[DOI] http://dx.doi.org/10.21487/jrm.2020.11.5.3.183

【연구경향】

Latent Class Modeling for Nested Data : Introduction to Multilevel Latent Class Model*

Jungkyu Park** · Changhwan Shin***

ABSTRACT

The fundamental assumption in any latent variable model is that observations are independent of one another, given the latent status. However, this assumption is often inadequate when observations are nested within higher-level units because such nested data structures induce dependencies in data. The nonparametric version of the multilevel latent class model (MLCM) is an extension of latent class model (LCM) in which the dependencies in data are accounted for by discrete latent variables. This paper aims to review models with discrete latent variables and introduce the MLCM which integrates LCM and random effect model. The model selection issue in the MLCM is also discussed with an empirical example.

Keywords: Multilevel Latent Class model, Multilevel modeling, Latent class analysis, Model selection

---

* This research was supported by Kyungpook National University Research Fund, 2018
** Assistant Professor, Department of Psychology, Kyungpook National University
** Corresponding Author: Associate Professor, Department of Social Welfare, Kyungpook National University

I. Introduction

Latent Class Model (LCM) (Goodman, 1974; Lazarsfeld and Henry, 1968) is a probabilistic classification technique that identifies hidden clusters arising dependency among observed responses. In LCM, the hidden clusters are represented by levels of a discrete latent variable and are called "latent classes". Since the latent variable assumes to be discrete, it is usually considered as not imposing strong assumptions on the distributions for observed responses. As the comments by McCutcheon (1987, p.79 "*The social world seems to have been created with less multivariate normality than many researchers are willing to assume*"), the discrete latent variable offers an alternative for the latent variables with assumption of normal distribution.

In social or behavioral science, the nested structure may also cause some dependency between responses of subjects due to belonging in a higher-level unit. For example, students nested within a same school might show similar scores in terms of academic performance comparing to the students from other schools because the students in same school might share same characteristics concerning academic performance.

Vermunt (2003) proposed Multilevel Latent Class model (MLCM) to take into account such dependency due to same membership of a group as well as the dependency among responses caused by hidden clusters at individual-level. The MLCM differentiates the nonparametric and parametric approaches to account for the dependency which differs in the specified distributions of the random effect at the higher-level (i.e., discrete or continuous latent variable) (Vermunt, 2003, 2008).

The purpose of this paper is to review models with discrete latent variables and introduce the MLCM that integrates LCM and random effect model. The model selection issue in the MLCM is also discussed with an empirical example.

Ⅱ. Model for Discrete Latent Variables—Latent Class Models

Many empirical researches in the field of social sciences assumes the presence of “latent variables”. The latent variables can be thought as representing unobservable constructs or hypothetical concepts such as intelligence, skills, attitudes or personality traits. These variables are not directly measurable but can be inferred from observed behaviors or responses. Over the last few decades, various latent variable modeling techniques have been developed in the field of social and behavioral sciences. However, the methodological developments have mainly focused on continuous latent variables that are assumed to follow a multivariate normal distribution. Examples of modeling techniques with continuous latent variables are structural equation models (SEMs; Bollen, 1989) and latent trait models, which are also called item response theory models (IRTs; Lord, 1980).

In contrast to continuous latent variables, discrete latent variables have received relatively less attention, although they provide easy and intuitive explanations for phenomena under certain circumstances (Aitkin, 1999; McCutcheon, 1987). As the name suggests, discrete latent variables consist of a finite number of levels, which can be used to capture or symbolize the latent categories of theoretical concepts, constructs, entities, or subgroups. Examples of discrete latent variables include Sternberg's (1998) profile of thinking styles, Ainsworth, Blehar, Waters, and Wall's (1978) three attachment styles (i.e., secure, anxious-resistant, and avoidant), Bennett and Jordan's (1975) teaching styles, Fischer and Fischer's (1979) styles of teaching and learning, and Jung's (1971) psychological types. These examples demonstrate that discrete latent variables provide a framework for interpreting constructs with parsimonious descriptions of underlying structures. Owing to such variables' usefulness, modeling techniques using discrete latent variables have grown in popularity over the last decade (e.g., McLachlan & Peel, 2000; Vermunt & Magidson, 2008).

The latent class model (LCM; Goodman, 1974; Lazarsfeld & Henry, 1968) is a classic analytic tool that analyzes categorically scored cross-sectional data by introducing discrete latent variables. In the LCM, discrete latent variables are composed of mutually exclusive and exhaustive latent classes that represent a small number of subpopulations (Clogg, 1995). The primary purpose of LCM is to identify the optimal number of classes in order to explain the dependency among responses properly and classes are characterized by parameters that allow to vary randomly across different latent classes.

The LCM can be defined as follows. Let $\mathbf{Y_i}$ is a response vector for subject i on a set of J items, where i =1,⋯,I and j = 1,⋯,J. The discrete latent variable (X) represents the discrete variable which consists of M latent classes with particular latent class membership m. The marginal density of response vector of subject $\mathbf{Y_i}$, is:

$$P(\mathbf{Y}_i) = \sum_{m=1}^{M} P(X_i = m) f(\mathbf{Y}_i = \mathbf{y_i} \mid X_i = m) \tag{1}$$

Equation (1) indicates that the probability of obtaining full response vector, Y_i, is a weighted average of M conditional responses probability, $f(Y_i=y_i|X_i=m)$ and the weight is the corresponding latent class probabilities, $P(X_i=m)$, which represents the prevalence or size of a particular latent class. The latent class probabilities have a constraint that the sum of the parameters over the M latent classes must be equal to one, since all subjects are assumed to belong to one and only one of a mutually exclusive and exhaustive latent classes. The latent class probabilities are usually assumed to follow a multinomial distribution as presented:

$$P(X_i = m) = \frac{\exp(\gamma_m)}{\sum_{m'=1}^{M} \exp(\gamma_{m'})} \tag{2}$$

The intercept parameters γ_m are subjected on identifying constraint, that is $\sum_{m=0}^{M} \gamma_m = 0$ for the case of effect coding, $\gamma_m = 0$ or $\gamma_1 = 0$ for the case of dummy coding.

