Chapter 12: Log-linear Model Methods

Stephen E. Fienberg and Daniel Manrique-Vallier

Abstract

The chapter introduces the structuring of categorical data in the form of contingency tables, and then turns to a brief introduction to log-linear models and methods for their analysis, followed by their application in the context of Customer Satisfaction Surveys (CSS). The focus is on the adaption of methods designed primarily for nominal data to the type of ordinal data gathered in the ABC Annual Customer Satisfaction Survey (ACSS). The chapter outlines some basic methodology based on maximum likelihood methods and related model search strategies, and then puts these methodological tools to work in the context of data extracted from ACSS.

Keywords and Phrases: Contingency tables, Likelihood equations, Maximum likelihood methods, Model search, Nominal data, Ordinal data.

1 Introduction

Categorical data are ubiquitous is virtually all branches of science, but especially in the social sciences and in marketing. A contingency table consists of counts of units of observation cross-classified according to the values of several categorical (nominal or ordinal) variables. Thus standard survey data, which are largely categorical in nature, are best thought of as forming a very large contingency table. If we have collected data from n individuals or respondents and there are p questions in the survey questionnaire, then the table is of dimension p and the counts total n.

Suppose there are $p = 16$ questions with binary response categories on a questionnaire. Then the corresponding contingency table would contain $2^{16} = 65,536$ cells. With a sample of $n = 1,000$ respondents, a value not atypical for a large customer satisfaction survey, we can expect to see few large counts since the average cell count would be 0.01! Thus methods for the analysis of categorical survey data need to be able to cope with such sparseness, especially since there is likely to be closer to 100 questions on the survey questionnaire, rather than 16. One strategy for coping involves latent variable structures with relatively few parameters, such at the Rasch model and its generalizations (see chapter 14, as well as some alternatives described briefly at the end of this chapter). Another strategy is to take small groups of variables and look more closely at their interrelationships. That is the strategy on which we focus here using log-linear models.

Hierarchical log-linear models provide an important and powerful approach to examining the dependence structure among categorical random variables. There are now many books describing these methods and computer programs in standard packages such as R and SAS for implementing them. See for example, the books by Agresti (2002, 2007, 2010), Christiensen (1997), Bishop, Fienberg & Holland (1975), Fienberg (1980), Goodman (1978), and Haberman (1974).

In the next section, we describe log-linear models and their interpretation and describe some basic methodology that allows for the analysis of survey data. Then we put the models and tools to work on an excerpt of the ABC Annual Customer Satisfaction Survey (ACSS) to illustrate their use and interpretation. We describe some related latent class models in a final section and explain their potential uses.

2 Overview of Log-linear Models and Methods

2.1 Two-Way Tables

Virtually all readers will be familiar with the usual chi-squared test for the independence of the row and column variables in a two-way contingency table first proposed by Pearson (1900) and which appears in most introductory statistics textbooks:

$$X^2 = \sum_{ij} \frac{(\text{Observed}(i,j) - \text{Expected}(i,j))^2}{\text{Expected}(i,j)}, \tag{1}$$

Suppose that the row variable is gender, male or female, and the column variable is product approval, yes or no. Then the X^2 test focuses on whether approval is independent of gender, where

$$\text{Expected Count}(i,j) = \frac{(\text{Row Total } i) \times (\text{Column Total } j)}{\text{Grand Total}}, \tag{2}$$

is the expected value of the (i,j) cell. We usually refer the value of X^2 to the tabulated upper-tail values of the chi-squared (χ^2) distribution on 1 degree of freedom and compute a p-value, e.g., a value of X^2 of 3.64 corresponds to a p-value of 5%.

Log-linear models can help extend analyses such as this, e.g., see Bishop, Fienberg & Holland (1975) and Fienberg (1980). If we let x_{ij} be the count in the (i,j) cell of the table, indexed by i for Gender and j for Product Approval, we display the 2×2 tables of cell counts and probabilities $p_{ij} = P[\text{Gender}= i,\text{Approval}= j]$ as

		Approval		
		Yes	No	Totals
Gender	Male	x_{11}	x_{12}	x_{1+}
	Female	x_{21}	x_{22}	x_{2+}
	Totals	x_{+1}	x_{+2}	n

