
EW	12			39.4	65.7
AW	8	21.5	20.1	8.92	16.7
A,EW	9	18.9	23.0	10.7	45.8
E,AW	7	15.5	12.6	4.41	13.6
W,AE	7	9.5*	23.2	14.3	61.2
AW,EW	6	15.5	10.8	1.83*	4.60*
AE,EW	6	8.6	13.8	9.70	45.1
AE,AW	4	5.3	5.80	3.53	11.6
AE,AW,EW	3	5.2	2.44*	1.04	3.89

A = age, *E* = education, *W* = desire for more children.

df = degrees of freedom corresponding to the model

* best model.

model (AW) which includes no effects of education should fit the data. Since the model (AW,EW) which includes education effects fits the data better than (AW) for both Fijians and Indians, the hypothesis is rejected for both Fijians and Indians, that is education does indeed appear to be related to Contraceptive Use, after controlling for A and W. However, the presence of the EW interaction in the best models indicates that the effect of education does not appear to be additive; that is, the nature of the effects depends whether or not women desire more children.

6.4 INSPECTION OF THE FITTED VALUES: LOGIT PLOTS

We noted in the previous section that non-additive effects such as this are best understood by reference to the fitted values from the model. In Figure 1 we have plotted the logits of the observed proportions against age for each combination of educational level and desire for more children. For clarity we have labelled these groups HY, HN, LY and LN where H stands for High Education (E=2), L stands for Low Education (E=1), Y stands for Desires More Children (W=1) and N stands for Desires No More Children (W=2). In Figure 2 we have plotted in the same way the logits of the fitted proportions based on the best model in each case.

Comparison of these plots shows the effect of fitting the models. The similarity in the positions of the corresponding observed and fitted logits reflects the fact that the models fit the data well; the parallel lines in the fitted logits reflect certain kinds of additivity in the models. For example, the model (AW,EW) for current use, Indians display a form of additivity because the AE interaction is missing, viz., within each level of W the effects of A and E are additive. This is reflected in the graph by two pairs of parallel lines for W=1 (Yes) and W=2 (No). The displacement between the lines in each pair represents the effect of education at the corresponding level of W.

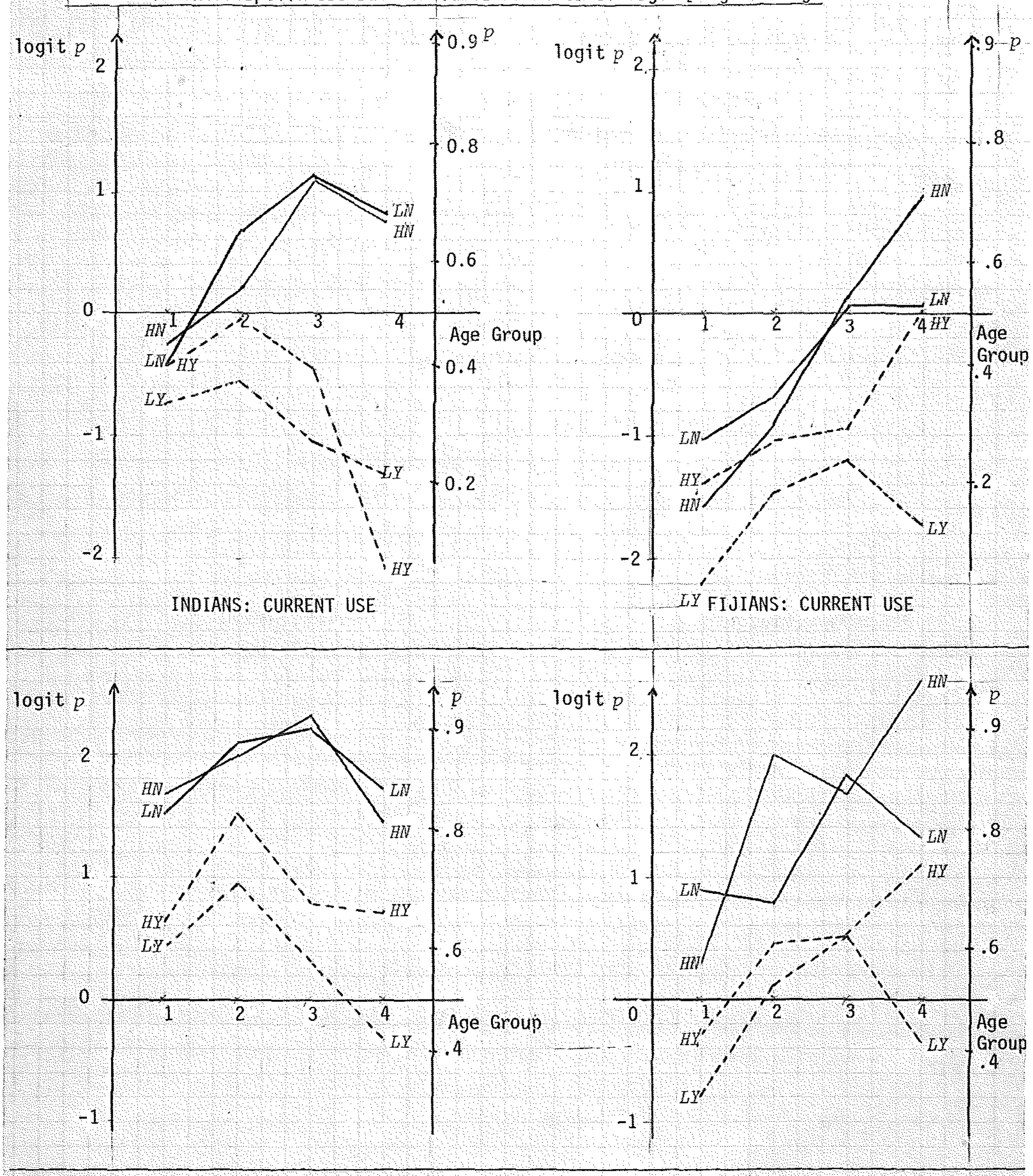
It is clear from the graph that the effect of education on current use for Indians is greater for women who desire more children than for women who do not desire more children (the same remark applies for Ever Use). Hence for Indian women, the impact of education seems to have increased contraceptive use among women who are delaying a desired future child, (spacers) more than among women who do not want another child.