An underlying assumption in LCM is that latent classes are internally homogeneous, which means that all subjects in the same latent class share the same probability of responding to a certain item, and the relationships among observed responses can be explained by these latent classes (Clogg, 1995). Another assumption in LCM is called "local independence" which states that the probability of responding to each item is mutually independent given the latent class (McCutcheon, 2002). Therefore, density of response vector, Y_i is easily calculated by multiplying the density of each variable directly. Suppose Y_{ij} denotes a response of i subject on an item j, the local independence assumption can be represented as:

$$P(\mathbf{Y}_i) = \sum_{m=1}^{M} P(X_i = m) \prod_{j=1}^{J} f(Y_{ij} = y_{ij} \mid X_i = m). \tag{3}$$

The conditional response probability, $f(Y_{ij}=y_{ij} \mid X_i=m)$ characterizes the nature of the discrete latent variable identified by a few latent classes. It can take the form of probability distributions depending on the types of observed responses. Usually, the form of a multinomial distribution is preferred for this density, so it can be parameterized as:

$$P(Y_{ij} = y_{ij} \mid X_i = m) = \frac{\exp(\beta_{0j} + \beta_{1jm})}{\sum_{m'=1}^{M} \exp(\beta_{0j} + \beta_{1jm'})} \tag{4}$$

where β_{0j} denotes item intercept and β_{1jm} indicates class-specific random effect of X on the items. Note that some restrictions should be imposed on parameters for identification purposes: dummy coding or effect coding schemes can be used as similar as latent class probabilities, $P(X_i=m)$. In the equation (4), the value of the random coefficient, β_{1jm} varies randomly across classes, representing the class-specific effects.

Because these latent classes are internally homogenous within the same unit, but distinct between different units in terms of response patterns, individuals in the same latent class share the same probability of responding to a certain observed variable. The LCM assumes individuals belong to one and only one latent class and observed responses are independent of each other given an individual's latent class membership (McCutcheon, 2002). These assumptions imply that relationships between the categorical variables can be accounted for by their membership in latent classes (Clogg, 1995; McCutcheon, 1987).

The LCM was initially introduced by Lazarsfeld and Henry (1968) as a way of deriving latent attitude variables measured by binary survey items. This basic idea was extended by Goodman (1974), who developed a simple algorithm for obtaining the maximum likelihood estimates of parameters and incorporated nominal variables in models. During the 1980s, LCMs became a general methodology for analyzing categorical variables, as it was placed within the framework of log-linear models (Formann, 1982, 1985; Haberman, 1979). More idea about the basic LCM were provided by McCutcheon (1987), Heinen (1993), Clogg (1995), and Hagenaars and McCutcheon (2002).

Over the past few decades, several extensions of LCMs have been proposed. For example, LCMs were extended to accommodate various observed variable scale types, such as ordinal (Clogg, 1988; Heinen, 1996; Uebersax, 1993), continuous (Fraley & Raftery, 1998; McLachlan & Basford, 1988; Wolfe, 1970), and combinations of different scale types (Lawrence & Krzanowski, 1996; Vermunt & Magidson, 2001). The extensions also involve the inclusion of covariates to predict latent class membership, impose constraints (i.e., equality, inequality, and specific values) on parameters (McCutcheon, 1987), and relax the local independence assumption to make the model more flexible (Hagenaars, 2002). More recently, LCMs have been extended to include more complex data, such as complex survey data (e.g., Patterson, Dayton, & Graubard, 2002), longitudinal data (e.g., Collins & Wugalter, 1992; Muthén & Shedden, 1999; Nagin, 2005; Welch, 2003; Zucchini & MacDonald, 2009), and hierarchically nested data (e.g., Asparouhov &

Muthén, 2008; Di & Bandeen-Roche, 2008; Henry & Muthén, 2010; Vermunt, 2003, 2004; Vermunt & Magidson, 2008).

The above-mentioned extensions in LCMs are implemented in several standard software packages, such as Mplus (Muthén & Muthén, 2012), Latent GOLD (Vermunt & Magidson, 2013), SAS (PROC LCA, and LTA; (Lanza, Collins, Lemmon, & Schafer, 2007; Lanza & Collins, 2008), R (polca; Linzer & Lewis, 2011), and Stata (Rabe-Hesketh, Skrondal, & Pickles, 2004). These software packages allow researchers to more easily apply LCMs to empirical studies.

Ⅲ. Models for Multilevel Data—Random Effects Models

In social or behavioral science, data are often collected from a hierarchical structure, such as repeated measurements nested within an individual or individuals clustered within a higher-level group. Examples in the literature include school effects on student learning (Wright, Horn, & Sanders, 1997), family effects on child development (McLoyd, 1998), and neighborhood effects on changes in children and adolescents' psychological and behavioral outcomes (Leventhal & Brooks-Gunn, 2000).

Data collected from such hierarchically nested structures are naturally correlated, that is; the observed responses of lower-level units (e.g., pupils, time points) belonging to the same higher-level unit (e.g., schools, individuals) tend to correlate more with each other than with those from a different higher-level unit. These dependencies can be attributed to commonalities that the lower-level units may share by belonging to the same higher-level unit (e.g., common environments, experiences, and interactions). The presence of such dependencies is a concern for researchers because it not only violates the independence assumption, but also leads to biased parameter estimates (Moerbeek, 2004). Therefore, the dependency should be taken into account when analyzing multilevel data.

