TWO

Basic Concepts and Procedures in Single- and Multiple-Group Latent Class Analysis

Allan L. McCutcheon

Latent class analysis is frequently used when the researcher has a set of categorically scored observed measures that are highly interrelated. The latent class model (LCM) – which is often characterized as the categorical data analog to factor analysis – is most appropriately used when the observed indicator variables are associated because of some underlying unobserved factor rather than being causally related.[1] For example, the correctness (incorrectness) of answers to questions on an exam may be highly interrelated as a result of mastery; those who have mastered the material will tend to answer items correctly, and those who have yet to master the material will tend to answer them incorrectly. Thus, in a sufficiently large sampling of student exams, we would anticipate that those who correctly answered question 1 would also be more likely to have correctly answered questions 2, 3, and so forth, yielding a clear association among the "variables" (exam questions). Frequent instances of this kind of association can be found in the social and behavioral sciences (e.g., self-esteem, religiosity, partisan identification, consumer loyalty).

Since the early 1990s, the LCM has emerged as a powerful new method for the analysis of categorically scored data. As is clear from the range of applications in this volume, the range of topics to which the LCM can be fruitfully applied is quite broad. A major reason for the utility of the LCM is that we can use two quite different, although highly interrelated and completely equivalent, parameterizations for this model: probabilistic and loglinear. In this chapter, these two parameterizations will be considered. In addition to exploring the differences between them, we will examine the isomorphisms of the two parameterizations and show how the selection of one may focus the analyst's attention on certain aspects of the model, whereas selection of the other may focus attention on a somewhat different aspect of the model.

Although the alternative parameterizations of the LCM are clearly a strength of the LCM and make the model highly flexible and useful across a wide range of research applications, they also present a set of special challenges. In particular, restrictions that can be readily imposed on one LCM parameterization may be difficult, or even impossible, to replicate in the other parameterization. Thus, although the basic, unrestricted latent class model is identical across the two parameterizations and yields identical "fits" to observed data, restricted models with one parameterization may not be readily translated into the alternative parameterization. Consequently, the researcher may need to consider both parameterizations in order to determine which of the two – the probabilistic or the loglinear – is most appropriate to the research problem.

In this chapter, we first examine the basic unrestricted latent class model in its probabilistic and loglinear parameterizations. This section focuses on model equivalence and interpretation across the two parameterizations. In the second section, model estimation issues are discussed. There are two main approaches to maximum-likelihood estimation (MLE) for LCMs: the expectation–maximization (EM) algorithm and the Newton–Raphson (NR) algorithm. These two iterative algorithms will be briefly introduced, along with some associated issues such as identification and problematic solutions. In the third section, model evaluation criteria are considered. These criteria are critical for guiding our decisions about the appropriateness of accepting a specific model as adequately characterizing the associations observed in the data. In the fourth section, we focus on the model restrictions for each of the parameterizations. Model restrictions are critical for establishing the equivalence of latent class structures when we wish to examine an identical set of categorically scored measures in two or more samples. We will discuss cases in which restrictions that are complicated under one parameterization are readily managed under the other. In the final section, we briefly examine the simultaneous latent class models (SLCMs) in which LCMs are compared in two or more populations. These SLCMs have proven highly useful in comparative research, as well as in the examination of cross-time trends.

1. PARAMETERIZATIONS OF THE BASIC LATENT CLASS MODEL

In this section, we examine the two parameterizations of the basic LCM and demonstrate the equivalence of these two parameterizations. We examine each of the parameterizations in some detail, showing how the common properties of the two parameterizations (e.g., local independence) are manifest in each.

A. Probabilistic Parameterization

Perhaps the most widely used and most intuitively grasped parameterization of the latent class model is the probabilistic parameterization. As Goodman (this volume) has discussed in detail, the probabilistic parameterization of the basic unrestricted LCM is characterized by two types of categorical variables – observed (manifest) indicator variables and unobserved (latent) variables – and two types of parameters: latent class and conditional probabilities. The LCM postulates that the relationship between any two manifest variables is accounted for by the latent variable; this is typically referred to as the *axiom of local independence*. Thus, for an LCM with a single latent variable (X) and four manifest variables (A, B, C, and D), we can formally express the basic LCM as the product of the latent class probabilities and conditional probabilities:

$$\pi_{ijklt}^{ABCDX} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}, \tag{1}$$

where the latent class probability (π_t^X) is the probability[2] that a randomly selected observation in the sample is located in latent class t, and the conditional probabilities (e.g., $\pi_{it}^{A|X}$) are the probabilities that a member of latent class t will be at a specified level of an observed indicator variable. For example, if our latent variable (X_t) is a measure of religiosity ($t = 1$, religious; $t = 2$, not religious), the first indicator variable (A_i) might be a self-report of church attendance in the previous week ($i = 1$, yes; $i = 2$, no). Thus, the conditional probability $\pi_{11}^{A|X}$ is the probability that a randomly selected religious (i.e., latent class 1) respondent would report having attended church in the previous week.

Within the LCM, hypotheses are tested by imposing restrictions and determining how these restrictions affect the fit of the model to the data. Goodman (1974a, 1974b) has shown that the LCM can be made identifiable by imposing a set of logical constraints (restrictions) on the basic LCM. Thus, for the basic LCM with a single latent variable (X_t) and four observed indicator variables (A_i, B_j, C_k, and D_l), we can express the restrictions as

$$\sum_t \pi_t^X = \sum_i \pi_{it}^{A|X} = \sum_j \pi_{jt}^{B|X} = \sum_k \pi_{kt}^{C|X} = \sum_l \pi_{lt}^{D|X} = 1.0. \tag{2}$$

The restriction that $\sum_t \pi_t^X = 1.0$ requires that the latent classes sum to

1.0 – that there is a latent class for each of the possible response patterns observed in the data. The remaining restrictions require that each of the indicator variables sum to one within each of the T classes.

Consider, for example, the data collected by Stouffer and Toby (1951) in their study of role conflict. As noted elsewhere in this volume (see Goodman), in February 1950 these researchers asked a group of Harvard and Radcliffe students about a set of situations "involving conflict between obligations to a friend and more general social obligations" (1951, p. 396). Students were presented with four scenarios in which either they or their friend confronted each of four role conflicts. For students responding to the friend's role conflict, each scenario was of the following type (item A):

> Your close friend is riding in a car that you are driving, and you hit a pedestrian. He knows that you were going at least 35 miles an hour in a 20-mile-an-hour zone. There are no other witnesses. Your lawyer says that if your friend testifies under oath that the speed was only 20 miles an hour, it may save you from serious consequences. What right do you have to expect him to protect you?

The other scenarios involve similar role conflicts for the respondent's close friend who is a drama critic (item B), an insurance doctor (item C), and a member of a board of directors (item D). An equal number of students were given scenarios with the same four situations, but modified so that the students themselves were in the role conflict position. The distributions of the responses for the four scenarios in these two situations are reported in Table 1.

Table 1. Responses to Four Role Conflict Scenarios for Ego and Ego's Close Friend

			Dilemma			
Items			**Ego Faces**		**Ego's Friend Faces**	
A	***B***	***C***	**Item *D* (+)**	**Item *D* (−)**	**Item *D* (+)**	**Item *D* (−)**
+	+	+	20	2	20	3
+	+	−	6	1	4	3
+	−	+	9	2	23	3
+	−	−	4	1	4	2
−	+	+	38	7	25	6
−	+	−	25	6	15	6
−	−	+	24	6	29	5
−	−	−	23	42	31	37

Source: Stouffer and Toby (1951).

As the data in Table 1 indicate, there are 16 (2^4) response patterns for each of the two dilemma situations (Ego and Ego's Friend). In general, if there are k dichotomous items, there are 2^k possible response patterns. Each of the 16 response patterns was selected by one or more of the student respondents for each of the dilemma situations. In this section, we focus only on the responses to the situation confronting Ego. We consider both situations later, when we take up the issue of multisample latent class analysis.