		Approval		
		Yes	No	Totals
Gender	Male	p_{11}	p_{12}	p_{1+}
	Female	p_{21}	p_{22}	p_{2+}
	Totals	p_{+1}	p_{+2}	1

The same notation works for more categories, i.e., for an $I \times J$ table. We can now write the log-linear model for the expected cell values, $\{m_{ij} = E[x_{ij}]\}$, in such tables as

$$\log m_{ij} = \log(np_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \tag{3}$$

where and $n = \sum_i \sum_j x_{ij}$, with constraints such as

$$\sum_{i=1}^{2} u_{1(i)} = \sum_{j=1}^{4} u_{2(j)} \sum_{i=1}^{2} u_{12(ij)} = \sum_{j=1}^{4} u_{12(ij)} = 0, \tag{4}$$

The version of the model in equation (3) has has the form of an analysis of variance model. When $u_{12(ij)} = 0$ for all i and j, we can rewrite equation (3) as $m_{ij} = m_{i+}m_{+j}/n$ and the maximum likelihood estimate of m_{ij} is simply the result of setting $m_{i+} = x_{i+}$ and $m_{+j} = x_{+j}$. In this notation, the X^2 statistic of expression (1),

$$X^2 = \sum_{ij} \frac{(x_{ij} - x_{i+}x_{+j}/n)^2}{x_{i+}x_{+j}/n}, \tag{5}$$

is equivalent to testing whether the interaction terms $u_{12(ij)}$ are all zero in the log-linear model. Alternatively, would could use the likelihood ratio statistic comparing the log-additive model ($u_{12(ij)} = 0$)

$$G^2 = 2 \sum_{i,j} x_{i,j} \log \left[\frac{x_{i,j}}{x_{i+}x_{+j}/n} \right], \tag{6}$$

due originally to Wilks (1935). The X^2 and G^2 statistics are usually close in value, differing mainly in the presence of small cell counts. Cressie and Read (1988) provide a detailed account of a class of goodness-of-fit measures that include both the X^2 and G^2 statistics as special cases and have the same asymptotic χ^2 distribution under the null hypothesis that the model of independence holds.

For a 2×2 table, a common measure of dependence between the row and column variables is the odds ratio introduced by Yule (1900):

$$OR = \frac{\text{odds}\,(X_2 = 2|X_1 = 2)}{\text{odds}\,(X_2 = 2|X_1 = 1)} = \frac{P[X_2 = 2|X_1 = 2]/(1 - P[X_2 = 2|X_1 = 2])}{P[X_2 = 2|X_1 = 1]/(1 - P[X_2 = 2|X_1 = 1])} = \frac{p_{11}p_{22}}{p_{12}p_{21}}.$$

We can estimate OR by

$$\widehat{OR} = \frac{x_{11}x_{22}}{x_{12}x_{21}}. \tag{7}$$

From (3) and (4), it is easy to show that

$$u_{12(11)} = \frac{1}{4}\log(OR). \tag{8}$$

Hence we can use the log-linear model in equation (3) to estimate the odds ratio. When $OR > 1$ ($u_{12(11)} > 0$), the variables in the table are positively associated; when $OR < 1$ ($u_{12(11)} < 0$) they are negatively associated. $OR = 1$ ($u_{12(11)} = 0$) corresponds to statistical independence.

Because for the 2×2 table there is only a single parameter associated with independence, i.e., either $u_{12(11)}$ or OR, we say that there is 1 degree of freedom (d.f.) associated with this model. Under the model of independence, both the X^2 and G^2 statistics have an asymptotic χ^2 distribution on 1 d.f. For the more general $I \times J$ contingency table, there are $(I-1)(J-1)$ such odds ratios of the form (7), and the test statistics have an asymptotic distribution that follows the χ^2 distribution on $(I-1)(J-1)$ d.f. under independence.

2.2 Hierarchical Log-Linear Models

All the ideas and notation in the preceding subsection carry over into the representation and analysis of higher-dimensional tables. Now we add doubly subscripted $u-$terms to capture higher-order interactions among sets of variables. Thus for a 3-way $I \times J \times K$ table, we write the general log-linear model as

$$\begin{aligned} \log m_{ijk} = \log E[x_{ijk}] \;\; = \;\; & u + u_{1(i)} + u_{2(j)} + u_{3(k)} \\ & + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}, \end{aligned} \tag{9}$$

with side-constraints that all doubly-subscripted $u$terms add to zero across any index. The model where we set all 3-factor terms $u_{123(ijk)}$ equal to zero is the "no-second-order interaction model" studied first by Bartlett (1935) and later by Birch (1963) and Goodman (1969).