A popular approach for analyzing multilevel data is the random effects approach (Bryk & Raudenbush, 1992; Goldstein, 1995; Snijders & Bosker, 1999). The key feature of this approach is incorporating random effects into the models to account for dependencies among lower-level units. In the literature, traditional linear regression models that include both fixed and random effects are referred to as random coefficient models, linear mixed models, hierarchical models, and multilevel models (Bryk & Raudenbush, 1992; Goldstein, 1995; Hox, 2002; Skrondal & Rabe-Hesketh, 2004; Snijders & Bosker, 1999). These models include parametric random effects (or continuous random effects) that originate from a multivariate normal distribution with estimated variance components. Such random effects can take the form of random intercepts, reflecting differences in the overall level of the dependent variable across higher-level units or random slopes, reflecting differences in the effects of predictors across higher-level units, or both.

The random coefficient model is based on the assumption that both random effects and random errors are normally distributed. However, these underlying assumptions may not be realistic, particularly when observed variables are categorical (McCutcheon, 1987). To overcome this problem, variants of random coefficient models have been developed for analyzing categorical data such as random effects logistic models and Poisson regression models (Hartzel, Agresti, & Caffo, 2001; Hedeker & Gibbons, 1996; Rabe-Hesketh et al., 2001; Wong & Mason, 1985). These models are referred to as nonlinear mixed models. They were integrated into a more general framework called the generalized linear mixed model (GLMM), which permits both fixed and random effects for various types of observed variables (Breslow & Clayton, 1993).

Although the most common specification of random effects is the adoption of a parametric normal distribution, it is possible to use a nonparametric specification of random effects (Aitkin, 1999). The nonparametric random effect (or discrete random effect) is characterized by a set of latent classes at the higher-level that follow a multinomial distribution (Heckman & Singer, 1982; Laird, 1978; Vermunt, 1997). The regression models, including such nonparametric random effects, are referred to as latent class regression or finite mixture regression models (Vermunt & Dijk, 2001; Wedel & DeSarbo, 1994). An important difference between continuous and discrete random effects is that the latter do not estimate a variance component; instead, they estimate regression coefficients for each latent class, relying on the assumption that each higher-level unit belongs to one of the higher-level latent classes.

Ⅳ. Integration of Two Models—Multilevel Latent Class Models

An important limitation of standard LCMs is that they do not provide information related to the higher-level structure beyond the individual-level. To overcome this difficulty, researchers have attempted to integrate random effects into LCMs to accommodate clustered multilevel data within the framework of LCMs (e.g., Rabe-Hesketh et al., 2004; Skrondal & Rabe-Hesketh, 2004; Vermunt & Dijk, 2001). The incorporation of random effects into LCMs enables researchers to account for dependencies due to the nested structure and to disentangle them from the lower-level. Currently, several different specifications of random effects in LCMs have been discussed. Vermunt (2003) introduced either continuous or discrete random effects in the model of interest. Hedeker (2003) and Asparouhov and Muthén (2008) introduced a factor analytic approach, which includes a common factor to reduce the dimensionality of continuous random effects. Further, Di and Bandeen-Roche (2011) provided a discussion about random effects following the Dirichlet distribution.

The multilevel latent class model (MLCM; Vermunt, 2003) is an extension of the LCM that accommodates the dependencies due to nested data structure. The basic idea of the MLCM is to include random effects, which are represented by randomly varying parameters across higher-level units to capture the dependencies. Vermunt (2003) discussed two versions of MLCMs; the parametric MLCM, which assumes a continuous random effect that follows a multivariate normal distribution with an estimated variance term so that the magnitude of variance estimates can be used as a measure of the higher-level effects (e.g., Hedeker, 2003). In contrast, the MLCM with a nonparametric specification (Aitkin, 1999; Laird, 1978) incorporates a discrete random effect at the higher-level, which can be represented by a set of latent classes following the multinomial distribution. The primary purpose of nonparametric MLCMs is to identify a potential class structure that accounts for the association among the observed responses, and to classify groups and individuals into latent classes at each level.

An advantage of using the nonparametric MLCM over the parametric one is that it provides classification information for both higher-level units and lower-level units. This additional information about the latent structure and classification of higher-level units provides more useful descriptions and explanations of the observed association than the typical variance decomposition approach (e.g., Bassi, 2009; Bijmolt, Paas, & Vermunt, 2004; da Costa & Dias, 2014; Finch & Marchant, 2013; Onwezen et al., 2012; Pirani, 2011; Rindskopf, 2006; Rüdiger & Hans-Dieter, 2013).

Another advantage of nonparametric MLCMs is that they are computationally less intensive than parametric MLCMs, especially when models contain more than two random effects. Previous studies have shown that estimation of continuous random effects in the parametric MLCM is computationally intensive and time-consuming because complicated integrals must be solved (Aitkin, 1999). The most common method for evaluating these complex integrals is the use of the Gauss-Hermite quadrature (Rabe-Hesketh et al., 2004, Skrondal & Rabe-Hesketh, 2004; Stroud & Secrest, 1966), in which the multivariate normal distribution is approximated by a limited number of quadrature points with a few mass points. The integral can be approximated to any practical degree of accuracy only when quadrature points are sufficiently large (Lesaffre & Spiessens, 2001); otherwise, it may perform poorly (Agresti, Booth, & Caffo, 2000; Rodriguez & Goldman, 1995, 2001). In contrast, the estimation of discrete random effects in nonparametric MLCMs does not require inappropriate and unverifiable assumptions about the distribution of the random effects, thus it could avoid bias due to misspecification of the distribution of the random effects (Vermunt & Van Dijk, 2001).