For Fijian women the (EA) interaction is included in the model, so that the effect of education on current use also depends on the age group of the respondents. The graph of fitted logits indicates that again the impact of education is mainly on women who desire more children, but that the impact is significantly higher for women in the fourth age group, that is, for women over 40 years old. In this age group a differential according to education emerges for the women who do not desire another child, that is, more educated women have a higher incidence of current use. However, this differential is still less marked than for the group who do want another child.

For both Fijians and Indians the effect of Desire for More Children on Current Use depends on both Educational Level and Age. The prevalence of Current Use among women who want no more children increases sharply with age. Also, for Indians, current use is highest in the 30-39 group for those who want no more children, and in the 25-29 group for those who want more children.

The magnitude of these effects can be estimated from the fitted values if required. For example, the difference in current use between High and Low education groups who want more children is about half on the logit scale, which corresponds to a difference of a half in the log-odds. Hence since $e^{\frac{1}{2}} = 1.65$, this corresponds to a 65 per cent increase in the odds in favour of current use. Once again a calculation of the standard error (not attempted here) would indicate that this increase is not well determined.

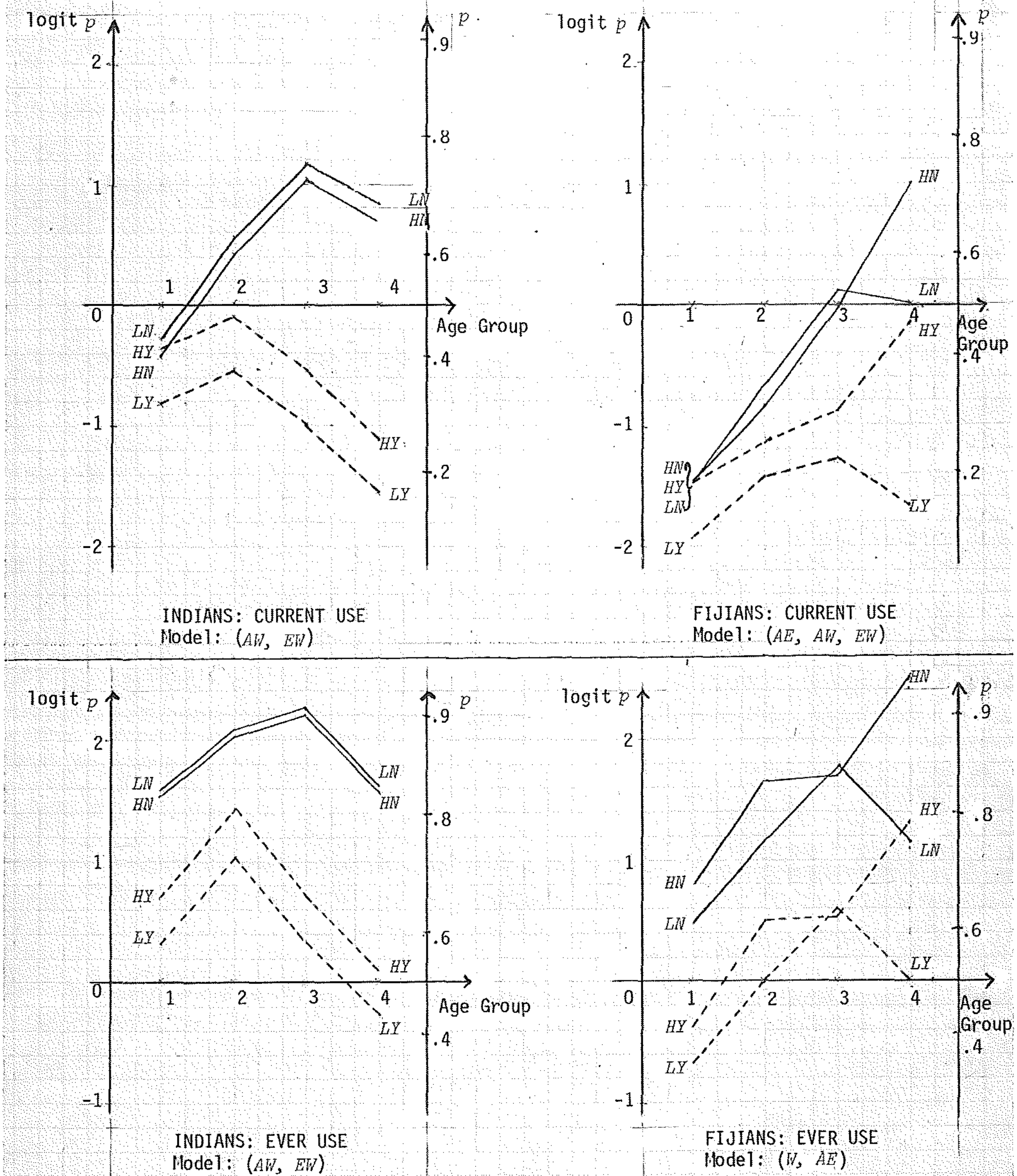
FIGURE 1. Contraceptive Use Data in Table 2. Plots of logit p against Age



Notes: Each graph shows 4 groups. H = high education, L = low education, Y = desires more children, N = desires no more children