Many authors use a form of shorthand notation to specify interpretable log-linear models. Here we use the notation [1][2] to refer to the additive model $u + u_{1(i)} + u_{2(j)}$, and the notation [12] to refer to the model with $u_{12(jk)}$ and all lower-oder terms: $u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$, exactly as we wrote in equation 3). We say that these models are hierarchical because the inclusion of interaction terms like [12] implies inclusion of all lower-order terms, i.e., [1] and [2]; this is known as the *hierarchy principle*. In general the quantities in square brackets in this notation then refer to the highest order $u-$terms in the model.

Thus for the no-second-order interaction model for the 3-way $I \times J \times K$ table we describe the model as [12][13][23]. The 3 components, or highest order u-terms correspond to the two-way marginal totals, $(\{m_{ij+}\}, \{m_{i+k}\}, \{m_{ij+}\})$, and the estimated expected are the solutions of the equations that set these equal to their expectations, i.e.,

$$
\begin{aligned}
m_{ij+} &= x_{ij+}, \text{ for } i = 1, 2, \ldots, I; j = 1, 2, \ldots, J, \\
m_{i+k} &= x_{i+k}, \text{ for } i = 1, 2, \ldots, I; k = 1, 2, \ldots, K, \\
m_{ij+} &= x_{ij+}, \text{ for } j = 1, 2, \ldots, J; k = 1, 2, \ldots, K.
\end{aligned}
$$

We solve these equations to get maximum likelihood estimates, $\{\hat{m}_{ijk}\}$, using some form of iterative method such as iterative proportional fitting (see Bishop, Fienberg & Holland, 1975), or compute them using a generalized linear model program.

All of these ideas and notation generalize to 4-way and higher-dimensional tables. For example, for a 4-way table, the model [134][234] includes the two three-way interaction terms $u_{134(ikl)}$ and $u_{234(jkl)}$, as well as all two-way interactions except for $u_{12(ij)}$ and all "main" effects, i.e.,

$$
\begin{aligned}
\log m_{ijkl} = \log E[x_{ijkl}] &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} \\
&+ u_{13(ik)} + u_{14(il)} + u_{34(kl)} + u_{23(jk)} + u_{24(jl)} \\
&+ u_{134(ikl)} + u_{234(jkl)}.
\end{aligned}
$$

If we hold variables 3 and 4 at fixed levels k_0 and l_0 in this model, most terms in the model are constant and the only terms that vary with variables 1 and 2 are additive, i.e.,

$$
\begin{aligned}
\log m_{ijk_0l_0} = \log E[x_{ijk_0l_0}] &= \{u + u_{3(k_0)} + u_{4(l_0)} + u_{34(k_0l_0)}\} \\
&+ \{u_{1(i)} + u_{13(ik_0)} + u_{14(il_0)} + u_{134(ik_0l_0)}\} \\
&+ \{u_{2(j)} + u_{23(jk_0)} + u_{24(jl_0)} + u_{234(jk_0l_0)}\}.
\end{aligned}
$$

This representation thus implies that variables 1 and 2 are conditionally independent given variables 3 and 4. Models where we set exactly one first order interaction term equal to zero, are often referred to as *partial association* models.

The computation of the degrees of freedom of general hierarchical log-linear models is analogous to the case of the $I \times J$ table, although more involved as we consider more complex models. Fortunately, most standard computer programs implementing log-linear methods, provide the degrees of freedom of the model as part as their standard output, so we seldom have to compute them directly. When tables are sparse, however, we may need to adjust them "by hand," as we see below.

We often represent conditional independence models graphically, as we do here in Figures 1 and 2, where each node is a variable in the model, and absence of edges represents conditional independence according to the graphical Markov property: If A, B and C are (possibly empty) subsets of nodes (variables), we say A and B are conditionally independent given C, if and only if deleting the nodes in C from the graph (and any edges incident to them) results in disjoint subgraphs, one containing the nodes in A and another the nodes in B. A hierarchical log-linear model whose conditional independence relations are precisely those that we can read off its graph in this way is known as a *graphical model*. For further details see Whitaker (1990) and Edwards (2000). Lauritzen (1996) provides a detailed theoretical presentation on graphical models more generally.