The MLCM incorporates latent variables at two layers. The latent variable H_g is the discrete latent variable at higher-level (groups) with L latent clusters, and X_{gi} denotes the discrete latent variables at lower-level (individuals) with M latent classes. Each outcome of discrete random variables can be conceptualized as a latent cluster/class consisting of groups/individuals that are homogenous within each cluster/class but are distinct between clusters/classes in the response patterns. The term "clusters" and "classes" are used to differentiate the higher and lower classes. Let Y_{gij} represent the response to the jth item of a subject (i) in a group (g), where g=1,⋯,G, i=1,⋯,ng, and j=1,⋯J. The vector $\mathbf{Y_{gi}}$ represents the J responses for a subject i nested in group g, and $\mathbf{Y_g}$ denotes the full vector of responses for all subjects in group g. A MLCM is defined by two separate equations for higher and lower levels.

The probability of observing a certain response pattern for all subjects nested in group g is:

$$P(\mathbf{Y}_g) = \sum_{l=1}^{L} P(H_g = l) \prod_{i=1}^{n_g} P(\mathbf{Y}_{gi} \mid H = l) \tag{5}$$

Equation (5) assumes that each group belongs to only one l (latent cluster), and conditional densities for each of the ng (individuals) within the g (group) are independent of each other given the latent cluster membership. The term, $P(H_g=l)$, is referred to as latent cluster probabilities with each element representing the probabilities of g being in the cluster l (l=1,⋯L). As the clusters are assumed to be mutually exhaustive and exclusive, elements of this vector can be conceptualized as cluster sizes, and thus the sum of this vector is one.

At the individual level, the probability of obtaining a certain response pattern for each subject is:

$$P(\mathbf{Y}_{gi}) = \sum_{m=1}^{M} P(X_{gi} = m \mid H_g = l) \prod_{j=1}^{J} f(Y_{gij} = y_{gij} \mid X_{gi} = m, H_g = l) \tag{6}$$

The term $P(X_{gi}=m|H_g=l)$ is the conditional latent class probabilities, which represents the distribution of latent class probabilities within a particular latent cluster. A $M\times L$ matrix denoted as ρml will be used to represent the conditional latent class probabilities. Columns in this matrix ρml represent the probability distributions of the latent classes for a specific latent cluster; thus, each column must have a sum of one.

The conditional response density, $f(Y_{gij}=y_{gij}|X_{gi}=m,\ H_g=l)$, is the probability of observing y_{gi} for variable j of individual i in group g given the latent cluster membership (l) and latent class membership (m). In most multilevel extensions of LCM (e.g., Asparouhov & Muthén, 2008; Vermunt, 2003, 2004; Vermunt & Magidson, 2008), a restricted model is proposed by imposing a constraint on the conditional density: $f(Y_{gij}=y_{gij}|X_{gi}=m,\ H_g=l)=f(Y_{gij}=y_{gij}|X_{gi}=m)$.

This constraint implies that the conditional response density is affected only by the latent class memberships; this not only simplifies the model but also aids the interpretation of results. By combining Equations (5) and (6) with the assumption of no effects of latent cluster membership on response probabilities and the assumption of local independence, the MLCM is:

$$P(\mathbf{Y}_g = \mathbf{y}_g) = \sum_{l=1}^{L} P(H = l) \left(\prod_{i=1}^{n_g} \sum_{m=1}^{M} P(X_{gi} = m \mid H_g = l) \prod_{j=1}^{J} f(Y_{gij} \mid X_{gi} = m) \right) \tag{7}$$

Density $f(Y_{gij}|X_{gi}=m)$ depends on the assumed distributions of responses. For binary items, the mth latent class density is given by $f(y_{gij}\,|\,X_{gi}=m)=\alpha_{mj}^{y_{gij}}(1-\alpha_{mj})^{1-y_{gij}}$, where α_{mj} denotes the probability of endorsing item j for an individual belonging to latent class m.

The EM algorithm (Dempster, Laird & Rubin, 1977) is an efficient iterative procedure to compute maximum likelihood estimate in the presence of incomplete or hidden data (Dempster et al, 1977). However, the standard EM algorithm is difficult to adjust in MLCM since the number of entries of joint posterior probabilities is huge making the method impractical. To avoid such problem, a variant of EM algorithm for MLCM called the "upward-downward" procedure was proposed. Details of this algorithm can be found in Vermunt (2003, 2004)

V. Model Selections for MLCM

There was extensive work on this issue of selecting number of classes for classical latent class models. There are two general approaches for selecting the number of latent classes in a LCM: hypothesis testing and information criteria. One common approach of statistical testing approach is to use likelihood ratio statistic. Many studies have discussed this approach in the context of LCM (e.g., Everitt & Hand, 1981; Everitt, 1988). Various information-based criteria were also studied for selecting number of classes for the classical LCM (e.g., Collins, Fidler, Wugalter, & Long, 1993; Lin & Dayton, 1997; Yang, 2006; Nylund, Asparouhov, & Muthén, 2007).

Even though the number of latent classes for LCM have been extensively studied in the literature, the MLCM with two levels of latent components makes the decisions on numbers of latent components at each level become more challenging and complex. This is because the task of model selection issue in the MLCM amounts to decide the optimal discrete levels of the latent variable in both higher-level and lower-level.

A common strategy to decide the discrete latent components in most applications is the stepwise decision. Lukočienė, Varriale, and Vermunt (2010) proposed a three-step approach. This approach first started with determining the optimal number of classes (m) ignoring the cluster (l) by assuming $l=1$. Subsequently, they chose the number of l by fixing the number of m obtained from the first step, and then re-determined the m by fixing the structure of l to the value chosen at the previous stage. An alternative model selection strategy is to choose m and l simultaneously. Specifically, a set of candidate models with combinations of l and m are fitted, and then the ICs obtained from the models are compared with each other to choose the optimal combination of l and m at once.