One rather obvious question that emerges from Table 1 is whether we must consider the 16 response patterns for each scenario as representing 16 distinct types, or whether some lesser number of patterns might account for the observed distribution of responses. If we allow for measurement error in each of the four indicator variables, we might view some of the responses as the result of misclassification caused by the measurement error, rather than true response types.

Setting aside, for a moment, the issue of how we determine if a particular latent class model is an adequate representation of the observed data, we can present the two-class LCM for the data in the 16 cells of the leftmost two columns in Table 1. The conditional probabilities (e.g., $\pi_{it}^{A|X}$) for respondents saying that their friend had a right to expect them to violate their role responsibilities (i.e., the positive or *particularistic* response), along with the latent class probabilities (π_t^X), for the probabilistic parameterization of the unrestricted basic LCM are reported in Table 2.

As the data in Table 2 indicate, approximately three-quarters (0.7206) of the sample are Class 1–type respondents and, by the restriction that $\sum_t \pi_t^X = 1.0$ specified in Equation (2), about one-quarter ($1.0 - 0.7206 = 0.2794$) of the sample are Class 2–type respondents. We can use the conditional probabilities to characterize the two latent classes in much the same way that factor loadings are used to characterize the factors in factor

Table 2. Latent Class and Conditional Probabilities for Ego's Dilemma

Indicator	Class 1	Class 2
A. Passenger Friend	0.286	0.007
B. Drama Critic Friend	0.646	0.074
C. Insurance Doctor Friend	0.670	0.060
D. Board of Directors Friend	0.868	0.231
Latent Class Probabilities	0.7206	0.2794

analysis.[3] Thus, we can see that Class 1–type respondents appear to have a consistently higher probability of giving the particularistic response – that is, of saying their friend had a right to ask them to violate their role obligation – than do Class 2–type respondents. Thus, we may wish to regard Class 1 respondents as the "particularistic" respondents and the Class 1 as the "universalistic" respondents (Stouffer and Toby, 1951, pp. 395–396; Parsons, 1949, Chapter 8).[4]

Before we turn to the loglinear representation of the LCM, it is worth noting that the restrictions on the conditional probabilities of the LCM expressed in Equation (2) mean that it is not necessary to report both the probability that the respondent would choose "my friend has a right" and the probability that he or she would choose the response "my friend does not have a right," because these two probabilities must sum to 1.0. Thus, for the dichotomous response options reported here (i.e., $I = J = K = L = 2$), we need report only the probability of a single response (e.g., the particularistic response), because the probability of giving the universalistic response is equal to the difference between 1.0 and the probability of giving the particularistic response (e.g., $\pi_{21}^{A|X} = 1.0 - 0.286 = 0.714$).

B. Loglinear Parameterization

It is also possible to represent the LCM as a loglinear model (Goodman, 1974a; Haberman, 1979, especially Chapter 10). The loglinear parameterization of the unrestricted basic LCM, however, differs from the usual loglinear model in two important ways. First, the loglinear model for the LCM includes an unobserved, latent variable (X_t). Second, only the two-variable parameters between the latent variable and each indicator variable are included – all higher-order terms involving combinations among the indicator variables of the loglinear LCM are set to zero. Thus, the loglinear LCM also exhibits local independence among the indicator variables. This basic model is expressed as

$$\ln\left(f_{ijklt}^{ABCDX}\right) = \lambda + \lambda_t^X + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{it}^{AX} + \lambda_{jt}^{BX} + \lambda_{kt}^{CX} + \lambda_{lt}^{DX}. \quad (3)$$

As Equation (3) indicates, the loglinear LCM includes only the single-variable lambda parameters along with the two-variable association parameters between each of the indicator variables and the latent variable.

Table 3. Loglinear Parameters for Ego's Dilemma Latent Class Model

	Parameter	
Indicator	**Single Variable**	**Two Variable**
A. Passenger Friend	−1.472	1.016
B. Drama Critic Friend	−0.483	0.784
C. Insurance Doctor Friend	−0.509	0.864
D. Board of Directors Friend	0.169	0.771
X. Latent Class Variable	0.474	—

As with the probabilistic parameterization (and with ordinary loglinear models), it is necessary to impose a set of identifying restrictions.

$$\begin{aligned}\sum_t \lambda_t^X &= \sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_k \lambda_k^C = \sum_l \lambda_l^D = \sum_i \lambda_{it}^{AX}\\ &= \sum_t \lambda_{it}^{AX} = \sum_j \lambda_{jt}^{BX} = \sum_t \lambda_{jt}^{BX} = \sum_k \lambda_{kt}^{CX}\\ &= \sum_t \lambda_{kt}^{CX} = \sum_l \lambda_{lt}^{DX} = \sum_t \lambda_{lt}^{DX} = 0. \end{aligned} \tag{4}$$

These restrictions to the loglinear model require that the product of odds and odds ratios are 1 (i.e., that the natural logarithm is 0). This, for example, means that if you have twice the likelihood of being at level 1 relative to level 2, you have the reciprocal likelihood (0.5) of being at level 2 relative to level 1.

The estimates of the loglinear parameters for the Stouffer and Toby study are reported in Table 3. Typically, of greatest interest are the two-variable association parameters that relate each of the indicator variables to the latent variable. Although not necessarily intuitive, the lambda parameters range from negative to positive infinity, with zero indicating complete independence. Thus, the positive values for each of the four two-variable parameters indicate that the student respondents' particularistic responses are positively associated with their location as Class 1–type respondents.

An alternative and somewhat more intuitive approach to interpreting the two-variable loglinear LCM parameters is to convert them to odds ratios. For example, we might note that the estimate of 1.016 for latent variable item A (passenger friend) corresponds to an estimated log cross-product ratio of

$$\tau^{AX} = 4\lambda^{AX} = 4.064$$

with an estimated cross-product ratio of

$$e^{4.064} = 58.21.$$

Thus, we would conclude from the estimates reported in Table 3 that the odds that a respondent with a Class 1–type of latent attitude will give a particularistic response to the passenger friend indicator are 58 times as great as the odds if he or she is a Class 2–type of respondent. The other two-variable lambda coefficients may be similarly interpreted.

Finally, it is important to note that the unrestricted, basic LCMs expressed in Equations (1) and (3) are essentially equivalent, requiring the estimation of an identical number of parameters and yielding identical expected values. This equivalence can be illustrated, in part, through the equivalence between the conditional probabilities of the probabilistic parameterization and the loglinear parameters (Haberman, 1979, p. 551):

$$\pi_{it}^{A|X} = \exp\left(\lambda_i^A + \lambda_{it}^{AX}\right) \Big/ \sum_i \exp\left(\lambda_i^A + \lambda_{it}^{AX}\right). \tag{5}$$

We can use the lambda coefficients reported in Table 3 to calculate the conditional probabilities reported in Table 2. For example, we can calculate the probability that a particularistic (latent Class 1)-type of respondent would give the particularistic response to the passenger friend indicator as

$$\begin{aligned}\pi_{11}^{A|X} &= \frac{\exp(-1.472 + 1.016)}{\exp(-1.472 + 1.016) + \exp(1.472 - 1.016)} \\ &= \frac{0.633560}{0.633560 + 1.578382} = 0.286427.\end{aligned}$$

Each of the other conditional probabilities in Table 2 can be derived in an analogous manner from the loglinear lambda coefficients presented in Table 3.

The probabilistic and loglinear parameterizations each permit the researcher to test a number of interesting hypotheses the researcher might wish to test by imposing restrictions on the model parameters. However, the two parameterizations lend themselves to somewhat different hypotheses. Before we consider restricted models, a brief discussion of model evaluation will be useful.