2.3 Model Search and Selection

We use hierarchical log-linear models to assess dependence among three or more variables in a contingency table. Special cases of these models, the log-linear graphical and decomposable models allow us to search large contingency tables efficiently for independence, conditional independence, and dependence relationships that are suggestive about the causal structure of the underlying data generating process (Edwards, 2000).

Extending the ideas from the two-way table case, we define the general G^2 statistic or *deviance* for a general hierarchical loglinear model, (a), as

$$G^2(a) = -2 \sum_i x_i \log \left[\frac{\hat{m}_i^{(a)}}{x_i} \right], \tag{10}$$

where $\hat{m}_i^{(a)}$ is the MLE of the expected value of the i-th cell, computed under model (a). The G^2 statistic is asymptotically χ^2 distributed, with the appropriate degrees of freedom for model (a) and can be used to test the goodness of fit of model (a).

We say that two hierarchical log linear models, (1) and (2), are nested when model (2) is a special case of model (1), obtained by setting some of the u-terms of (1) to zero, respecting the hierarchy principle—e.g. [1][2][3] is nested within [13][23], which is in turn nested within [12][23][13]. When this is the case, we can compute the G^2 of model (2), conditional on model (1) being correct, as

$$\begin{aligned} \Delta G^2 = G^2(2|1) &= -2 \sum_i \hat{m}_i^{(1)} \log \left[\frac{\hat{m}_i^{(2)}}{\hat{m}_i^{(1)}} \right] \\ &= G^2(2) - G^2(1), \end{aligned} \tag{11}$$

which we can use to formally test the difference between models (1) and (2). If models (1) and (2) have respectively ν_1 and ν_2 degrees of freedom, then $G^2(2|1)$ is asymptotically distributed as χ^2 with $\nu_2 - \nu_1$ degrees of freedom. Although conditional tests based on G^2 statistics are a powerful tool when comparing competing models, we should always keep in mind that the theoretical results are only valid for nested models, thus they cannot guide us when comparing models that do not belong to the same hierarchy—e.g. [1][23] and [2][13].

As a general principle, more complex models will always fit the data better than simpler models (less bias), but at the price of less precise estimates (more variance), while simpler models allow for more precise estimates at the price of higher bias. A popular technique to balance fit and parsimony is to penalize the G^2 with some measure that expresses complexity. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), defined as

$$AIC = G^2 - 2 \cdot df. \tag{12}$$

$$BIC = G^2 - \log(N) \cdot df, \tag{13}$$

where N is the sample size of the contingency table and *df* the degrees of freedom implied by the model, are two common choices and can be used to compare non nested models. The AIC criterion tends to favor more complex models than BIC.

2.4 Sparseness in Contingency Tables and Its Implications

As we consider larger number of variables, the number of elementary cells in the corresponding contingency tables increases exponentially and with it, the probability of having cells with zero counts. One powerful feature of log-linear models is that parsimonious models can yield positive expected values for elementary cells, even if the observed counts are zero. However, if some of those observed zeros result in a pattern that produces a zero in a margin corresponding to a minimal configuration required by a model, the corresponding parameter cannot be estimated. For example, if $x_{+01} = 0$, then $u_{23(01)}$ cannot be estimated from the data.

We can still use log-linear models to in these cases, provided that we take the precaution of excluding the problematic zero elementary cells from analysis, by setting their expected values to zero a priori, or equivalently, conditioning on the fact that the marginal counts are zero and thus that the cell values that add to them must be zero as well. Corresponding to these zero cells tvalues are non-estimable parameters in the log-linear model and these essentially are excluded from the model or set equal to zero. The resulting test statistics, X^2 and G^2, will still be asymptotically as a χ^2 random quantity, but we must adjust the degrees of freedom to account for the loss of data and parameters, using the formula

$$df' = df - z_e + z_p, \tag{14}$$

where *df* are the degrees of freedom implied by the original model, z_e the number of cells with zero estimates and z_p the number of parameters that cannot be estimated (see Bishop, Fienberg & Holland (1975) and Fienberg (1980) for details). Most of the times the adjustment will have to be done by hand, as most computer programs do not yet do this automatically.