1. Information Criteria

Information criteria for model selection were originally derived by Akaike (1973) who used the Kullback-Leibler information measure to discriminate between competing models. Schwarz (1978) derived another major class of information criteria by using Bayesian statistics. Since then, many modified information criteria have been derived or proposed. Using the notation from Sclove (1987), information criteria (IC) can be summarized in the following form,

$$IC = -2\log(\max L(k)) + a(n)m(k) + b(k,n) \tag{8}$$

Where n is the sample size, "log" denotes the natural logarithm, max $L(k)$ denotes the maximum of the likelihood over the parameters, and $m(k)$ is the number of independent parameters in the kth model. For a given model, $a(n)$ is the cost of fitting an additional parameter and $b(k,n)$ is an additional term depending upon the criterion and the model k.

Therefore, the IC aim to find a good balance between model fitness (trying to maximize the likelihood function) and parsimony (penalizing additional complexity). A model is good if it gives a small value of IC relative to the values given by competing models. Using this notation, various information criteria can be summarized

〈Table 1〉 Summary of information criteria

Criterion	Definition	Reference
AIC	$-2LL+2P$	Akaike, 1973
AIC3	$-2LL+3P$	Bozdogan, 1993
CAIC	$-2LL+(1+\log(n))P$	Bozdogan, 1987
BIC	$-2LL+\log(n)P$	Schwarz, 1978
Adjusted BIC (ABIC)	$-2LL+\log((n+2)/24)P$	Sclove, 1987

Ⅵ. Empirical Example

Data from study 2 of the Motivated Identity Construction in Cultural Context 2008-2011 (Vignoles & Brown, 2011) is used to illustrate the model selection by IC. Study 2 (N=8,652) of this data set investigate the contextualism across 35 nations in predicting in-group favoritism, corruption, and differential trust of in-group and out-group members. Further information of this study is available on the webpage of this study.

The contextualism is shown to be an important part of cultural collectivism. The six-item contextualism scale (listed in Table 2) developed by Owe at el. (2012) is used in this illustration. Three items of this scale required reverse coding. The original scale is a six-point Likert scale: "Completely disagree," "Moderately disagree," "Slightly disagree," "Slightly agree," "Moderately agree," and "Completely agree." Responses were dichotomized into "disagree" and "agree" to be consistent with what is specified in the simulation studies. We sample 50 participants for each participated nation and results to the final data set of 1700 participants from 34 nations.[1)]

〈Table 2〉 Items descriptions of the contextualism scale

1	To understand a person well, it is essential to know about which social groups he/she is a member of.
2	One can understand a person well without knowing about his/her family [reversed].
3	To understand a person well, it is essential to know about the place he/she comes from.
4	One can understand a person well without knowing about his/her social position [reversed].
5	One can understand a person well without knowing about the place he/she comes from [reversed].
6	To understand a person well, it is essential to know about his/her family

1)One country was dropped due to an insufficient number of participants.

〈Table 3〉 Likelihood values, numbers of parameters, and fit indices for data analyzed by MLCM

Model	H	L	LL	Npar	AIC	AIC3	CAIC	BIC	ABIC
M1	2	2	-6386	15	12802	12817	12898	12883	12836
M2	2	3	-6255	23	12556	12579	12704	12681	12608
M3	2	4	-6208	31	12478	12509	12677	12646	12548
M4	2	5	-6187	39	12451	12490	12702	12663	12539
M5	2	6	-6171	47	12435	12482	12738	12691	12542
M6	3	2	-6382	17	12797	12814	12906	12889	12835
M7	3	3	-6244	26	12540	12566	12707	12681	12599
M8	3	4	-6195	35	12460	12495	12685	12650	12539
M9	3	5	-6169	44	12427	12471	12710	12666	12526
M10	3	6	-6150	53	12405	12458	12746	12693	12525
M11	4	2	-6382	19	12801	12820	12923	12904	12844
M12	4	3	-6239	29	12535	12564	12722	12693	12601
M13	4	4	-6184	39	12446	12485	12697	12658	12535
M14	4	5	-6155	49	12408	12457	12724	12675	12519
M15	4	6	-6141	59	12399	12458	12779	12720	12533
M16	5	2	-6382	21	12805	12826	12940	12919	12852
M17	5	3	-6236	32	12536	12568	12742	12710	12609
M18	5	4	-6178	43	12442	12485	12719	12676	12539
M19	5	5	-6151	54	12409	12463	12757	12703	12531
M20	5	6	-6130	65	12389	12454	12808	12743	12536

The sampled data were fitted to 20 MLCMs with latent structure with 2-5 clusters in combination of 2-6 classes. The log-likelihood values, numbers of parameters, and IC under comparisons are presented in

Table 3. Among the five IC, the CAIC and BIC preferred Model 3, ABIC preferred Model 14, and AIC and AIC3 selected the most complicated model (Model 20). When using G and ng for calculating IC, CAIC and BIC picked the four-cluster and five-class model (Model 14) and ABIC preferred the most complicated Model 20.

Both Model 14 and Model 20 are very complex for assuming many classes and clusters; unless the hypotheses suggested such a latent structure, researchers generally would prefer a simpler model. Moreover, since the estimated parameters for these two models do not give a clear pattern in interpreting the latent structure, the estimated parameters of Model 3, is the model preferred by CAIC and BIC, are presented in Table 4. The four classes have relatively equal size, though the Class 1 is a little larger. The four classes demonstrated four distinctive patterns: participants of Class 4 generally agree with all six items, while participants in Class 2 tend to disagree with these items. Class 1 and Class 3 show reverse patterns on these items, except item 2.