Left hand scale is logit p. Right hand scale is p.

FIGURE 2: Contraceptive Use Data in Table 2. Plots of fitted logits against Age



Notes: Groups as for Figure 1. Fitted values are from best logit-linear models, as in Table 6.

The interpretation of the models for Ever Use are left to the reader; in a sense they are of limited substantive interest since the relationship between the historical measure of contraceptive use (Ever Use) and the current measure of attitude towards child-bearing (W) is rather obscure.

6.5 SUMMARY

We have illustrated the use of logit-linear models for cross-classified proportions on data on contraceptive use. The best fitting model is selected by methods similar to those described in Section 4. Here however, non-additive models are required, and these models are interpreted from plots of the fitted values.

REFERENCES

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, 1975.

Davis, J.A., (1974) "Hierarchical Models for Significance Tests in Multivariate Contingency Tables: An Exegesis of Goodman's Recent Papers". *Sociological Methodology*, Herbert L. Costner ed., Jossey-Bass, San Francisco, Washington, London, pp. 189-231.

Goodman, L.A. (1972) *A Model for the Analysis of Surveys*. Am.J.Soc.77, 1035-1086.

Kendall, M.G. and O'Muircheartaigh, C.A. (1976) *Path Analysis and Model Building*. Technical Bulletin Series, No.2. International Statistical Institute, Voorburg, Netherlands, 1977.

Nelder, J.A. (1974) "Log-linear Models for Contingency Tables: A Generalization of Least Squares", *Applied Statistics* 23, 323.

Nelder, J.A. and Wedderburn, R.W.M. (1972) *Generalized Linear Models*. J.Roy,Statis.Soc.Ser. A 135, 370-383.

Pullum, T.W. (1976) *Standardization*. Technical Bulletin Series.No. 3, International Statistical Institute, Voorburg, Netherlands, 1978.