2. MODEL ESTIMATION

There are two main alternatives to estimating the parameters of the LCM – the expectation–maximization and the Newton–Raphson algorithms – both of which are iterative, maximum-likelihood estimation approaches. As iterative approaches, both algorithms begin with a set of "start values" and proceed with a series of steps of parameter estimation and reestimation iterations until some designated criterion is reached. Usually, the "stop" criteria focuses on convergence – each additional iteration in the parameter reestimation procedure finally approaches some predesignated "small" change, and the procedure stops. As a number of excellent expositions of the EM and NR algorithms and their variants now exist (see, e.g., Everitt, 1987; McLachlan and Krishnan, 1997; see also Vermunt, 1997a, and Wedel and Kamakura, 1998), only a brief discussion of these approaches is considered here.

The EM algorithm has become one of the most widely used approaches to LCM estimation. Certainly among the contributing factors to the popularity of the EM algorithm are its robustness with respect to the initial (start) values – these can be quite distant from the final estimates and will still reach at least a local maximum – and its relative ease to program. Two of the more frequently mentioned disadvantages of the EM algorithm are that this approach may require a large number of iterations to reach a final solution and that the EM algorithm does not directly provide estimates of the standard errors. The rapid increase in computational speed has reduced the first of these problems.

The EM algorithm consists of two steps. In the first step – the expectation, or E, step – the expected value of the log of the likelihood function is computed, conditional on the observed data and the initial parameter estimates. In the second step – the maximization, or M, step – the function is maximized in order to give updated values of the parameter estimates. These new estimates of the parameters replace the initial estimates and the algorithm returns to the E step. The algorithm continues in this iterative manner until the changes in either the parameter estimates or the changes in the likelihood function (or its logarithm) reach some predefined level of precision, at which point the iterative process halts.[5]

The NR method also uses an iterative approach to produce maximize likelihood estimates of the LCM parameters. The NR method has the advantage of being relatively fast and producing the standard errors of the parameter estimates as a by-product of parameter estimation. Among the disadvantages of this method are the need to invert the Hessian matrix at every iteration and the requirement that the initial estimates must be close to the final estimates or the matrix may become negative definite and thus cannot be inverted. Consequently, when the start values used with this method are too far from the final solution, the NR algorithm may not converge on a set of final estimates.

The NR method begins with a set of initial parameter values (θ) and improves on these values by modifying them by the product of the inverse of the Hessian matrix (H) and the gradient vector (g) of derivatives of the log-likelihood function:

$$\theta_{i+1} = \theta_i - H_i^{-1} g_i. \tag{6}$$

The parameter estimates are updated at the end of each iteration. As noted earlier, the convergence of this method is quite fast when the initial parameter estimates (θ) are near the ML estimates.

An important problem with likelihood approaches is that they can have several local optima. That is, both the EM and NR estimation methods may converge to local optima and not the true (global) maximum of the likelihood function. To determine whether this has occurred, the researcher can repeat the procedure by using different start values. If the results of the repeated estimation nets different parameter estimates and a higher likelihood (i.e., a lower value for the test statistic G^2; see the next section), then the result with the highest likelihood (the lowest G^2 value) is the MLE.

Figure 1 is a hypothetical illustration of the local optima issue for a parameter space from the probabilistic parameterization. Because probabilities are bounded by zero and one, no parameter estimate may lie

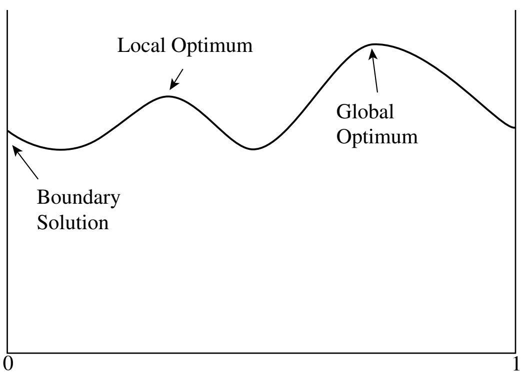

Figure 1. Maxima and boundaries.

"outside" of those values. In Figure 1 we see that there are three maxima – one at 0, one at about 0.4 and one at about 0.8 – although only one of these is the global maximum (at about 0.8). Thus, if the start value for the parameter estimate is too close to the boundary 0, the iterative process will move this particular parameter estimate to the local maximum (0), but it will not provide the true (global) maximum estimate. Thus, it is important to note that a consideration of the start values and the final estimates – especially if one or more of these estimated parameter values lie on the boundary of the parameter space – may be of critical importance, and should be given considerable attention.[6]

A final cautionary note should be made with respect to model estimation. As will be discussed in the next section, the issue of model identification – whether there is sufficient information in the observed cross-tabulation to estimate the parameters of the proposed model – is of crucial importance in latent class analysis. One practical approach to exploring whether the model is identified is to begin with quite different start values for the each of the model parameters and estimate the same model several times. If the final estimates of the model parameters are quite different for the several analyses but the estimated frequencies and the chi squares are the same for each of the analyses, it is a sure sign that the specified model is not identified. A more complete discussion of model identification issues is presented by Goodman (1974a, 1974b) and Clogg (1981; see also McCutcheon, 1987).

3. MODEL EVALUATION

Several model evaluation criteria have become more or less standard in the evaluation of LCMs. All of these criteria are evaluations of how well the expected cell counts under the model hypothesis replicate the originally observed cell counts. Four in particular – the Pearson chi square (X^2), the likelihood ratio chi square (L^2), the Akaike information criteria (AIC), and the Baysian information criteria (BIC) – have become widely used and appear throughout this volume. In this section, we briefly consider the basics of these four criteria.

Each of the four evaluation criteria relies upon a comparison between the expected cell frequency count (f_{ijkl}) given by the estimated LCM parameters and the actual (observed) cell frequency count (F_{ijkl}) found in the sample data. LCMs that lead to expected cell frequencies that are too far from the observed cell frequencies are deemed unacceptable or implausible, whereas models that yield expected cell counts that are similar to what has actually been observed are believed to be more plausible or acceptable. Models with more parameters (e.g., a model with more latent classes) usually provide a better "fit" with the data; that is, the expected cell frequencies are typically closer to the observed cell frequencies for models with more parameters. More parsimonious models tend to have a somewhat poorer fit. Thus, the usual task is to find the most parsimonious model – one with the fewest parameters[7] – that has an acceptable fit to the observed data.

As is clear from the two LCM parameterizations expressed in Equations (1) and (3), however, neither directly yields expected frequency counts for the cells of the observed contingency table. Instead, each of the outcomes on the left side of Equations (1) and (2) include the latent variable (X). By summing over the T classes of the latent variable, however, we can obtain the LCM's expected frequency counts for the observed table. For example, for Equation (1), we see that

$$\pi_{ijkl}^{ABCD} = \sum_t \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}. \tag{7}$$

By multiplying the expected joint probability of being at level *ijkl* in the *ABCD* contingency table (π^{ABCD}) by the sample size (here, $N = 216$), we obtain the expected cell count (f_{ijkl}).

These expected cell count values can be compared with the observed cell counts (F_{ijkl}) in order to evaluate the model fit. For example, the expected values can be used with the Pearson chi-square statistic to evaluate how well the expected cell counts from the specified latent class model compare to the observed distribution.

$$X^2 = \sum_{ijkl} \frac{(F_{ijkl} - f_{ijkl})^2}{f_{ijkl}}. \tag{8}$$

The degrees of freedom (df) for the chi-square statistic of the unrestricted LCM are typically calculated as

$$\text{df} = (IJKL - 1) - [T(I + J + K + L + d) - 1], \tag{9}$$

where I, J, K, and L represent the number of levels for the respective indicator variables. In the case of four dichotomies, $I = J = K = L = 2$, T indicates the number of latent classes in the LCM, and d is the number of indicator variables minus 1. In the current example, we obtain $15 - [2(5) - 1] = 6$ df for the model test. Although Equation (9) works in nearly all instances, a necessary and sufficient condition for determining the local identifiability of an LCM involves determining the rank of the

Hessian matrix of second-order partial derivatives. As mentioned earlier, more complete discussions of model identification issues are presented by Goodman (1974a, 1974b) and Clogg (1981; see also McCutcheon, 1987).