Unfortunately, depending on the model, there exist patterns of zeros that do not result in marginal tables with zero entries but also lead to the non existence of the MLEs. For example in a $2 \times 2 \times 2$ table with zeros in the (1,1,1) and (2,2,2) cells, i.e.,

0	

	0

under the no-2nd-order interaction model [12][13][23], the two cells with zero entries are constrained to be zero. The problem associated with zeros depends only on the *location* of the zeros, and not on the magnitude of the non-zero cells. There are many more complex examples of these kinds of 'zero cell" problems but no trivial characterization exists. Tools from algebraic geometry do, however, provide a vehicle for understanding and dealing with the problem, e.g., see Rinaldo, (2005), Ericksen et al. (2006), and Dobra et al. (2008).

2.5 Computer Programs for Log-linear Model Analysis

Most standard statistical packages have one or more programs or sets of routines that are useful for implementing the methodology described above. In particular SAS, STATA, and R both have generalized linear model routines, and R also has multiple versions of iterative proportional fitting and separate programs for assessing goodness-of-fit.

MIM is a software system for graphical modeling written for the windows operating system, available from http://www.hypergraph.dk/, and it is compatible with graphical search methods described in Edwards (2000). MIM is usefully for generating the graphs that go with the models it fits. A version of MIM is also available in a special R library, but again it runs only under the Windows operating system.

As we noted in the previous subsection, none of the existing log-linear model or generalized linear model programs deals especially well with the problems of sparsity, i.e., the identification of problematic zero cells and the attendant corrections required for computing degrees of freedom. Fienberg and Rinaldo (2007) give further details of what happens in such circumstances.

3 Application to ABC Company Survey Data

To illustrate the log-linear methodology summarized in the preceding section, we turn to the second section of the ABC 2010 Customer Satisfaction Survey. This section asks customers their evaluation (1 to 5) about a series of characteristics related to their level of satisfaction with the equipment provided by the company (Table 1). The data form a 5^5 contingency table, with a total of 3,125 cells but with only n=220 observations—a sparse table.

In order to reduce the sparsity of the resulting table we recoded the answers as binary variables by grouping the first three levels (1 to 3) and the last two (2,3) into single categories. Table 2 shows the resulting 5-way contingency table. Most of the alternative choices for collapsing categories did not adequately address the sparsity issue.

Label	Description
A	The equipment's features and capabilities meet your needs
B	Improvements and upgrades provide value
C	Output quality meets or exceeds expectations
D	Uptime is acceptable
E	Overall satisfaction with the equipment

Table 1: Variables in section Equipment

		B	No				Yes			
		C	No		Yes		No		Yes	
		D	No	Yes	No	Yes	No	Yes	No	Yes
E	A									
Low	No		27	10	6	2	4	8	1	3
	Yes		7	1	4	3	4	1	2	2
High	No		0	1	0	1	0	4	2	5
	Yes		3	10	4	21	8	11	7	65

Table 2: Data from Equipment section

We begin our search for a descriptive model by fitting each of the log-linear models that include all the uniform order terms (Table 7), as well as all of the partial association models (Table 4). We note that two of the 4-way margins contain one zero ([ACDE] and [ABDE]), thus we must be very careful with models that include terms that require those configurations.

In the absence of zeros, all the partial association models in Table 4 would have 8 degrees of freedom. However, six of them depend on either margin [ACDE] or [ABDE] (models 1, 2, 6, 7, 8 an 9) and one (model 5) on both. Thus we need to adjust the degrees of freedom to assess the significance of any goodness-of-fit test statistic. For the first group, the marginal zero prevents us from estimating one of the parameters, while forcing 2 elementary cells to have zero expected counts; thus we have to adjust the degrees of freedom by subtracting 1. In the case of model [ACDE][ABDE], the marginal zeros prevents the estimation of 2 parameters, while producing 3 expected elementary counts with zeros; then we have to adjust the degrees of freedom again by subtracting 1, producing 7 degrees of freedom.