APPENDIX I

ADDITIONAL ANALYSIS BASED ON WITHIN-CELL SAMPLE VARIANCES

In section 1, we emphasized the need to collect the sample variances $\{s_c^2\}$ for each cell c of the table. In this appendix, we indicate how these data can be used to help determine the error structure of the model.

For a dichotomous response, leading to cross-classifications of proportions, the sample variance is completely determined by the sample proportion. (If the sample proportion is $\bar{y}$ and simple random sampling is assumed then the sample variance is $\frac{\bar{y}(1-\bar{y})n}{n-1}$). Consequently, we shall not consider these tables or the Binomial error structure associated with them.

For models considered in this paper we are left with two possible error structures, the Normal and the Poisson. For simplicity, let us assume a Normal error structure for linear models and a Poisson error structure for log-linear models. The sample variance can be used (a) to determine the multipliers k_c in the normal error variance, if linear models are fitted, or (b) to check the Poisson error assumption, if log-linear models are fitted. We shall illustrate both cases using the within-cell variances for the data in Table 3, given in Table A1. In that case, the log-linear models were clearly more appropriate than the linear models, but the data also provide a good illustration of procedures when (a) applies.

NORMAL ERROR

Recall that the normal error structure assumes that the variance of the mean $\bar{y}_c$ for a typical cell c is given by

$$\text{var } (\bar{y}_c) = \sigma^2 . k_c / n_c$$

and the corresponding within-cell variance by var $(Y_c) = \sigma^2 k_c$. Since s_c^2, the sample variance for cell c, estimates var (Y_c), and obvious procedure would be to set $k_c = s_c^2$ and $\sigma^2 = 1$. However, in general this is not recommended, since the sample variances are not determined for cells with less than two observations, and are themselves subject to considerable sampling variation for cells with small sample sizes. A logical procedure would be to smooth these estimates by fitting a model to the within-cell variances, but this is perhaps a little elaborate for this secondary type of analysis.

A simple suggestion is to calculate averages of the within-cell variances for each level of a factor and then to look for obvious relationships. In our example, the variance is clearly related to marital duration. The mean sample variances (calculated as weighted means with weights for cell c proportional to $n_c - 1$) for different levels of D are shown in the last column of Table A1, and show a clear increase in the variance as marital duration increases.

The error structure for the fitted Linear Models in Table 4 assumed

$$\text{var } Y_{tde} = d\sigma^2 \text{ for all } t,\ d \text{ and } e \qquad \text{(A1.2)}$$

If $d\sigma^2$, the variance for row d, is estimated as the weighted average of the within-cell variances for that row, then the following estimates of σ^2 are obtained:

	Marital Duration							
	1	2	3	4	5	6	mean	
Estimated σ^2	0.77	0.70	0.98	0.92	1.25	1.79	1.07	(A1.3)

There is some evidence that these variances still increase with D, that is that the assumption (A1.2) underestimates the rate of increase of the variance with marital duration. However, (A1.2) clearly improves on the assumption of homoscedasticity (that is, equal within-cell variances).

If (A1.2) is accepted and the variances in (A1.3) are taken as estimates of σ^2 then they can be compared directly with estimates of σ^2 derived from the unsaturated models, that is, the mean deviances. From Table 4, the mean deviance for the largest unsaturated linear model (TD,TE,DE) is 1.42, which is higher than the variances in (A1.3), except for the value for D=6. This suggests that the mean deviance from this model overestimates the within-cell variance, which implies that the assumption that the 3-factor effects of $T \times D \times E$ are zero is questionable. Hence, inspection of the within-cell varances leads to the suspicion that none of the unsaturated linear models are entirely satisfactory for these data, and adds to the evidence in favour of another choice of link function.

POISSON ERROR

For the Normal Error distribution we suggested plots of the sample variances against the factors. A basic property of the Poisson distribution, that the variance of Y is equal to the mean, indicates a plot of the sample variances against the sample means.

We suggest plotting the log sample variances against the log sample means, for the following reason. A useful form for the relationship between m_c, the expected parity of an individual in cell c, and the within-cell variance is

$$\text{var } Y_c = k m_c{}^a ,$$

where k and a are constants. The special case k=1, a=1 corresponds to the Poisson variance; a=0 corresponds to the assumption of homo-cedasticity. Taking logarithms gives

$$\log \text{var } Y_c = \log k + a \log m_c$$

Thus in a plot of log variances against log means, the Poisson error variance corresponds to a straight line through the origin (k=1, log k=0) with slope 1.