The likelihood ratio chi-square test (G^2) provides a general-purpose test for evaluating models and, as Goodman (1968, 1970) and Agresti (1990) note, is of special importance for comparing alternative models. This latter point will be of considerable interest as we take up the issues of hypothesis testing and model restrictions in the following sections. The likelihood ratio chi-square statistic is a function of the ratio of the observed to expected cell counts:

$$G^2 = 2 \sum_{ijkl} F_{ijkl} \ln(F_{ijkl}/f_{ijkl}) \tag{10}$$

As with the Pearson chi-square statistic, the likelihood ratio chi-square statistic has asymptotic chi-square distributions with respect to the degrees of freedom, and thus the probability of acceptance of the alternative hypothesis – the LCM model – can be determined. Unlike with the Pearson chi-square statistic, however, when one of two nested models is true, the difference between the likelihood ratio chi-square statistics for the two models can expressed explicitly as the conditional likelihood ratio chi-square statistic with degrees of freedom equal to the difference between the degrees of freedom for the two models, thus allowing statistical comparisons of successive models. The conditional likelihood (G^2) test statistic follows a chi-square distribution when the "baseline" (i.e., less restricted) model is acceptable and when the more restricted model is nested within the less restricted model. This *partitioning* of the likelihood ratio chi square has contributed to its widespread adoption in statistical modeling.

Although the Pearson chi-square and the likelihood ratio chi-square statistics are used throughout this volume and throughout the latent class analysis literature, alternative model evaluation criteria for LCMs have been identified in recent years – notably, information criteria. These alternative criteria avoid some limitations of the traditional X^2. First, the chi-square statistics tend to be conservative when sample sizes are large; that is, it is difficult to reject the significance of even quite modest parameters when the sample size is large. Second, LCMs can require the estimation of a rather large number of parameters even for models of modest size.

Information criteria approaches penalize the likelihood for the increased number of parameters required to estimate more complex (i.e., less parsimonious) models. Because more parameters (i.e., more complex models) yield a greater likelihood, each of the information criteria penalizes the likelihood by reducing it by a function of the increased number of estimated parameters. The two most widely used of the information evaluation criteria are the AIC (Akaike, 1974) and the BIC (see Rafftery, 1995):

$$\text{AIC} = G^2 - 2\text{df}$$
$$\text{BIC} = G^2 - \text{df} * [\ln(\text{N})],$$

where df is the number of degrees of freedom and *N* is the sample size. Thus, we see that the AIC penalizes the G^2 by the total number of parameters required for model estimation (by subtracting two times the number of degrees of freedom), and the BIC penalizes the G^2 by both the total number of parameters required for model fit and the total sample size (by subtracting the natural log of the sample *N* times the number of degrees of freedom). Consequently, models with lower AICs and BICs are preferred to those with higher values for these criteria.

In Table 4, these four evaluation criteria are presented for the LCM in the current example. As these data indicate, the independence model – that is, the model in which the four indicator variables are independent of one another – is clearly an unacceptable hypothesis; all of the evaluation criteria are unequivocal with respect to this conclusion. The unrestricted two-class LCM – that is, the model in which these four indicator variables are independent of one another within the two classes of the latent variable – is clearly an acceptable solution; all of the evaluation criteria support this conclusion.

As nearly all of the advances to the basic LCM presented in this volume demonstrate, there are a wide range of interesting variations on the basic LCM that make this modeling approach a powerful tool for data analysis and research. Among the many modifications are restrictions on the parameters of the LCM. In the next section, we examine some of the basics of these restrictions.

Table 4. LCM Evaluation Criteria for Ego's Dilemma Data

Model	X^2	G^2	AIC	BIC	df
Independence	104.11	81.08	59.08	21.96	11
Two-Class LCM	2.72	2.72	−9.28	−29.53	6

4. RESTRICTED LATENT CLASS MODELS

A variety of hypotheses may be tested by restricting the parameters of the LCM. Because the two parameterizations are essentially equivalent, a number of parameter restrictions are equivalent across the two parameterizations. However, there are also several somewhat substantively different parameter restrictions in the loglinear parameterization compared with the probabilistic parameterization. Thus, we will examine the two parameterizations separately, beginning with the probabilistic parameterization.

There are two basic types of restrictions that we can impose on the parameters of the probabilistic parameterization of the LCM: deterministic and equality restrictions. When these restrictions are applied to the conditional probabilities across classes and across indicators, as well as to the latent class probabilities across classes, these two general types of restrictions yield a number of different hypotheses we might wish to test. Moreover, as we will note later, imposing restrictions on parameters "frees up" additional model parameters, which may allow us to fit LCMs with more latent classes than are permitted according to Equation (9). We must proceed with caution when imposing model restrictions because by imposing certain restrictions we may actually turn an identified model into an unidentified model.

Consider the use of equality restrictions in the case of Ego's dilemma presented earlier. We may wish to test the *parallel* indicators hypothesis (Goodman, 1974a, 1974b; Hagenaars, 1990), in which we hypothesize that two or more of the indicator variables have identical error rates with respect to each of the latent classes. As with all of the parameter restrictions, we should suggest our hypotheses prior to reviewing the outcome of prior model estimation because post-hoc hypothesizing capitalizes on chance findings. For purposes of illustration, however, let us assume that we had earlier hypothesized that indicators B and C (Drama Critic Friend and Insurance Doctor Friend) were parallel indicators – that is, that these two indicators have identical error rates with respect to the two latent classes. We can formally test this hypothesis by imposing the following restrictions:

$$\pi_{11}^{B|X} = \pi_{11}^{C|X}, \qquad \pi_{12}^{B|X} = \pi_{12}^{C|X}. \tag{11}$$

By imposing these two restrictions, we reduce by two the number of parameters that must be estimated for the model. Thus, this model will yield two additional degrees of freedom relative to the unrestricted two-class model. The evaluation criteria for the estimated LCM with the parallel indicators restriction are reported in Table 5.

Table 5. LCM Evaluation Criteria for Ego's Dilemma Data

Model	X^2	G^2	**AIC**	**BIC**	**df**
H_1: two-class LCM	2.72	2.72	−9.28	−29.53	6
H_2: $H_1 + B$ & C parallel indicators	2.84	2.89	−13.11	−40.12	8
H_3: $H_2 + D$ equal error rate	3.60	3.65	−14.35	−44.73	9
H_4: $H_3 + A$ as perfect indicator for class 2	3.61	3.66	−16.34	−50.09	10

The data in Table 5 indicate that we can accept the hypothesis that indicator variables B and C are parallel indicators (H_2); that is, that latent Class 1– and latent Class 2–type of respondents are equally likely to err in their responses with respect to the Drama Critic Friend and Insurance Doctor Friend indicators in their respective classes. Both the Pearson (2.84) and the likelihood ratio (2.89) chi squares (with 8 df) indicate that H_2 is a plausible model. Moreover, because H_2 is nested within H_1, we can use the conditional likelihood ratio X-square test to examine whether the newly imposed restrictions are acceptable, or whether they result in an unacceptable erosion of fit to the observed data. In the current instance, the conditional difference between the two model G^2s ($2.89 - 2.72 = 0.17$, with $8 - 6 = 2$ df) clearly indicates that the newly hypothesized model (H_2) produces only a very modest erosion of fit and is, consequently, preferred to the less parsimonious, unrestricted model represented by H_1. In addition to these X-square tests, the greater negative values for both the AIC and BIC indicate that model H_2 is preferred to H_1. Thus, we accept H_2 as our new "baseline" model.