Inspecting the results reported in Table 7, we see that we need at least some 1st-order interactions to obtain a well fitting model. Results in (Table 4) suggest to start with a model that incorporates the two-way interactions u_{AE}, u_{BE}, u_{DE}, u_{AC}, and u_{CD}. The minimal model that includes those interactions, [AE][BE][DE][AC][CD], does not fit the data very well by itself ($G^2 = 33.92, d.f. = 21$), so we will use it as a starting point for a model search.

Model	G^2	d.f.	p-value
All 1-way	278.0422	26	0
All 2-way	18.96407	16	0.2705308
All 3-way	4.41867	6	0.6202117
All 4-way	0	0	–

Table 3: All k-way interactions models for the Equipment data. (Note that model with all 4-way interactions is saturated, with 0 degrees of freedom, due to constrains induced by zeros in the margins corresponding to the minimal configurations).

	Model	Test	G^2	d.f.	p-value
1	$[BCDE][ACDE]$	$u_{AB} = 0$	6.913625	7	0.4379284
2	$[BCDE][ABDE]$	$u_{AC} = 0$	14.84056	7	0.03809831
3	$[BCDE][ABCE]$	$u_{AD} = 0$	11.58159	8	0.1708711
4	$[BCDE][ABCD]$	$u_{AE} = 0$	60.14289	8	4.36966×10^{-10}
5	$[ACDE][ABDE]$	$u_{BC} = 0$	4.828682	7	0.6808607
6	$[ACDE][ABCE]$	$u_{BD} = 0$	11.54538	7	0.1165359
7	$[ACDE][ABCD]$	$u_{BE} = 0$	15.69821	7	0.02802116
8	$[ABDE][ABCE]$	$u_{CD} = 0$	14.83598	7	0.03816039
9	$[ABDE][ABCD]$	$u_{CE} = 0$	8.978162	7	0.2542277
10	$[ABCE][ABCD]$	$u_{DE} = 0$	27.40129	8	0.0006025608

Table 4: All partial association models for the Equipment data.

We used the computer program MIM and its bidirectional stepwise procedure based on conditional G^2 tests to search in the space of graphical models. We find a good fitting model in just one step, by adding the term corresponding to the configuration [BC]. The final model, [AE][BE][BC][DE][AC][CD], provides a good fit to the data, with $G^2 = 28.42$ and 20 degrees of freedom. The corresponding graph is shown in Figure 1.

We can repeat the model selection procedure using AIC as our criterion for assessing fit, and we get the sequence of models $[AE][BE][DE][AC][CD] \rightarrow [CDE][BE][ACE] \rightarrow [CDE][BDE][ACE]$. The selected model also fits the data very well ($G^2 = 17.63$, df=16), but it is much more complex, thus we choose to retain the simpler model. A stepwise search based on the BIC criterion results in the same model selected by conditional G^2 tests.

Since our final model,[AE][BE][BC][DE][AC][CD], is graphical we can read some of interesting structural features directly from the graph in Figure 1. First, we see that given variables C and E, all the rest are mutually independent. This means that for each combination of the overall satisfaction with the equipment and whether or not the output quality meets the client's expectations, the rest of the characteristics measured by the other ques-

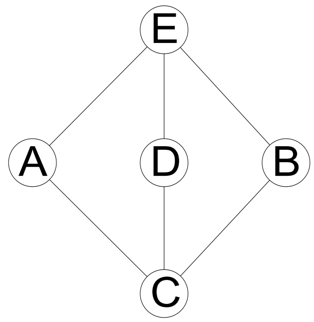

Figure 1: Graphical representation of the final model, [AE][BE][BC][DE][AC][CD], for the equipment data.

tions are mutually independent. Second given the answers to questions A, B and D, the overall level of satisfaction with the equipment turns out to be independent of whether or not the output quality meets the client's expectations.

As a further illustration, we analyze the overall satisfaction questions together. Table 5 show the selected variables that result in a $5^6 = 15,625$ contingency table. Again, because this is even sparser than in the previous illustrative extract, we dichotomize the variables by grouping the first three levels (1,2,3=Low) and last two levels (4 and 5 = High). Table 6 shows the resulting cross-classification.

Even the collapsed 2^6 table presents challenges for our analysis, with 3 out of the 20 3-way margins containing zero counts (configurations [ADE], [BDF] and [DEF]), 10 out of the 15 4-ways margins with zeros, and all of the 5-way margins containing zeros. As a consequence, we have to be extra careful when fitting high-order models. For this same reason, this time we do not try to fit the partial association models to begin our model search.