In Figure A1, the $\log_e$ (variances) are plotted against the log (means) for the data in Table 3; the straight line corresponds to the Poisson variance. It appears that the variances are less than Poisson for low mean parities, and more than Poisson for high mean parities, an intuitively reasonable result. The Poisson error assumption seems acceptable as an overall approximation.

In certain situations this plot may reveal that the variance is proportional to the mean (a=1), but with a constant of proportionality k not equal to one. Then k can be estimated as the weighted average of s_c^2/m_c,

$$\hat{k} = \sum_{\text{cells, } c} (n_c-1)s_c^2/m_c \Big/ \sum_{\text{cells, } c} (n_c-1)$$

where cells with a small value of n_c (say n_c< 20) are excluded from the

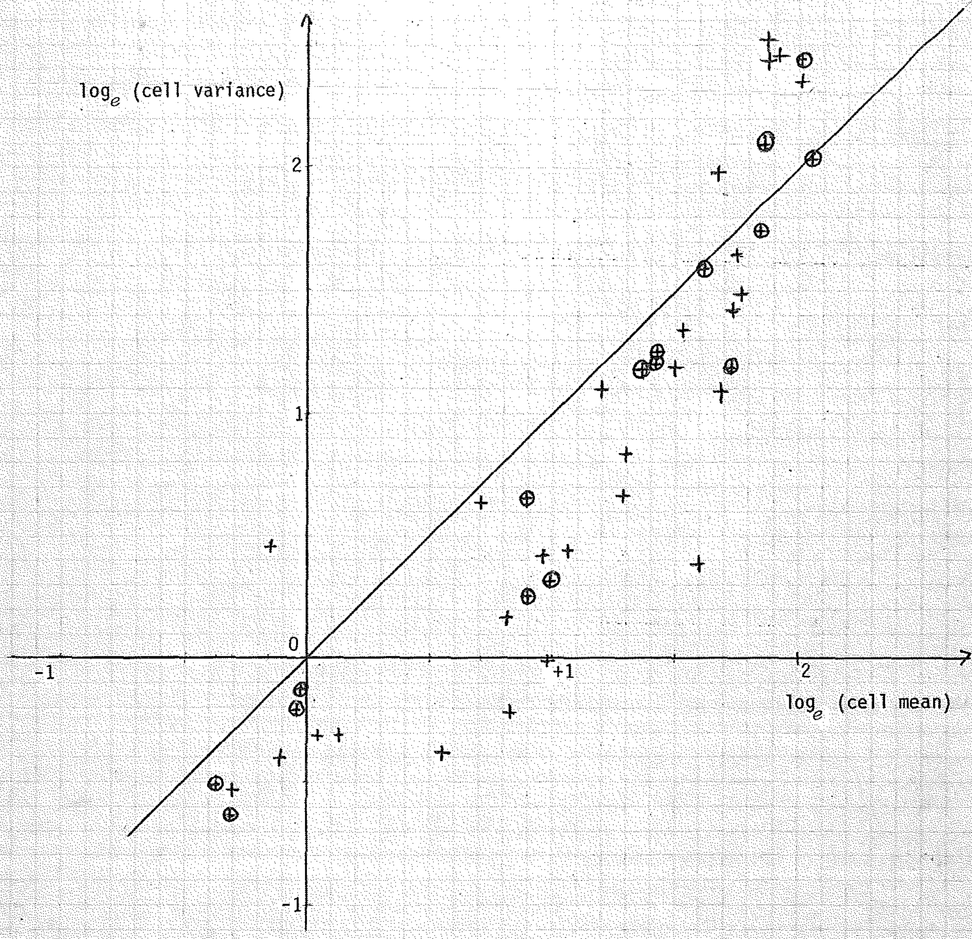
TABLE A1

Indians: Within Cell Variances for data in Table 3

Years Since First Marriage	TYPE OF PLACE: SUVA Education 1	SUVA Education 2	SUVA Education 3	SUVA Education 4	URBAN Education 1	URBAN Education 2	URBAN Education 3	URBAN Education 4	RURAL Education 1	RURAL Education 2	RURAL Education 3	RURAL Education 4	Mean[a]
<5	1.14	.73	.67	.48	1.06	1.59	.73	.54	.88	.81	.80	.59	.77
5 - 9	1.66	.99	1.87	.68	3.44	1.51	.97	.81	1.93	1.36	1.30	1.19	1.41
10 - 14	1.72	2.31	1.57	1.82	2.97	2.99	1.96	1.52	3.52	3.31	3.28	2.50	2.94
15 - 19	2.03	1.46	.81	.92	7.40	2.97	3.83	.70	4.91	3.23	3.29	-	3.69
20 - 24	4.15	4.64	4.08	4.30	7.19	4.44	4.33	.33	8.20	5.72	5.20	.50	6.26
25+	12.46	11.66	4.27	-	11.45	10.53	12.60	-	11.34	7.57	7.07	-	10.75

a) Weighted mean with cell c given weight $n_c - 1$, where n_c is the sample size.
Blanks denote sample sizes of zero or one, for which sample variances cannot be calculated.