Another equality model restriction that we might wish to examine with the probabilistic parameterization of the LCM is the *equal error rate* hypothesis (Goodman, 1974a, 1974b; Hagenaars, 1990). The equal error rate hypothesis suggests that we can impose equality constraints on the conditional probabilities to test whether an indicator variable has the same error rate across the two classes. For example, we might wish to test that indicator variable D (Board of Directors Friend) has an equal error rate for Classes 1 and 2.[8] This hypothesis can be formally stated as

$$\pi_{21}^{D|X} = \pi_{12}^{D|X}. \tag{12}$$

This hypothesis indicates that the likelihood of a particularistic-type (Class 1) respondent giving a universalistic ($l = 2$) response is equal to a universalistic-type (Class 2) respondent giving a particularistic ($l = 1$) response.

Once again, the data in Table 5 support the acceptance of this hypothesis. All of the evaluation criteria for H_3 indicate acceptance. With

9 df for this model, both the Pearson (3.60) and likelihood ratio chi squares (3.65) indicate acceptance. Also, the conditional G^2 (0.76, 1 df) indicates that the additional restriction is acceptable. Finally, the two information criteria values (−14.35 and −44.73) also indicate that H_3 is preferred over H_2. Thus, we accept the hypothesis that indicator D has equal error rates for Classes 1 and 2.

A second type of hypothesis that we can consider with respect to the LCM's conditional probabilities is that of a *deterministic model restriction*. A deterministic restriction on a conditional probability tests the hypothesis that the conditional probability equals a specific value – usually, this value is either 1.0 or 0. For example, we might wish to test the hypothesis that indicator A (Passenger Friend) is a perfect indicator of Class 2; that universalistic-type (Class 2) respondents will have a zero probability of giving a particularistic response to question A. This restriction can be formally stated as

$$\pi_{12}^{A|X} = 0. \tag{13}$$

Because the model with this restriction requires the estimation of one fewer parameter, the model will have an additional degree of freedom.

As the data in Table 5 indicate, model H_4 provides an acceptable fit to the data. The Pearson (3.61) and likelihood ratio (3.66) chi squares (10 df) are clearly acceptable. Also, the conditional G^2 (0.01 with 1 df) clearly indicates acceptance of this model. The AIC (−16.34) and the BIC (−50.09) indicate that this model is a preferable model to all of those presented in Table 5. The parameter estimates for the probabilistic LCMs of H_4 are presented in Table 6.

The parameter estimates in Table 6 reflect the three types of model restrictions on the conditional probabilities. We see, for example, that for Class 1 and Class 2 the conditional probabilities are identical for indicator

Table 6. Latent Class and Conditional Probabilities for the Restricted Ego's Dilemma LCM

Indicator	Class 1	Class 2
A. Passenger Friend	0.275	0.000[b]
B. Drama Critic Friend	0.636[a]	0.046[a]
C. Insurance Doctor Friend	0.636[a]	0.046[a]
D. Board of Directors Friend	0.852[a]	0.148[a]
Latent Class Probabilities	0.7574	0.2426

[a] Equality restriction.
[b] Deterministic restriction.

variables B and C; thus, we can conclude that these two are parallel indicators for our model. Also, we see that indicator D has equal error rates for both classes; the probability of giving a level 2 response to D among Class 1 respondents is $(1.0 - 0.852) = 0.148$, which is identical to the probability of giving a level 1 response among Class 2 respondents. Finally, we see that Class 2–type respondents have a zero probability of responding that they have a right to expect their friend to falsely report that the respondent was driving within the speed limit.

A third type of restriction that may be of interest to researchers is the *inequality restriction* on conditional probabilities (see Croon, this volume). For example, inequality restrictions could be useful when a three-class solution is found and the researcher wishes to test whether the three classes lie on a continuum, with one of the classes being intermediate to the other two "end" classes. If this were the case using the current data, for example, we may wish to test the hypothesis that

$$\pi_{11}^{A|X} \geq \pi_{12}^{A|X} \geq \pi_{13}^{A|X}, \tag{14}$$

which indicates that Class 1 respondents are the most likely to respond at level 1 of A; Class 2 respondents are more likely than Class 3 but less likely than Class 1 to respond at level 1 of A; and Class 3 are the least likely to respond at level 1 of A. As Croon notes, however, there are problems associated with the determination of the degrees of freedom for these model tests.

In principle, each of these types of probabilistic parameter restrictions – equality, deterministic and inequality – may also be imposed on the latent class probabilities. In practice, however, deterministic restrictions are rarely used; unless there exists some a priori hypothesis regarding the exact size of a class, there would be no practical use for a deterministic restriction on a latent class probability. In our current example, an equality restriction on the two latent class probabilities – although possible – appears implausible.

A. Restricted Loglinear Models

A brief consideration of Equation (5) indicates that we can replicate in the loglinear parameterization each of the equality restrictions we imposed on the conditional probabilities of the LCM. For example, the parallel indicator restriction imposed on the conditional probabilities of indicators B and C can be replicated by imposing the restrictions

$$\lambda_1^B = \lambda_1^C, \qquad \lambda_{11}^{BX} = \lambda_{11}^{CX}. \tag{15}$$

The equal error rate hypothesis imposed on the conditional probabilities of indicator D can be replicated by imposing the restriction

$$\lambda_1^D = 0. \tag{16}$$

By imposing these three restrictions on the loglinear parameterization, we obtain the model specified as H_3 in Table 5.

Imposing the deterministic restriction that the Class 2 conditional probabilities of indicator item A equals 0 and 1, however, is problematic in the loglinear parameterization. For Equation (5) to yield a conditional probability of 0 requires that the numerator on the right side of the equation also equal 0. This, of course, means that the sum of the two lambdas in the numerator would have to equal negative infinity. Thus, to test this particular deterministic restriction the loglinear parameterization, our hypothesis must include one or more *structural zeros*.

In contrast to the probabilistic parameterization, the loglinear parameterization tends to focus attention on the association between the indicator variables and the latent variable. For example, with the loglinear restriction, we can impose an equal association restriction to test the hypothesis that two or more indicator items have an identical pattern of association with the latent variable. When either the indicator items or the latent variable has more than two levels, the loglinear parameterization permits linear restrictions on the two-variable association parameters.

As Clogg noted (1988, 1995; see also McCutcheon, 1996; Heinen, 1996), the linear restrictions that were first applied to the usual loglinear models in which all variables are observed (Haberman, 1974; Goodman, 1979; Clogg and Shihadeh, 1994) can also be applied to the loglinear parameterization of the LCM. For an example of this approach, we consider a set of four dichotomous items about the approval of social reasons for abortion taken from the 1982 General Social Survey (Davis and Smith, 1982). In this survey, a sample of American respondents were asked about their approval of legal abortions for a woman if she is single and does not wish to marry the man (A), if the woman is married and too poor to have any more children (B), if the woman wants an abortion for any reason (C), and if the woman is married and does not want any more children (D). The observed cell counts for the responses to these items are reported in Table 7.

We begin by fitting an unrestricted two-class latent model to these data, as this is the simplest and most parsimonious model. As the model evaluation criteria reported in Table 8 indicate, however, the unrestricted two-class LCM does not provide an acceptable fit to the observed data.

Table 7. Responses to Four Social Reasons for Abortion Items (1982 General Social Survey)

Items			Approval of Legal Abortion	
A	***B***	***C***	**Item *D*(+)**	**Item *D*(−)**
+	+	+	567	11
+	+	−	62	32
+	−	+	17	10
+	−	−	22	38
−	+	+	18	13
−	+	−	42	62
−	−	+	9	11
−	−	−	28	719

Thus, we must reject this model as implausible. The next model we might wish to test is the unrestricted three-class model. As Goodman (1974a) notes, however, this model is not identified, even though Equation (9) indicates that this model would be identified with a single degree of freedom for the model test.