Label	Description
A	Overall satisfaction level with equipment
B	Overall satisfaction level with sales support
C	Overall satisfaction level with technical support
D	Overall satisfaction level with training
E	Overall satisfaction level with supplies
F	Overall satisfaction level with terms conditions and pricing

Table 5: Overall satisfaction variables

By fitting all the models that include all the same order terms, we can start our search for a good model by examining models that include 2-way interactions.

			D	High				Low			
			E	High		Low		High		Low	
			F	High	Low	High	Low	High	Low	High	Low
A	B	C									
High	High	High		22	9	6	2	0	2	1	3
		Low		2	1	0	3	0	1	1	3
	Low	High		2	8	3	7	0	1	0	5
		Low		2	2	0	2	0	0	0	3
Low	High	High		2	5	0	7	0	0	1	1
		Low		0	1	0	2	0	0	0	1
	Low	High		0	1	0	8	0	0	0	6
		Low		2	4	1	10	0	0	0	12

Table 6: Data from Overall satisfaction questions

k	Deviance	d.f.	p-value
1	187.7329	57	7.771561e-16
2	42.2594	42	0.4597639
3	7.085046	9*	0.6283

(*) adjusted for sparsity

Table 7: All k-way interactions models for Overall Satisfaction data, for $k = 1, 2$ and 3.

Starting from the independence model ([A][B][C][D][E][F]), we used MIM to perform a bidirectional stepwise search using conditional likelihood ratio tests, and the AIC and BIC criteria. The final models are summarized in Table 8. Note that again we have had to adjust the degrees of freedom due to the presence of a zero in the [DEF] margin. This time, stepwise based on conditional tests of the form ΔG^2 and AIC select the same model, while BIC selects a different, more parsimonious one, but whose fit is not so good. Furthermore, these two models are nested and the conditional test between them gives conditional test value of $\Delta G^2 = 24.368$ with 13 degrees of freedom, suggesting that the inclusion of the extra terms has a significant effect. Thus we select model [DEF][BEF][AEF][AC] and display its independence graph in Figure 2

We can read some interesting structural features of the data directly from the corresponding independence graph. First, given satisfaction with the equipment (A), the overall satisfaction with technical support (C) is independent from all other items. Second, joint knowledge of the satisfaction level with supplies (E) and conditions and pricing (F) makes satisfaction with sales support (B) and satisfaction with training (D) independent from all other items. Furthermore, it also makes the pair of variables, A and C, independent from the rest of the variables.

Criterion	Final Model	G^2	d.f.
ΔG^2 and AIC	[DEF][BEF][AEF][AC]	54.910	39*
BIC	[EF][DE][BF][AF][AC]	79.278	52

(*) adjusted for sparsity

Table 8: Final models selected by stepwise search using conditional likelihood ratio tests, as well as the AIC and BIC criteria for the overall satisfaction data.

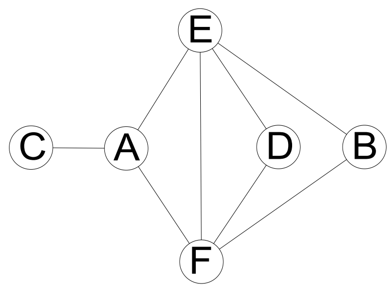

Figure 2: Graphical representation of the final model, [DEF][BEF][AEF][AC], for the overall satisfaction data.

4 Related Methodologies

As we noted in the introduction, while log-linear models are often useful in exploring relatively large sparse contingency tables, they typically can only cope with subsets of variables from lengthy survey questionnaires. A powerful extension to these models involves the introduction of latent categorical variables, and then use the model of conditional independence of the observed (or manifest variables) given the latent variable. These *latent class models* often provide a more parsimonious interpretation for the relationships observed in the contingency table, and with relatively small numbers of parameters. Goodman (1974,1979, 1984) is largely responsible for the basic methodology used with latent class models, but their likelihood functions involve complexities such as multiple modes and some identification problems (see Fienberg et al., 2010).

Another related class of models introduced more recently utilizes a different for of latent structure through the notion of mixed membership of individuals in idealized classes, or grade of membership, and has been applied to relatively large contingency table problems by Erosheva, Fienberg, and Joutard (2007) and Manrique-Vallier and Fienberg (2010).