FIGURE A1: Plot of $\log_e$ (cell means) vs $\log_e$ (cell variances) for data in Table 3



Note: Cells = with less than twenty observations ignored.

+ = 20 - 49 observations

⊕ = 50+ observations

Straight line with slope one corresponds to Poisson Variance assumption (Variance = Mean) calculation. The deviances from log-linear models should then be adjusted by dividing by $\hat{k}$ before comparing them as chi-squared deviates, and standard errors from the models should be multiplied by the factor $\sqrt{\hat{k}}$.

SUMMARY

The within-cell sample variances should be used to check the error structure of linear and log-linear models. For linear models, these quantities can be used to estimate the underlying variance $\sigma^2$and the factors k_c which characterize departures from homoscedasticity. For log-linear models, the Poisson Error assumption can be examined by plots of the log sample means against the log sample variances.

APPENDIX II

COMPUTING NOTES

Methods of analysis presented here are based on maximum likelihood estimation for the general linear model. The theory is described in Nelder and Wedderburn (1972), and the practice is facilitated by the computer package GLIM*. Before presenting material on GLIM (Appendix III) we shall discuss comparable systems of analysis available in common statistical computing packages, such as SPSS and BMD. This note is based partly on the paper by Francis and Williams (1976).

The traditional method of analysis of cross-classified data is analysis of variance, where each effect (main, two-factor or higher order) is assigned a sum of squares which represents its contribution to the explained variance of the response. See, for example SPSS subprogram ANOVA and BMD program P2V. For balanced data (that is, equal-cell sample sizes) the effects are orthogonal, which means that the decomposition of the sum of squares is essentially unique; for unbalanced data this uniqueness is lost, because the sum of the squares for a particular effect depends on whether other factors are included, or excluded, from the model.

The underlying models for analysis of variance are hierarchical linear models with normal error and constant within-cell variance (that is, $k_c = 1$ in the notation of section 3), and the sums of squares produced by analysis of variance are simply differences of deviances from those models. Hence, analysis of variance provides the same kind of information as that provided by deviances, that is, it allows the analyst to compare the fit of hierarchical models and hence to assess the overall effect of factors, but it does not provide fitted values and parameter estimates for the analysis of individual effects as discussed in sections 5 and 6.

* References for the computer packages mentioned here appear at the end of the appendix.

To illustrate the connection between the sums of squares of analysis of variance and deviances from normal linear models, consider the two-way table with hierarchical models ($\emptyset$), (A), (B), (A, B), and (AB). If $SS_\emptyset$, SS_A, SS_B, $SS_{A,B}$ and SS_{AB} are the corresponding deviances, then the analysis of variance sums of squares are constructed as follows:

Source	Sum of squares
(1) main effect of A	$SS_A - SS_\emptyset$
(2) main effect of B	$SS_B - SS_\emptyset$
(3) main effect of A, adjusted for B	$SS_{A,B} - SS_B$
(4) main effect of B, adjusted for A	$SS_{A,B} - SS_A$
(5) two factor effect of $A \times B$	$SS_{AB} - SS_{A,B}$
(6) total	$SS_{AB} - SS_\emptyset$

The two factor effect of $A \times B$ (also called the two-way interaction) is calculated adjusting for the main effects of A and B ; the sum of squares for the main effects depend on whether the other effect is controlled (lines (3) and (4)) or not controlled (lines (1) and (2)). For balanced data line (1) = line (3) and line (2) = line (4), but this is not the case for unequal cell sample sizes. In the analysis of variance programs, the output is controlled by options chosen by the user. Thus one option might give lines (3) - (6), another option lines (1), (4), (5) and (6) and a further option lines (2), (3), (5) and (6).

A special case of analysis of variance is Multiple Classification Analysis (MCA), which assumes an additive linear model. The computations are equivalent to an analysis of variance, but programs also provide estimated of effects in the form of deviations from the mean, and related statistics.