An alternative to the unrestricted three-class model is a three-class model in which the three-classes are linearly ordered from more approving to less approving. We can estimate this model by modifying Equation (3) to include a set of linear restrictions on the two-variable indicator-by-latent variable parameters:

$$\ln\left(f_{ijklt}^{ABCDX}\right) = \lambda + \lambda_t^X + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + g_t\lambda_{i*}^{AX} + g_t\lambda_{j*}^{BX} + g_t\lambda_{k*}^{CX} + g_t\lambda_{l*}^{DX}. \quad (17)$$

The single variable parameters of this loglinear parameterization are subject to the restrictions expressed in Equation (4). The two-variable parameters in Equation (16) are now restricted only with respect to summation

Table 8. LCM Evaluation Criteria for Abortion Approval Data

Model	X^2	G^2	AIC	BIC	df
H_1: two-class LCM	216.45	182.13	170.13	137.64	6
H_2: three-class model with linear restrictions	2.76	2.75	−7.25	−34.32	5
H_3: $H_2 + A, B, C$ restricted to equal association	4.52	4.57	−9.43	−47.32	7

over the levels of the indicator variable

$$\sum_i g_t \lambda_{i^*}^{AX} = \sum_j g_t \lambda_{j^*}^{BX} = \sum_k g_t \lambda_{k^*}^{CX} = \sum_l g_t \lambda_{l^*}^{DX} = 0, \tag{18}$$

because there is only a single estimate of each two-variable lambda for all T levels of the latent variable. The linear coefficient, g_t, takes on the values −1, 0, and 1 for the three levels of the latent variable.[9]

As the evaluation criteria in Table 8 indicate, the three-class linearly restricted model provides a good fit to the observed data. Both the Pearson (2.76) and likelihood ratio (2.75) chi squares (5 df) indicate a good fit to the data, as do the AIC (−7.25) and the BIC (−34.32). Because H_1 is a two-class model and H_2 is a three-class model, these are not nested; thus, no conditional G^2 test can be made. Given the four evaluation criteria listed in Table 8, we can conclude that the loglinear parameters for the four indicator variables can be linearly restricted across the three classes of the latent variable.

Although we have linearly restricted the dichotomously scored variables in the abortion attitudes example across the three classes of the latent variable, it is important to note that linear restrictions may also be imposed on the indicator variables when they have three or more ordered categories. Also, linear-by-linear restrictions can be imposed when both the indicator and latent variables have three or more ordered levels. McCutcheon (1996) has also shown that such restrictions are also possible on parameters related to a grouping variable when it has three or more ordered levels.

Our final hypothesis for these data tests the equality of linear association across the three indicator variables A, B, and C. As the evaluation criteria in Table 8 indicate, this hypothesis (H_3) is clearly acceptable, with a X^2 test statistic of 4.52 and a G^2 test statistic of 4.57 (7 df). Because H_3 is nested in H_2, we can use the conditional G^2 to evaluate the improvement of fit (1.82, 2 df) and find that, like the AIC (−9.43) and the BIC (−47.32), the conditional G^2 recommends acceptance of the equal linear association hypothesis for indicator variables A, B, and C.

As the parameters in Table 9 indicate, the indicator variables are very highly related to the three levels of the latent abortion attitudes variable in this 1982 sample of American adults. Moreover, the fourth indicator variable – whether married women who want to have no more children should be able to obtain a legal abortion – appears to be the indicator most highly related to the unobserved latent classes.

Table 9. Linearly Restricted Loglinear Parameters for Abortion Attitudes LCM (H_3)

	Parameter	
Indicator	**Single Variable**	**Two Variables**
A. Single woman	−0.095	4.151[a]
B. Poor married woman	0.124	4.151[a]
C. Any reason	−0.629	4.151[a]
D. Wants no more children	−0.084	6.917
X. Latent class variable		
($t = 1$)	0.141	—
($t = 2$)	−0.441	—

[a] Equality restrictions imposed.

5. MULTI-SAMPLE LATENT CLASS MODELS

Often researchers are confronted with a set of responses to identical indicator items in sample data from two or more populations. These samples may be from different social, cultural, or economic groups (McCutcheon, 1987); different regions, states, or nations (see, e.g., McCutcheon and Nawojczk, 1995; McCutcheon and Hagenaars, 1997); or the samples may be from the same group at two or more points in time (see, e.g., McCutcheon, 1986, 1996). Indeed, the samples may be from any mutually exclusive groups. When we are in such a situation, we can use a multi-sample, or simultaneous, LCM (SLCM) to compare the latent structures in the samples (Clogg and Goodman, 1984, 1985, 1986).

We begin our consideration of the multisample LCM by noting that our original parameterizations, shown in Equations (1) and (3), must be modified to reflect the addition of a grouping variable that reflects the populations from which we have samples. We designate that variable as G with S samples. Thus, if we have samples from four nations (i.e., $S = 4$), for example, we would have G_1, G_2, G_3, and G_4. The formal multisample probabilistic parameterization of the LCM can be expressed as

$$\pi_{ijklts}^{ABCDXG} = \pi_{s}^{G}\pi_{ts}^{X|G}\pi_{its}^{A|XG}\pi_{jts}^{B|XG}\pi_{kts}^{C|XG}\pi_{lts}^{D|XG}. \qquad (19)$$

We must also note that an inclusion of a grouping variable in this parameterization results in changes in the model restrictions:

$$\sum_{t}\pi_{ts}^{X|G} = \sum_{i}\pi_{its}^{A|XG} = \sum_{j}\pi_{jts}^{B|XG} = \sum_{k}\pi_{kts}^{C|XG} = \sum_{l}\pi_{lts}^{D|XG} = 1.0. \qquad (20)$$

We must also add similar modifications to the loglinear parameterizations of Equations (3):

$$\begin{aligned}\ln\left(f_{ijklts}^{ABCDXG}\right) = {} & \lambda + \lambda_s^G + \lambda_t^X + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{it}^{AX} \\ & + \lambda_{jt}^{BX} + \lambda_{kt}^{CX} + \lambda_{lt}^{DX} + \lambda_{ts}^{XG} + \lambda_{is}^{AG} + \lambda_{js}^{BG} \\ & + \lambda_{ks}^{CG} + \lambda_{ls}^{DG} + \lambda_{its}^{AXG} + \lambda_{jts}^{BXG} + \lambda_{kts}^{CXG} + \lambda_{lts}^{DXG}.\end{aligned} \tag{21}$$

Similarly, we must modify the model restrictions for Equation (21). In addition to the restrictions noted in Equation (4), we must add

$$\begin{aligned}\sum_s \lambda_s^G &= \sum_i \lambda_{is}^{AG} = \sum_j \lambda_{js}^{BG} = \sum_k \lambda_{ks}^{CG} = \sum_l \lambda_{ls}^{DG} = \sum_s \lambda_{is}^{AG} \\ &= \sum_s \lambda_{js}^{BG} = \sum_s \lambda_{ks}^{CG} = \sum_s \lambda_{ls}^{DG} = \sum_i \lambda_{its}^{AXG} = \sum_t \lambda_{its}^{AXG} \\ &= \sum_s \lambda_{its}^{AXG} = \sum_j \lambda_{jts}^{BXG} = \sum_t \lambda_{jts}^{BXG} = \sum_s \lambda_{jts}^{BXG} \\ &= \sum_k \lambda_{kts}^{CXG} = \sum_t \lambda_{kts}^{CXG} = \sum_s \lambda_{kts}^{CXG} = \sum_l \lambda_{lts}^{DXG} \\ &= \sum_t \lambda_{lts}^{DXG} = \sum_s \lambda_{lts}^{DXG} = 0.\end{aligned} \tag{22}$$

As with the single-sample LCM, the two parameterizations of the SLCM are equivalent. Thus, for convenience, we will adopt the usage of referring to the specific SLCM by the marginals that correspond to the highest-order interaction terms because in hierarchical loglinear models these terms represent the minimal sufficient statistics (Goodman, 1978; Agresti, 1990; Hagenaars, 1990). Because of the axiom of local independence, the SLCM is one of the few instances when we hypothesize a model with three-variable interactions. Consequently, we can refer to the model specified in Equations (19) and (21) as the {*AXG*, *BXG*, *CXG*, *DXG*} SLCM.