Space limitations do not permit the exploration of these approaches in the context of customer satisfaction surveys here.

References

[1] Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., Wiley, New York.

[2] Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd ed., Wiley, New York.

[3] Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, 2nd ed., Wiley, New York.

[4] Bartlett, M. S. (1935). Contingency table interactions, *Supplement to the Journal of the Royal Statistical Society*, **2**, 248–252.

[5] Birch, M. W. (1963). Maximum likelihood in three-way contingency tables, *Journal of the Royal Statistical Society*, **25**, 220–233.

[6] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. with contributions by R.J. Light and F. Mosteller (1975). *Discrete Mulivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA. Reprinted by Springer (2007).

[7] Christensen, R. (1997). *Log-linear Models and Logistic Regression. 2nd Ed.* Springer, New York.

[8] Cressie,N. A. C., and Read, T. R. C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer, New York.

[9] Dobra, A., Fienberg, S. E., Rinaldo, A., Slavkovic, A. B., and Zhou, Y. (2008). Algebraic statistics and contingency table problems: Log-linear models, likelihood estimation, and disclosure limitation. In M. Putinar and S. Sullivant, editors, *Emerging Applications of Algebraic Geometry*, IMA Series in Applied Mathematics, Springer, New York, 63–88.

[10] Edwards, D. (2000). *Introduction to Graphical Modelling*, 2nd ed., Springer, New York.

[11] Eriksson, N., Fienberg, S. E., Rinaldo, A., and Sullivant, S. (2006). Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models. *Journal of Symbolic Computation*, **41**, 222–233.

[12] Erosheva, E. A., Fienberg, S. E., and Joutard, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *The Annals of Applied Statistics*, **1**, 502–537.

[13] Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data*, MIT Press, Cambridge, MA, 2nd ed., Reprinted by Springer (2007).

[14] Fienberg, S. E., Hersh, P., Rinaldo, A., and Zhou, Y. (2010). Maximum likelihood estimation in latent class models for contingency table data. In P. Gibilisco, E. Riccomagno, M. P. Rogantin, and H. P. Wynn, eds., *Algebraic and Geometric Methods in Statistics*, Cambridge University Press, New York, 31–66.

[15] Fienberg, S. E. and Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, **137**, 3430–3445.

[16] Goodman L. A. (1969). On partitioning χ^2 and detecting partial association in three-way contingency tables, *Journal of the Royal Statistical Society*, Series B **31**, 486–498.

[17] Goodman L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika*, **61**, 215–231.

[18] Goodman L. A. (1979). On the estimation of parameters in latent structure analysis, *Psychometrika*, **44**, 123–128.

[19] Goodman, L. A. with contributions by C. C. Clogg and a foreword by O. D. Duncan (1984). *The Analysis of Cross-classified Data Having Ordered Categories*, Harvard University Press, Cambridge, MA.

[20] Goodman, L. A. with contributions by J. A. Davis and J. Magidson (1978). *Analyzing Qualitative/Categorical Data: Log-linear Models and Latent-structure Analysis*, Abt Books, Cambridge, MA.

[21] Haberman, S. J. (1974). *The Analysis of Frequency Data*, University of Chicago Press, Chicago, IL.

[22] Lauritzen, S. F. (1996). *Graphical Models*, Oxford University Press, New York.

[23] Manrique-Vallier, D., and Fienberg, S. E. (2010). Longitudinal mixed-membership models for survey data on disability, *Longitudinal Surveys: from Design to Analysis, XXV International Methodology Symposium*, Statistics Canada.

[24] Pearson, K. (1900). On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling, *Philosophical Magazine*, **50**, 59–73.

[25] Rinaldo, A. (2005). *Maximum likelihood estimates in large sparse contingency tables*, Ph.D. thesis, Department of Statistics, Carnegie Mellon University.

[26] Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley, New York, 1990.

[27] Wilks, S. S. (1935). The likelihood test of independence in contingency tables, *Annals of Mathematical Statistics*, **6**, 190–196.

[28] Yule, G. U. (1900). On the association of attributes in statistics: with illustration from the material of the childhood society, &c., *Philosophical Transaction of the Royal Society*, **194**, 257–319.