Logit-linear models for proportions are equivalent to certain log-linear models for contingency tables, as discussed in section 3. These can be fitted using various contingency table programs (for example, BMDP3F*,

* Available in the 1977 version of BMDP.

ECTA, C-TAB and CONTAB) which use margin fitting algorithms. They can fit any hierarchical logit-linear model and provide deviances, fitted values and estimates of effects. They do not provide standard errors* of effects, but have the advantage of requiring less computer space than GLIM and hence they can be more economical to run and place less restrictions on the maximum dimensions of the table.

The computer package GLIM treats a particular class of non-linear regression models and computes parameters by an iteratively reweighted least squares algorithm. An alternative procedure is to use a general non-linear least squares program, such as BMDP3R. Another system which uses the same computational algorithm is GENCAT. These are powerful programs but they require considerable statistical expertise to apply them to the models considered here. In particular, a knowledge of maximum likelihood estimation and of the regression formulation of analysis of variance models is necessary. Interested readers are referred to the documentation of these programs, and to Appendix C13 of the 1977 BMDP manual.

The author feels that the class of models fitted by GLIM are sufficiently general for most data encountered in World Fertility Surveys, and the package is particularly simple to use. An example of the output it produces is given in Appendix III.

---

* However, BMDP3F calculates standard errors for certain models, and failing this gives upper bounds.

REFERENCES TO STATISTICAL PACKAGES

GLIM (A Fortran Program using Iterative Weighted Least Squares) Distributed by Numerical Algorithms Group,13 Banbury Road, Oxford OX2 6NN, U.K. (Can be used as both a batch and an interactive program).

BMDP: Biomedical Programs. P. Series. W.J. Dixon, editor. University of California Press, Berkeley, U.S.A., 1977.

SPSS Manual, 2nd Edition. Norman H. Nie, C. Hadlai Hull, Jean G. Jenkins, Karen Steinbrenner and Dale H. Brent. McGraw Hill, New York, U.S.A., 1975.

ECTA (Everyman's Contingency Table Analysis : Parameter Estimates and Tests). For information, write to Leo A. Goodman, Department of Statistics, University of Chicago, Illinois, 60637, U.S.A.

C-TAB Distributed by International Educational Services, P.O. Box A3650, Chicago, Illinois, 60690, U.S.A., (available for IBM and CDC machines).

CONTAB (Zahn, DA). For information contact Department of Statistics and Statistical Consulting Center. Florida State University, Tallahassee, Florida 32306, U.S.A.

GENCAT. For details see Biostatistics Technical Report No. 8, Department of Biostatistics, University of Michigan, Ann Arbor, Mi. 48109, U.S.A.

Reference

Francis, I. and Williams, K. (1976). A look at Programs for the Analysis of Multiway Tables. American Statistical Association, Proceedings of the Statistical Computing Section.

APPENDIX III

AN EXAMPLE OF OUTPUT FROM GLIM

An example of the output obtained from a typical GLIM run is appended. The system is designed to run interactively (although it can also be run in batch mode) ; once the elements of the system have been learnt it is very easy to use.

Although no attempt is made to explain the system (the interested reader is referred to the GLIM manual), we include some comments to clarify the output. Words beginning with a $ sign are GLIM commands. Lines starting with $C are comments. The commands $DATA, $READ, $FACTor and $CALCulate are used to define the data and the factors, $TERms creates the space for the largest model to be fitted, $YVA specifies the response, $ERRor, $WEIght and $FIT specify the model to be fitted, and $DISPlay controls the output.