Typically, the researcher's first interest in SLCM is to establish *structural equivalence* – that is, to establish that the indicator variables are independent of the grouping variable. This means that the preferred instance is one in which the model specified in Equations (19) and (21) can be reduced to {*XG*, *AX*, *BX*, *CX*, *DX*}. In this instance, we can represent the latent variable as *structurally homogeneous* in the two or more samples, meaning that the associations between the latent variable and each of the indicator variables are identical across all of the samples. In this instance, we are confident that we are measuring the identical latent variable in each of the samples.

The less our data conform to this structural equivalence ideal, the less our certainty that we have truly measured the identical phenomenon in each of the samples. Divergences from complete structural equivalence in the several samples must be evaluated by the researcher. In some instances, divergences may be readily explained by factors such as differing historical, political, cultural, social, and economic circumstances (see, e.g., McCutcheon and Nawojczk, 1995).

Consider the data presented in Table 1. You will recall that in this example, students were asked about four scenarios involving role conflict. One group of 216 students was asked about the situations in which they personally were confronted with the dilemmas, and another group of 216 was asked about the situations in which their close friend was confronted by the dilemmas. Recall that a two-class model fit the Ego's Dilemma data well. Thus, the first model we consider is the two-class model for each of these two samples. The evaluation criteria presented in Table 10 suggest that the unrestricted two-class model (H_1) fits the data well.

As we see from these evaluation criteria, the two-class per sample model appears to provide an acceptable characterization of these two samples of students. Both the X^2 (9.06) and the G^2 (8.25) with 12 df support this conclusion. Moreover, both of the information criteria statistics also support the characterization of the two samples as each having two classes (H_1).

The next hypothesis we test (H_2) is the two-class *structural equivalence* model. This model tests the hypothesis that the two latent classes for the Ego's Dilemma sample are the same two classes as in the Friend's Dilemma sample. As the evaluation criteria indicate, this too is an acceptable hypothesis: the X^2 (24.78) and L^2 (23.47) on 20 df support this conclusion. Because H_2 is nested within H_1, we can also examine the conditional G^2 test (15.22 with 8 df), which also supports the structural equivalence hypothesis at the 0.05 alpha level ($p = .0550$). Finally, both of the information criteria statistics – the AIC (−16.53) and the BIC (−97.90) – indicate a preference for H_2 over H_1.

Table 10. Simultaneous LCM Evaluation Criteria for Ego's Dilemma and Friend's Dilemma Data: Two-Latent-Class Solution ($T = 2$)

Model	X^2	G^2	AIC	BIC	df
H_1: $\{AXG, BXG, CXG, DXG\}$	9.06	8.25	−15.75	−64.57	12
H_2: $\{XG, AX, BX, CX, DX\}$	24.78	23.47	−16.53	−97.90	20
H_3: $\{G, AX, BX, CX, DX\}$	24.82	23.48	−18.52	−103.96	21

The final hypothesis we might wish to examine is the *complete homogeneity* hypothesis (see H_3). The complete homogeneity hypothesis tests whether the latent structure is identical in both (all) samples *and* that the distribution of the latent variable is identical in both (all) samples. Because these two samples were both from the same population and the questionnaires with Ego's and Friend's Dilemmas were randomly assigned, in this particular instance the complete homogeneity hypothesis allows us to assess whether problem framing (Ego vs. Friend) appears to have a significant influence on the outcome of the observed response patterns (beyond what might occur by chance variation).

As the data in the last row of Table 10 indicate, the complete homogeneity hypothesis is clearly preferred to the other two hypotheses. Both of the chi-square statistics support this conclusion – the X^2 (24.82) and L^2 (23.48) on 21 df. Because H_3 is nested within H_2, we can examine the conditional G^2 test (0.01 with 1 df), which indicates virtually no increase in the likelihood ratio chi square from imposing this constraint. Finally, the two information criteria statistics – the AIC (−18.52) and the BIC (−103.96) – clearly indicate a preference for H_3 over H_2. Thus, we might conclude from these data that the manner in which one frames these scenarios – whether as Ego's Dilemma or as Friend's Dilemma – has no consequences for the pattern of responses; at least, not among the population of Harvard and Radcliffe social science students of 1950.

As the conditional probabilities and latent class probabilities reported in Table 11 indicate, there are no differences between the models for the two question framings. In every instance, the conditional probabilities for group 1 (Ego's Dilemma) equals the conditional probabilities for group 2 (Friend's Dilemma). Moreover, the distribution of the two classes of respondents is also identical.

Table 11. Latent Class and Conditional Probabilities for the SLCM for Ego's and Friend's Dilemmas: Complete Homogeneity (H_3)

	Ego's Dilemma		Friend's Dilemma	
Indicator	**Class 1**	**Class 2**	**Class 1**	**Class 2**
A. Passenger Friend	0.354	0.010	0.354	0.010
B. Drama Critic Friend	0.567	0.108	0.567	0.108
C. Insurance Doctor Friend	0.717	0.021	0.717	0.021
D. Board of Directors Friend	0.849	0.319	0.849	0.319
Latent Class Probabilities	0.7083	0.2917	0.7083	0.2917

Table 12. Loglinear Parameters for the SLCM for Ego's and Friend's Dilemmas: Complete Homogeneity (H_3)

Indicator	Single-Var. Parameters	Two-Var. Parameters (With *X*)	Two-Var. Parameters (With *G*)	Three-Var. Parameters (With *XG*)
A. Passenger Friend	−1.296	0.976	0.000	0.000
B. Drama Critic Friend	−0.461	0.596	0.000	0.000
C. Insurance Doctor Friend	−0.731	1.195	0.000	0.000
D. Board of Directors Friend	0.242	0.622	0.000	0.000
X. Latent Class Variable	0.443	—	0.000	—
G. Grouping Variable	0.000	0.000	—	—

Finally, as the loglinear coefficients in Table 12 clearly indicate, the complete homogeneity hypothesis involves restricting to zero all parameters that include the grouping variable (G). As a consequence, the two rightmost columns of Table 12 are all zero, as are all of the entries in the final row of Table 12. Only those loglinear coefficients that exclude the grouping variable are significantly different from zero.

Before this section is completed, it is important to note that all of the restrictions that are possible with the single-sample model are also possible in the multisample instance (McCutcheon and Hagenaars, 1997). Equality and deterministic restrictions may be used in both instances, as can linear restrictions on the loglinear parameters (McCutcheon, 1996). As with the single-sample instance, however, one must be cautious about imposing the restrictions, as these may influence the identification of the model. Moreover, the usual variant of the EM algorithm for the estimation of LCMs does not always work properly when models are restricted (Mooijaart and Van der Heijden, 1992).[10]

6. CONCLUSION

In this brief overview of the LCM, it has been possible to illustrate only a few of the reasons for its growing popularity as a research tool. A number of popular software programs are making LCMs easily accessible to social and behavioral researchers worldwide. Clearly, the two parameterizations make this approach a highly flexible and attractive method for the analysis of categorical data. The ability to impose a variety of restrictions on these parameterizations and the range of hypotheses that can be explicitly tested have also played an important part. Also, the extension of the LCM to the multisample instance has played an important role in its application in comparative research. As the remaining chapters of this volume clearly illustrate, the extensions of this powerful model continue to push it to the forefront in new areas of the analysis of categorical data.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.

Akaike, H. (1974). "A new look at statistical model identification," *IEEE Transactions on Automatic Control*, **AC-19**, 716–23.

Clogg, C. C. (1981). "New developments in latent structure analysis." In D. M. Jackson & E. F. Borgatta (eds.), *Factor Analysis and Measurement in Sociological Research*. Beverly Hills: Sage, pp. 214–80.