At the nations-level, a larger cluster (A) consisted of 80% of the nations, and the smaller cluster (B) consisted of 20% of the nations. Based on the responses of each nation, we can assign nations into the two clusters based on the posterior latent cluster probability. The six nations categorized in Cluster B are: China, Ethiopia, Philippines, Singapore, Turkey, and Thailand. The majorities of nations were grouped in Cluster A. The differences between the two clusters are clear. The two clusters have comparable class sizes for class 1 and 3. Cluster A has a larger Class 2 (all-low) and a smaller Class 4 (all-high), but Cluster B has a dominant Class 4 (all-high) and a very small Class 2 (all-low). Since the six countries categorized into Cluster B are mainly in Asia, as the dominant Class 4 has high loading on all six items on the contextualism scale, this cluster is consistent with the cultural trait of collectivism (Triandis, 1995). On the other hand, the pattern in Cluster A reflects the typical individualism of the non-Asian countries in the study.

〈Table 4〉 The estimated model parameters of the 2-clusters and 4-classes MLCM (Model 3)

	Items	Class1	Class2	Class3	Class4
1	Know social group	0.48	0.07	0.14	0.30
2	Without Knowing family	0.26	0.10	0.26	0.38
3	Know place come from	0.48	0.03	0.16	0.34
4	Without knowing social position	0.21	0.09	0.30	0.41
5	Without knowing come from	0.14	0.03	0.38	0.45
6	Know family	0.47	0.11	0.13	0.29
		Two clusters and four classes			
	Latent cluster probabilities	$P(H_g = l) = \begin{bmatrix} .80 \\ .20 \end{bmatrix}$			
	Conditional latent class probabilities	$P(X \mid H) = \begin{bmatrix} .34 & .31 \\ .30 & .05 \\ .22 & .26 \\ .14 & .38 \end{bmatrix}$			

Ⅶ. Conclusion

LCMs have become a standard tool in many applied research fields. For many years, this basic methodology has been extended to accommodate multilevel nested data. The MLCM is an extension of

LCMs that accounts for the additional dependency in multilevel data by introducing random effects in the model. The nonparametric version of the MLCM includes the discrete random effect at the higher-level, relying on the assumption that higher-level units belong to homogeneous latent classes. In the MLCM, higher-level units can be classified into a small number of latent classes, providing useful descriptions of latent structures with categorical or typological natures. This paper introduced the nonparametric MLCM that combined LCM and random effect model and addressed model selection issue related to the use of MLCMs as tools for analyzing multilevel data.

Ⅷ. Reference

Agresti, A., Booth, J. G., and Caffo, B. 2000. "Random-effects Modelling of Categorical Response Data," *Sociological Methodology* 30: 27-80.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., and Wall, S. 1978. *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Aitkin, M. 1999. "A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models." *Biometrics* 55: 117-128.

Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle" edited by B. N. Petrov, and F. Caski, 267-281. Paper presented at the Second International Symposium on Information Theory, Budapest.

Asparouhov, T., and Muthén, B. 2008. "Multilevel mixture models." In *Advances in Latent Variable Mixture Models,* edited by G. R. Hancock & K. M. Samuelson, 27–51. Charlotte, NC: Information Age.

Bassi, F. 2009. "Latent Class Models for Marketing Strategies: an Application to the Italian Pharmaceutical Market." *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 5: 40–45.

Bennett, N., and Jordan, J. 1975. "A Typology of Teaching Styles in Primary Schools." *British Journal of Educational Psychology* 45: 20–28.

Bijmolt, T. H. A., Paas, L. J., and Vermunt, J. K. 2004. "Country and Consumer Segmentation: Multi-level Latent Class Analysis of

Financial Product Ownership." *International Journal of Research in Marketing* 21: 323-334.

Bollen, K. A. 1989. *Structural Equations with Latent Variables.* New York, NY: Wiley.

Bozdogan, H. 1987. "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical extensions." *Psychometrika* 52: 345-370.

Bozdogan, H. 1993. "Choosing the Number of Component Clusters in the Mixture-model Using a New Informational Complexity Criterion of the Inverse-fisher In-formation Matrix." In *Information and Classification*, edited by O. Opitz, B. Lausen, and R. Klar, 218-234. Heidelberg: Springer.

Breslow, N. E., and Clayton, D. G. 1993. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88: 9-25.

Bryk, A. S., and Raudenbush, S. W. 1992. "Application of Hierarchical Linear Models to Assessing Change." *Psychological Bulletin* 101: 147-158.

Clogg, C. C. 1995. "Latent Class Models." In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by G. Arminger, C. C. Clogg, and M. E. New York: Plenum.

Collins, L. M., Fidler, P. L., Wugalter, S. E., and Long, J. D. 1993. "Goodness-of-fit Testing for Latent Class Models." *Multivariate Behavioral Research* 28: 375-389.

Collins, L. M., and Wugalter, S. E. 1992. "Latent Class Models for Stage-sequential Dynamic Latent Variables." *Multivariate Behavioral Research* 27: 131-157.

da Costa, L. P., and Dias, J. G. 2014. "What Do Europeans Believe to Be the Causes of Poverty? A Multilevel Analysis of Heterogeneity Within and between Countries." *Social Indicators Research* 122: 1–20.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B* 39: 1–38.

Di, C. Z., and Bandeen-Roche, K. 2011. "Multilevel Latent Class Models with Dirichlet Mixing Distribution." *Biometrics* 67: 86–96.

Everitt, B. S. 1981. "A Monte Carlo Investigation of the Likelihood Ratio Test for Number of Components in a Mixture of Normal Distributions." *Multivariate Behavioral Research* 16: 171–180.

Everitt, B. S. 1988. "A Monte Carlo Investigation of the Likelihood Ratio Test for Number of Classes in Latent Classes Analysis." *Multivariate Behavioral Research* 23: 531–538.