The example fits the log-linear additive model (T, D, E) to the data in Table 3.

```
/GET,TAPE1=IND
/-GLIM

                 GLIM RELEASE 2A
                 ***************

    GENERAL LINEAR INTERACTIVE MODELLING
    ------------------------------------


? $C
? $C FIRST THE DATA IN TABLE 3 IS READ FROM TAPE 1.
? $C 72 OBSERVATIONS ON P=MEAN PARITY,N=SAMPLE SIZE.
? $C
? $INP 1 $
? $C
? $C THE FUNCTION <GL(A,B)>GENERATES LEVELS FOR A FACTOR BY
? $C ASSIGNING TO IT THE INTEGERS 1 TO A IN BLOCKS OF B.
? $C
? $FACT T 3 D 6 E 4
? $CALC T=<GL(3,4) & D=<GL(6,12) & E=<GL(4,1)
? $C
? $C NOW CREATE SPACE FOR LARGEST.MODEL AND SPECIFY DEPENDENT
? $C VARIABLE = P.
? $C
? $TERMS T*D,D*E,T*E,P $YVA P $
? $C
? $C SPECIFY POISSON ERROR DISTRIBUTION (P).THIS GIVES
? $C LOG-LINEAR MODELS BY DEFAULT.
? $C
? $ERR P $
? $C
? $C SPECIFY WEIGHTS PROPORTIONAL TO N, THE CELL SAMPLE SIZE.
? $C
? $WEIGHT N $
? $C
? $C FIT THE NULL MODEL. GLIM  RETURNS THE DEVIANCE AND DEGREES
? $C OF FREEDOM.(CF TABLE 4)
? $C
? $FIT $
   69 DF
 DEVIANCE  CYCLE
   3732.     4
? $C
? $C FIT (TD,TE,DE) AND (T,D,E)
? $C
? $C FIT T*D,T*E,D*E $
   28 DF
 DEVIANCE  CYCLE
   30.95     4
? $FIT T,D,E $
   59 DF
 DEVIANCE  CYCLE
   70.65     4
```

```
? $C
? $C FROM THESE AND OTHER MODELS,THE ADDITIVE MODEL (T,D,E) FITS
? $C BEST. HENCE PRINT OUT PARAMETER ESTIMATES,FITTED VALUES AND
? $C STANDARD ERRORS OF DIFFERENCES OF PARAMETER ESTIMATES FOR
? $C THIS MODEL
? $C
? $DISP E R S $

 ERROR POISSON LINK LOG
 Y-VARIATE  P

     ESTIMATE        S.E.     PARAMETER
 1   1.701E+00     5.81E-02   GM.
 2  -1.512E-01     2.88E-02   T 1
 3  -3.896E-02     2.46E-02   T 2
 4  -1.977E+00     5.00E-02   D 1
 5  -9.791E-01     3.51E-02   D 2                       [Cf. Table 5]
 6  -6.063E-01     3.09E-02   D 3
 7  -3.626E-01     2.93E-02   D 4
 8  -1.913E-01     2.90E-02   D 5
 9   3.096E-01     5.52E-02   E 1
10   3.327E-01     5.39E-02   E 2
11   2.079E-01     5.61E-02   E 3
S.E.S ASSUME MEAN DEVIANCE OF 1

 UNIT    OBS      FITTED   RESIDUAL      WEIGHT    LIN.PRED
    1      1         .89      -1.17    7.11E+00   -1.17E-01
    2      1         .91       1.10    1.91E+01   -9.432-02
    3      1         .80        .70    3.37E+01   -2.19E-01
    4      1         .65        .69    3.33E+01   -4.27E-01
    5      1         .99        .61    1.19E+01   -5.08E-03    [Cf. Table 3]
    .      .           .          .           .           .
    .      .           .          .           .           .
    .      .           .          .           .           .
   70      8        7.64        .46    4.51E+02    2.03E+00
   71      6        6.75      -1.15    6.75E+01    1.91E+00
   72      0        5.48       0.00    0.          1.70E+00

S.E. OF DIFFERENCES
 1   0.
 2   7.056E-02 0.
 3   6.837E-02 3.249E-02 0.
 4   8.895E-02 5.632E-02 5.431E -02 0.                          [Cf. Table 5]
 5   7.757E-02 4.363E-02 4.142E-02 5.275E-02 0.
 6   7.402E-02 4.039E-02 3.819E-02 5.108E-02 3.719E-02 0.
 7   7.130E-02 3.983E-02 3.756E-02 5.121E-02 3.698E-02 3.336E-02 0.
 8   7.115E-02 3.962E-02 3.680E-02 5.122E-02 3.704E-02 3.340E-02 3.234E-02 0.
 9   1.115E-01 5.784E-02 5.688E-02 6.377E-02 5.838E-02 5.836E-02 6.025E-02
X.34E-02 0.
10   1.092E-01 5.765E-02 5.705E-02 6.495E-02 5.982E-02 5.971E-02 6.073E-02
X.XX4E-02 2.266E-02 0.
11   1.089E-01 6.044E-02 5.986E-02 6.970E-02 6.400E-02 6.267E-02 6.317E-02
X.XX3E-02 3.099E-02 3.000E-02 0.
          1         2         3         4         5         6         7
     8         9        10        11
S.E.S ASSUME MEAN DEVIANCE OF 1


? $STOP
EXIT
/BYE
```