Clogg, C. C. (1988). "Latent class models for measuring." In R. Langeheine & J. Rost (eds.), *Latent Trait and Latent Class Models*. New York: Plenum, pp. 173–205.

Clogg, C. C. (1995). "Latent class models." In G. Arminger, C.C. Clogg, & M. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum, pp. 311–59.

Clogg, C. C., & Sawyer, D. O. (1981). "A comparison of alternative models for analyzing the scalability of response patterns." In S. Leinhart (ed.), *Sociological Methodology, 1981*. San Francisco: Jossey–Bass, pp. 240–80.

Clogg, C. C., & Goodman, L. A. (1984). "Latent structure analysis of a set of multidimensional contingency tables," *Journal of the American Statistical Society*, **79**, 762–71.

Clogg, C. C., & Goodman, L. A. (1985). "Simultaneous latent structure analysis in several groups." In N. B. Tuma (ed.), *Sociological Methodology*. San Francisco: Jossey–Bass, pp. xxx–xxx.

Clogg, C. C., & Goodman, L. A. (1986). "On scaling models applied to data from several groups," *Psychometrika*, **51**, 123–35.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical Models for Ordinal Variables*. Newbury Park, CA: Sage.

Davis, J. A., & Smith, T. W. (1982). *General Social Surveys, 1972–1982*. [Machine-Readable Data File]. Chicago: National Opinion Research Center.

Everitt, B. S. (1987). *Introduction to Optimization Methods and Their Application in Statistics*. London: Chapman & Hall.

Everitt, B. S., & Hand, D. J. (1981). *Finite Mixture Models*. New York: Chapman & Hall.

Goodman, L. A. (1968). "The analysis of cross-classified data: independence, quasi-independence and interactions in contingency tables with or without missing entries," *Journal of the American Statistical Association*, **63**, 1091–131.

Goodman, L. A. (1970). "The multivariate analysis of qualitative data: interactions among multiple classifications," *Journal of the American Statistical Association*, **65**, 225–56.

Goodman, L. A. (1974a). "The analysis of systems of qualitative variables when some of the variables are unobservable. Part I – A modified latent structure approach," *American Journal of Sociology*, **79**, 1197–59.

Goodman, L. A. (1974b). "Exploratory latent structure analysis using both identifiable and unidentifiable models," *Biometrika*, **61**, 215–31.

Goodman, L. A. (1978). *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis*. Jay Magidson (ed.). Cambridge, MA: Abt Books.

Goodman, L. A. (1979). "Simple models for the analysis of association in cross-classifications having ordered categories," *Journal of the American Statistical Association*, **74**, 537–52.

Haberman, S. J. (1974). "Log-linear models for frequency tables with ordered classification." *Biometrics*, **30**, 589–600.

Haberman, S. J. (1979). *Analysis of Qualitative Data, Vol. 2, New Developments*. New York: Academic Press.

Hagenaars, J. A. (1990). *Categorical Longitudinal Data: Log-Linear Panel, Trend, and Cohort Analysis*. Newbury Park, CA: Sage.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: Sage.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. Hayward, CA: Institute of Mathematical Science.

McCutcheon, A. L. (1986). "Sexual morality, pro-life values and attitudes toward abortion," *Sociological Methods and Research*, **16**, 256–75.

McCutcheon, A. L. (1987). *Latent Class Analysis*. Newbury Park, CA: Sage.

McCutcheon, A. L. (1996). "Multiple group association models with latent variables: an analysis of secular trends in abortion attitudes, 1972–1988." In A. Raftery (ed.), *Sociological Methodology, 1996*. Cambridge, MA: Blackwell, pp. xxx–xx.

McCutcheon, A. L., & Hagenaars, J. A. (1997). "Simultaneous latent class models for comparative social research." In R. Langeheine & J. Rost (eds.), *Applications of Latent Trait and Latent Class Models*. New York: Waxman, pp. 266–77.

McCutcheon, A. L., & Nawojczk, M. (1995). "Making the break: popular sentiment toward legalized abortion among Polish and American Catholic laities," *International Journal for Public Opinion Research*, **7**, 232–52.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

McLachlan, G. J., & Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

Mooijaart, A., & Van der Heijden, P. G. M. (1992). "The EM algorithm for latent class models with constraints," *Psychometrica*, **57**, 261–71.

Parsons, T. (1949). *Essays in Sociological Theory, Pure and Applied*. Glencoe, IL: Free Press.

Raftery, A. E. (1995). "Baysian model selection in social research." In P. V. Marsden (ed.), *Sociological Methodology, 1995*. Cambridge, MA: Blackwell, pp. xxx–xx.

Stouffer, S. A., & Toby, J. (1951). "Role conflict and personality," *American Journal of Sociology*, **89**, 395–406.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

Vermunt, J. K. (1997). *Log-Liner Models for Event Histories*. Thousand Oaks, CA: Sage.

Vermunt, J. K. (1997). *LEM: A General Program for the Analysis of Categorical Data* (users' manual). Tilburg, The Netherlands: Tilburg University.

Wedel, M., & Kamakura, W. A. (1998). *Market Segmentation: Conceptual and Methodological Foundations*. Boston: Kluwer Academic.

NOTES

1. In some instances, it is plausible that the observed associations reflect cause–effect relationships. In such instances, latent class analysis is inappropriate. As Goodman (1974a) and Hagenaars (1990, this volume) show, however, the LCM can be integrated into models that have such causal associations.
2. The LCM is part of the larger family of mixture models (see, e.g., McLaughlin and Peel, 2000; Lindsay, 1995; McLachlan and Basford, 1988; Titterington et al., 1985; Everitt and Hand, 1981). Within the area of mixture models, the latent class probability is often referred to as the *mixing proportion*.
3. We must be cautious in drawing the analogy between latent class probabilities and factor loadings because, unlike factor loadings, the interpretation of conditional probabilities depends on their size relative to those in other latent classes. For example, if Class 1 has a conditional probability of .70 of getting a right answer for item *A*, .70 is high if for Class 2 the conditional probability of getting it right is .25, but it is low if the conditional probability in Class 2 is .99.
4. It is important to note that an absolute 0.000 or 1.000 estimate of a conditional probability for an indicator is referred to as a *boundary estimate* because probabilities are bounded by 0.00 and 1.00. Boundary estimates are problematic within latent class analysis. During the iterative process for finding the parameter estimates, if one or more of the estimates goes to a boundary (i.e., either 0 or 1), all of the other parameters are maximum-likelihood solutions only if the parameters estimated to be at the boundary of the parameter space are truly zero or one in the population. Fortunately, most of the available software programs permit the researcher to specify multiple start values for the iterative process; it is highly recommended that this option be used in any instance in which a boundary value is estimated for an LCM.
5. An alternative approach is for the iterations to end after some prespecified number have been completed. This criterion is risky, however, as substantial changes in the parameter estimates may still occur after a very large number of iterations.
6. An additional concern regarding boundary values has to do with the determination of the degrees of freedom. When the boundary value is established as an a priori hypothesis (i.e., as a deterministic restriction), no parameter is

estimated. When the boundary value is the result of the estimation procedure, however, there is some concern about the number of degrees of freedom to associate with the model.

7. An alternative view of parsimony is to define the simplest model in terms of the logic of the model instead of the number of model parameters. Within the context of LCMs, for example, this approach is used to examine the scalability of a set of indicator variables (see, e.g., Clogg and Sawyer, 1981; Clogg and Goodman, 1986).
8. Once again, we must note that this approach to post-hoc hypothesis testing is for illustrative purposes only.
9. In general, if $j = 1, \ldots, J$, then each of the linear coefficients has the values that are computed as $j - (J + 1)/2$.
10. The LEM program (Vermunt, 1997) was used to estimate all of the models presented in this chapter. LEM corrects for the errors common to problems discussed by Mooijaart and van der Heijden (1992).