Finch, W. H., and Marchant, G. J. 2013. "Application of Multilevel Latent Class Analysis to Identify Achievement and Socio-economic Typologies in the 20 Wealthiest Countries." *Journal of Educational and Developmental Psychology* 3: 201–221.

Fischer, B. B., and Fischer, L. 1979. "Styles in Teaching and Learning." *Educational Leaderships* 36: 245–254.

Fraley, C., and Raftery, A. E. 1998. "How Many Clusters? Which Clustering Methods? Answers via Model-based Cluster Analysis." *Computer Journal* 41: 578–588.

Formann, A. K. 1982. "Linear Logistic Latent Class Analysis." *Biometrical Journal* 24: 171–190.

Formann, A. K. 1985. "Constrained Latent Class Models: Theory and Applications." *British Journal of Mathematical and Statistical Psychology* 38: 87-111.

Goldstein, H. 1995. *Multilevel Statistical Models.* New York, NY: Halsted.

Goodman, L. A. 1974. "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: 1179-1259.

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

Hagenaars, J. A., and McCutcheon, A. 2002. *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press.

Hartzel, J., Agresti, A., and Caffo, B. 2001. "Multinomial Logit Random Effects Models." *Statistical Modelling* 1: 81-102.

Heckman, J., and Singer, B. 1982. "Population Heterogeneity in Demographic Models." In *Multidimensional Mathematical Demography*, edited by A. Rodgers and K. Land. New York: Academic Press.

Hedeker, D. 2003. "A Mixed-effects Multinomial Logistic Regression Model." *Statistics in Medicine* 22: 1433-1446.

Hedeker, D., and Gibbons, R. 1996. "MIXOR: A Computer Program for Mixed-effects Ordinal Regression Analysis." *Computer Methods and Programs in Biomedicine* 49: 157-176.

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

Henry, K. L., and Muthén, B. 2010. "Multilevel Latent Class Analysis: An Application of Adolescent Smoking Typologies with Individual and

Contextual Predictors." *Structural Equation Modeling* 17: 193-215.

Hox, J. 2002. *Multilevel Analysis: Techniques and Applications.* Mahwah, NJ: Lawrence Erlbaum.

Jung, C. G. 1971. *Psychological Types (Collected works of C. G. Jung) (Vol. 6),* edited by R. F. Hull. London: Routledge and Kegan Paul.

Laird, N. M. 1978 "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution." *Journal of the American Statistical Association* 73: 805-811.

Lawrence C. J., and Krzanowski, W. J. 1996. "Mixture Separation for Mixed-mode Data." *Statistics and Computing* 6: 85-92.

Lazarsfeld, P. F., and Henry, N. W. 1968. *Latent Structure Analysis.* Boston, MA: Houghton Mifflin.

Lanza, S. T., and Collins, L. M. 2008. "A New SAS Procedure for Latent Transition Analysis: Transitions in Dating and Sexual Behavior." *Developmental Psychology* 42: 446-456.

Lanza, S. T., Collins, L. M., Lemmon, D., & Schafer, J. 2007. "PROC LCA: A SAS Procedure for Latent Class Analysis." *Structural Equation Equation Modeling* 14: 671-694.

Lin, T. H., and Dayton, C. M. 1997. "Model Selection Information Criteria for Nonnested Latent Class Models." *Journal of Educational and Behavioral Statistics* 22: 249-264.

Linzer, D. A., and Lewis, J. B. 2011. "poLCA: An R Package for Polytomous Variable Latent Class Analysis." *Journal of Statistical Software* 42: 1-29.

Lesaffre, E., and Spiessens, B. 2001. "On the Effect of the Number of Quadrature Points in a Logistic Random-effects Model: an Example." *Applied Statistics* 50: 325-335.

Leventhal, T., and Brooks-Gunn, J. 2000. "The Neighborhoods they Live in: The Effects of Neighborhood Residence upon Child and Adolescent Outcomes." *Psychological Bulletin* 126: 309-337.

Lord, F. M. 1980. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Erlbaum.

Lukočienė, O., Varriale, R., and Vermunt, J. K. 2010. "The Simultaneous Decision(s) about the Number of Lower- and Higher-level Classes in Multilevel Latent Class Analysis." *Sociological Methodology* 40: 247-283.

McCutcheon, A. L. 1987. *Latent Class Analysis.* London, UK: Sage.

McCutcheon, A. L. 2002. "Basic Concepts and Procedures in Single-and multiple-group Latent Class Analysis." In *Applied Latent Class Analysis*, edited by J. A. Hagenaars and A. L. McCutcheon, 57-88. Cambridge, UK: Cambridge University Press.

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

McLachlan, G. J., and Peel, D. 2000. *Finite Mixture Models.* Wiley: New York.

McLoyd, V. C. 1998. "Socioeconomic Disadvantage and Child Development." *American Psychologist* 53: 185-204.

Moerbeek, M. 2004. "The Consequence of Ignoring a Level of Nesting in Multilevel Analysis." *Multivariate Behavioral Research* 39: 129-149.

Muthén, B., and Shedden, K. 1999. "Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm." *Biometrics* 55: 463-469.

Nagin, D. S. 2005. *Group-based Modeling of Development.* Cambridge, MA: Harvard University Press.

Nylund, K. L., Asparouhov, T., and Muthén, B. O. 2007. "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study." *Structural Equation Modeling* 14: 535-569.

Onwezen, M. C., Reinders, M. J., Lans, V. D. I., Sijtsema, S. J., Jasiulewicz, A., Guardia, M. D., and Guerrero, L. 2012. "A Cross-national Consumer Segmentation Based on Contextual Differences in Food Choice Benefits." *Food Quality and Preference* 24: 276-286.

Owe, E., Vignoles, V. L., Becker, M., Brown, R., Smith, P. B., Lee, S. W. S., and Jalal, B. 2013. "Contextualism as an Important Facet of Individualism-collectivism: Personhood Beliefs Across 37 National Groups." *Journal of Cross-Cultural Psychology* 44: 24-45.

Patterson, B., Dayton, C. M., and Graubard, B. 2002. "Latent Class Analysis of Complex Survey Data: Application to Dietary Data." *Journal of the American Statistical Association* 97: 721 – 729 .

Pirani, E. 2011. "Evaluating Contemporary Social Exclusion in Europe: A Hierarchical Latent Class Approach." *Quality and Quantity* 47: 923-941.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. 2001. "GLLAMM: A General Class of Multilevel Models and a Stata Program." *Multilevel Modelling Newsletter* 13: 17-23.

Rindskopf, D. 2006. "Heavy Alcohol Use in the "Fighting Back" Survey Sample: Separating Individual and Community Level Influences Using Multilevel Latent Class Analysis." *Journal of Drug Issues* 36: 441-462.

Rodriguez, G., and Goldman, N. 1995. "An Assessment of Estimation Procedures for Multilevel Models with Binary Responses." *Journal of the Royal Statistical Society, Series A* 158: 73-89.

Rodriguez, G., and Goldman, N. 2001. "Improved Estimation Procedures for Multilevel Models with Binary Response: a Case Study." *Journal of the Royal Statistical Society, Series A* 164: 339-355.

Rüdiger, M., and Hans-Dieter, D. 2013. "University and Student Segmentation: Multilevel Latent Class Analysis of Students' Attitudes Toward Research Methods and Statistics." *British Journal of Educational Psychology* 83: 280-304.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461-464.

Sclove, S. L. 1987. "Application of Model-selection Criteria to Some Problems in Multivariate Analysis." *Psychometrika* 52: 333-343.

Skrondal, A., and Rabe-Hesketh, S. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*." Boca Raton, FL: Chapman & Hall/CRC.

Snijders, T. A. B., and Bosker, R. J. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Sternberg, R. J. 1998. "Mental Self-government: A Theory of Intellectual Styles and their Development." *Human Development* 31: 197-224.

Stroud, A. H., and Secrest, D. 1966. *Gaussian Quadrature Formulas.* Englewood Cliffs, NJ: Prentice-Hall.

Uebersax, J. S. 1993. "Statistical Modeling of Expert Ratings on Medical Treatment Appropriateness." *Journal of the American Statistical Association* 88: 421-427.

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

Vermunt, J. K. 2003. "Multilevel Latent Class Models." Sociological *Methodology* 33: 213-239.

Vermunt, J. K. 2004. "An EM Algorithm for the Estimation of Parametric and Nonparametric Hierarchical Nonlinear Models." *Statistica Neerlandica* 58: 220-233.

Vermunt, J. K., and van Dijk, L. A. 2001. "A Nonparametric Random Coefficient Approach: The Latent Class Regression Model." *Multilevel Modelling Newsletter* 13: 6-13.

Vermunt, J. K., and Magidson, J. 2008. *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module.* Belmont, MA: Statistical Innovations.

Vignoles, V., and Brown, R. 2011. *Motivated Identity Construction in Cultural Context, 2008- 2011 (Study No. 6877)* [Computer file]. Colchester, Essex: UK Data Archive [Distributor].

Wedel, M., and DeSarbo, W. S. 1994. "A Review of Latent Class Regression Models and their Applications." In *Advanced Methods for Marketing Research*, edited R. P. Bagozzi, 353-388. Cambridge, MA: Blackwell.

Welch, L. R. 2003. "Hidden Markov Models and the Baum-Welch Algorithm." *IEEE Information Theory Society Newsletter* 53: 1-13.

Wolfe, J. H. 1970. "Pattern Clustering by Multivariate Cluster Analysis." *Multivariate Behavioral Research* 5: 329-350.

Wong, G. Y., and Mason, W. M. 1985. "The hierarchical logistic-regression model for multilevel analysis." *Journal of the American Statistical Association* 80: 513-524.

Wright, S. P., Horn, S. P., and Sanders, W. L. 1997. "Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation." *Journal of Personnel Evaluation in Education* 11: 57-67.

Yang, C. C. 2006. "Evaluating Latent Class Analyses in Qualitative Phenotype Identification." *Computational Statistics & Data Analysis* 50: 1090-1104.

Zucchini, W., and MacDonald, I. L. 2009. *Hidden Markov Models for Time Series: An Introduction Using R*. Boca Raton: Chapman

다층자료 분석을 위한 잠재계층분석: 다층 잠재계층모형*

박중규** · 신창환***

논문요약

잠재변수를 포함한 통계모형의 기본적인 가정은 개인의 응답이 서로 독립적이라는 것이다. 그러나 관찰치가 상위 수준의 집단에 속한 다층 구조에서 수집된 자료의 경우에는 동일한 집단에 속한 관찰치의 응답 간 상관이 발생하므로 독립성 가정이 충족되지 않는다. 비모수 다층잠재계층 모형은 집단수준의 범주형 잠재변인을 가정하여 동일한 집단에 속하여 발생하는 응답 간 상관을 설명하는 모형이다. 본 논문은 잠재계층분석 모형과 무선효과 모형을 통합한 비모수적 다층잠재계층 모형을 소개하고 실제 자료를 바탕으로 개인과 집단수준에서 최적의 잠재계층의 수를 결정하는 과정에 대해서 논의한다.

주제어: 다층 잠재계층 모형, 다층 모형, 잠재계층분석, 모형 선택

투고일: 2020.10.06. 심사일: 2020.10.16. 게재확정일: 2020.11.06.

* 이 논문은 2018학년도 신임교수정착연구비에 의하여 연구되었음

** 경북대학교 심리학과 조교수, 주저자.
*** 경북대학교 사회복지학과 부교수, 교신저자.